

A Statistical Model of Microscope Resolution



Dissertation
zur Erlangung des mathematisch-naturwissenschaftlichen
Doktorgrades
“Doctor rerum naturalium”
der Georg-August-Universität Göttingen
im Promotionsprogramm
“PhD School of Mathematical Sciences (SMS)”
der Georg-August University School of Science (GAUSS)

vorgelegt von
Gytis Kulaitis
aus Vilnius, Litauen

Göttingen, 2020

Betreuungsausschuss:

Prof. Dr. Axel Munk

Institut für Mathematische Stochastik, Universität Göttingen

Prof. Dr. Tatyana Krivobokova

Institut für Mathematische Stochastik, Universität Göttingen

Mitglieder der Prüfungskommission:

Referent:

Prof. Dr. Axel Munk

Institut für Mathematische Stochastik, Universität Göttingen

Korreferentin:

Prof. Dr. Tatyana Krivobokova

Institut für Mathematische Stochastik, Universität Göttingen

Weitere Mitglieder der Prüfungskommission:

Prof. Dr. Alexander Egner

Laser-Laboratorium Göttingen, Universität Göttingen

Prof. Dr. Gerlind Plonka-Hoch

Institut für Numerische und Angewandte Mathematik, Universität
Göttingen

Prof. Dr. Dominic Schuhmacher

Institut für Mathematische Stochastik, Universität Göttingen

Dr. Frank Werner

Institut für Mathematische Stochastik, Universität Göttingen

Tag der mündlichen Prüfung: 21.02.2020

Acknowledgments

First and foremost I would like to thank Prof. Axel Munk for trusting an outsider to do a PhD in Statistics and for his constant guidance. Next I would like to thank Dr. Frank Werner for coming up with the topic and being very helpful when I had any questions.

I am grateful to the members of the committee for agreeing to evaluate my thesis, in particular, to Prof. Tatyana Krivobokova for taking up the duty of Korreferentin.

I find myself lucky to have colleagues at the IMS. I think that the institute is a friendly, supportive and an inspiring place to work.

A big thanks goes to my office mates: Florian Pein, Marco Seiler and Miguel del Álamo Ruiz for helping me with various (non)mathematical issues. Especially when I started and when I was sick. I would like to thank Natalia Khizanishvili and Heiner Keilholz for the same.

My sincere gratitude goes to the RTG 2088 for organizing nice workshops, lecture series and conference funding. I would also like to thank the DFG, which partially supported my PhD through CRC 755.

I would also like to acknowledge many great teachers outside of Göttingen: Virginija Ulinskaitė, Laima Visminienė, Ala Tiščenko, Dr. Rimantas Raudonis, Ovidijus Kavaliauskas, Inga Miliukaitė, Rimantė Bukauskienė, Dr. Jonathan Keeling, Dr. Antje Kohnle, Dr. Christopher Hooley, Prof. Lars Olsen, Prof. Martin Schottenloher and others, who I might have forgotten to mention. I think that any project such as this comes to existence only because of years of hard work and great teachers who make sure that the student does not become yet another “not good at math”.

For proofreading my thesis, I thank Ieva Kazlauskaitė.

My family.

Last but not least, my girlfriend Harithaa.

Preface

A general rule of thumb in imaging is that the resolution of a light microscope depends linearly on the full width at half maximum (FWHM) of its point spread function (psf). In the present work we carefully define a statistical model of resolution by introducing a notion of discernability based on statistical testing whether one or two objects with the same total intensity are present. We consider four common ways of modeling photons detected in a microscopy experiment: as binomial, Poisson, variance stabilized Gaussian (VSG) or homogeneous Gaussian (HG) independent random variables. We show that under the binomial, Poisson and VSG photon models the resolution indeed depends linearly on the FWHM. However, under the HG model, the resolution depends on the FWHM to the power of $5/4$. Thus, at least for microscopy the HG model is too simple and in most experiments the Poisson or the VSG model is preferred, since they are easier to tackle than the binomial model, yet still capture the dependence on the FWHM correctly.

The rest of this work is arranged as follows:

- In Chapter 1 we provide a short introduction to modern microscopy (so-called nanoscopy or super-resolution) together with a concise history of resolution criteria that are still in use.
- In Chapter 2 we specify our modeling and give a statistical definition of resolution. We also state our main theorem on asymptotic statistical resolution and present its experimental implications.
- In Chapter 3 we compare our results to others', most notably, to Helstrom's more general quantum optical approach (Helstrom, 1973) that shows the limitations of our modeling, and to Acuña and Horowitz's (Acuña and Horowitz, 1997) demonstrating the applicability of our modeling to telescopes.
- Chapter 4 is devoted to assess the finite sample validity of the asymptotic theory. Reassuringly, even with low number of photons and coarse detector discretization, the slopes of the simulation fits can be approximated well by the theoretical ones.

For larger number of photons and finer discretization, the asymptotic formulas become good approximations to simulations.

- We split the proofs into two chapters: proofs that we think are necessary in the first reading are presented in Chapter 5 and the rest of the proofs in Appendix A.
- Finally, in Chapter 6 we discuss our results in a broader context and indicate possible future research directions.

With the exception of multiple dimensions, the vast majority of this thesis results out of a collaborative research effort among Prof. Axel Munk, Dr. Frank Werner and I (Kulaitis et al., 2020).

Contents

1	Introduction	1
1.1	Lens optics and diffraction	1
1.2	From microscopy to nanoscopy	3
2	Model and main results	11
2.1	Statistical model	11
2.2	Statistical testing problem	13
2.3	Main theorem	18
2.4	Implications for experiments	20
3	Comparisons with previous work	25
3.1	Overview	25
3.2	Abbe and Rayleigh	27
3.3	Milanfar and coauthors' work	28
3.4	Acuña and Horowitz's work	30
3.5	Helstrom's work	32
4	Simulations	35
4.1	One dimension	36
4.1.1	Symmetric $q = 1/2$ case	36
4.1.2	Asymmetric $q = 1/5$ case	38
4.2	Two and three dimensions	39
5	Proof of the main theorem	43
5.1	Homogeneous Gaussian model	43
5.2	Variance stabilized Gaussian model	46
5.3	Poisson and binomial models	48
5.3.1	Analysis in the central limit theorem regime	48
5.3.2	1D Poisson model analysis in the asymptotic equivalence regime	55

6	Discussion and future work	59
A	Other proofs	61
A.1	An integral approximation	61
A.2	Which alternative is the most difficult	62
A.3	Proof of the binomial CLT	64
	Bibliography	71

CHAPTER 1

Introduction

1.1 Lens optics and diffraction

According to geometrical optics, an ideal light microscope would be able to distinguish two points in space being arbitrarily close. However, in 1873 Abbe formulated (Abbe, 1873) what later became known as the *Abbe diffraction limit* (Figure 1.1C), namely, that two points can be resolved only if their distance d in space is at least

$$d = \frac{\lambda}{2 \text{NA}}, \quad (1.1)$$

where λ is the wavelength of incoming light and NA is the numerical aperture of the microscope. The numerical aperture is equal to the product of the refractive index of the medium (1 for vacuum, ≈ 1 for air) and the sine of one-half of the angle of the cone of light that can enter the microscope. Abbe argued in (Abbe, 1873) that diffraction and interference of light have to be taken into account when distances in the order of the wavelength of the illumination light are considered. See (Cremer and Masters, 2013) and references therein for a modern summary of Abbe's work. This paradigm has limited light microscopy for more than a century until the advent of super-resolution microscopy (Hell and Wichmann, 1994), see Section 1.2. For the following, it is beneficial to recall the basic physics tailored to our needs.

Given a specimen under the microscope f , due to diffraction (and the resulting interference) the imaging system causes a blur so that we do not simply observe an M times magnified image of f . This blur is usually modeled by first calculating or estimating the blur pattern of a single point – the point spread function (psf) h . For an incoherent imaging system, e.g. a fluorescence microscope, using Huygens's principle, the image of the specimen then can be obtained by summing up the blurred images of

the points constituting the sample. In other words we get a convolution

$$g(x) = \int_O h(x - Mx') f(x') dx', \quad (1.2)$$

where O is the space containing the specimen – the object space – and $f : O \rightarrow \mathbb{R}$. The space consisting of magnified points Mx' is called the image space I and $g : I \rightarrow \mathbb{R}$ is the image of the specimen.

If the microscope was perfect and there was no blur, then the psf h would simply correspond to a delta function $\delta_{x-Mx'}$, so that $g(x) = f(x/M)$. In general, the psf h can be computed explicitly by scalar diffraction theory. Under the assumption of circular aperture and using the paraxial approximation (Born and Wolf, 1999; Orfanidis, 2016), h becomes proportional to the Airy pattern (Airy, 1835) (Figure 1.1A)

$$h(x) \propto \left| 2A \left(\frac{2\pi \text{NA}}{\lambda} \frac{\text{NA}}{M} \|x\|_2 \right) \right|^2, \quad (1.3)$$

where λ is the illumination wavelength and $\|\cdot\|_2$ is the Euclidean norm. The function A in (1.3) is given by $A(u) = J_1(u)/u$, where J_1 is the Bessel function of the first kind.

Independently of Abbe, Lord Rayleigh formulated in 1879 a resolution criterion for spectroscopes (Strutt, 1879). Applied to microscopes Rayleigh's criterion reads that two point sources at x_1 and x_2 having equal intensity can just be resolved if the central maximum of the first psf centered at x_1 coincides with the first minimum of the second psf. The first zero of the Bessel function J_1 is at $x \approx 3.8317$ and hence $x/2\pi \approx 0.6098$. Thus, in the case of circular aperture the Rayleigh criterion reads

$$d = 0.61 \frac{\lambda}{\text{NA}}. \quad (1.4)$$

Note that this is slightly more conservative than Abbe's result. See Figure 1.1C and D for a comparison.

The resolution criteria (1.1) and (1.4) can be understood in terms of the full width at half maximum (FWHM) of the (effective) psf, see Figure 1.1B, where $\text{FWHM} = |x_2 - x_1|$. More precisely, the FWHM is defined as the width of the psf when its intensity is half of its maximal intensity. The ability to state both Abbe and Rayleigh criteria in terms of the FWHM has led to the common understanding that two point sources in space can be resolved by a light microscope as soon as their distance is larger than roughly the FWHM of the psf h . Usage of the FWHM as a resolution criterion dates back to at least the 1927 paper by Houston (Houston, 1927) and is still popular today (Egner et al., 2020). The FWHM criterion is particularly well-suited if the psf is approximated by a Gaussian kernel as shown in Figure 1.1B, since this function

does not have any local minima. Note that the approximation of the psf by a Gaussian is very common, see e.g. (von Diezmann et al., 2017). For an Airy pattern (1.3), the FWHM can be computed by first calculating the FWHM of $A(u)^2 = (J_1(u)/u)^2$. Due to $\max_u (A(u)^2) = A(0)^2 = 1$, this is equivalent to solving $J_1(u) = \pm u/\sqrt{2}$ resulting in the FWHM of 3.232 for A^2 . Hence, taking the additional scaling factors in (1.3) into account together with $Mx' = x$, we get the FWHM resolution criterion in its most common form

$$d = \text{FWHM} = 0.51 \frac{\lambda}{\text{NA}}. \quad (1.5)$$

Thus, the FWHM limit is almost equal to the Abbe resolution limit (1.1) and somewhat below the Rayleigh resolution limit (1.4).

Due to their generality, the above resolution criteria are not confined to microscopes and can also be applied to telescopes, see Section 3.4, or imaging in general. We stress that there are many other resolution criteria, such as the recently repurposed Fourier ring correlation criterion (Banterle et al., 2013), which will not be discussed in this thesis.

From Equations (1.1), (1.4) and (1.5) it seems that there are only two possible ways to improve the resolution: either the wavelength has to be decreased, or the numerical aperture increased. Since the wavelength λ is inversely proportional to the energy, decreasing the wavelength might damage the sample, a major issue in living cell microscopy, and hence visible light (380 nm – 760 nm) is preferred for most applications. Concerning the second option, the numerical aperture of a modern lens is around 1.3 – 1.5 (von Diezmann et al., 2017), and this value has not improved substantially during the last decades. In fact, Abbe’s resolution limit has been standing as a paradigm for more than a hundred years, limiting conventional light microscopes to about 250 nm lateral and 500 nm axial resolution (Hell, 2007; Cremer and Masters, 2013; Heintzmann and Ficz, 2013).

1.2 From microscopy to nanoscopy

One important idea to improve on Abbe’s resolution limit is confocal microscopy suggested by Minsky (Minsky, 1961; Pawley, 2006) in 1961. Here only a small spot of the object is illuminated at any given time, and non-focused light is blocked by a pinhole. Moving the pinhole over the sample (scanning) creates multiple images which are then combined to produce the full image. Clearly, the smaller the pinhole, the more the resolution is increased. On the other hand, a smaller pinhole decreases the overall image intensity. Theoretically confocal microscopy increases the resolution by $\sqrt{2}$, see e.g. (Egner et al., 2020) or (Hell, 2007), but due to these competing effects practical increase is lower. Consequently, although providing some improvement, confocal microscopy on

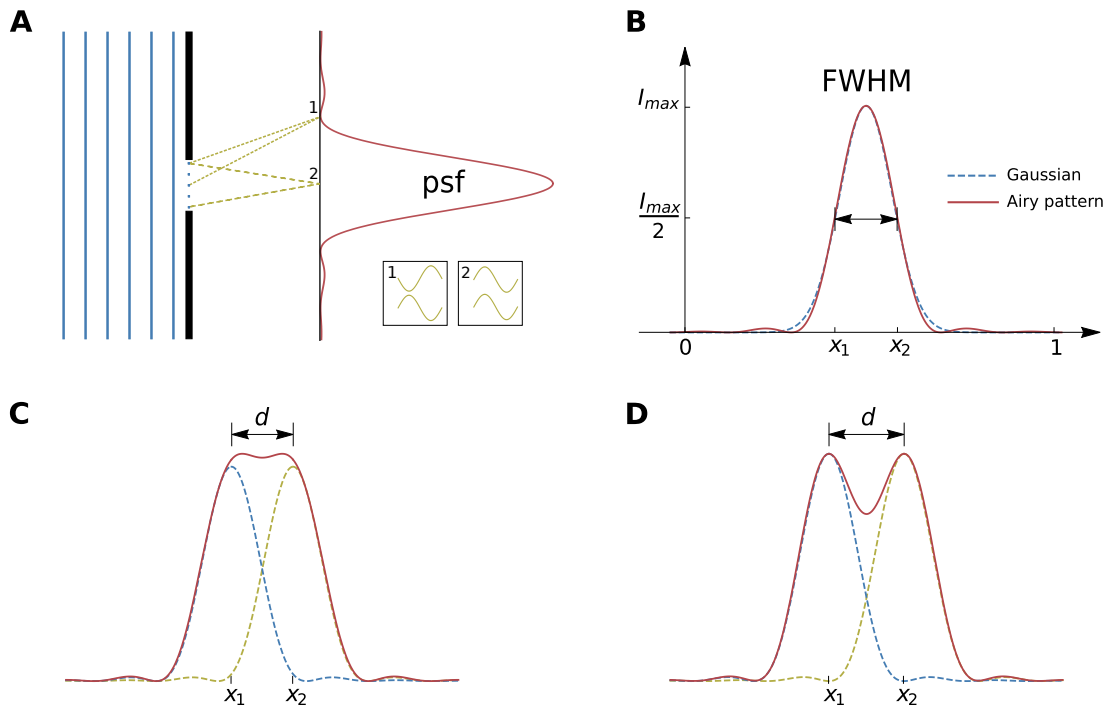


Figure 1.1: (A) 1D view of a 2D wave traveling through a circular aperture of width on the same order as the wavelength. By Huygen's principle each point on a wavefront acts as a point source (5 points shown). Due to diffraction and interference an Airy pattern is formed—where the light interferes constructively/destructively we get (local) maxima/minima in the intensity pattern. If the distance between the aperture and the screen is much larger than the wavelength, the slit acts as a point light source. (B) Approximation of an Airy pattern centered at $\frac{1}{2}(x_1 + x_2)$ by a Gaussian profile matching the maxima with the FWHM indicated. (C)/(D) Two Airy patterns centered at x_1 and x_2 , distance (1.1)/(1.4) apart, and their superposition (solid red).

its own cannot break the resolution barrier (Aspelmeier et al., 2015).

An early approach to overcome Abbe's resolution limit relies on the fact that both limits in Equations (1.1) and (1.4) are only valid in the *far-field*, i.e. when sample and microscope are sufficiently far apart. Similarly, the regime when the sample and the microscope are less than a wavelength apart is called *near-field*. In this case, the size of the aperture and not the wavelength determines the resolution (Courjon, 2003). In 1972 Ash and Nicholls (Ash and Nicholls, 1972) went below Abbe's diffraction limit in the near-field. Using 3 cm wavelength they achieved a resolution of $\lambda/60$. Current experiments are able to achieve a lateral resolution of 20 nm and a vertical resolution of 2–5 nm (Dürig et al., 1986; Oshikane et al., 2007). Although impressive, near-field microscopes have certain disadvantages, the most obvious being that the specimen must be very near the microscope and hence are mostly limited to surface measurements. Moreover, they are unsuitable for transparent objects which excludes many biological samples.

Breaking Abbe's diffraction limit using far-field microscopy is intimately related to the development of photoswitchable fluorophores (Aspelmeier et al., 2015; Huang et al., 2009) which can be switched on and off in a statistically controlled manner. After a laser excitation they emit light of higher wavelength (less energy) than absorbed due to rotational and vibrational losses. Exploiting this, the sample is scanned along a grid by illuminating it with a (pulsed) excitation beam focused at the current grid point and only the resulting fluorescence is measured. Using a dichroic beamsplitter, it is ensured that only the fluoresced light is detected at the detector. On each grid point this procedure is repeated for a fixed time (the pixel dwell time) t or equivalently for a fixed number of pulses (also denoted by t). Therefore, one is able to image specific predefined structures, instead of observing a superposition of the whole sample. This methodology led to the 2014 Nobel prize in Chemistry being awarded to Eric Betzig, Stefan W. Hell and William E. Moerner "for the development of super-resolved fluorescence microscopy" (Ehrenberg, 2014), where the term *super-resolution* refers to any technique, which is able to break Abbe's diffraction limit in the far field.

Nowadays there exist two main approaches to photoswitching:

Scanning mode Exploiting non-linearity of the response to excitation, dyes in a pre-defined region are shut off to enhance resolution.

Stochastic mode Exploiting chemical complexity of dyes, they can be forced to emit light at separate times making them resolvable in time.

In our mathematical treatment *we will focus exclusively on the scanning mode*, which makes our modeling more transparent by forgoing time dependency. However,

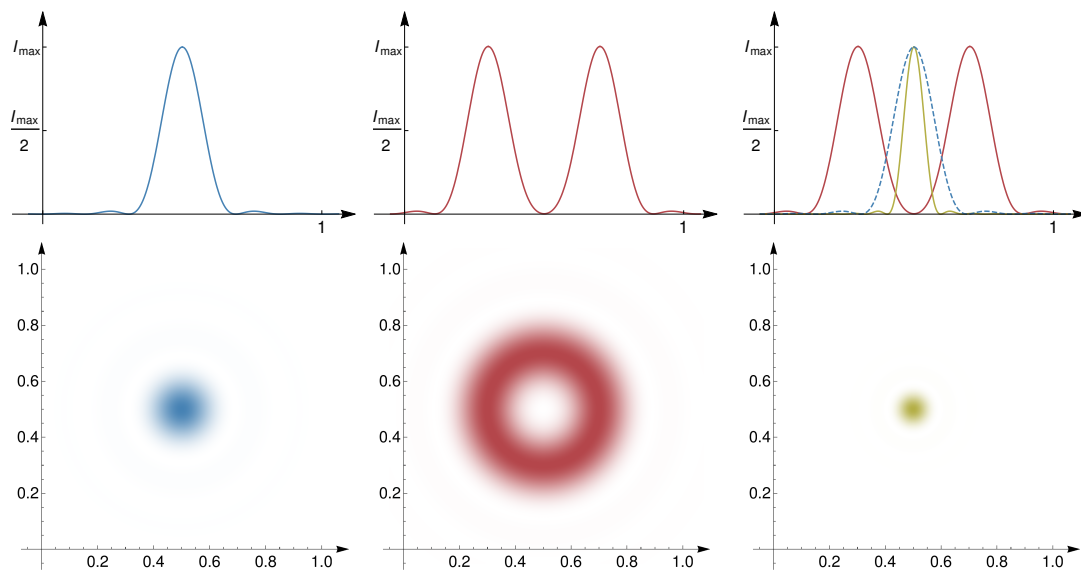


Figure 1.2: STED microscopy. Column I: Original psf (blue), Column II: Depletion psf (red), Column III: effective psf (solid beige). The top row shows psfs in 1D, the bottom row in 2D.

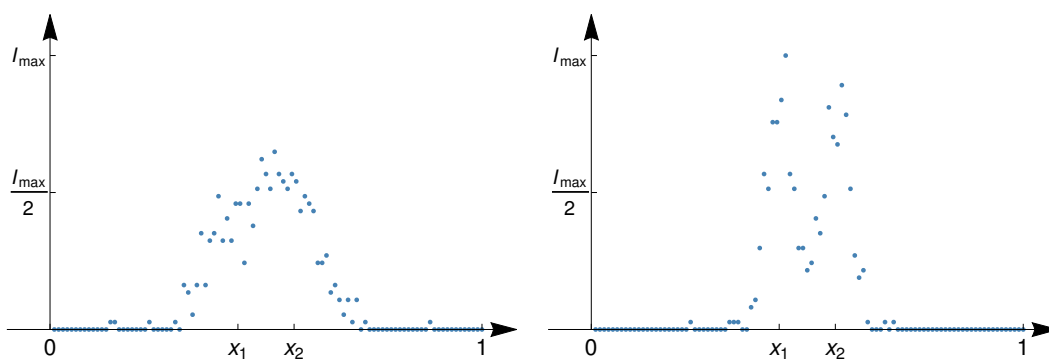


Figure 1.3: Two point sources at x_1 and x_2 that are difficult to distinguish with the original Airy psf (left), but are easily distinguishable with narrower Airy psf after STED (right).

if a summation over time is applied (possibly causing a loss of information), then the subsequent analysis could also be applied to the stochastic mode. To get the gist of these two types of approaches to photoswitching, we now briefly describe one scanning and one stochastic mode technique.

Arguably the state-of-the-art scanning mode super-resolution technique is Stimulated Emission Depletion (STED) (Hell and Wichmann, 1994; Klar et al., 2000), where the fluorescent dyes are only excited in the center of a torus shaped region and are actively depleted inside the torus, see Figures 1.2 and 1.3. The resolution under a STED microscope is described by a modified Abbe formula (Westphal and Hell, 2005; Hell, 2007; Reuss et al., 2010)

$$d = \frac{\lambda}{2 \text{NA} \sqrt{1 + \frac{I_{max}}{I_{sat}}}}, \quad (1.6)$$

where I_{max} is the maximum intensity of the depletion psf (2nd column of Figure 1.2) and I_{sat} is usually on the order of $1 - 10 \text{ MW/cm}^2$. Defining $\lambda' = \lambda / \sqrt{1 + \frac{I_{max}}{I_{sat}}}$, or, equivalently, using (1.5) and setting $\text{FWHM}' = \text{FWHM} / \sqrt{1 + \frac{I_{max}}{I_{sat}}}$ we see that previous resolution considerations still hold, but now we have to use an *effective* psf having FWHM' (3rd column of Figure 1.2). In principle, in STED microscopy the resolution can be increased indefinitely by increasing I_{max}/I_{sat} . In practice, resolutions of around 2.4 nm have been achieved this way, see (Rittweger et al., 2009).

As an example of stochastic mode photoswitching, we mention Single Marker Switching (SMS) nanoscopy in its various variants (Betzig et al., 2006; Rust et al., 2006; Hess et al., 2006; Heilemann et al., 2008; Egner et al., 2007), see also (Staudt et al., 2020) for a survey from a statistical perspective and (Du and Kou, 2020) for a survey on single-molecule techniques. Here one excites only a few dyes per pulse by using only a small illumination intensity. Hence, only single dyes which are spatially well separated are excited in each pulse with high probability. Consequently, there is no need to distinguish between two or more point sources, and thus from this point of view the resolution is arbitrarily small. However, the actual limitation is given by the localization accuracy when estimating the position of each fluorophore by the center of the observed psf (without any need for deconvolution). This can be understood from a statistical point of view as estimating the mean μ of a distribution by its empirical mean. Let N be the random number of photons observed in a small region of space and denote by X_1, \dots, X_N their spatial positions. Note that N depends on the illumination time $t > 0$, which can be chosen in the experimental setup, and $\mathbb{E}[N] \sim t$ (in our model to be introduced below, we in fact have $\mathbb{E}[N] = t$, see Section 2.1). Then in two dimensions

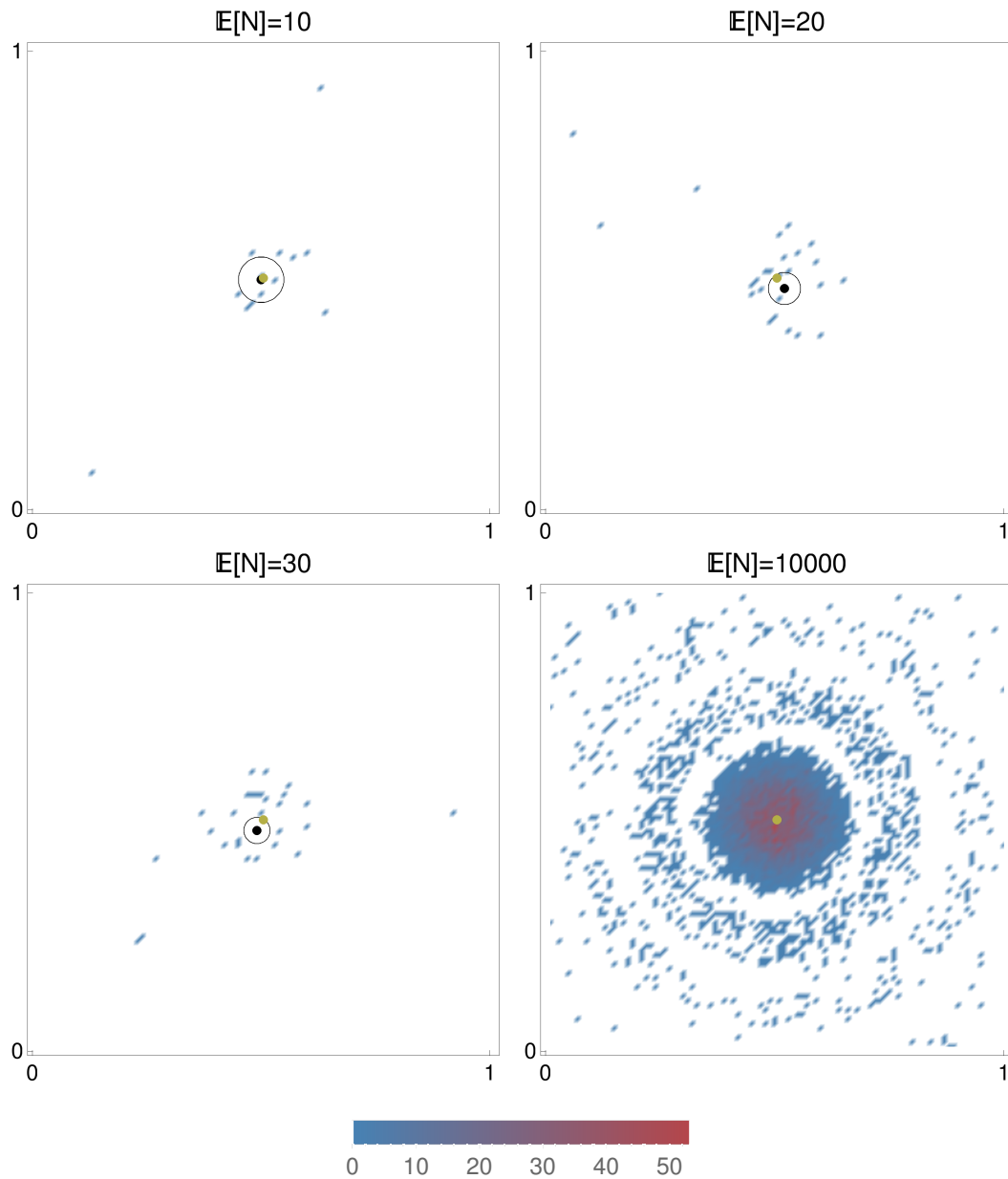


Figure 1.4: Illustration of localization error in SMS microscopy (1.8). Here the psf h is the Airy pattern (1.3). The beige dots mark the center of the Airy distribution $(0.5, 0.5)$ and the black dots the empirical means. The black circles correspond to the 90% confidence circles under the CLT (1.7).

the CLT yields

$$\sqrt{N} \left(\frac{1}{N} \sum_{i=1}^N X_i - \mu \right) \rightarrow^{\mathcal{D}} \mathcal{N}_2(0, \Sigma) \quad \text{as } t \rightarrow \infty, \quad \text{a.s.} \quad (1.7)$$

with a covariance matrix Σ given in terms of the psf h . Thus, neglecting the background and pixelation noise, the position of the sample's center can be estimated as the average of the fluorophore centers leading to the localization error

$$d \sim \frac{1}{\sqrt{\mathbb{E}[N]}}, \quad (1.8)$$

see (Thompson et al., 2002). This can be made more precise in terms of confidence circles for the true position of the dye as shown in Figure 1.4. Note that although (1.8) suggests that the resolution can be increased indefinitely, in practice the number of observable photons is limited due to the dyes suffering from photodamage which causes them to bleach and hence to lose the ability to fluoresce.

Comparing any of the FWHM based criteria (1.1), (1.4), (1.5) to (1.8) reveals a gap in the general understanding of *resolution* and *localization accuracy*, namely, that both the experimental setup *and* the statistical error should play a role in the actual resolution of a microscope. In any real world experiment, noise plays a central role in the actual ability to distinguish two point sources, and thus the noise level (e.g. the observed number of photons) should also play a role in Equations (1.1) and (1.4). This becomes more severe as the resolution increases. In addition, the effective psf should also affect the localization accuracy in (1.8).

Given the vast applications of microscopy and rapid progress of super-resolution, a refined understanding of fundamental principles governing resolution is of immense importance. However, as far as we know, a mathematically rigorous treatment of statistical resolution is still lacking. The current work aims to bridge this gap by presenting a statistical model including both the influence of the psf and the noise, and defining resolution rigorously in terms of statistical hypothesis testing that could also be used to quantify localization accuracy.

CHAPTER 2

Model and main results

2.1 Statistical model

To derive a mathematically rigorous formulation of the resolution of a (fluorescence) microscope with psf h , we start with modeling the observations. In practice, the physical space O is scanned bin-wise or sampled at once by a CCD camera or another detection device. We will assume that the image space I , the space of magnified points, is a cube $[0, 1]^m$ (typically $m = 2, 3$) and that it inherits the binning from the object space, i.e. that I consists of bins B_i , $i \in \{1, \dots, n\}^m$, with each B_i just an m -dimensional cube of volume $(1/n)^m$. Scanning at a bin B_i now means to center the psf at the center of B_i . Each bin is either illuminated $t \in \mathbb{N}$ times by a short excitation pulse (pulsed illumination) or illuminated continuously for some time t (continuous illumination), which we may also assume to be an integer due to time discretization (e.g. t can denote time in pico- or nanoseconds). We denote the number of detected photons in the i th bin by $Y_i \in \mathbb{N}$. Clearly, Y_i is a random quantity, but according to the above reasoning, we may assume that

$$\mathbb{E}[Y_i] = t \int_{B_i} g(x) dx, \quad (2.1)$$

where g is the image of the specimen as defined in (1.2). In the following we assume that the measurements at B_i and B_j are independent if $i \neq j$, which is experimentally confirmed in many different settings, see e.g. (Aspelmeier et al., 2015; Hohage and Werner, 2016). Consequently, we observe an m -dimensional field $(Y_i)_{i \in \{1, \dots, n\}^m}$ of independent random variables in the mean value parametrization

$$Y_i \stackrel{\text{indep.}}{\sim} F_{t, \int_{B_i} g(x) dx}, \quad i \in \{1, \dots, n\}^m \quad (2.2)$$

with a family of distributions $F_{t, \theta}$ for parameters $t \in \mathbb{N}$, $\theta \in (0, 1)$. Note that although the illumination time t enters the mean in (2.1) only as a factor, the distribution of Y_i

might be affected differently. Thus, we separate t in (2.2) from

$$\int_{B_i} g(x) dx. \quad (2.3)$$

The specific choice of $F_{t,\theta}$ depends fundamentally on the imaging setup and on the number of photons collected. We consider the following scenarios here:

Binomial model (B)

In case of pulsed illumination, we can think of detected photons as independent particles. Thus, the number of photons is distributed binomially

$$F_{t,\theta} = \text{Bin}(t, \theta). \quad (2.4)$$

This seems to be the most accurate model for microscopy, see e.g. (Aspelmeier et al., 2015).

Poisson model (P)

In case of continuous illumination, it follows from elementary properties of Poisson processes that a Poisson model

$$F_{t,\theta} = \text{Poi}(t\theta) \quad (2.5)$$

is appropriate (Munk et al., 2020). This model can also be derived from the binomial by the law of small numbers if t is large and θ is small.

Variance stabilized Gaussian model (VSG)

Due to the central limit theorem, for sufficiently large t also normal models appear a reasonable approximation. Following the previous reasoning, this then leads either to $\mathcal{N}(t\theta, t\theta(1 - \theta))$ or $\mathcal{N}(t\theta, t\theta)$ if we start with binomial or Poisson distribution, respectively. Since in our asymptotic analysis we let $t, n \rightarrow \infty$ and $\theta = O(1/n)$, we choose the simpler model $\mathcal{N}(t\theta, t\theta)$. Applying the variance stabilizing transform $f(x) = 2\sqrt{x}$, we thus analyze

$$F_{2\sqrt{t}, \sqrt{\theta}} = \mathcal{N}(2\sqrt{t\theta}, 1). \quad (2.6)$$

Homogeneous Gaussian model (HG)

The simplest model to assume in this situation is the homogeneous Gaussian model $\mathcal{N}(\mu, \sigma^2)$ for some general mean $\mu = t\theta$ and some constant variance σ^2 . In particular, many algorithms for recovery assume this model, see e.g. (Bertero et al., 2009;

Hohage and Werner, 2016) for further discussion. After re-normalizing the mean μ by σ , we can w.l.o.g. set $\sigma = 1$ and consider the model

$$F_{t,\theta} = \mathcal{N}(t\theta, 1). \quad (2.7)$$

For a comprehensive discussion and more details on the modeling see e.g. (Munk et al., 2020). We emphasize that the homogeneous Gaussian model is commonly used as a proxy for “microscopy with noise” and has been investigated in many studies. We will, however, show that it is misleading in the present context. In contrast, we will show that the other 3 models (asymptotically) lead to the same resolution which scales linearly with the FWHM in agreement with the experimental evidence.

Remark 2.1.1. We consider photons, but treat them as classical particles. In the case of Poisson model, our modeling as given in (2.1) and (2.2) corresponds to the so-called *semiclassical detection model*, see e.g. Chapter 9 of (Goodman, 1985). This model is an approximation and follows from the general theory of light and matter interactions – quantum electrodynamics (QED), see e.g. (Leonhardt, 2010) and in particular its Appendix B.

2.2 Statistical testing problem

In the following, we will describe the resolution of a microscope with a psf $h \geq 0$ as a detection problem. We consider general psfs and provide a mathematically rigorous (asymptotic) statistical testing theory for resolution. We test the hypothesis that there is one point source at x'_0 against the alternative that there are two point sources at x'_1 and x'_2 of brightness q and $(1 - q)$, respectively, with $q \in (0, 1)$. The symmetric detection problem $q = 1/2$ is well-known and was considered by many authors from different perspectives, see e.g. (den Dekker and van den Bos, 1997) for an overview of different resolution criteria. Taking into account the previous considerations on diffraction, in particular (1.2), and setting $x_i = Mx'_i$ for $i \in \{0, 1, 2\}$, we define our testing problem as

$$H_0 : g(x) = h(x - x_0) \quad (2.8a)$$

against the alternative

$$H_1 : g(x) = q h(x - x_1) + (1 - q) h(x - x_2), \quad (2.8b)$$

see Figure 2.1 for an illustration. The factors of q and $1 - q$ in the alternative ensure that the test function g has the same intensity under H_0 and H_1 . We always assume that

q and x_0 are fixed and known. For each particular alternative, we also assume that x_1 and x_2 are fixed and known as well. However, in the asymptotic analysis we will let $d = \|x_1 - x_2\| \rightarrow 0^1$.

Without loss of generality, we let

$$g : I = [0, 1]^m \rightarrow [0, +\infty) \quad (2.9)$$

and be normalized to have volume 1. Normalization will allow us to interpret integrals of the form (2.3) as probabilities in the binomial model (2.4), whereas (2.9) together with (2.8) allow us to interpret $h(\cdot - x_i)$ as functions with domain I for $i \in \{0, 1, 2\}$. We also assume that the psf h is even, so that $h(\cdot - x_i)$ is symmetric around x_i . This is a reasonable assumption since many experimental psfs are not only even, but also rotationally invariant, see e.g. the Airy pattern (Figures 1.4 and 1.2). Mathematically this allows us to define the center of h as $\mathbf{0}$. Experimentalists are often interested in resolution in any of the m coordinates. Thus, in the following w.l.o.g. we will always assume that our statistical test is used to determine the resolution in the first coordinate, i.e. we set

$$d = \|x_1 - x_2\| = |x_{11} - x_{21}| \quad \text{and} \quad x_{0i} = x_{1i} = x_{2i} \quad \text{for } i \in \{2, 3, \dots, m\}. \quad (2.10)$$

A (*randomized*) statistical test for this problem is a measurable map $\Phi_{t,n,d} : (\mathbb{R}^n)^m \rightarrow [0, 1]$, where $t, n \in \mathbb{N}$, $d \in \mathbb{R}_{\geq 0}$ and $\Phi_{t,n,d} = p$ means that we reject the null hypothesis with probability p . Each statistical test can make a *type I error* when the hypothesis is falsely rejected with probability

$$\mathbb{E}_{H_0} \Phi_{t,n,d}(Y),$$

and a *type II error* when the hypothesis is falsely accepted with probability

$$1 - \mathbb{E}_{H_1} \Phi_{t,n,d}(Y).$$

For each fixed number of random variables n^m , we have a simple hypothesis (one psf) against a simple alternative (two psfs) testing problem, when the locations x_0, x_1, x_2 and the asymmetry parameter q are fixed, see (2.8). Thus, according to the Neyman-Pearson lemma (Lehmann and Romano, 2005) in a given dimension m , for a fixed one dimensional discretization n and a fixed significance level α , the *likelihood ratio test* (LRT) for H_0 vs H_1 is uniformly most powerful, i.e. no other statistical test can perform

¹In our analysis we will couple all parameters to the illumination time t . However, for ease of readability we omit the subscripts t , i.e. we write $n = n_t$ and $d = d_t$ throughout.

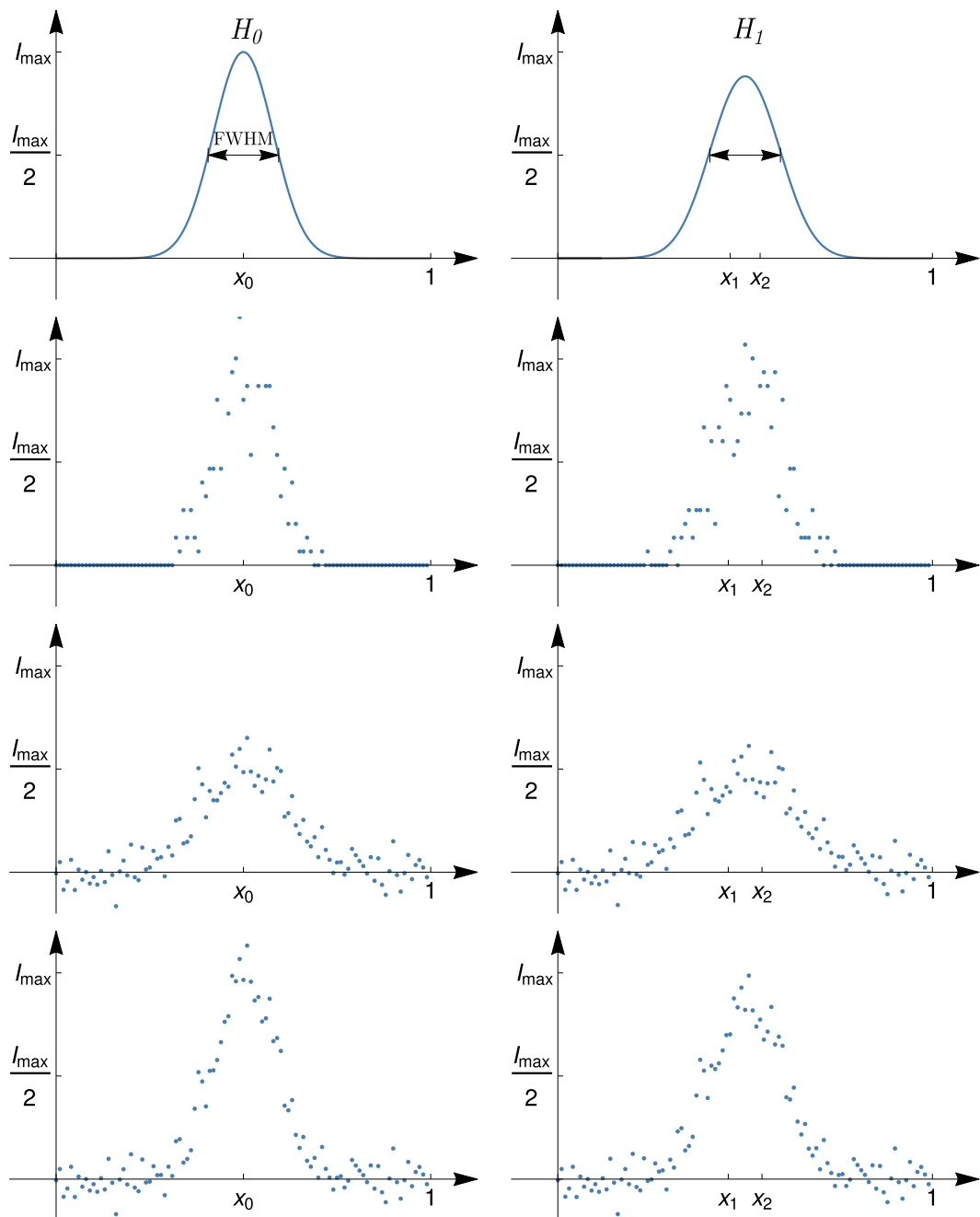


Figure 2.1: Resolution as a statistical testing problem in one dimension. First row: On the left hand side the hypothesis with the psf centered at x_0 , on the right hand side the alternative with two psfs centered at x_1 and x_2 , distance $d < \text{FWHM}$ apart. Second, third and fourth rows: The corresponding observational data generated according to the Poisson, VSG and HG models, respectively.

better. In our case (2.2), the LRT $\Phi_{t,n,d}: (\mathbb{R}^n)^m \rightarrow [0, 1]$ takes the form

$$\Phi_{t,n,d}(Y) = \begin{cases} 1 & \text{if } T_n(Y) > q_{\alpha,t,n,d}^*, \\ \gamma & \text{if } T_n(Y) = q_{\alpha,t,n,d}^*, \\ 0 & \text{if } T_n(Y) < q_{\alpha,t,n,d}^*, \end{cases} \quad (2.11)$$

with the log likelihood ratio statistic $T_{t,n,d}(Y)$ given in terms of the probability mass functions or densities $f_{t,\theta}$ of $F_{t,\theta}$ by

$$T_{t,n,d}(Y) = \log \left(\frac{\prod_{i \in [n]^m} f_{t,p_{1i}}}{\prod_{i \in [n]^m} f_{t,p_{0i}}} \right) = \sum_{i \in [n]^m} \log \left(\frac{f_{t,p_{1i}}}{f_{t,p_{0i}}} \right). \quad (2.12)$$

Here and in what follows we use the notation

$$\prod_{i \in [n]^m} := \prod_{i_1=1}^n \cdots \prod_{i_m=1}^n, \quad \sum_{i \in [n]^m} := \sum_{i_1=1}^n \cdots \sum_{i_m=1}^n,$$

and abbreviate the detection probabilities in the i th bin by

$$p_{0i} := \int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} h(x - x_0) dx \quad (2.13)$$

under the hypothesis H_0 and

$$p_{1i} := q \int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} h(x - x_1) dx + (1-q) \int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} h(x - x_2) dx \quad (2.14)$$

under the alternative H_1 .

Given a significance level $\alpha \in (0, 1)$, the threshold $q_{\alpha,t,n,d}^*$ and the constant γ in (2.11) have to be chosen such that $\mathbb{E}_{H_0} \Phi_{t,n,d}(Y) = \mathbb{P}_{H_0}(T_{t,n,d}(Y) > q_{\alpha,t,n,d}^*) + \gamma \mathbb{P}_{H_0}(T_{t,n,d}(Y) = q_{\alpha,t,n,d}^*) = \alpha$, as this ensures α to be the level (i.e. the probability of the type I error) of the test.

In the following we adopt a minimax testing point of view. To this end, we begin by determining which choice of x_1 and x_2 in (2.8) is the most difficult to detect.

Proposition 2.2.1. *Consider the testing problem (2.8) in the setup (2.10) for $x_0 = (1/2, \dots, 1/2)$. Assume that the psf h is even. Let $0 < \alpha < 1/2$ and consider the asymptotic regime with $n \rightarrow \infty$ and $d \rightarrow 0$. Then for each of the four models defined in Equation (2.4) the uniformly most powerful test Ψ^* (and hence the LRT) for (2.8) with*

level $\mathbb{E}_{H_0} \Psi^*(Y) \rightarrow \alpha$ has asymptotically the smallest power $\mathbb{E}_{H_1} \Psi^*(Y)$ when

$$x_{01} = qx_{11} + (1 - q)x_{21},$$

i.e. when x_0 is the center of mass of the two psfs in the alternative.

The proof is provided in the appendix A.2, since it is not necessary for the first reading.

We are now ready to define the resolution of a microscope statistically.

Definition 2.2.2 (Statistical resolution). Let $(Y_i)_{i \in \{1, \dots, n\}^m}$ be as in (2.2) and let h be the point spread function of the microscope under investigation. Choose one of the four models (2.4). Let $0 < \alpha, \beta < 1/2$, $x_0 \in (0, 1)^m$, $t \in \mathbb{N}$ and $n \in \mathbb{N}$ be fixed. We define the microscope's *statistical resolution (in the first coordinate) at point x_0 , asymmetry parameter q , discretization n , exposure time t , type I error α and type II error β under the prescribed model* as the unique value $d \in (0, 1)$ such that the uniformly most powerful test (and hence the LRT (2.11)) Ψ^* for (2.8) with x_1 and x_2 chosen such that $d = \|x_1 - x_2\| = |x_{11} - x_{21}|$, $x_{01} = qx_{11} + (1 - q)x_{21}$ and $x_{0i} = x_{1i} = x_{2i}$ for $i \in \{2, 3, \dots, m\}$ has exactly level α and power $1 - \beta$, i.e. the most powerful test satisfies

$$\mathbb{E}_{H_0} \Psi^*(Y) = \alpha \quad \text{and} \quad \mathbb{E}_{H_1} \Psi^*(Y) = 1 - \beta.$$

In other words, if the distance d between the two sources x_1 and x_2 in (2.8) satisfies $|x_{11} - x_{21}| = d$, the statistical resolution is determined by the best possible test with detection power $1 - \beta$ while the error of incorrectly assigning two sources (when only one is valid) is controlled by α . It is immediately clear that a larger value of d will result in larger power, and a smaller value of d will result in smaller power, i.e. for x_1 and x_2 with $|x_{11} - x_{21}| \leq d$ no level α test is able to distinguish H_0 and H_1 with power $\geq 1 - \beta$. Thus, the sum of errors is bounded by $\alpha + \beta$, which is why we restrict ourselves to the case $\alpha, \beta \in (0, \frac{1}{2})$. Consequently, if $\alpha = \beta = \frac{1}{2}$, then $\Psi \sim \text{Bin}(1, \frac{1}{2})$, and hence we expect the resolution to be 0 which corresponds to the information of a coin flip to decide between H_0 and H_1 .

One might wonder what about the case when $x_1 \rightarrow x_2$, but $x_1, x_2 \rightarrow x'_0 \neq x_0$? In this case asymptotically as $n \rightarrow \infty$ both type I and type II errors *always* tend to 0. In other words, asymptotically the problem is trivial. To see this, take any non-trivial test that counts the photons in the interval $[0, x_0]$ and compares the corresponding counts under the H_0 and H_1 .

The aim of this thesis is to study the *asymptotic* behavior (as $n, t \rightarrow \infty$ and $d \rightarrow 0$) of the statistical resolution d in the four models from Equation (2.4). We will see that the

(asymptotic) behavior of d serves as a good proxy in finite sample situations whenever n and t are sufficiently large and d is sufficiently small. This is investigated in simulations presented in Chapter 4.

2.3 Main theorem

To derive the asymptotic behavior of the statistical resolution d of a given microscope, we have to pose smoothness assumptions on its psf h depending on the employed model. In the HG model we require the following.

Assumption 2.3.1 (HG model). Suppose that the psf h is even and non-constant. Furthermore let $h \geq 0$ and $h(\cdot - x_i) \in C^2 [0, 1]^m$ for all $i \in \{0, 1, 2\}$.

The requirement that $h \geq 0$ is natural in view of h being an intensity. The differentiability condition is rather mild and clearly satisfied for the Airy pattern in (1.3) and its most common approximation by a Gaussian.

In case of the binomial, Poisson and VSG models we need a stronger condition.

Assumption 2.3.2 (B, P and VSG models). Suppose that the psf h is even and non-constant. Furthermore let $h > 0$ and $h(\cdot - x_i) \in C^4 [0, 1]^m$ for all $i \in \{0, 1, 2\}$.

Note that due to compactness of $[0, 1]^m$, Assumption 2.3.2 implies that $h \geq c > 0$.

Remark 2.3.3. We emphasize that the Airy pattern in (1.3) does not satisfy $h > 0$. However, in accordance with many models considered in the literature it is pertinent to include so-called background contributions, i.e. photons arising from other sources than the psf. Examples of such modeling include (Acuña and Horowitz, 1997) and (von Diezmann et al., 2017), which in the notation of (2.2) would correspond to $Y_i \sim F_t \int_{B_i} g(x) dx + \gamma/n$ with a positive constant γ and g given by (2.8). If we were to incorporate this background noise into the psf h and hence due to (1.2) into the image g , we would obtain (2.2) with $\tilde{g} = g + \gamma > 0$. From this point of view, the assumption $h > 0$ corresponds to the natural requirement that photons can be detected everywhere. We also note that a Gaussian psf on $[0, 1]^m$ (2.18), which is the most commonly used approximation to the Airy pattern (see e.g. (von Diezmann et al., 2017) or Figure 1.1B, clearly satisfies Assumption 2.3.2.

For two sequences $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ we write $a_n \asymp b_n$, $a_n \ll b_n$, $a_n \gg b_n$ and $a_n \sim b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 1$, $\lim_{n \rightarrow \infty} a_n/b_n = 0$, $\lim_{n \rightarrow \infty} b_n/a_n = 0$ and $\lim_{n \rightarrow \infty} a_n/b_n = c$ for some constant $c > 0$, respectively. Note that, due to asymptotic considerations, we may restrict to non-randomized tests in what follows, i.e. to set

$\gamma = 0$ in (2.11). Recall that we consider asymptotics as $d \rightarrow 0$ and $n, t \rightarrow \infty$. We are now ready to state our main result on the asymptotic behavior of d .

Theorem 2.3.4. *Assume model (2.2) with fixed dimension m and consider the testing problem (2.8) with $x_0 \in (0, 1)^m$ and $x_1, x_2 \in (0, 1)^m$ such that $x_{01} = qx_{11} + (1 - q)x_{21}$ and $x_{0i} = x_{1i} = x_{2i}$ for $i \in \{2, 3, \dots, m\}$. Let $0 < \alpha, \beta < 1/2$ be type I and II errors, respectively. For $0 < \gamma < 1$ denote by q_γ the γ quantile of the standard normal distribution $\mathcal{N}(0, 1)$.*

Binomial model

Let the distribution in (2.2) be given by $F_{t,\theta} = \text{Bin}(t, \theta)$, the psf h satisfy the Assumption 2.3.2 and $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$. Then the statistical resolution d of the corresponding microscope is

$$d \asymp \frac{\sqrt{2}}{\sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \left(\int_I \frac{h''(x - x_0)^2}{h(x - x_0)} dx \right)^{-1/4} t^{-1/4}. \quad (2.15)$$

Poisson model

Let the distribution in (2.2) be given by $F_{t,\theta} = \text{Poi}(t\theta)$ and the psf h satisfy the Assumption 2.3.2 and $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$. Then the statistical resolution d of the corresponding microscope satisfies (2.15). If $m = 1$, then the Assumption 2.3.2 suffices for (2.15) to hold, and the coupling between t and d can be arbitrary.

Variance stabilized Gaussian model

Let the distribution in (2.2) be given by $F_{t,\theta} = \mathcal{N}(2\sqrt{t\theta}, 1)$ and the psf h satisfy the Assumption 2.3.2. Then the statistical resolution d of the corresponding microscope also satisfies (2.15).

Homogeneous Gaussian model

Let the distribution in (2.2) be given by $F_{t,\theta} = \mathcal{N}(t\theta, 1)$, $n^m = o(t^2)$ and the psf h satisfy the Assumption 2.3.1. Then the statistical resolution d of the corresponding microscope is

$$d \asymp \frac{\sqrt{2}}{\sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \left(\int_I h''(y - x_0)^2 dy \right)^{-1/4} t^{-1/2} n^{m/4}. \quad (2.16)$$

Here and in what follows we set $'$ to denote the *partial* derivative in the first coordinate, i.e.

$$f'(y - x_0) := \frac{\partial f(y - x_0)}{\partial y_1} \quad (2.17)$$

for $f(\cdot - x_0) : C^1[0, 1]^m \rightarrow \mathbb{R}$. We also follow this convention for higher order derivatives.

Let us briefly comment on the techniques employed in the proof of Theorem 2.3.4. In case of the VSG and the HG models, type I and type II errors of the LRT can be computed explicitly provided that $d \searrow 0$. The formulas (2.15) and (2.16) are then derived by straightforward approximations of integrals by sums. In case of the Poisson and binomial models, the analysis is more difficult. In these models the LRT statistic consists of n weighted Poisson / binomially distributed random variables of varying intensity tending to any value in $[0, +\infty]$. The particular values depend on the asymptotic relation between n and t . If $t = c'n^{2m-\delta'}$ for some $c' > 0$ and $2m > \delta' > 0$, then the sum grows not too fast and hence the LRTs satisfy CLTs (Theorems 5.3.1 and 5.3.3).

In one dimension, provided that $t \gg \sqrt{n} \log^8 n$, we can apply recent results from (Ray and Schmidt-Hieber, 2018) on asymptotic equivalence in the Le Cam sense, to prove that the Poisson model is asymptotically equivalent to the VSG model. Combining asymptotic equivalence with the CLT, we see that (2.15) holds true asymptotically for any coupling between t and n .

2.4 Implications for experiments

To understand the experimental implications of Theorem 2.3.4, recall that for many microscopes the psf can be well approximated by a Gaussian kernel

$$h(x - x_0) = \prod_{i=1}^m \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{1}{2\sigma_i^2} \left(x_i - \frac{1}{2}\right)^2\right) \quad (2.18)$$

centered at $1/2$ with variance $\sigma_i^2 > 0$, see Figure 1.1B for an illustration. In this case,

$$\text{FWHM}_i = 2\sqrt{2 \log 2} \sigma_i \approx 2.355 \sigma_i. \quad (2.19)$$

To ensure that our psfs are contained in the unit interval under the alternative (2.8b), we have chosen values of $\text{FWHM}_i \leq 0.25$, equivalently, $\sigma_i \leq 0.107$ in the analysis below and simulations of Chapter 4. This allows us to skip normalizing (2.18), since $\int_{[0,1]^3} h(x - x_0) dx \approx 0.999994$ and in dimensions one and two the integral is even closer to 1. We have that

$$\int_I h''(x - x_0)^2 dx = \left(\frac{6\sqrt{\pi} \operatorname{erf}\left(\frac{1}{2\sigma_1}\right) \sigma_1^5 + e^{-\frac{1}{4\sigma_1^2}} (2\sigma_1^2 - 1) \sigma_1^2}{8\sqrt{\pi} 2^m \pi^{m/2} \sigma_1^{10}} \right) \prod_{i=2}^m \left(\frac{\operatorname{erf}\left(\frac{1}{2\sigma_i}\right)}{\sigma_i} \right)$$

$$= \frac{3}{2^{m+2} \pi^{m/2}} \sigma_1^{-5} \prod_{i=2}^m \sigma_i^{-1} + o\left(\sigma_1^{-5} \prod_{i=2}^m \sigma_i^{-1}\right)$$

and

$$\begin{aligned} \int_I \frac{h''(x-x_0)^2}{h(x-x_0)} dx &= \left(\frac{2 \operatorname{erf}\left(\frac{1}{2\sqrt{2}\sigma_1}\right)}{\sigma_1^4} - \frac{e^{-\frac{1}{8\sigma_1^2}}(4\sigma_1^2+1)}{4\sqrt{2\pi}\sigma_1^7} \right) \prod_{i=2}^m \operatorname{erf}\left(\frac{1}{2\sqrt{2}\sigma_i}\right) \\ &= 2\sigma_1^{-4} + o(\sigma_1^{-4}), \end{aligned}$$

where

$$\operatorname{erf}(x) = \frac{1}{\sqrt{\pi}} \int_{-x}^x e^{-t^2} dt = 2\Phi(\sqrt{2}x) - 1,$$

and we have used Bürmann's series

$$\operatorname{erf}(x) = \sqrt{1-e^{-x^2}} + \sqrt{1-e^{-x^2}} \frac{2}{\sqrt{\pi}} \sum_{k=1}^{\infty} c_k e^{-kx^2}$$

valid for all $x > 0$ with $c_k \in \mathbb{R}$ some constants. Thus, according to (2.16) we obtain in the **homogeneous Gaussian model** that asymptotically the statistical resolution satisfies

$$\begin{aligned} d &\asymp \frac{2^{m/4+1} \pi^{m/8}}{3^{1/4} \sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \frac{n^{m/4}}{\sqrt{t}} \sigma_1^{5/4} \prod_{i=2}^m \sigma_i^{1/4} \\ &= \frac{2^{m/4} \pi^{m/8}}{3^{1/4} 2^{7/8} (\log(2))^{5/8} \sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \frac{n^{m/4}}{\sqrt{t}} \operatorname{FWHM}_1^{5/4} \prod_{i=2}^m \frac{\operatorname{FWHM}_i^{1/4}}{2^{3/8} (\log(2))^{1/8}}. \end{aligned} \quad (2.20)$$

Note that this is *not* in agreement with the previously discussed FWHM resolution criterion, which postulates a linear dependency of d on the FWHM_1 , see also (Egner et al., 2020) or (den Dekker and van den Bos, 1997). From this point of view it becomes evident that the homogeneous Gaussian model is too simple to capture the fine details of actual experiments. However, in the **variance stabilized Gaussian, Poisson** and **binomial** models we have

$$\begin{aligned} d &\asymp \frac{2^{1/4}}{\sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} t^{-1/4} \sigma_1 \\ &= \frac{1}{2^{5/4} \sqrt{\log 2} \sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} t^{-1/4} \operatorname{FWHM}_1, \end{aligned} \quad (2.21)$$

i.e. d depends linearly on the FWHM_1 in agreement with experiments and FWHM based resolution criteria (1.1), (1.4), (1.5).

We summarize these results in Table 2.1. We have set $q = 1/2$ because it is standard in two point resolution criteria, see e.g. (van den Bos and den Dekker, 2001). From (2.16) and (2.15) we see that this is the *easiest* case to distinguish, i.e. for given parameter values the resolution is the *smallest* when $q = 1/2$. Thus, the choice $q = 1/2$ corresponds to the common interpretation of resolution as the smallest distance that allows different objects to be distinguished. We have also set $\beta = \alpha$, since in most microscopy experiments type I and type II errors are of equal importance.

Focusing on 2D and on the FWHM_1 values in the range $[0.15, 0.25]$ with fixed $\text{FWHM}_2 = 0.2$, the ratio between the homogeneous Gaussian model resolution of (2.20) and the other models' (2.21) lies in the interval $[0.195 n^{1/4} t^{-1/4}, 0.222 n^{1/4} t^{-1/4}]$. Therefore, if $t = n$, then the other models' resolution is approximately five times greater than the homogeneous Gaussian. The range is wider if the discretization n is greater than the illumination time t and vice versa.

Error $\alpha = \beta$	0.01	0.05	0.1
Model			
Homogeneous Gaussian 1D	3.08 $\text{FWHM}^{5/4} t^{-1/2} n^{1/4}$	2.59 $\text{FWHM}^{5/4} t^{-1/2} n^{1/4}$	2.29 $\text{FWHM}^{5/4} t^{-1/2} n^{1/4}$
Homogeneous Gaussian 2D	0.68 $\text{FWHM}^{5/4} t^{-1/2} n^{1/2}$	0.57 $\text{FWHM}^{5/4} t^{-1/2} n^{1/2}$	0.51 $\text{FWHM}^{5/4} t^{-1/2} n^{1/2}$
Homogeneous Gaussian 3D	0.15 $\text{FWHM}^{5/4} t^{-1/2} n^{3/4}$	0.13 $\text{FWHM}^{5/4} t^{-1/2} n^{3/4}$	0.11 $\text{FWHM}^{5/4} t^{-1/2} n^{3/4}$
VSG / Poisson / Binomial	$2.18 t^{-1/4}$ FWHM	$1.83 t^{-1/4}$ FWHM	$1.62 t^{-1/4}$ FWHM

Table 2.1: Limiting asymptotic statistical resolution as given by Theorem 2.3.4 for the Gaussian psf (2.18). Here we have set $q = 1/2$, $\beta = \alpha$, $\text{FWHM}_1 = \text{FWHM}$ and $\text{FWHM}_i = 0.2$ for $i \in \{2, \dots, m\}$. For general expressions see Equations (2.20) and (2.21).

Let us also comment on the dependency of the constants in (2.21) and (2.20) on the type I error α and the type II error β . From the Table 2.1, we have that $d(\alpha = \beta = 0.1)/d(\alpha = \beta = 0.01) = 0.51/0.68 = 0.75$ for the 2D HG and ≈ 0.74 for the other models, i.e. depending on the acceptable errors the resolution might decrease approximately by a quarter. More generally, if we increase α , then $-q_\alpha$ decreases ($\alpha < 1/2$) and hence d becomes smaller. This is due to the fact that a larger value of α implies a higher probability to falsely reject the hypothesis that there is only one object. Similarly for β . In the limiting case $\alpha = \beta = 1/2$, we reject with probability $1/2$ the hypothesis even though it is correct and likewise with probability $1/2$ we accept the hypothesis under the alternative. Thus, in such case our test is, as expected, equivalent to tossing a fair coin to determine whether we have one or two psfs and hence the resolution is perfect (with probability $1/2$), i.e. $d = 0$. Similarly, if $\alpha = 0$ or $\beta = 0$, then the resolution is infinite, i.e. we cannot distinguish two point sources flawlessly – the type I or the type II error must be greater than zero.

To summarize, our results imply that when the dependency on the resolution is important, the homogeneous Gaussian model seems to be too simple and hence other models should be preferred. Notably, the Poisson and the VSG models already correctly capture the dependency and are simpler than the binomial model.

CHAPTER 3

Comparisons with previous work

3.1 Overview

Investigation of resolution in a statistical setting is not new. The HG model (and variations) was considered in (Harris, 1964; Milanfar and Shakouri, 2002; Shahram and Milanfar, 2004; Shahram, 2005; Shahram and Milanfar, 2006) and the Poisson model (and variations) in (Helstrom, 1964, 1973; Acuña and Horowitz, 1997). However, with the exception of (Acuña and Horowitz, 1997), most of these works lack mathematical rigor, whereas (Acuña and Horowitz, 1997) instead of defining resolution statistically suggest a redefinition in terms of the power function (3.4) and do not work out the dependency on the FWHM, see below for more details.

Already in the 1960s, resolution has been investigated from a decision theoretic point of view in signal processing theory. Early references include Harris (Harris, 1964) for the homogeneous Gaussian model and Helstrom (Helstrom, 1964, 1965) for the Poisson model. In (Helstrom, 1964, 1965) Helstrom considered signals consisting of different wavelengths varying in space, noting that using Reiffen and Sherman's paper (Reiffen and Sherman, 1963) on optimum demodulation for time-varying Poisson processes one could consider a signal varying in both space and time. For ease of understanding, we assumed that our psf intensity does not vary with time and is monochromatic, see (1.3). Harris (Harris, 1964) only calculated the probability of a correct decision (power) without any consideration of the level. Helstrom (Helstrom, 1964) assumed a CLT and basically obtained type I error and power expressions in the CLT regime 5.3.1 for our Poisson model in his Equation (15). To see this, we have to set $g_0 = q_{\alpha,t,n,d}^* := q_{1-\alpha}\sqrt{\mathbb{V}_{H_0}T_n} + \mathbb{E}_{H_0}T_n$ (5.36) as the threshold in Helstrom's theory (which is not specified there), $M_0(x) = p_{0i}$, $M_1(x) = p_{1i}$, where $M.(x)$ is the effective photon count rate density at $x \in [-1/2, 1/2]^2$, and change integrals in his work to sums.

In (Helstrom, 1973) Helstrom went even further than in (Helstrom, 1964) and considered (2.8) in the context of quantum information theory, following the statistical

paradigm originally set out by Middleton (Middleton, 1953). Among other things, Helstrom found out that P_e , the average of type I and type II errors, converges to $1/2 \exp(-t)$ with increasing distance d . Here t is interpreted as the average number of photons. As expected, the bound tends to zero in the classical regime as $t \rightarrow \infty$. Reassuringly, the form of his combined error probability P_e becomes the same as ours with increasing t . However, Helstrom's results cannot be transferred to our case due to the quantum information theoretic setting, and his proofs are not mathematically rigorous. Notably, he found that P_e is very close to its asymptotic minimum $1/2 \exp(-t)$ whenever d approximately equals twice the Rayleigh criterion, which led him to define the resolution as twice the Rayleigh limit. Much of the current research on resolution in quantum information theory revolves around trying to design different measurement techniques (Tsang et al., 2016a,b; Nair and Tsang, 2016; Lu et al., 2018) which would allow to experimentally come as close as possible to the theoretical limits calculated by Helstrom (Helstrom, 1973). Some of these measurement techniques have been already confirmed by proof of principle experiments, see e.g. (Tham et al., 2017), others even applied to biological imaging (Tenne et al., 2019). We emphasize that our theory is designed to describe everyday microscopy experiments with rather many photons so that Helstrom's limit $1/2 \exp(-t)$ can be safely disregarded. Even though the mathematical treatment of quantum optics experiments is beyond the scope of this paper, we think that it is a fruitful research direction also for statisticians (see e.g. (Yamagata et al., 2013), where the authors have defined a *quantum* likelihood ratio).

We also mention contributions from the field of modern signal processing and engineering, namely the works by Milanfar and collaborators (Milanfar and Shakouri, 2002; Shahram and Milanfar, 2004, 2006), see also (Shahram, 2005) for an overview. These authors also investigate resolution in terms of statistical measurement errors, and they derive a dependency of the resolution on the inverse fourth root of the so-called *measurement* signal-to-noise-ratio. Note that this has some similarity with the dependency on t in (2.15). However, even though resolution is treated as a statistical testing problem, in all these papers a homogeneous Gaussian model (which is challenged by our analysis) is assumed and they lack some mathematical rigor as well. The same can be said of Terebizh (Terebizh, 1995) who suggested a statistical definition of resolution for extended objects.

Closest to our paper is the work (Acuña and Horowitz, 1997) by Acuña and Horowitz on telescope resolution. There, the testing problem $H_0: d = 0$ vs. $H_1: d > 0$ in a 2D model on a line is considered. This corresponds to our Poisson model, but with explicit constant background noise. Their main quantity of interest is p_{1i} (2.14) considered as a function of d . Under assumptions on p_{1i} 's roughly corresponding to our assumptions on

the psf h , they analyzed the likelihood ratio test in the regime where $t \rightarrow \infty$, but kept the number of measurements (discretization) n fixed and finite. Clearly, a finite value of n will at some point restrict the resolution to be of the order $1/n$, as no information finer than the bin-size can be obtained. Moreover, the mathematical treatment of this regime is substantially simpler, as the LRT statistic is given by a finite sum of independent weighted Poisson random variables, whose intensity tends to ∞ , and hence one obtains a CLT trivially. Acuña and Horowitz (Acuña and Horowitz, 1997) also note that there is a different regime with finite fixed t and $n \rightarrow \infty$, but do not treat this. All of our results except for asymptotic equivalence also hold in this regime: See Remarks 5.1.1 and 5.2.1, and note that the relation between t and n necessary for Theorem 5.3.1 is trivially satisfied for constant t . The authors define resolution as the (asymptotic) power function of the likelihood ratio test rather than as a single number, which in some sense, is close to our Definition 2.2.2. However, we believe that it is not intuitive for practitioners to define the resolution as a probability, since they are used to thinking of resolution as a distance. The main result of (Acuña and Horowitz, 1997) is the calculation of this power function in the regime $t \rightarrow \infty, n = \text{const}$, which we can reproduce asymptotically for large n and t from our more general results (up to dimension and the explicit constant background noise) if we keep a sum instead of the integral in (2.15), see Remark 5.2.1. Note furthermore that the power expression of (Acuña and Horowitz, 1997) is only valid if $d = \text{const} \times t^{-1/4}$ in accordance with our result (2.15). We stress that our results give an explicit dependency on the FWHM.

Finally we mention, that the term ‘super-resolution’ is used in mathematical and statistical communities also in a different context, see (Donoho, 1992; Morgenshtern and Candès, 2016; Candès and Fernandez-Granda, 2013, 2014; Fernandez-Granda, 2015). There super-resolution addresses the ways to localize signals with (un)known amplitudes by observing their (noisy) Fourier samples, i.e. samples in the frequency domain. The domain is always assumed to have some cut-off frequency f_c corresponding to the inverse Abbe limit in our context. In contrast, in this paper we assume that the locations of our signals are always *known*, i.e. we will follow the experimentalists’ terminology.

In the next sections of this chapter we will concentrate on some of the relevant parts of the works above and compare their results to ours in more detail.

3.2 Abbe and Rayleigh

Once the value of t has been fixed, the asymptotic statistical resolution (2.21) allows us to compare our results to the classical resolution limits by Abbe (1.1) and Rayleigh (1.4). Since in most microscopy experiments type I and type II errors are equally important,

we set $\alpha = \beta$ in (2.21). Recall that the FWHM of the Airy pattern is $0.51\lambda/\text{NA}$, and hence both criteria can be read as $c \cdot \text{FWHM}$ with a constant $c > 0$. Consequently, we can compute the corresponding value of α such that the right-hand side of (2.21) equals $c \cdot \text{FWHM}$. The results are shown in Table 3.1. A possible interpretation is the following: if $t = 10$, then a microscope achieves the resolution equal to the Abbe criterion with type I and II errors being equal to 6.8%, whereas if we say that the resolution is equal to the one given by the Rayleigh criterion (larger than Abbe), then $\alpha = \beta = 1.3\%$. The values in Table 3.1 are rather small, but we have neglected any background noise in the choice of our psf (2.18). Thus, we believe that in actual microscopy experiments more photons are necessary to achieve such low type I and II errors like in Table 3.1.

Error $\alpha = \beta$	$\mathbb{E}[N] = t$	10	20	30	40	50
Abbe criterion		6.81%	1.76%	0.494%	0.144%	0.0432%
Rayleigh criterion		1.33%	0.0857%	0.00614%	$4.61 \cdot 10^{-4}\%$	$3.56 \cdot 10^{-5}\%$

Table 3.1: The type I and II errors ($\alpha = \beta$) such that Abbe or Rayleigh criterion is fulfilled for the VSG, Poisson and binomial models for different values of the expected number of photons t in any dimension in the symmetric $q = 1/2$ (2.8b) case. Here we have assumed a Gaussian psf (2.18), and thus the expression (2.21) can be simply inverted to calculate α .

3.3 Milanfar and coauthors' work

As mentioned in the overview, the problem of resolution was also investigated in signal processing and decision theory, most notably by Milanfar and collaborators (Milanfar and Shakouri, 2002), (Shahram and Milanfar, 2004), (Shahram, 2005), (Shahram and Milanfar, 2006). The authors considered the same statistical testing problem as we (2.8) in one and two dimensions, but only under the HG model. According to the authors, this model is well justified by the properties of signal detectors. The researchers defined the resolution as the minimum detectable point separation and considered its dependence on the so-called *measured* signal-to-noise-ratio (SNR_m).

In our case, for even, symmetrically placed psfs under the alternative ($q = 1/2$ in (2.8b)), we have in one dimension that

$$\text{SNR}_m = \frac{1}{\sigma^2} \left\| h + d^2 \frac{h''}{8} \right\|^2 = \frac{1}{\sigma^2} \left(\int_0^1 h^2 + \frac{d^4}{64} \int_0^1 (h'')^2 + \frac{d^2}{4} \int_0^1 hh'' \right). \quad (3.1)$$

Here σ^2 is the variance of the homogeneous Gaussian model $\mathcal{N}(\mu, \sigma^2)$ (2.7) *without*

renormalization (2.7), i.e. without rescaling and setting $\sigma = 1$. In this case the resolution from Equation (2.16) becomes

$$d \asymp 2\sqrt{2} \sqrt{q_{1-\beta} - q_\alpha} \left(\int_0^1 h''(x - x_0)^2 dx \right)^{-1/4} t^{-1/2} n^{1/4} \sqrt{\sigma}. \quad (3.2)$$

As usual, assuming that (3.2) holds also non-asymptotically, we can rewrite the SNR_m as

$$\begin{aligned} \text{SNR}_m &= \frac{64 (q_{1-\beta} - q_\alpha)^2 n}{d^4 \int_0^1 (h'')^2 t^2} \left(\int_0^1 h^2 + \frac{d^2}{4} \int_0^1 h h'' + \frac{d^4}{64} \int_0^1 (h'')^2 \right) \\ &= \frac{(q_{1-\beta} - q_\alpha)^2 n}{t^2} \left(64 \frac{\int_0^1 h^2}{\int_0^1 (h'')^2} d^{-4} + 16 \frac{\int_0^1 h h''}{\int_0^1 (h'')^2} d^{-2} + 1 \right). \end{aligned} \quad (3.3)$$

Our SNR_m is the same like in Equation (2.31) of (Shahram, 2005) if we require that $h = 0$ or $h' = 0$ at the boundaries (also assumed in (Shahram, 2005)), use integration by parts, set $n = t$ and interpret t as Nyquist rate.

Note that considering d as a function of SNR_m is somewhat inconvenient mathematically, since $d(\text{SNR}_m)$ is multivalued for small values of SNR_m , see Figure 3.1 (b).

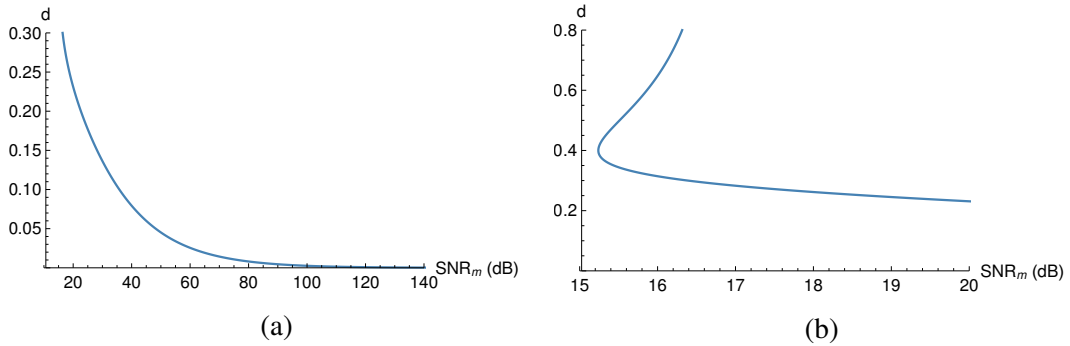


Figure 3.1: Plots of Equation (3.3). Resolution d vs SNR_m in dB for a Gaussian psf (2.18) and $n = t$ at the Nyquist rate, i.e. d vs $10 \log_{10}(\text{SNR}_m)$ at $t = 1$. Here $\beta = 0.01$ and $\alpha = 10^{-6}$.

(a) For large SNR_m the d^{-4} term dominates (cf. (Shahram, 2005) Fig. 2.1, where the psf is $\sin(\pi x)/(\pi x)$).

(b) For smaller values of SNR_m the resolution d is a multivalued function of SNR_m .

3.4 Acuña and Horowitz's work

As previously mentioned, the work closest to ours is (Acuña and Horowitz, 1997). There the authors mathematically rigorously investigated a model of resolution for *telescopes* corresponding to our 2D Poisson model in the regime $t \rightarrow \infty$, $d \rightarrow 0$, but with constant background noise. The main quantity of their analysis is the power function

$$p(t, d) = \Phi \left(q_\alpha + \sigma_0 \frac{d^2 \sqrt{t}}{8} \right) \quad (3.4)$$

with t now interpreted as the telescope exposure time and σ_0 a constant depending on the optical system and discretization (Acuña and Horowitz kept n fixed and finite). Neglecting the constant background noise and using the integration shorthand (5.1) the constant σ_0 can be written as

$$\sqrt{\kappa \sum_{i \in [n]^2} \left(\int_i h \right)^{-1} \left(\frac{\partial^2}{\partial \left(\frac{d}{2} \right)^2} \Big|_{d=0} \left(\frac{1}{2} \int_i h \left(\cdot - \frac{d}{2} \right) + h \left(\cdot + \frac{d}{2} \right) \right) \right)^2}$$

with constant $\kappa > 0$ describing the total intensity of the star in question. As usual assuming that our asymptotic formulas are also valid for finite samples, we can rewrite (2.15) to get the expression for power

$$p(t, d) = \Phi \left(q_\alpha + \sqrt{\int_I \frac{h''(x - x_0)^2}{h(x - x_0)} dx} \frac{d^2 \sqrt{t}}{8} \right). \quad (3.5)$$

Reassuringly, the expression for σ_0 coincides (up to κ) for large n with our factor $\sqrt{\sum_{i \in [n]^2} \frac{(\int_i h'')^2}{\int_i h}}$ in 2D by the mean value theorem, i.e. we are able to reproduce the main result of Acuña and Horowitz (1997) from ours.

(Acuña and Horowitz, 1997) plot some plausible power vs distance between the two stars plots. It is interesting to compare them and find a value of a Gaussian psf corresponding to their astronomy-experiment-inspired values (mostly Hubble's telescope), even though Acuña and Horowitz stated that the psf of the Hubble space telescope is clearly non-Gaussian, see Figure 3.2.

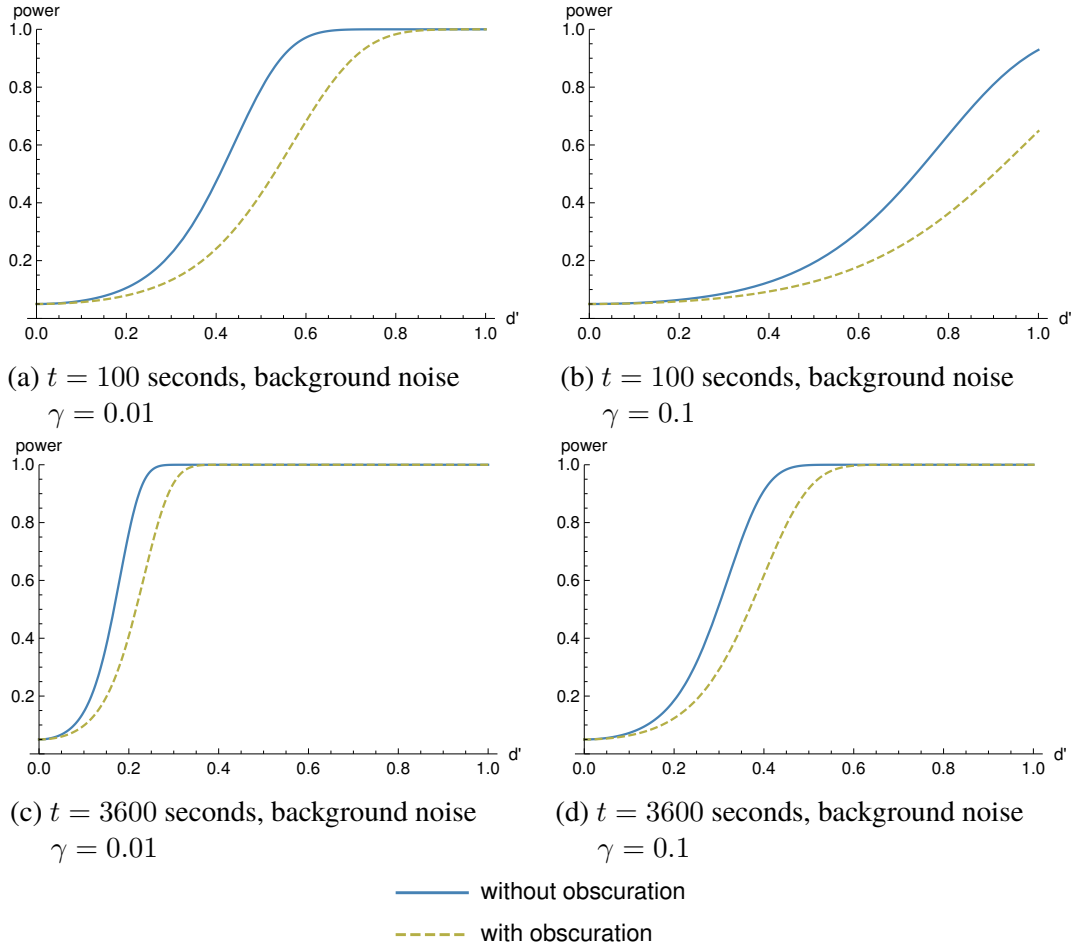


Figure 3.2: Reproduction of Figure 6 from (Acuña and Horowitz, 1997). The plots show the power dependence on $d' = d/R_{limit}$ (3.4) for an ideal Hubble space telescope. Here R_{limit} stands for the Rayleigh limit. The total intensity is $\kappa = 2.82 \times 10^{-6}$ and the level $\alpha = 0.05$. For the ideal Hubble space telescope with obscuration (dashed beige line) the optical parameters are $\sigma_0(\gamma = 0.01) = 0.0224 \mu m^{-2} s^{-1/2}$ and $\sigma_0(\gamma = 0.1) = 0.0071 \mu m^{-2} s^{-1/2}$, and $R_{limit} = 7.2468 \mu m$. Here γ is the constant background noise parameter. For the ideal Hubble space telescope without obscuration (thick blue line) $\sigma_0(\gamma = 0.01) = 0.0304 \mu m^{-2} s^{-1/2}$ and $\sigma_0(\gamma = 0.1) = 0.0096 \mu m^{-2} s^{-1/2}$, and $R_{limit} = 8.052 \mu m$. See also (Schroeder, 1987) Chapter 10 for more details on the experimental setup.

The corresponding plots according to our theory (3.5) with the Gaussian psf (2.18) have FWHMs in range from ≈ 0.03 ((c) without obscuration) to ≈ 0.18 ((b) with obscuration). Due to (1.4) and (1.5) we used the ratio $FWHM = 51R_{limit}/61$ when converting between different length scales.

3.5 Helstrom's work

The previous two sections have shown that our work agrees well with the previous works on the subject. However, the comparison in this section shows our theory's expected limitations, when the number of detected photons is very low.

In a series of papers Helstrom considered our testing problem (2.8) with $q = 1/2$ in the Poisson model as well as related ones at the level of rigor of theoretical physics. His earlier work is well summarized in his book (Helstrom, 1968). Later, Helstrom went even further and solved many of the same problems using quantum information theory, which resulted in another book (Helstrom, 1976).

Most of Helstrom's works treat problems in Bayesian and Neyman-Pearson way with the latter corresponding to our analysis. One important quantity in Helstrom's results, when treated the Neyman-Pearson way, is the average of type I and II errors

$$P_e = \frac{1}{2} \text{type I error} + \frac{1}{2} \text{type II error} . \quad (3.6)$$

As mentioned in Section 3.1, his classical treatment got basically our results in a model corresponding to our Poisson model (Helstrom, 1964). However, in (Helstrom, 1973) the researcher found out that the P_e for the quantum theoretical Poisson model (2.8) converges to around $\exp(-t)/2$ with t interpreted as the average of detected photons. The convergence is oscillatory around $\exp(-t)/2$ with the limit reached first when the distance between the two points in the alternative is approximately 2 Rayleigh distances, leading Helstrom to define 2 Rayleigh distances as the resolution for this problem. In our case, if the model is Poisson, binomial or VSG and as usual assuming that the asymptotic relations are valid also for finite quantities (see Chapter 4 for validity of this claim), the P_e is

$$P_e = 1 - \Phi \left(\sqrt{\int_I \frac{h''(x - x_0)}{h(x - x_0)} dx} \frac{d^2 \sqrt{t}}{16} \right) . \quad (3.7)$$

In particular, for a fixed t it tends to $1/2$ as $d \rightarrow 0$ just like in (Helstrom, 1973), but it tends to zero as $d \rightarrow \infty$ for all $t \in \mathbb{N}$.

In Figure 3.3 we have plotted our P_e together with Helstrom's. In part (a) of the figure ($t = 1, \dots, 5$), we can clearly see the difference, however, as expected by the correspondence principal, the difference diminishes as the number of photons t increases in the sense that it becomes significantly harder to distinguish the form of our P_e from Helstrom's already when $t = 20$ Figure 3.3 (b). The only noticeable difference that persists is that for a given distance d , Helstrom's average error seems to be significantly lower than ours. This is to be expected as Helstrom's P_e is achieved by a quantum

mechanically optimal measurement strategy, which has not been achieved experimentally so far, see modern works on quantum information (Tsang et al., 2016a), (Tsang et al., 2016b), (Nair and Tsang, 2016), (Lu et al., 2018), (Tenne et al., 2019).

To produce the above plot of d vs P_e from Helstrom's results is difficult, since the expression for P_e is an infinite series (Equation (39) of (Helstrom, 1973)) and we are not aware of a closed form solution, nor does the author provide any information about the series convergence. To circumvent this problem, we have simply summed a finite number of terms from the series, ending when the final result stopped changing significantly. It seems that one has to take more and more terms as t increases and hence the duration of the above procedure increases as well. Therefore, the quantum theoretical solution is not all-encompassing and our model seems to be easier applicable to modern microscopy.

We note in passing that in a related problem of one point source at x_1 vs another at x_2 by the same reasoning Helstrom was lead to define resolution as equal to the Rayleigh distance, see Chapter VII, Section 6 of (Helstrom, 1976). This problem seems to be more relevant to the Stochastic mode microscopy, see Section 1.2.

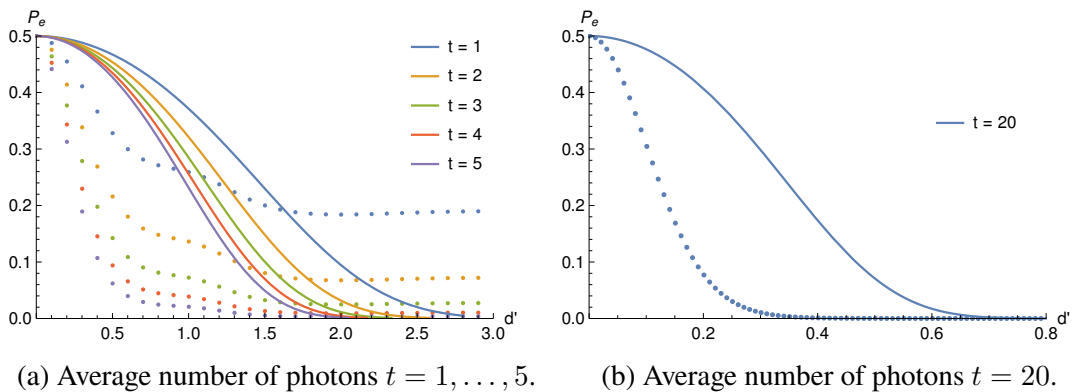


Figure 3.3: Comparison of the average error probabilities P_e vs $d' = d/R_{limit}$. Here R_{limit} stands for the Rayleigh limit (3.6). Due to (1.4) and (1.5) we have used the ratio $\text{FWHM} = 51R_{limit}/61$ when converting between different length scales. Dotted lines are the Helstrom's quantum mechanical P_e 's for the Poisson model with circular aperture (Airy psf) (Helstrom, 1973). Solid lines are the VSG, Poisson and binomial P_e 's (3.7) for a Gaussian psf (2.18). Each line is indexed by the mean number of detected photons. (a) Each of the Helstrom's P_e 's converges to $\exp(-t)/2$ as $d \rightarrow \infty$. This value is first reached at $d \approx 2R_{limit}$ and as d continues to grow, P_e oscillates around $\exp(-t)/2$ with diminishing oscillations. (b) As t increases, the form of Helstrom's P_e plot becomes indistinguishable from the VSG, Poisson and binomial P_e form (3.7).

CHAPTER 4

Simulations

To investigate the finite sample validity of our asymptotic theory, we performed simulations with all 4 different models defined in Section 2.1 in one, two and three dimensions. We checked the dependence of the (asymptotic) resolution d on the illumination time t , FWHM and discretization n . We then applied linear log-log fits. The slopes of the fits are close to the theoretical slopes already at small values of n and t , especially in one dimension (Table 4.1). As t and n increase, the theoretical formulas become good approximations of the simulation fits in one dimension (2nd column of Figure 4.1).

In all simulations we have used the Gaussian psf (2.18), set level $\alpha = 0.1$ and have tried to get the type II error to be in the range $\beta \in [0.95\alpha, 1.05\alpha)$. For simplicity, we only describe the one dimensional ($m = 1$) simulation d vs FWHM from Figure 4.1 (a) in detail, others were conducted similarly. We set the level to be $\alpha = 0.1$, the discretization $n = 20$, the illumination time $t = 20$ and $d = \text{FWHM}$ as the starting distance between the peaks in the alternative (2.8b). Then for 10 000 times we generated n independent random variables following the corresponding model (2.2) under the alternative. Afterwards we calculated the type II error and used the bisection method to advance d until the type II error became between 0.95α and 1.05α (aiming for $\beta = \alpha$). We performed the test for the FWHM range $0.15, 0.16, \dots, 0.25$. We also chose to cap our simulations at 20 steps per single FWHM value to avoid excessively long runtimes or getting trapped at parameter values, where the algorithm does not converge (usually due to the resolution d being too large to fit both psfs under the alternative fully inside the unit interval, see Figure 2.1). The runtime of this simulation is of the order of

$$O(20 \cdot n^m \cdot \text{reps} \cdot \# \text{FWHM}) \tag{4.1}$$

with $\text{reps} = 10\,000$ and $\# \text{FWHM} = 11$. Note that the dependence of the runtime on the dimension m is exponential.

4.1 One dimension

1D simulation fits have slopes approximately equal to the ones calculated with the asymptotical theory already at $t = n = 20$ (Table 4.1). As t and n increase, the asymptotical theory approximates the simulation fits really well (2nd column of Figure 4.1 and Figure 4.2). Moreover, the asymmetry parameter q (2.8b) seems to have the same effect already for finite t , n and d as predicted asymptotically (2.16), (2.15) – the values of d in the subfigures of Figure 4.2 with $q = 1/5$ are around 25% larger than in the respective subfigures of Figure 4.1 with $q = 1/2$.

4.1.1 Symmetric $q = 1/2$ case

Model	$d(\text{FWHM})_{sim}$	$d(t)_{sim}$	$d(n)_{sim}$
	$d(\text{FWHM})_{th}$	$d(t)_{th}$	$d(n)_{th}$
HG	$1.23 \text{ FWHM}^{1.26}$	$1.17 t^{-0.665}$	$0.0502 n^{0.368}$
	$1.08 \text{ FWHM}^{5/4}$	$0.647 t^{-1/2}$	$0.0685 n^{1/4}$
Binomial	$0.828 \text{ FWHM}^{0.911}$	$0.474 t^{-0.306}$	$0.188 n^{-0.00777}$
Poisson	$0.879 \text{ FWHM}^{0.979}$	$0.519 t^{-0.352}$	$0.177 n^{0.00274}$
VSG	$0.873 \text{ FWHM}^{0.975}$	$0.495 t^{-0.336}$	$0.183 n^{-0.00464}$
	0.765 FWHM	$0.323 t^{-1/4}$	0.153

Table 4.1: Comparison of the theoretical formulas from Table 2.1 to the fits of simulations. The entries in $d(\text{FWHM})_{sim}$ correspond to the fits of simulated data points in Figure 4.1 (a), in $d(t)_{sim}$ to Figure 4.1 (c) and in $d(n)_{sim}$ to Figure 4.1 (e).

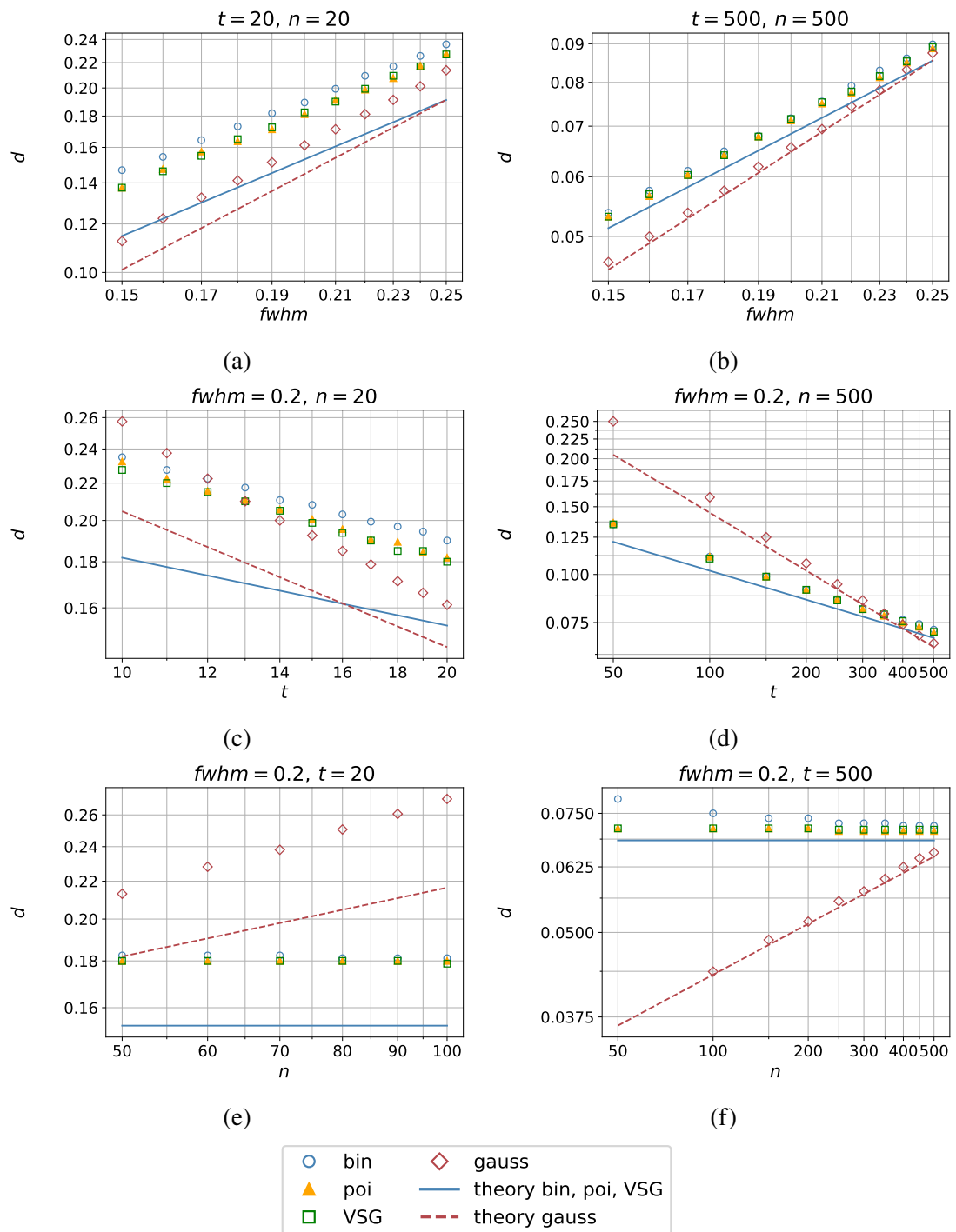


Figure 4.1: Log-log plots of 1D simulations assessing the finite sample validity of the asymptotic relations $d = 2.29 t^{-1/2} n^{1/4} \text{FWHM}^{5/4}$ (Table 2.1 first row, third column) for the homogeneous Gaussian model and $d = 1.62 t^{-1/4} \text{FWHM}$ (Table 2.1 fourth row, third column) for the VSG, poisson and binomial models, see (2.2). The slopes of theoretical formulas are close to the slopes of fits already at low values of illumination time t and discretization n (left column, see also Table 4.1). As t and n increase, the theoretical formulas become reasonable approximations of the simulation fits (right column).

4.1.2 Asymmetric $q = 1/5$ case

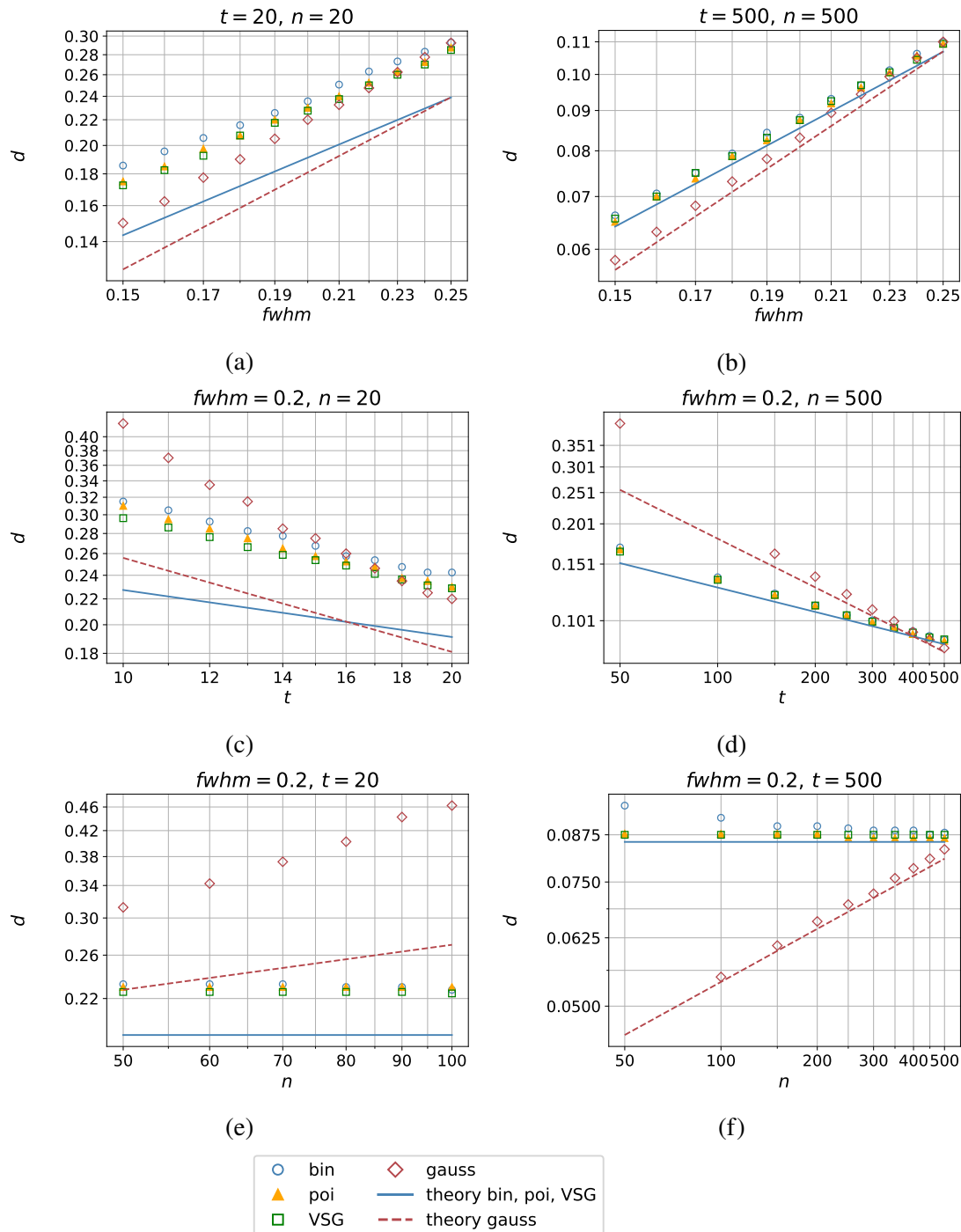


Figure 4.2: Log-log plots of 1D simulations assessing the finite sample validity of the asymptotic relations $d = 2.86 t^{-1/2} n^{1/4} FWHM^{5/4}$ for the homogeneous Gaussian model and $d = 2.03 t^{-1/4} FWHM$ for the VSG, poisson and binomial models with the asymmetry parameter set to $q = 1/5$ in the alternative (2.8b). According to the Theorem 2.3.4 the resolution of the $q = 1/5$ case with respect to the symmetric case $q = 1/2$ (Figure 4.1) is larger by 25% and this is confirmed by simulations as well.

4.2 Two and three dimensions

We focus only on the symmetric case $q = 1/2$ in higher dimensions. To make the runtime (4.1) shorter, we have set $reps = 1000$ in higher dimensions. This does not impact the accuracy significantly (not shown).

The simulations in two and three dimensions are similar to the ones in 1D (Section 4.1): already at low parameter values $t = n = 20$ the slopes of the simulation fits coincide with the theoretical ones as shown in Figure 4.3 (a) and (b), and as t and n increase all simulations converge to our asymptotic theoretical results (Figure 4.3 (c), (d), (e) and (f)). The discrepancy for the HG model in Figure 4.3 (e) and (f) arises because our HG theory is only valid if $n = o(t)$ in 2D by Theorem 2.3.4. The relationship between the resolution d and other parameters holds also as expected, see Figure 4.4.

The only significant difference to 1D is in the HG model. In m dimensions it holds that $d \sim \text{FWHM}_1^{5/4} \prod_{i=2}^m \text{FWHM}_i^{1/4}$, whereas for other models $d \sim \text{FWHM}_1$. By choosing directions in which FWHM is varied or kept constant, we can achieve many different combinations. We chose to plot the most representative ones:

- $d \sim \text{FWHM}_1^{5/4}$ if $\text{FWHM}_i = \text{const.}$ for all $i \in \{2, \dots, m\}$ as shown in Figure 4.3 Column I for 2D and Figure 4.5 (a) for 3D,
- $d \sim \text{FWHM}^{3/2}$ if $\text{FWHM}_1 = \text{FWHM}_2 = \text{FWHM}$ as shown in Figure 4.3 Column II for 2D,
- $d \sim \text{FWHM}^{7/4}$ if $\text{FWHM}_1 = \text{FWHM}_2 = \text{FWHM}_3 = \text{FWHM}$ as shown in Figure 4.5 (d) for 3D,
- $d \sim \text{FWHM}_j^{1/4}$ for some $j \in \{2, \dots, m\}$ and $\text{FWHM}_i = \text{const.}$ for all $i \neq j$ as shown in Figure 4.4 (c) for 2D, and Figure 4.5 (b) and (c) for 3D.

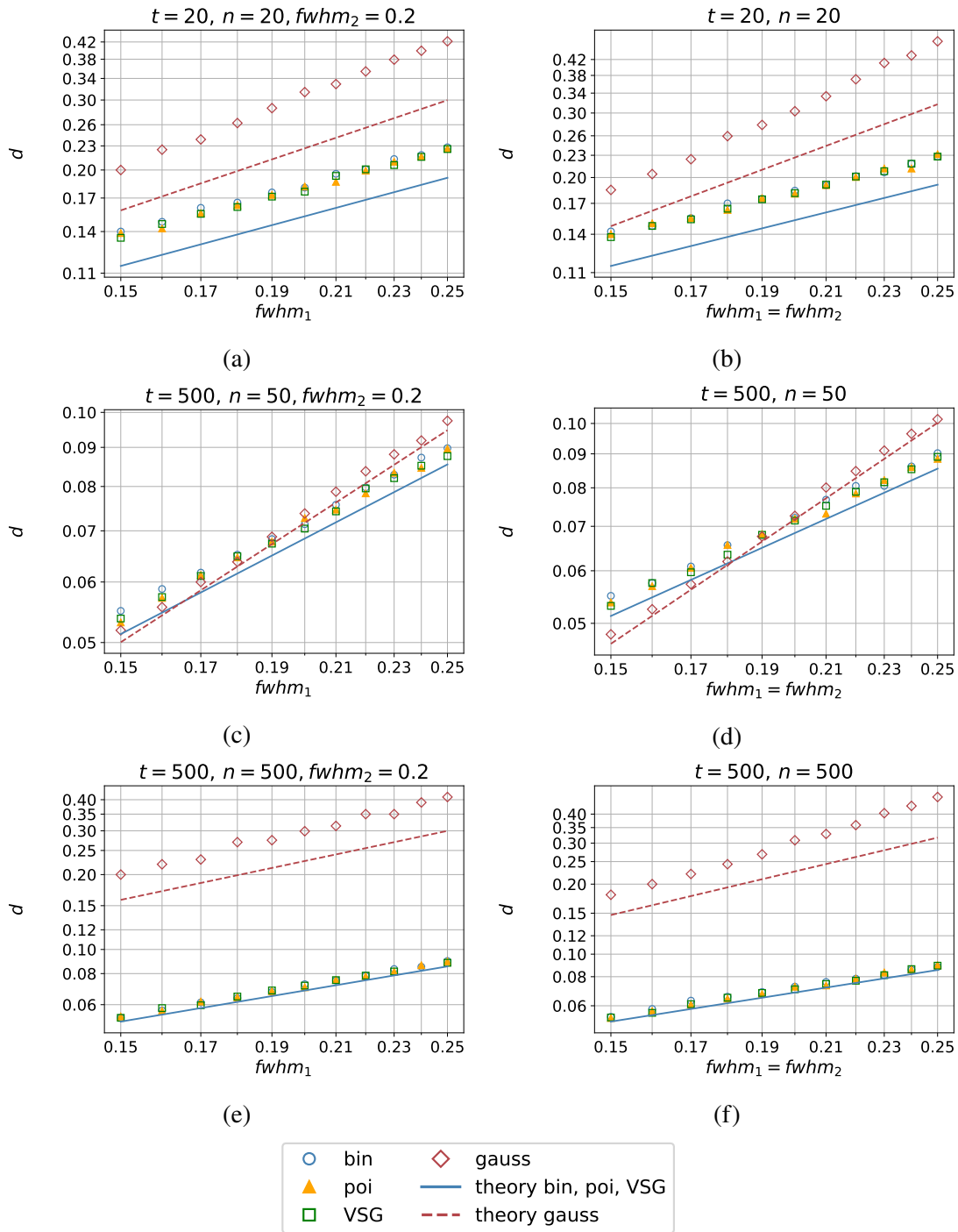


Figure 4.3: Log-log plots of 2D simulations investigating the dependence of resolution d on FWHM: Column I looks at d vs $FWHM_1$ with $FWHM_2 = 0.2$ (constant) and Column II at d vs $FWHM_1 = FWHM_2$. For the HG model we assessed if $d = 2.54 FWHM_1^{5/4} FWHM_2^{1/4} n^{1/2} t^{-1/2}$ (2.20) holds and for the other models if $d = 1.62 t^{-1/4} FWHM_1$ (2.21) holds. In all models, the theory becomes a good approximation of the simulations as t and n increase. The higher discrepancy between the theory and simulations for the HG model in plots (a), (b), (e) and (f) is due to (2.20) being valid only if $n = o(t)$ by Theorem 2.3.4.

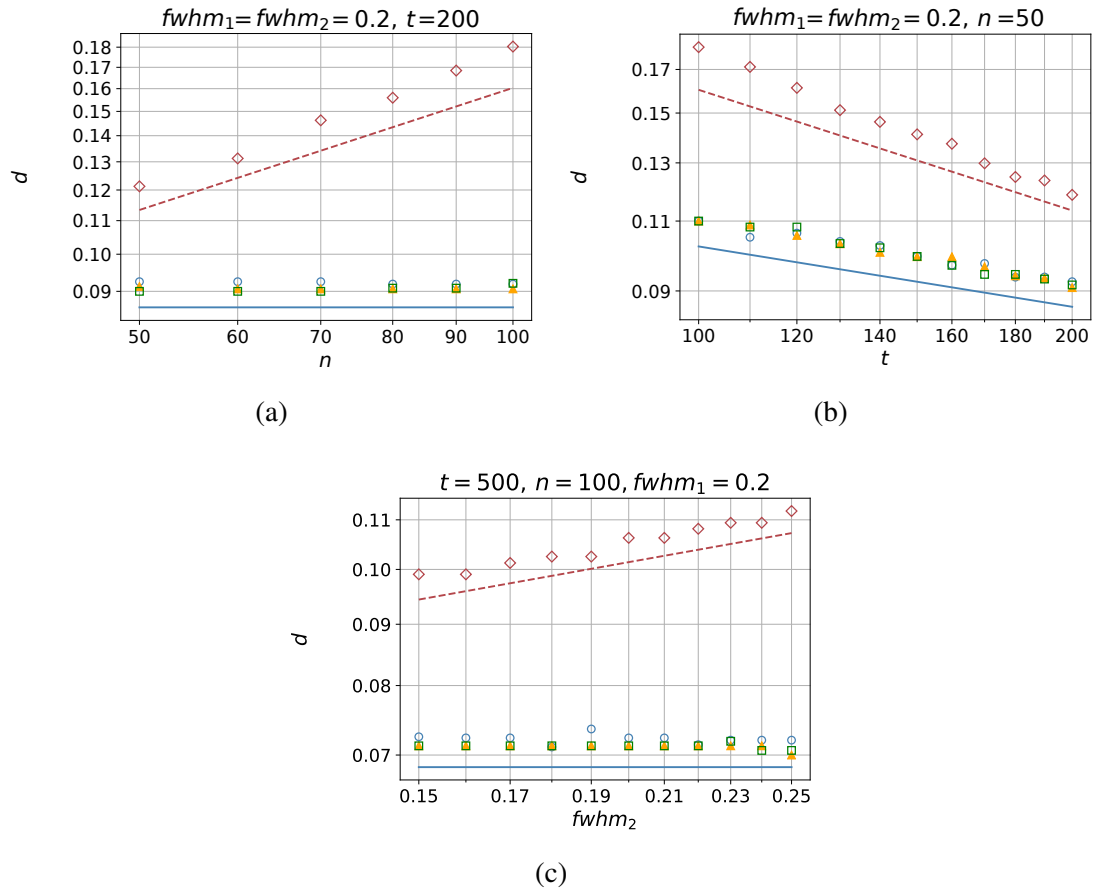


Figure 4.4: Log-log plots of 2D simulations investigating the dependence of resolution d on secondary parameters. For the HG model we assessed if $d = 2.54 \text{FWHM}_1^{5/4} \text{FWHM}_2^{1/4} n^{1/2} t^{-1/2}$ (2.20) holds and for the other models if $d = 1.62 t^{-1/4} \text{FWHM}_1$ (2.21) holds.

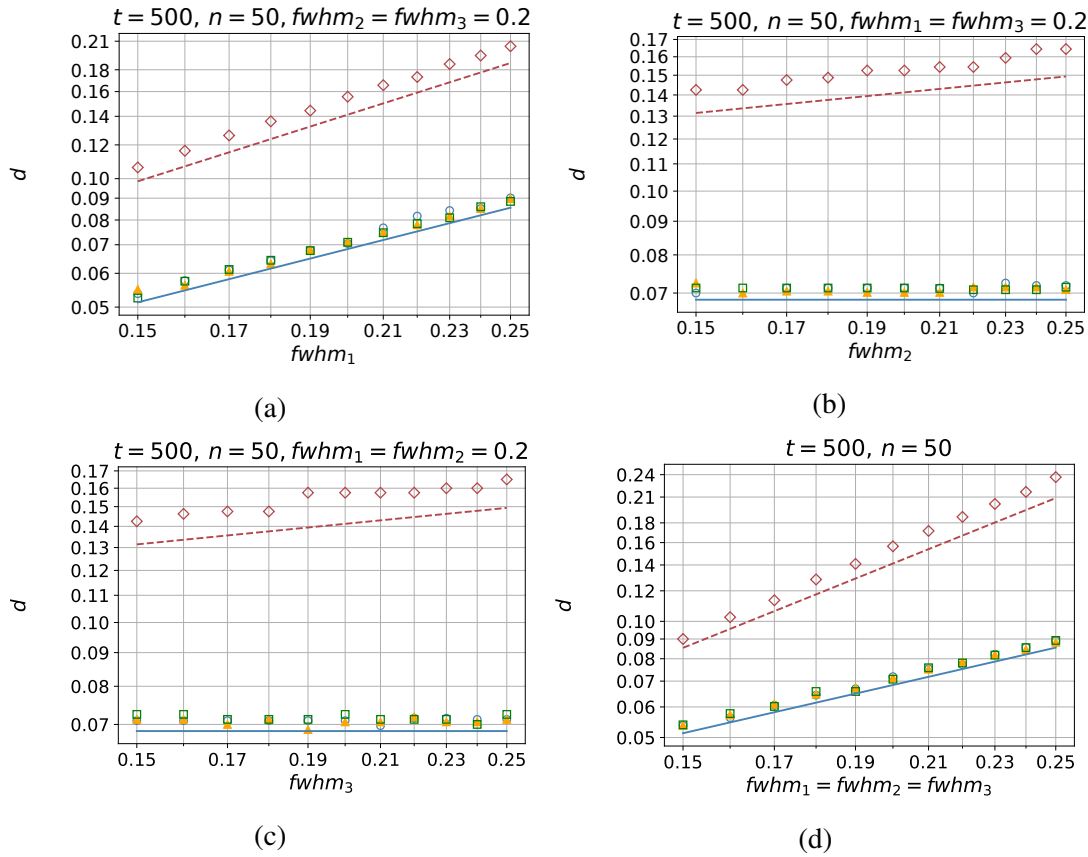


Figure 4.5: Log-log plots of 3D simulations investigating the dependence on FWHM. For the HG model we assessed if $d = 2.81 \text{FWHM}_1^{5/4} \text{FWHM}_2^{1/4} \text{FWHM}_3^{1/4} n^{3/4} t^{-1/2}$ (2.20) holds and for the other models if $d = 1.62 t^{-1/4} \text{FWHM}_1$ (2.21) holds. Plot (a) shows what happens if only FWHM_1 is varied and others are kept constant, (b) if only FWHM_2 , (c) if only FWHM_3 and (d) if all three are varied and set equal to each other.

CHAPTER 5

Proof of the main theorem

In this chapter we will prove Theorem 2.3.4. We will do this separately for the four models defined in Section 2.1. We start with the homogeneous and variance stabilized Gaussian models because the proofs of the binomial and Poisson models rely on them. In both Gaussian models the proofs follow by direct calculation.

In contrast, in binomial and Poisson models the proofs are more involved. If $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$, we prove CLTs. In one dimension, provided that $t \gg \sqrt{n} \log^8 n$, we can also find the statistical resolution d for the Poisson model using asymptotic equivalence. Since $n = ct^{1/2+\delta}$ is equivalent to $c'n^{2-\delta'} = t$ for some $c' > 0$ and $2 > \delta' > 0$, we thus find statistical resolution for all couplings of t and n in the one dimensional Poisson model.

Before we start, let us introduce some notation. For $f(\cdot - x_0) \in L^1[0, 1]^m$ functions we define the integral operators $\int_i, \int_I: L^1[0, 1]^m \rightarrow \mathbb{R}$ as

$$\int_i f := \int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} f(x - x_0) dx \quad \text{and} \quad \int_I f := \int_I f(x - x_0) dx. \quad (5.1)$$

Here, as before, we let $I = [0, 1]^m$. Mostly we will use this notation for the psf h and its derivatives. Note that $p_{0i} = \int_i h$.

5.1 Homogeneous Gaussian model

Proof of Theorem 2.3.4 for the HG model. As $F_{t,\theta} = \mathcal{N}(t\theta, 1)$, the LRT statistic in (2.12) becomes

$$T_{t,n,d}(Y) = \log \left(\frac{\varphi(Y | H_1)}{\varphi(Y | H_0)} \right) = \frac{1}{2} \sum_{i \in [n]^m} (t^2 p_{0i}^2 - t^2 p_{1i}^2 + 2Y_i t (p_{1i} - p_{0i}))$$

with φ the density of a standard normal variate. The corresponding likelihood ratio test (2.11) is given by

$$\Phi_{t,n,d}(Y) := \begin{cases} 1 & \text{if } T_{t,n,d}(Y) > q_{\alpha,t,n,d}^*, \\ 0 & \text{otherwise,} \end{cases} \quad (5.2)$$

where $q_{\alpha,t,n,d}^* = \sqrt{2\mu_{t,n,d}q_{1-\alpha}} - \mu_{t,n,d}$ with $q_{1-\alpha}$ the $1 - \alpha$ quantile of $\mathcal{N}(0, 1)$ and

$$\mu_{t,n,d} = \frac{t^2}{2} \sum_{i \in [n]^m} (p_{1i} - p_{0i})^2. \quad (5.3)$$

For ease of readability, we write only the dependence on n in the following.

It holds that

$$\text{under } H_0 : T_n(Y) \sim \mathcal{N}(-\mu_n, 2\mu_n), \quad (5.4a)$$

$$\text{under } H_1 : T_n(Y) \sim \mathcal{N}(\mu_n, 2\mu_n). \quad (5.4b)$$

We calculate

$$\begin{aligned} \mathbb{P}_{H_0}(\mathbf{reject}) &= \mathbb{P}_{H_0}(T_n(Y) > q_{\alpha,n}^*) \\ &= \mathbb{P}\left(-\mu_n + \sqrt{2\mu_n}W > q_{\alpha,n}^*\right) = 1 - \mathbb{P}\left(W \leq \frac{q_{\alpha,n}^* + \mu_n}{\sqrt{2\mu_n}}\right) = \alpha, \end{aligned}$$

where $W \sim \mathcal{N}(0, 1)$. Thus, the test is indeed a level α test.

We want the type II error to be equal to β . Thus, we require

$$\begin{aligned} \beta = \mathbb{P}_{H_1}(\mathbf{accept}) &= \mathbb{P}_{H_1}(T_n(Y) \leq q_{\alpha,n}^*) \\ &= \mathbb{P}\left(\mu_n + \sqrt{2\mu_n}W \leq q_{\alpha,n}^*\right) = \mathbb{P}\left(W \leq q_{1-\alpha} - \sqrt{2\mu_n}\right), \end{aligned} \quad (5.5)$$

where again $W \sim \mathcal{N}(0, 1)$. This implies that

$$\mu_n = (q_{1-\alpha} - q_\beta)^2/2 = (q_{1-\beta} - q_\alpha)^2/2, \quad (5.6)$$

since $q_{1-\gamma} = -q_\gamma$ for quantiles of $\mathcal{N}(0, 1)$. By definition of μ_n we have

$$\mu_n = \frac{t^2}{2} \sum_{i \in [n]^m} (p_{1i} - p_{0i})^2 = \frac{t^2}{2} \sum_{i \in [n]^m} \left(\int_i \Delta\right)^2, \quad (5.7)$$

where

$$\Delta(x - x_0) := qh(x - x_1) + (1 - q)h(x - x_2) - h(x - x_0) \quad (5.8)$$

is the difference between the psfs under H_1 and H_0 . Since $h \in C^2[0, 1]^m$ and

$$d = \|x_1 - x_2\| = |x_{11} - x_{21}|, \quad (5.9)$$

we have

$$h(x - x_i) = \sum_{j=0}^k \frac{\partial_1^j h(x - x_0)}{j!} (x_{01} - x_{i1})^j + o((x_{01} - x_{i1})^k), \quad (5.10)$$

where $\partial_1 f(y - x_0) = \partial f(y - x_0)/\partial y_1$. Assume w.l.o.g. that $x_{21} \geq x_{11}$ so that $d = x_{21} - x_{11}$. Note that for $x_{01} = qx_{11} + (1-q)x_{21}$, it holds that $x_{01} - x_{11} = (1-q)d$ and $x_{01} - x_{21} = -qd$. Hence, for $d \rightarrow 0$ we have

$$\begin{aligned} \Delta(x - x_0) &= qh(x - x_1) + (1-q)h(x - x_2) - h(x - x_0) \\ &= q \sum_{j=0}^k \frac{\partial_1^j h(x - x_0)}{j!} (x_{01} - x_{11})^j + (1-q) \sum_{j=0}^k \frac{\partial_1^j h(x - x_0)}{j!} (x_{01} - x_{21})^j \\ &\quad - h(x - x_0) + o((x_{01} - x_{11})^k + (x_{01} - x_{21})^k) \end{aligned} \quad (5.11)$$

$$= \sum_{j=2}^k \frac{\partial_1^j h(x - x_0)}{j!} (q((1-q)d)^j + (-1)^j(1-q)(qd)^j) + o(d^k). \quad (5.12)$$

Thus, setting $k = 2$ and using the notation

$$h''(y - x_0) = \frac{\partial^2 h(y - x_0)}{\partial y_1^2} \quad (5.13)$$

as before (2.17), we get

$$\begin{aligned} \mu_n &= \frac{t^2}{2} \sum_{i \in [n]^m} (p_{1i} - p_{0i})^2 = \frac{t^2}{2} \sum_{i \in [n]^m} \left(\int_i \Delta \right)^2 \\ &= \frac{t^2}{2} \sum_{i \in [n]^m} \left(\frac{q(1-q)}{2} d^2 \int_i h'' + o\left(\frac{d^2}{n^m}\right) \right)^2 \\ &= \frac{t^2}{2} \sum_{i \in [n]^m} \left(\frac{q^2(1-q)^2}{4} d^4 \left(\int_i h'' \right)^2 + o\left(\frac{d^4}{n^{2m}}\right) \right) \\ &= \frac{t^2}{2} \sum_{i \in [n]^m} \frac{q^2(1-q)^2}{4} d^4 \left(\int_i h'' \right)^2 + o\left(\frac{t^2 d^4}{n^m}\right) \end{aligned} \quad (5.14)$$

$$= \frac{q^2(1-q)^2}{8} \frac{t^2 d^4}{n^m} \int_I (h'')^2 + o\left(\frac{t^2 d^4}{n^m}\right), \quad (5.15)$$

applying Lemma A.1.1 from the Appendix. Rearranging (5.15) for d and using (5.6) we get the desired relation (2.16). Since we need $d \rightarrow 0$ for the above to hold, we have to require that $n^m = o(t^2)$. \square

Remark 5.1.1. As long as asymptotically

$$\mu_{t,n,d} = \frac{(q_{1-\beta} - q_\alpha)^2}{2}$$

and $d \searrow 0$, the above proof still holds. Thus, using (5.14) for finite n we get

$$d \asymp \frac{\sqrt{2}}{\sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \left(\sum_{i \in [n]^m} \left(\int_i h'' \right)^2 \right)^{-1/4} t^{-1/2}.$$

If $d \not\rightarrow 0$, i.e. if $n^m = o(t^2)$ does *not* hold, then we can find the corresponding resolution d numerically using the bisection method from

$$\frac{(q_{1-\beta} - q_\alpha)^2}{t^2} = \sum_{i \in [n]^m} \left(\int_i \Delta \right)^2 \quad (5.16)$$

and Equations (5.8) and (5.9). To derive (5.16), combine different expressions for μ_n from Equations (5.6) and (5.7). Another method is to use simulations, see Chapter 4.

5.2 Variance stabilized Gaussian model

Let $F_{2\sqrt{t}, \sqrt{\theta}} = \mathcal{N}(2\sqrt{t\theta}, 1)$, i.e.

$$Y_i \stackrel{\text{indep.}}{\sim} \mathcal{N} \left(2 \left(t \int_{(i-1)/n}^{i/n} g(x - x_0) dx \right)^{1/2}, 1 \right). \quad (5.17)$$

Then the log-likelihood function is

$$T_{t,n,d}(Y) = \log \left(\frac{\varphi(Y | H_1)}{\varphi(Y | H_0)} \right) = \sum_{i \in [n]^m} \left[2t(p_{0i} - p_{1i}) + 2Y_i \sqrt{t} (\sqrt{p_{1i}} - \sqrt{p_{0i}}) \right]$$

with p_i defined in Equations (2.13) and (2.14). We define the corresponding likelihood ratio test as in (5.2), but this time we set $q_{\alpha,t,n,d}^* = \sqrt{2\nu_{t,n,d}q_{1-\alpha}} - \nu_{t,n,d}$ with

$$\nu_{t,n,d} = 2t \sum_{i \in [n]^m} (\sqrt{p_{1i}} - \sqrt{p_{0i}})^2.$$

Proof. The proof is similar to the proof of the homogeneous Gaussian model in Section 5.1. We skip the indices t and d in what follows.

We have under $H_0 : T_n(Y) \sim \mathcal{N}(-\nu_n, 2\nu_n)$ and under $H_1 : T_n(Y) \sim \mathcal{N}(\nu_n, 2\nu_n)$. We calculate

$$\begin{aligned} \mathbb{P}_{H_0}(\mathbf{reject}) &= \mathbb{P}_{H_0}(T_n(Y) > q_{\alpha,n}^*) \\ &= \mathbb{P}(-\nu_n + \sqrt{2\nu_n}W > q_{\alpha,n}^*) = 1 - \mathbb{P}\left(W \leq \frac{q_{\alpha,n}^* + \nu_n}{\sqrt{2\nu_n}}\right) = \alpha, \end{aligned}$$

where as previously $W \sim \mathcal{N}(0, 1)$. Thus, the test is indeed a level α test.

We want the type II error to be equal to β . Thus, we require

$$\begin{aligned} \beta = \mathbb{P}_{H_1}(\mathbf{accept}) &= \mathbb{P}_{H_1}(T_n(Y) \leq q_{\alpha,n}^*) \\ &= \mathbb{P}(\nu_n + \sqrt{2\nu_n}W \leq \sqrt{2\nu_n}q_{1-\alpha} - \nu_n) = \mathbb{P}(W \leq q_{1-\alpha} - \sqrt{2\nu_n}). \end{aligned}$$

This implies that

$$\nu_n = (q_{1-\alpha} - q_\beta)^2/2 = (q_{1-\beta} - q_\alpha)^2/2, \quad (5.18)$$

since $q_{1-\gamma} = -q_\gamma$ for quantiles of $\mathcal{N}(0, 1)$. On the other hand, by definition of ν_n we have

$$\nu_n = 2t \sum_{i \in [n]^m} (\sqrt{p_{1i}} - \sqrt{p_{0i}})^2. \quad (5.19)$$

Using the Taylor series expansion (5.10) as $d \rightarrow 0$ we get

$$\begin{aligned} (\sqrt{p_{1i}} - \sqrt{p_{0i}})^2 &= \left(\sqrt{\int_i h + \frac{q(1-q)d^2}{2} \int_i h'' + o\left(\frac{d^2}{n^m}\right)} - \sqrt{\int_i h} \right)^2 \\ &= \left(\sqrt{\int_i h} \sqrt{1 + \frac{q(1-q)d^2}{2} \frac{\int_i h''}{\int_i h} + o(d^2)} - \sqrt{\int_i h} \right)^2 \\ &= \left(\frac{q(1-q)d^2}{4} \frac{\int_i h''}{\sqrt{\int_i h}} + o\left(\frac{d^2}{n^{m/2}}\right) \right)^2 \\ &= \frac{q^2(1-q)^2 d^4}{16} \frac{(\int_i h'')^2}{\int_i h} + o\left(\frac{d^4}{n^m}\right), \end{aligned} \quad (5.20)$$

where as before the acute accent $'$ denotes the partial derivative with respect to the first coordinate (5.13). Thus,

$$\nu_n = 2t \sum_{i \in [n]^m} \left(\frac{q^2(1-q)^2 d^4}{16} \frac{(\int_i h'')^2}{\int_i h} + o\left(\frac{d^4}{n^m}\right) \right)$$

$$= \frac{q^2(1-q)^2 td^4}{8} \int_I \frac{(h'')^2}{h} + o(td^4) \quad (5.21)$$

as $n \rightarrow \infty$ by Lemma A.1.1 with $f(x) = h''(x-x_0)$ and $g(x) = h(x-x_0)$. Rearranging the last equation for d together with (5.18) gives (2.15), as required. \square

Remark 5.2.1. Just like in the homogeneous Gaussian model (Remark 5.1.1), we can keep n finite in the above proof, provided that the resolution satisfies

$$d \asymp \frac{\sqrt{2}}{\sqrt{q(1-q)}} \sqrt{q_{1-\beta} - q_\alpha} \left(\sum_{i \in [n]^m} \frac{(\int_i h'')^2}{\int_i h} \right)^{-1/4} t^{-1/2}.$$

5.3 Poisson and binomial models

5.3.1 Analysis in the central limit theorem regime

5.3.1.1 Poisson model

Here we have (recall (2.2))

$$F_{t,\theta} = \text{Poi}(t\theta),$$

or more explicitly

$$Y_i \stackrel{\text{indep.}}{\sim} \text{Poi}(\lambda_i) \quad \text{with} \quad \lambda_i = t \int_i g. \quad (5.22)$$

The likelihood ratio statistic for (2.8) under the model (5.22) is

$$T_{t,n,d}(Y) = \log \left(\prod_{i \in [n]^m} e^{-(\lambda_{1i} - \lambda_{0i})} \left(\frac{\lambda_{1i}}{\lambda_{0i}} \right)^{Y_i} \right) = \sum_{i \in [n]^m} Y_i \log \left(\frac{\lambda_{1i}}{\lambda_{0i}} \right). \quad (5.23)$$

We have

$$\mathbb{E}[T_{t,n,d}] = \sum_{i \in [n]^m} \log \left(\frac{\lambda_{1i}}{\lambda_{0i}} \right) \lambda_i \quad \text{and} \quad \mathbb{V}[T_{t,n,d}] = \sum_{i \in [n]^m} \log \left(\frac{\lambda_{1i}}{\lambda_{0i}} \right)^2 \lambda_i.$$

Here $\lambda_{1i} = tp_{1i}$ and $\lambda_{0i} = tp_{0i}$.

Theorem 5.3.1 (CLT for Poisson LRT). *Assume the point spread function h satisfies Assumption 2.3.2 and that $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$. Then a CLT holds for $T_n(Y)$ under the hypothesis (2.8a) and the alternative (2.8b) as $t, n \rightarrow \infty$ and $d \rightarrow 0$, i.e.*

$$\frac{T_{t,n,d} - \mathbb{E}[T_{t,n,d}]}{\sqrt{\mathbb{V}[T_{t,n,d}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

as $t, n \rightarrow \infty$ and $d \rightarrow 0$.

Proof of Theorem 5.3.1. We apply the Lindeberg-Feller CLT for triangular arrays (see (Billingsley, 1986)). For ease of readability, we skip indices t and d , and only give the full proof in the symmetric $q = 1/2$ case in what follows. See the end of the proof (5.34) for the outline of the general $q \in (0, 1)$ case.

Let

$$X_{ni} = a_i Y_i,$$

so that

$$T_n(Y) = \sum_{i \in [n]^m} X_{ni},$$

where as before

$$Y_i \stackrel{\text{indep.}}{\sim} \text{Poi}(\lambda_i) \quad \text{and} \quad a_i := \log \left(\frac{\lambda_{1i}}{\lambda_{0i}} \right). \quad (5.24)$$

Note that λ_i and a_i depend on n as well. We also set $\mu_{ni} = \mathbb{E}[X_{ni}]$, $\sigma_{ni}^2 = \mathbb{V}[X_{ni}]$ and let $\tau_n^2 = \sum_{i=1}^n \sigma_{ni}^2$. We need to show that $\sigma_{ni}^2 < \infty$ and that for all $\varepsilon > 0$ we have

$$L_n(\varepsilon) = \frac{1}{\tau_n^2} \sum_{i \in [n]^m} \int (x - \mu_{ni})^2 \mathbf{1}_{\{|x - \mu_{ni}| > \varepsilon \tau_n\}} d\mathbb{P}_{X_{ni}}(x) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We use the Taylor approximation $\log(1 + y_i) = \sum_{k=0}^2 (-1)^k / (k+1) y_i^{k+1} + o(y_i^3)$ with

$$y_i = \frac{\lambda_{1i}}{\lambda_{0i}} - 1 = \frac{\lambda_{1i} - \lambda_{0i}}{\lambda_{0i}} = \frac{\int_i \Delta}{\int_i h} = \frac{1}{8} \frac{\int_i h''}{\int_i h} d^2 + \frac{1}{384} \frac{\int_i h''''}{\int_i h} d^4 + o(d^4), \quad (5.25)$$

where we have used Equation (5.11) for $\Delta(x - x_0)$ with $k = 4$ and $q = 1/2$. Under the hypothesis H_0 it holds

$$\begin{aligned} \mu_{ni} &= \mathbb{E}_{H_0} X_{ni} = a_i \mathbb{E}_{H_0} Y_i = a_i \lambda_{0i} = a_i t \int_i h = \log(1 + y_i) t \int_i h \\ &= \frac{td^2}{8} \int_i h'' + td^4 \left(-\frac{1}{128} \frac{(\int_i h'')^2}{\int_i h} + \frac{1}{384} \int_i h'''' \right) + O\left(\frac{td^6}{n^m}\right), \end{aligned}$$

$$\begin{aligned} \nu_n &= \sum_{i \in [n]^m} \mu_{ni} = \mathbb{E}_{H_0} T_n(Y) = t \sum_{i \in [n]^m} a_i \int_i h \\ &= \frac{td^2}{8} \int_I h'' + td^4 \left(-\frac{1}{128} \rho_n + \frac{1}{384} \int_I h'''' \right) + O(td^6), \end{aligned}$$

$$\sigma_{ni}^2 = \mathbb{V}_{H_0} X_{ni} = a_i^2 \mathbb{V}_{H_0} Y_i = a_i^2 \lambda_{0i} = t \log(1 + y_i)^2 \int_i h = \frac{td^4}{64} \frac{(\int_i h'')^2}{\int_i h} + O\left(\frac{td^6}{n^m}\right),$$

$$\tau_n^2 = \sum_{i \in [n]^m} \sigma_{ni}^2 = \mathbb{V}_{H_0} T_n(Y) = t \sum_{i \in [n]^m} a_i^2 \int_i h = t \sum_{i \in [n]^m} \log(1 + y_i)^2 \int_i h$$

$$= \frac{td^4}{64} \rho_n + O(td^6)$$

with

$$\rho_n := \sum_{i \in [n]^m} \frac{\left(\int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} h''(x-x_0) dx \right)^2}{\int_{(i_1-1)/n}^{i_1/n} \cdots \int_{(i_m-1)/n}^{i_m/n} h(x-x_0) dx}. \quad (5.26)$$

Clearly, it holds that $\sigma_{ni}^2 < \infty$. Applying Lemma A.1.1 with $f(x) = h''(x-x_0)$ and $g(x) = h(x-x_0)$ we see that

$$\rho_n = \int_I \frac{h''(x-x_0)^2}{h(x-x_0)} dx + o(1) < \infty$$

and hence

$$\tau_n^2 = \frac{td^4}{64} \int_I \frac{h''(x-x_0)^2}{h(x-x_0)} dx + O(td^6).$$

We consider

$$\begin{aligned} L_n(\varepsilon) &= \frac{1}{\tau_n^2} \sum_{i \in [n]^m} \sum_{\substack{k \in a_i \mathbb{N}_0 \\ |k - \mu_{ni}| > \varepsilon \tau_n}} (k - \mu_{ni})^2 \mathbb{P}_{H_0}(X_{ni} = k) \\ &= \frac{1}{\tau_n^2} \sum_{i \in [n]^m} \sum_{\substack{l \in \mathbb{N}_0 \\ |a_i l - \mu_{ni}| > \varepsilon \tau_n}} (a_i l - \mu_{ni})^2 \mathbb{P}_{H_0}(Y_i = l). \end{aligned} \quad (5.27)$$

Note that if $a_i = 0$, then $|a_i l - \mu_{ni}| = 0$. If $a_i \neq 0$, then the condition $|a_i l - \mu_{ni}| > \varepsilon \tau_n$ on l is equivalent to $l \in \mathcal{D}_{\varepsilon, i}$, where $\mathcal{D}_{\varepsilon, i}$ is the set consisting of all $l \in \mathbb{N}_0$ satisfying

$$\begin{cases} l > \varepsilon \sqrt{t} \frac{\sqrt{\sum_{i \in [n]^m} a_i^2 \int_i h}}{|a_i|} + t \int_i h \\ l < -\varepsilon \sqrt{t} \frac{\sqrt{\sum_{i \in [n]^m} a_i^2 \int_i h}}{|a_i|} + t \int_i h. \end{cases} \quad (5.28)$$

It holds that

$$a_i = \log(1 + y_i) = y_i + O(y_i^2) = \frac{\int_i h'' d^2}{\int_i h} \frac{d^2}{8} + o(d^2),$$

$$\sum_{i \in [n]^m} a_i^2 \int_i h = \sum_{i \in [n]^m} (y_i^2 + O(y_i^3)) \int_i h = \sum_{i \in [n]^m} \frac{(\int_i h'')^2 d^4}{\int_i h} \frac{d^4}{64} + o(d^4) \quad (5.29)$$

and thus

$$R := \min_{j \in [n]^m} \frac{\sqrt{\sum_{i \in [n]^m} a_i^2 \int_i h}}{|a_j|} = O(1). \quad (5.30)$$

Hence, the domain $\mathcal{D}_{\varepsilon,i}$ is a subset of those indices $l \in \mathbb{N}_0$ that satisfy

$$\begin{cases} \varepsilon\sqrt{t}R + \frac{t}{n^m}q < l < \infty \\ 0 \leq l < -\varepsilon\sqrt{t}R + \frac{t}{n^m}\bar{q}, \end{cases} \quad (5.31)$$

where

$$q = \min_{x \in I} h(x - x_0) \quad \text{and} \quad \bar{q} = \max_{x \in I} h(x - x_0) < \infty, \quad (5.32)$$

since $h(\cdot - x_0) \in C^4[0, 1]^m$.

For $n^m = ct^{1/2+\delta}$ there are no l 's satisfying the second inequality of (5.31) for sufficiently large t . W.l.o.g. we set $c = 1$ in what follows. Setting $l_0 = \lceil \sqrt{t}(\varepsilon R + t^{-\delta}q) \rceil$ it holds that

$$\begin{aligned} L_n(\varepsilon) &\leq \frac{1}{\tau_n^2} \sum_{i \in [n]^m} \sum_{l=l_0}^{\infty} a_i^2 \left(l - t \int_i h \right)^2 e^{-\lambda_{i0}} \frac{\lambda_{i0}^l}{l!} \\ &= \frac{t^2}{\tau_n^2} \sum_{i \in [n]^m} a_i^2 \sum_{l=l_0}^{\infty} \left(\frac{1}{t} - \frac{\int_i h}{l} \right)^2 e^{-\lambda_{i0}} \frac{\lambda_{i0}^l l}{(l-1)!}. \end{aligned} \quad (5.33)$$

Moreover,

$$\left(\frac{1}{t} - \frac{\int_i h}{l} \right)^2 \leq \frac{1}{t^2} - 2 \frac{q}{tl n^m} + \frac{\bar{q}^2}{l^2 n^{2m}} = o(1).$$

Note that (5.29) also implies that

$$\sum_{i \in [n]^m} a_i^2 = O(n^m d^4).$$

Thus, we have that

$$\begin{aligned} L_n(\varepsilon) &\leq C' \frac{t^2}{\tau_n^2} \sum_{i \in [n]^m} a_i^2 \sum_{l=l_0}^{\infty} e^{-\lambda_{i0}} \frac{\lambda_{i0}^l l}{(l-1)!} \leq C n^m t \sum_{l=l_0}^{\infty} \frac{l}{(l-1)!} \left(\frac{t}{n^m \bar{q}} \right)^l \\ &= C t^{3/2+\delta} \sum_{l=l_0}^{\infty} \frac{l}{(l-1)!} \left(\frac{t}{n^m \bar{q}} \right)^l, \end{aligned}$$

for some constants $C, C' > 0$.

Consider

$$\sum_{l=a}^{\infty} \frac{(t^{1/2-\delta} \bar{q})^l}{l!} = \frac{(t^{1/2-\delta} \bar{q})^a}{a!} \left(1 + \sum_{l=a+1}^{\infty} \frac{(t^{1/2-\delta} \bar{q})^l}{l!} \frac{a!}{(t^{1/2-\delta} \bar{q})^a} \right)$$

$$= \frac{(t^{1/2-\delta}\bar{q})^a}{a!} \left(1 + \frac{t^{1/2-\delta}\bar{q}}{a+1} + \frac{(t^{1/2-\delta}\bar{q})^2}{(a+1)(a+2)} + \dots \right).$$

Setting $a = \lceil \sqrt{t}(\varepsilon + t^{-\delta}) \rceil$, second and further terms in the brackets are of the order $t^{-\delta k} \varepsilon^{-k}$ and so we get

$$1 + \sum_{k=1}^{\infty} (\varepsilon^{-1} t^{-\delta})^k = \frac{1}{1 - \varepsilon^{-1} t^{-\delta}} = O(1),$$

as $t \rightarrow \infty$. Using Stirling's approximation

$$\log k! = k \log k - k + O(\log k)$$

we have that

$$\begin{aligned} \frac{(t^{1/2-\delta})^a}{a!} &= \frac{\exp(a \log(t^{1/2-\delta}))}{\exp(a \log a - a + O(\log a))} = \exp(a (\log t^{1/2-\delta} - \log a + 1) + O(\log a)) \\ &= \exp\left(\sqrt{t}(\varepsilon + t^{-\delta}) (-\log(t^\delta \varepsilon + 1) + 1 + O(1))\right) = O\left((t^\delta \varepsilon + 1)^{-\sqrt{t}(\varepsilon + t^{-\delta})}\right). \end{aligned}$$

In our case the terms are of the form

$$t^{1-2\delta} \bar{q}^2 \sum_{k=a}^{\infty} \frac{(t^{1/2-\delta} \bar{q})^k}{k!} \frac{k+2}{k+1},$$

with $a = \lceil \varepsilon \sqrt{t} R + t^{1/2-\delta} q \rceil - 2 \sim (\lceil \sqrt{t}(\varepsilon + t^{-\delta}) \rceil)$ and $(k+2)/(k+1) \rightarrow 1$ as $t \rightarrow \infty$. Thus, the above considerations apply and all together we get

$$L_n(\varepsilon) \leq O\left(t^{3/2+\delta} (t^\delta \varepsilon + 1)^{-\sqrt{t}(\varepsilon + t^{-\delta})}\right) \rightarrow 0 \quad \text{as } t, n \rightarrow \infty, d \rightarrow 0.$$

Under the hypothesis H_1 we have

$$\begin{aligned} \mu_{ni} &= \mathbb{E}_{H_1} X_{ni} = a_i \lambda_{1i} = a_i (1 + y_i) \lambda_{0i} \\ &= \frac{td^2}{8} \int_i h'' + td^4 \left(\frac{1}{128} \frac{(\int_i h'')^2}{\int_i h} + \frac{1}{384} \int_i h'''' \right) + O\left(\frac{td^6}{n^m}\right), \\ \nu_n &= \sum_{i \in [n]^m} \mu_{ni} = \mathbb{E}_{H_1} T_n(Y) = \sum_{i \in [n]^m} a_i (1 + y_i) \lambda_{0i} \\ &= \frac{td^2}{8} \int_I h'' + td^4 \left(\frac{1}{128} \rho_n + \frac{1}{384} \int_I h'''' \right) + O(td^6), \\ \sigma_{ni}^2 &= \mathbb{V}_{H_1} X_{ni} = a_i^2 \lambda_{1i} = a_i^2 (1 + y_i) \lambda_{0i} = \frac{td^4}{64} \frac{(\int_i h'')^2}{\int_i h} + O\left(\frac{td^6}{n^m}\right), \end{aligned}$$

$$\tau_n^2 = \sum_{i \in [n]^m} \sigma_{ni} = \mathbb{V}_{H_1} T_n(Y) = \sum_{i \in [n]^m} a_i^2 (1 + y_i) \lambda_{0i} = \frac{td^4}{64} \rho_n + O(td^6)$$

and hence similar considerations prove Lindeberg's condition in this case.

In the general case with the asymmetry parameter $q \in (0, 1)$, Equation (5.25) becomes

$$\begin{aligned} y_i = \frac{\lambda_{1i}}{\lambda_{0i}} - 1 &= \frac{\lambda_{1i} - \lambda_{0i}}{\lambda_{0i}} = \frac{\int_i \Delta}{\int_i h} = \frac{q(1-q)}{2} \frac{\int_i h''}{\int_i h} d^2 + \frac{q(1-q)(1-2q)}{6} \frac{\int_i h'''}{\int_i h} d^3 \\ &+ \frac{q(1-q)((1-q)^2 - q(1-2q))}{24} \frac{\int_i h''''}{\int_i h} d^4 + o(d^4) \end{aligned} \quad (5.34)$$

and the following terms $\mathbb{E}_{H_0} T_n$, $\mathbb{V}_{H_0} T_n$, $\mathbb{E}_{H_1} T_n$ and $\mathbb{V}_{H_1} T_n$ change accordingly. We skip these expressions due to their length and because they are not particularly insightful. Since the asymmetric factor $q \in (0, 1)$ does not change the asymptotic behavior, it is clear that the CLTs under H_0 and H_1 hold just like in the symmetrically placed $q = 1/2$ case. \square

Remark 5.3.2. Due to

$$\frac{\sigma_{ni}^2}{\tau_n^2} \rightarrow 0$$

for all $i \in \{1, \dots, n\}$, Lindeberg's condition is necessary for the CLTs above to hold, see e.g. (Petrov, 1975) or (Billingsley, 1986).

Now we can analyze the Poisson LRT

$$\Phi_{t,n,d}(Y) := \begin{cases} 1 & \text{if } T_{t,n,d}(Y) > q_{\alpha,t,n,d}^* \\ 0 & \text{otherwise,} \end{cases} \quad (5.35)$$

in the CLT regime above. Here

$$q_{\alpha,t,n,d}^* := q_{1-\alpha} \sqrt{\mathbb{V}_{H_0} T_{t,n,d}} + \mathbb{E}_{H_0} T_{t,n,d} \quad (5.36)$$

with the $1 - \alpha$ quantile $q_{1-\alpha}$ of $\mathcal{N}(0, 1)$.

Proof of Theorem 2.3.4 Poisson model in the CLT regime. As usual, we skip the indices of t and d .

We want to find such $q_{\alpha,n}^*$ that

$$\mathbb{P}_{H_0}(\text{reject}) = \mathbb{P}_{H_0}(T_n(Y) > q_{\alpha,n}^*) = \alpha \quad (5.37)$$

and

$$\mathbb{P}_{H_1}(\mathbf{accept}) = \mathbb{P}_{H_1}(T_n(Y) \leq q_{\alpha,n}^*) = \beta \quad (5.38)$$

hold. If we take $q_{\alpha,n}^* := \sqrt{\mathbb{V}_{H_0} T_n} q_{1-\alpha} + \mathbb{E}_{H_0} T_n$, then by the CLT 5.3.1, Equation (5.37) holds asymptotically, i.e. for sufficiently large t, n and sufficiently small d , (5.37) holds exactly with some $\tilde{q}_{\alpha,n}^* = q_{\alpha,n}^* + o(1)$. Similarly, by the CLT 5.3.1 under H_1 we get Equation (5.38) with $q_{\alpha,n}^* := \sqrt{\mathbb{V}_{H_1} T_n} q_{\beta} + \mathbb{E}_{H_1} T_n$. For the quantile to be well-defined, we need to figure out when

$$\sqrt{\mathbb{V}_{H_1} T_n} q_{\beta} + \mathbb{E}_{H_1} T_n = \sqrt{\mathbb{V}_{H_0} T_n} q_{1-\alpha} + \mathbb{E}_{H_0} T_n + o(1).$$

Using previous calculations it holds that

$$\mathbb{E}_{H_1} T_n - \mathbb{E}_{H_0} T_n = \sum_{i \in [n]^m} y_i \log(1 + y_i) \lambda_{0i} = \sum_{i \in [n]^m} \lambda_{0i} (y_i^2 + O(y_i^3)) \quad (5.39)$$

and

$$\sqrt{\mathbb{V}_{H_0} T_n} = \sqrt{\sum_{i \in [n]^m} \lambda_{0i} (y_i^2 + O(y_i^3))} = \sqrt{\mathbb{V}_{H_1} T_n}. \quad (5.40)$$

Thus, the quantile is well-defined if

$$q_{1-\alpha} \sqrt{\mathbb{V}_{H_0} T_n} - q_{\beta} \sqrt{\mathbb{V}_{H_1} T_n} = \mathbb{E}_{H_1} T_n - \mathbb{E}_{H_0} T_n + o(1) \iff (5.41)$$

$$\sqrt{\sum_{i \in [n]^m} \lambda_{0i} (y_i^2 + O(y_i^3))} = q_{1-\alpha} - q_{\beta} \iff (5.42)$$

$$\frac{q(1-q)\sqrt{t}d^2}{2} \sqrt{\int_I \frac{h''(x-x_0)^2}{h(x-x_0)} dx} + o(d^4) = q_{1-\alpha} - q_{\beta} = q_{1-\beta} - q_{\alpha}. \quad (5.43)$$

Solving for d , we get the desired resolution relation (2.15). \square

5.3.1.2 Binomial model

In the binomial model we have

$$F_{t,x} = \text{Bin}(t, x)$$

or more explicitly

$$Y_i \stackrel{\text{indep.}}{\sim} \text{Bin}(t, p_i) \quad (5.44)$$

where p_i are defined in Equations (2.13) and (2.14), and the rest of the setup is the same as in the Poisson model. The likelihood ratio statistic is

$$\begin{aligned} \log \left(\frac{\varphi(Y|H_1; Y_i)}{\varphi(Y|H_0; Y_i)} \right) &= \log \left(\frac{\prod_{i=1}^n \binom{t}{Y_i} p_{1i}^{Y_i} (1-p_{1i})^{t-Y_i}}{\prod_{i=1}^n \binom{t}{Y_i} p_{0i}^{Y_i} (1-p_{0i})^{t-Y_i}} \right) \\ &= \sum_{i=1}^n Y_i \log \left(\frac{p_{1i}}{p_{0i}} \right) + \sum_{i=1}^n (t - Y_i) \log \left(\frac{1-p_{1i}}{1-p_{0i}} \right) \\ &= \sum_{i=1}^n Y_i \log \left(\frac{p_{1i}(1-p_{0i})}{p_{0i}(1-p_{1i})} \right) + \sum_{i=1}^n t \log \left(\frac{1-p_{1i}}{1-p_{0i}} \right), \end{aligned}$$

where the last term is deterministic so we set

$$T_{t,n,d}(Y) = \sum_{i=1}^n Y_i \log \left(\frac{p_{1i}(1-p_{0i})}{p_{0i}(1-p_{1i})} \right) \quad (5.45)$$

as our test statistic.

Similarly as before, we define the LRT as in (5.2) and set

$$q_{\alpha,t,n,d}^* := q_{1-\alpha} \sqrt{\mathbb{V}_{H_0} T_{t,n,d}} + \mathbb{E}_{H_0} T_{t,n,d}. \quad (5.46)$$

Here $q_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution.

Theorem 5.3.3 (CLT for binomial LRT). *Assume a psf h satisfies Assumption 2.3.2 and that $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$. Then a CLT holds for $T_{t,n,d}(Y)$ in the hypothesis and alternative as $t, n \rightarrow \infty$ and $d \rightarrow 0$, i.e.*

$$\frac{T_{t,n,d} - \mathbb{E}[T_{t,n,d}]}{\sqrt{\mathbb{V}[T_{t,n,d}]}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

The proof is similar to the Poisson model CLT proof and thus is postponed to Appendix A.3.

The rest of the proof of Theorem 2.3.4 for the binomial model in the CLT regime follows the same lines as the Poisson proof starting with Equation (5.37) and is therefore omitted.

5.3.2 1D Poisson model analysis in the asymptotic equivalence regime

We complete the Theorem 2.3.4 Poisson model proof by showing that it is asymptotically equivalent to the VSG model provided that $t \gg \sqrt{n} \log^8 n$.

We briefly recall the theory of asymptotic equivalence developed by Le Cam (Le Cam, 1986), (Le Cam and Yang, 2000). We mostly follow the presentation of (Grama

and Nussbaum, 2002). In our context we consider a *statistical experiment*—a set

$$\mathcal{E} = (X, \mathcal{X}, \{P_\theta : \theta \in \Theta\}),$$

where (X, \mathcal{X}) is a measurable space with the parameter set $\Theta \subset \mathbb{R}$, a possibly unbounded interval, and P_θ is an absolutely continuous probability measure with respect to some dominating σ -finite measure μ . Consider a second, possibly easier to tackle, experiment $\mathcal{G} = (Y, \mathcal{Y}, \{Q_\theta : \theta \in \Theta\})$ over the same parameter space Θ . Let further (D, \mathcal{D}) be a measurable space of possible decisions. Then the set of Markov kernels $\kappa : (X, \mathcal{X}) \rightarrow (D, \mathcal{D})$ is the set of randomized decision procedures for the experiment \mathcal{E} . We denote it by $\Pi(\mathcal{E})$. We let $\mathcal{L}(D, \mathcal{D})$ to be the set of all loss functions $L : \Theta \times D \rightarrow [0, \infty)$ such that $0 \leq L(\theta, z) \leq 1$ for all $\theta \in \Theta$ and $z \in D$. Given a decision procedure $\kappa \in \Pi(\mathcal{E})$, the true value $\theta \in \Theta$ and a loss function $L \in \mathcal{L}(D, \mathcal{D})$, the risk is

$$R(\mathcal{E}, \kappa, L, \theta) = \int_X \int_D L(\theta, z) \kappa(x, dz) P_\theta(dx).$$

We define the deficiency as

$$\delta(\mathcal{E}, \mathcal{G}) := \sup_{L \in \mathcal{L}(D, \mathcal{D})} \sup_{\kappa_1 \in \Pi(\mathcal{E})} \inf_{\kappa_2 \in \Pi(\mathcal{G})} \sup_{\theta \in \Theta} |R(\mathcal{E}, \kappa_1, L, \theta) - R(\mathcal{G}, \kappa_2, L, \theta)|$$

with the first supremum ranging over all possible decision spaces (D, \mathcal{D}) . Since deficiency is asymmetric, we define the Le Cam (pseudo) distance as

$$\Delta(\mathcal{E}, \mathcal{G}) := \max\{\delta(\mathcal{E}, \mathcal{G}), \delta(\mathcal{G}, \mathcal{E})\}.$$

Definition 5.3.4. Two sequences of statistical experiments \mathcal{E}^n and \mathcal{G}^n , $n \in \mathbb{N}$, are *asymptotically equivalent* if

$$\Delta(\mathcal{E}^n, \mathcal{G}^n) \rightarrow 0.$$

We can summarize the implications of the above definition for our analysis in the following proposition.

Proposition 5.3.5. *Let \mathcal{E}_1^n and \mathcal{E}_2^n , $n \in \mathbb{N}$, be two sequences of statistical experiments that are asymptotically equivalent, and let Ψ_1^n and Ψ_2^n be the corresponding optimal tests. Then we have*

$$\mathbb{E}_{H_0} \Psi_1^n \rightarrow \alpha \quad \text{and} \quad \mathbb{E}_{H_1} \Psi_1^n \rightarrow 1 - \beta \quad \iff \quad \mathbb{E}_{H_0} \Psi_2^n \rightarrow \alpha \quad \text{and} \quad \mathbb{E}_{H_1} \Psi_2^n \rightarrow 1 - \beta,$$

i.e. the type I error of Ψ_1^n converges to α and the type II error to β if and only if the type I error of Ψ_2^n converges to α and type II error to β . Thus, an asymptotic resolution

sequence in the sense of Definition 2.2.2 for the first sequence of experiments will also be an asymptotic resolution for the second sequence.

The above proposition allows us to transfer the VSG result to the Poisson model in the asymptotic equivalence regime:

Corollary 5.3.6. Let $0 < \alpha, \beta < 1/2$ be type I and II errors, respectively. Assume that $\sqrt{n} \log^8 n = o(t)$ and Assumption 2.3.2 are valid. Then Theorem 2.3.4 (a) holds.

Proof. Our VSG model can be viewed as a Gaussian model

$$Y_i \stackrel{\text{indep.}}{\sim} \mathcal{N}\left(2\sqrt{f_n(i/n)}, 1\right)$$

with

$$f_n(x) = t \int_{x-1/n}^x g(y) dy \quad (5.47)$$

for $x \in [1/n, 1]$. According to Example 4.2 of (Grama and Nussbaum, 2002), a sequence of n Poisson observations

$$X_i \stackrel{\text{indep.}}{\sim} \text{Poi}(f(i/n))$$

is asymptotically equivalent to the above Gaussian model with some *fixed function* $f : [0, 1] \rightarrow \mathbb{R}$ provided that f is bounded $c_1 \leq f(x) \leq c_2$ by some absolute constants $c_1, c_2 > 0$ and it is Hölder with exponent $\beta > 1/2$. This result was extended in Theorem 4 of (Ray and Schmidt-Hieber, 2018) to include functions f which are not bounded away from zero: functions $f = f_n$ that may depend on $n \in \mathbb{N}$, satisfy

$$\inf_x f_n(x) \gg n^{-\beta/(\beta+1)} \log^8 n \quad (5.48)$$

and f_n are Hölder with $1/2 < \beta \leq 1$. Thus, we only need to extend our f_n 's to functions on $[0, 1]$ and prove that they satisfy the assumptions of Theorem 4 of (Ray and Schmidt-Hieber, 2018) to complete the proof.

As a first step, extend the image function g to a function on $C^2[-1/n, 1]$ such that

$$n^{-1/2} \log^{8+\delta} n \leq t \int_{-1/n}^0 g(y) dy \leq t, \quad (5.49)$$

for some $\delta > 0$. Then we can extend f_n 's in (5.47) to $f_n : [0, 1] \rightarrow \mathbb{R}$.

We have that $f_n \leq t$ since $\int_0^1 g = 1$ and $g > 0$, and $f_n \in C^3[0, 1]$ since $g \in C^2[-1/n, 1]$. Hence, f_n is Hölder with $\beta = 1$. Due to the psf h being fixed, our testing problem (2.8) and Assumption 2.3.2, for all $x \in [1/n, 1]$ it holds that

$$\int_{x-1/n}^x g(y) dy \geq \frac{\min_{x \in [0, 1]} g(x)}{n}.$$

Due to the continuity of g and compactness of $[0, 1]$, $\min_{x \in [0, 1]} g(x) \geq c$ for some constant $c > 0$. Thus,

$$\inf_{x \in [0, 1]} f_n = \inf_{x \in [0, 1]} t \int_{x-1/n}^x g(y) \, dy \geq \min \left\{ c \frac{t}{n}, n^{-1/2} \log^{8+\delta} n \right\} \gg n^{-\frac{1}{2}} \log^8(n),$$

by (5.49), and hence our assumption $\sqrt{n} \log^8 n = o(t)$ implies (5.48). Therefore, in this case the Poisson model is equivalent to the VSG model for which Theorem 2.3.4 (a) holds by the above proof. \square

CHAPTER 6

Discussion and future work

The present thesis investigated the classical two point resolution problem in a statistical setting. As far as we know, this is the first work providing a mathematically rigorous proof that for even psfs asymptotically the resolution depends linearly on the FWHM for the binomial, Poisson and variance stabilized Gaussian models in any dimension m , subject to the illumination time t and discretization n satisfying $n^m = ct^{1/2+\delta}$ for arbitrary constants $c, \delta > 0$. Moreover, we showed that for the 1D Poisson model the relationship between t and n can be arbitrary, as long as $t, n \rightarrow \infty$ and $d \rightarrow 0$.

Another important finding is that for the homogeneous Gaussian model the resolution asymptotically depends on the FWHM to the power of $5/4$. Furthermore, unlike in the other models, in the higher dimensional HG model the resolution also depends on the FWHM, perpendicular to the line connecting the two psfs in the alternative, to the power of $1/4$. These observations show that either the HG model is too simple for describing resolution, or that these distinctions could be used to determine if the given experimental setup is fundamentally different from the ones usually encountered in microscopy, where the resolution depends linearly on the FWHM.

We have also provided full proofs showing that the most difficult statistical testing problem is when the psfs are put in such a way that the psf under the hypothesis is centered at the center of mass of the two psfs under the alternative, and that the testing problem with the symmetric placement of the psfs under the alternative is the easiest.

Finally, in simulations we have addressed our results' applicability to finite samples. In one, two and three dimensions the simulated resolution seems to be approximated really well by our theory at t and n values that seem plausible for modern experiments.

Just like any other research endeavor, the present one raises many questions that require future work. Here we list some of them:

- Are the Poisson and binomial CLTs 5.3.1 and 5.3.3 also valid for other combinations of t and n ? Due to Remark 5.3.2, the CLTs are only valid if and only if the respective Lindeberg's conditions hold.

- Can the results of (Ray and Schmidt-Hieber, 2018) on the 1D Poisson intensity estimation be extended to higher dimensional Poisson experiments? Would these results be valid at least in the whole complement of the CLT validity domain, thereby proving (2.15) for any relationship between t and n , just like in 1D? Similarly, what are the respective results for the binomial experiments?
- Throughout the thesis we have assumed that the psf is even and time invariant, the locations of the centers of the psfs are known and that the asymmetry parameter q from (2.8) is also known. Experimentally this is not always the case and thus it would be interesting to relax some or all of these assumptions.
- In Section 3.5 we showed that our results do not agree with Helstrom's quantum optical results (Helstrom, 1973) at low number of photons. This is expected, since Helstrom's solution is optimal from the point of view of quantum information theory. It would be interesting to see if it is possible to prove Helstrom's results rigorously, but this might fall into the general problem of putting quantum field theory on rigorous mathematical footing (Jaffe and Witten, 2000), but see e.g. (Yamagata et al., 2013), where the authors define a *quantum* log-likelihood ratio.
- Modern super-resolution imaging is truly remarkable – the researchers are able to measure distances of the order of tens of nanometers very accurately using the so-called DNA origami nanorulers (Raab et al., 2018). We hope that such experiments could confirm our theoretical findings.

CHAPTER A

Other proofs

A.1 An integral approximation

The following lemma is used throughout to prove the main theorems.

Lemma A.1.1. *Let $f : [0, 1]^m \rightarrow \mathbb{R}$ and $g : [0, 1] \rightarrow \mathbb{R}_{>0}$ be two absolutely continuous functions. Then*

$$\sum_{i \in [n]^m} \frac{\left(\int_{B_i} f(x) \, dx \right)^2}{\int_{B_i} g(x) \, dx} \xrightarrow{n \rightarrow \infty} \int_I \frac{f(x)^2}{g(x)} \, dx < \infty,$$

where $B_i = \left[\frac{i_1-1}{n}, \frac{i_1}{n} \right] \times \dots \times \left[\frac{i_m-1}{n}, \frac{i_m}{n} \right]$.

Proof. Note that $f(x)^2/g(x)$ is absolutely continuous, and thus Riemann integrable. Using the mean value theorem we get

$$\begin{aligned} \sum_{i \in [n]^m} \frac{\left(\int_{B_i} f(x) \, dx \right)^2}{\int_{B_i} g(x) \, dx} &= \frac{1}{n^m} \sum_{i \in [n]^m} \frac{f(\xi'_i)^2}{g(\xi_i)} = \frac{1}{n^m} \sum_{i \in [n]^m} \frac{(f(\xi'_i) - f(\xi_i) + f(\xi_i))^2}{g(\xi_i)} \\ &= \frac{1}{n^m} \sum_{i \in [n]^m} \frac{(f(\xi'_i) - f(\xi_i))^2 + 2(f(\xi'_i) - f(\xi_i))f(\xi_i) + f(\xi_i)^2}{g(\xi_i)} \end{aligned}$$

with $\xi_i, \xi'_i \in B_i$. Now by continuity of f it holds for all $i \in [n]^m$ that

$$|f(\xi_i) - f(\xi'_i)| \leq \max_{x \in B_i} f(x) - \min_{x \in B_i} f(x) \rightarrow 0,$$

as $n \rightarrow \infty$. Thus, by Riemann integrability

$$\sum_{i \in [n]^m} \frac{\left(\int_{B_i} f(x) \, dx \right)^2}{\int_{B_i} g(x) \, dx} = \frac{1}{n^m} \sum_{i \in [n]^m} \frac{f(\xi_i)^2}{g(\xi_i)} + o(1) \rightarrow \int_I \frac{f(x)^2}{g(x)} \, dx,$$

as $n \rightarrow \infty$. □

A.2 Which alternative is the most difficult

Proof of Proposition 2.2.1. As usual, we skip indices t and d in the proof.

We first prove the theorem for the **homogeneous Gaussian model** (2.7). Let

$$\lambda := |x_{01} - (qx_{11} + (1-q)x_{21})|, \quad (\text{A.1})$$

i.e. unlike in the previous proofs x_{01} is not necessarily equal to $qx_{11} + (1-q)x_{21}$. Using this notation, (5.11) can be written as

$$\Delta(x - x_0) = \lambda h'(x - x_0) + \frac{h''(x - x_0)}{2} (\lambda^2 + (1-q)qd^2) + O(\text{higher order terms}). \quad (\text{A.2})$$

Thus, μ_n as defined in (5.3) is equal to

$$\begin{aligned} \mu_n &= \frac{t^2}{2} \sum_{i \in [n]^m} \left(\int_i \Delta \right)^2 \\ &= \frac{t^2}{2} \sum_{i \in [n]^m} \left(\lambda \int_i h' + \frac{\int_i h''}{2} (\lambda^2 + (1-q)qd^2) + O(\text{higher order terms}) \right)^2 \\ &= \frac{t^2}{2n} \left(\frac{\lambda^4}{4} \int_I (h'')^2 + \frac{q(1-q)d^2\lambda^2}{2} \int_I (h'')^2 + \frac{q^2(1-q)^2d^4}{4} \int_I (h'')^2 + \lambda^2 \int_I (h')^2 \right. \\ &\quad \left. + (\lambda^2 + (1-q)qd^2) \lambda \int_I h'' h' + O(\text{higher order terms}) \right), \end{aligned}$$

where in the last step we have used Lemma A.1.1. Since h is even, $\int_I h' h'' = 0$. Considering μ_n as a function of λ , we find its minimum at $\lambda = 0 + O(d^2)$. Since by (5.4) under $H_0 : T_n(Y) \sim \mathcal{N}(-\mu_n, 2\mu_n)$ and under $H_1 : T_n(Y) \sim \mathcal{N}(\mu_n, 2\mu_n)$, the testing problem becomes the most difficult when μ_n is minimal. Thus, we see that the case $x_{01} = qx_{11} + (1-q)x_{21}$ is indeed the most difficult to distinguish.

The proof in the **variance stabilized Gaussian model** (2.6) follows the same lines and is therefore omitted.

For **Poisson** and **binomial** models we have two cases to consider. In one dimension, whenever the asymptotic equivalence holds, the proposition follows from the variance stabilized Gaussian model. In contrast, the CLT holds in all dimensions, provided that $n^m = ct^{1/2+\delta}$ for some arbitrary constants $c, \delta > 0$. We go over the main steps of the statistical resolution proof in the CLT regime and show that they still hold. Just like in

the models above, we show that the center of mass placement, i.e. $\lambda = 0$ (A.1), is the most difficult. We only consider the Poisson model; the proof for the binomial model is similar.

The proof of the CLT clearly holds as before, but with more terms in the Taylor expansions. The only tricky step is to make sure that (cf. (5.30))

$$R := \min_{j \in [n]^m} \frac{\sqrt{\sum_{i \in [n]^m} a_i^2 \int_i h}}{|a_j|}$$

is finite. Using (A.2) we have (cf. (5.25))

$$y_i = \frac{\lambda_{1i} - \lambda_{0i}}{\lambda_{0i}} = \frac{\int_i \Delta}{\int_i h} = \frac{\int_i h'}{\int_i h} \lambda + \frac{1}{2} \frac{\int_i h''}{\int_i h} (\lambda^2 + (1-q)q d^2) + o(\lambda^2) + o(d^2), \quad (\text{A.3})$$

$$a_i = \log\left(\frac{\lambda_{1i}}{\lambda_{0i}}\right) = \log(1+y_i) = y_i + O(y_i^2) = \frac{\int_i h'}{\int_i h} \lambda + \frac{(1-q)q d^2}{2} \frac{\int_i h''}{\int_i h} + O(\lambda^2) + o(d^2),$$

$$\begin{aligned} \sqrt{\sum_{i \in [n]^m} a_i^2 \int_i h} &= \sqrt{\sum_{i \in [n]^m} \log(1+y_i)^2 \int_i h} = \sqrt{\sum_{i \in [n]^m} (y_i^2 + O(y_i^3)) \int_i h} \\ &= \left\{ \sum_{i \in [n]^m} \left(\frac{(\int_i h')^2}{\int_i h} \lambda^2 + \frac{\int_i h' \int_i h''}{\int_i h} (1-q)q \lambda d^2 + \frac{(1-q)^2 q^2 (\int_i h'')^2}{4 \int_i h} d^4 \right) \right. \\ &\quad \left. + O(\lambda^3) + o(d^4) + o(\lambda d^2) \right\}^{1/2} \end{aligned}$$

and thus

$$R = O(1),$$

as before. The rest of the CLT proof follows same lines as before (Theorem 5.3.1) and is therefore omitted.

The calculation of the statistical resolution in Poisson and binomial models essentially boils down to Equation (5.42) stated here once more for convenience

$$\begin{aligned} q_{1-\alpha} \sqrt{\mathbb{V}_{H_0} T_n} - q_\beta \sqrt{\mathbb{V}_{H_1} T_n} &= \mathbb{E}_{H_1} T_n - \mathbb{E}_{H_0} T_n + o(1) \iff \\ \sqrt{\sum_{i \in [n]^m} \lambda_{0i} (y_i^2 + O(y_i^3))} &= q_{1-\alpha} - q_\beta = q_{1-\beta} - q_\alpha. \end{aligned}$$

Thus, using the general expression for y_i (A.3) the above is equivalent to

$$t \left(\lambda^2 \int_I \frac{(h')^2}{h} + \frac{(1-q)^2 q^2 d^4}{4} \int_I \frac{(h'')^2}{h} + \lambda \left((1-q) q d^2 + \lambda^2 \right) \int_I \frac{h' h''}{h} + o(\lambda^3) + o(d^2 \lambda) + o(d^4) \right) = (q_{1-\beta} - q_\alpha)^2.$$

If the psf h is even, then h' is odd and h'' is even. Hence,

$$\int_I \frac{h' h''}{h} = 0$$

and thus the left hand side considered as a function of λ attains its minimum at $\lambda = 0$. This implies that for given values of α , t and d , the power $1 - \beta$ is the smallest when $\lambda = 0$, i.e. $x_0 = qx_1 + (1 - q)x_2$, is the most difficult alternative. \square

A.3 Proof of the binomial CLT

Proof of Theorem 5.3.3. For ease of readability, we skip indices t and d in what follows. Just like in the Poisson case, we will give full details for the symmetric $q = 1/2$ problem and only outline the general $q \in (0, 1)$ proof at the end.

Recall that by (5.45) the test statistic in the binomial model is

$$T_n(Y) = \sum_{i \in [n]^m} \left(\log \left(\frac{p_{1i}}{p_{0i}} \right) + \log \left(\frac{1 - p_{0i}}{1 - p_{1i}} \right) \right) Y_i = \sum_{i \in [n]^m} (a_i - \bar{a}_i) Y_i$$

where as before

$$Y_i \stackrel{\text{indep.}}{\sim} \text{Bin}(t, p_i).$$

We use the Taylor series $\log(1 + y_i) = \sum_{k=0}^2 (-1)^k / (k + 1) y_i^{k+1} + o(y_i^3)$ for each term separately. Because the weight a_i is the same as in the Poisson case (5.24), the expression for y_i is also the same (5.25) and $a_i \asymp y_i \sim d^2$. For the second weight \bar{a}_i we use the log expansion with

$$y_i = -\frac{\int_i h''}{8(1 - \int_i h)} d^2 - \frac{\int_i h''''}{384(1 - \int_i h)} d^4 + o\left(\frac{d^4}{n^m}\right). \quad (\text{A.4})$$

Thus, asymptotically $a_i \gg \bar{a}_i$ as $n \rightarrow \infty$ and w.l.o.g. we can consider $\tilde{T}_n(Y) = \sum_{i \in [n]^m} a_i Y_i = \sum_{i \in [n]^m} X_{ni}$ with $X_{ni} := a_i Y_i$.

This leads to

$$\begin{aligned}
\mu_{ni} &= \mathbb{E}_{H_0} X_{ni} = a_i \mathbb{E}_{H_0} Y_i = a_i t p_{0i} = a_i t \int_i h = \log(1 + y_i) t \int_i h \\
&= \frac{td^2}{8} \int_i h'' + td^4 \left(-\frac{1}{128} \frac{(\int_i h'')^2}{\int_i h} + \frac{1}{384} \int_i h'''' \right) + O\left(\frac{td^6}{n^m}\right), \\
\nu_n &= \sum_{i \in [n]^m} \mu_{ni} = \mathbb{E}_{H_0} \tilde{T}_n(Y) = t \sum_{i \in [n]^m} a_i \int_i h \\
&= \frac{td^2}{8} \int_I h'' + td^4 \left(-\frac{1}{128} \rho_n + \frac{1}{384} \int_I h'''' \right) + O(td^6), \\
\sigma_{ni}^2 &= \mathbb{V}_{H_0} X_{ni} = a_i^2 \mathbb{V}_{H_0} Y_i = a_i^2 t p_{0i} (1 - p_{0i}) = \frac{td^4}{64} \frac{(\int_i h'')^2}{\int_i h} + O\left(\frac{td^4}{n^{2m}}\right) + O\left(\frac{td^6}{n^m}\right), \\
\tau_n^2 &= \sum_{i \in [n]^m} \sigma_{ni}^2 = \mathbb{V}_{H_0} \tilde{T}_n(Y) = t \sum_{i \in [n]^m} a_i^2 \int_i h \left(1 - \int_i h\right) \\
&= \frac{td^4}{64} \rho_n + O\left(\frac{td^4}{n^m}\right) + O(td^6),
\end{aligned}$$

where ρ_n was defined in (5.26). Applying Lemma A.1.1 with $f(x) = h''(x - x_0)$ and $g(x) = h(x - x_0)$ we get

$$\rho_n = \int_I \frac{h''(x - x_0)^2}{h(x - x_0)} dx + o(1)$$

and

$$\tau_n^2 = \frac{td^4}{64} \int_I \frac{h''(x - x_0)^2}{h(x - x_0)} dx + o(td^4). \quad (\text{A.5})$$

Note that this also implies that

$$\sum_{i \in [n]^m} a_i^2 = O(n^m d^4). \quad (\text{A.6})$$

Let

$$\mathcal{D}_{\varepsilon, i} = \left\{ l \in \mathbb{N}_0 \mid 0 \leq l \leq t, |a_i l - a_i t p_{0i}| > \varepsilon \sqrt{\sum_{i \in [n]^m} a_i^2 t p_{0i} (1 - p_{0i})} \right\}.$$

Just as in the Poisson proof (Remark 5.3.2), Lindeberg's condition is necessary here as well. Let

$$q = \min_{x \in I} h(x - x_0) \quad \text{and} \quad \bar{q} = \max_{x \in I} h(x - x_0). \quad (\text{A.7})$$

We have

$$\begin{aligned}
L_n(\varepsilon) &= \frac{1}{\tau_n^2} \sum_{i \in [n]^m} \sum_{\substack{k \in a_i \{0,1,\dots,t\} \\ |k - \mu_{ni}| > \varepsilon \tau_n}} (k - \mu_{ni})^2 \mathbb{P}_{H_0}(X_{ni} = k) \\
&= \frac{1}{\tau_n^2} \sum_{i \in [n]^m} a_i^2 \sum_{l \in \mathcal{D}_{\varepsilon,i}} (l - tp_{0i})^2 \binom{t}{l} \left(\int_i h \right)^l \left(1 - \int_i h \right)^{t-l} \\
&\leq \frac{t^2}{\tau_n^2} \sum_{i \in [n]^m} a_i^2 \sum_{l \in \mathcal{D}_{\varepsilon,i}} \left(\frac{1}{t} - \frac{\int_i h}{l} \right)^2 \binom{t}{l} l^2 \left(\frac{\bar{q}}{n^m} \right)^l \left(1 - \frac{q}{n^m} \right)^{t-l} \\
&\leq C n^m t \sum_{l \in \mathcal{D}_{\varepsilon,i}} \binom{t}{l} l^2 \left(\frac{\bar{q}}{n^m} \right)^l \left(1 - \frac{q}{n^m} \right)^{t-l},
\end{aligned} \tag{A.8}$$

where $C > 0$ is some constant and we have used (A.5), (A.6) and that

$$\left(\frac{1}{t} - \frac{\int_i h}{l} \right)^2 \leq \frac{1}{t^2} - 2 \frac{q}{tl n^m} + \frac{\bar{q}^2}{l^2 n^{2m}} = o(1),$$

for all $i \in [n]^m$, independent of n . Using (A.6), it holds that

$$R := \min_{j \in [n]^m} \frac{\sqrt{\sum_{i \in [n]^m} a_i^2 t p_{0i} (1 - p_{0i})}}{|a_j|} = O(1).$$

Hence, the domain $\mathcal{D}_{\varepsilon,i}$ is a subset of the set containing all $l \in \mathbb{N}_0$ such that

$$\begin{cases} \varepsilon \sqrt{t} R + \frac{t}{n^m} q < l \leq t \\ 0 \leq l < -\varepsilon \sqrt{t} R + \frac{t}{n^m} \bar{q}. \end{cases} \tag{A.9}$$

Let $c, \delta > 0$. For $n^m = ct^{1/2+\delta}$ and t sufficiently large there are no l 's satisfying the second inequality. Thus,

$$L_n(\varepsilon) \leq C n^m t \sum_{l=\lceil \sqrt{t}(\varepsilon R + t^{-\delta} q) \rceil}^t \binom{t}{l} l^2 \left(\frac{\bar{q}}{n^m} \right)^l \left(1 - \frac{q}{n^m} \right)^{t-l}. \tag{A.10}$$

W.l.o.g we set $c = 1$ in what follows. Consider

$$\begin{aligned}
&\sum_{l=a}^t \binom{t}{l} \left(\frac{\bar{q}}{n^m} \right)^l \left(1 - \frac{q}{n^m} \right)^{t-l} \\
&= \binom{t}{a} \left(\frac{\bar{q}}{n^m} \right)^a \left(1 - \frac{q}{n^m} \right)^{t-a} \left(1 + \sum_{l=a+1}^t \frac{\binom{t}{l} \left(\frac{\bar{q}}{n^m} \right)^l \left(1 - \frac{q}{n^m} \right)^{t-l}}{\binom{t}{a} \left(\frac{\bar{q}}{n^m} \right)^a \left(1 - \frac{q}{n^m} \right)^{t-a}} \right)
\end{aligned} \tag{A.11}$$

$$\begin{aligned}
&= \binom{t}{a} \left(\frac{\bar{q}}{n^m}\right)^a \left(1 - \frac{q}{n^m}\right)^{t-a} \left(1 + \sum_{l=a+1}^t \frac{a!(t-a)!}{l!(t-l)!} \left(\frac{\bar{q}}{n^m - q}\right)^{l-a}\right) \\
&= \binom{t}{a} \left(\frac{\bar{q}}{n^m}\right)^a \left(1 - \frac{q}{n^m}\right)^{t-a} \\
&\quad \left\{ 1 + \frac{t-a}{a+1} \cdot \frac{\bar{q}}{n^m - q} + \frac{(t-a)(t-a-1)}{(a+2)(a+1)} \left(\frac{\bar{q}}{n^m - q}\right)^2 + \dots \right. \\
&\quad \left. + \frac{a!(t-a)!}{t!} \cdot \left(\frac{\bar{q}}{n^m - q}\right)^{t-a} \right\}.
\end{aligned}$$

Furthermore, it holds

$$\left\lceil \sqrt{t} (\varepsilon R + t^{-\delta} q) \right\rceil \sim \sqrt{t}.$$

Hence, we can set $a = \sqrt{t}$ and using Stirling's approximation

$$\log k! = k \log k - k + O(\log k)$$

we have

$$\begin{aligned}
\binom{t}{a} &= \frac{t!}{a!(t-a)!} = \frac{\exp(\log t!)}{\exp(\log a!) \exp(\log(t-a)!)} \\
&= \frac{\exp(t \log t - t + O(\log t))}{\exp(a \log a - a + O(\log a)) \exp((t-a) \log(t-a) - (t-a) + O(\log(t-a)))} \\
&= \exp(t \log t + O(\log t) - a \log a + O(\log a) - (t-a) \log(t-a) + O(\log(t-a))) \\
&= \exp\left(t \log t - \sqrt{t} \log \sqrt{t} - (t - \sqrt{t}) \log(t - \sqrt{t}) + O(\log t)\right) \\
&= \exp\left(\log \frac{t^t}{t^{\sqrt{t}/2} (t - \sqrt{t})^{t - \sqrt{t}}} + O(\log t)\right) \sim \exp\left(\log \frac{t^{t - \sqrt{t}/2}}{t^{t - \sqrt{t}}}\right) \\
&\sim t^{\sqrt{t}/2}.
\end{aligned}$$

Moreover, it holds

$$\begin{aligned}
\left(\frac{q}{n^m}\right)^a &= O\left(\frac{q}{t^{1/2+\delta}}\right)^{\sqrt{t}} = O\left(t^{-\sqrt{t}/2 - \sqrt{t}\delta}\right) \\
\left(1 - \frac{q}{n^m}\right)^t &= \left(1 - \frac{q}{t^{1/2+\delta}}\right)^t = \left(1 - \frac{q\sqrt{t}t^{-\delta}}{t}\right)^t = O\left(\exp\left(-q\frac{\sqrt{t}}{t^\delta}\right)\right) \\
\left(1 - \frac{q}{n^m}\right)^a &= \left(1 - \frac{q}{t^{1/2+\delta}}\right)^{\lceil \sqrt{t}(\varepsilon + c/t^\delta) \rceil} = O\left(\left(1 - \frac{qt^{-\delta}\varepsilon}{\varepsilon t^{1/2}}\right)^{\lceil \sqrt{t}\varepsilon \rceil}\right) \\
&= O\left(\exp(-qt^{-\delta}\varepsilon)\right) \rightarrow O(1).
\end{aligned}$$

Each term in the sum of (A.11) is of the form

$$\left(\frac{t - \sqrt{t}}{\sqrt{t}} \frac{q}{t^{1/2+\delta} - q} \right)^k = O(t^{-k\delta})$$

and hence the whole sum is of the order

$$1 + \sum_{k=1}^t t^{-\delta k} = \frac{1}{1 - t^{-\delta}} = O(1).$$

All in all, we get

$$L_n(\varepsilon) = n^m t O(t^{-\sqrt{t}\delta}) = O(t^{-\sqrt{t}\delta + 3/2 + \delta}) \rightarrow 0.$$

Under the alternative H_1 we have

$$\begin{aligned} \mu_{ni} &= \mathbb{E}_{H_1} X_{ni} = a_i t p_{1i} = a_i (1 + y_i) t p_{0i} \\ &= \frac{td^2}{8} \int_i h'' + td^4 \left(\frac{1}{128} \frac{(\int_i h'')^2}{\int_i h} + \frac{1}{384} \int_i h'''' \right) + O\left(\frac{td^6}{n^m}\right), \\ \nu_n &= \sum_{i \in [n]^m} \mu_{ni} = \mathbb{E}_{H_1} \tilde{T}_n(Y) = \sum_{i \in [n]^m} a_i (1 + y_i) t p_{0i} \\ &= \frac{td^2}{8} \int_I h'' + td^4 \left(\frac{1}{128} \rho_n + \frac{1}{384} \int_I h'''' \right) + O(td^6), \\ \sigma_{ni}^2 &= \mathbb{V}_{H_1} X_{ni} = a_i^2 t p_{1i} (1 - p_{1i}) = a_i^2 (1 + y_i) t p_{0i} (1 - p_{1i}) \\ &= \frac{td^4}{64} \frac{(\int_i h'')^2}{\int_i h} + O\left(\frac{td^4}{n^{2m}}\right) + O\left(\frac{td^6}{n^m}\right), \\ \tau_n^2 &= \sum_{i \in [n]^m} \sigma_{ni}^2 = \mathbb{V}_{H_1} \tilde{T}_n(Y) = \sum_{i \in [n]^m} a_i^2 (1 + y_i) t p_{0i} (1 - p_{1i}) \\ &= \frac{td^4}{64} \rho_n + O\left(\frac{td^4}{n^m}\right) + O(td^6). \end{aligned}$$

hence Lindeberg's condition holds just like for H_0 .

For the general $q \in (0, 1)$ case, note that we write the LR test statistic as

$$T_n(Y) = \sum_{i \in [n]^m} \left(\log \left(\frac{p_{1i}}{p_{0i}} \right) - \log \left(\frac{1 - p_{1i}}{1 - p_{0i}} \right) \right) \sum_{i \in [n]^m} (a_i - \bar{a}_i) Y_i,$$

where $Y_i \stackrel{\text{indep.}}{\sim} \text{Bin}(t, p_i)$. Now for the first coefficient we use the Taylor series $a_i =$

$\log(1 + y_i)$ with y_i given in (5.34). For the second, we get

$$y_i = -\frac{q(1-q)}{2} \frac{\int_i h''}{1 - \int_i h} d^2 + o\left(\frac{d^2}{n^m}\right)$$

which is a generalization of (A.4). Therefore, just like in the $q = 1/2$ proof, $a_i \gg \bar{a}_i$ and thus it suffices to consider the test statistic

$$\tilde{T}_n(Y) = \sum_{i \in [n]^m} a_i Y_i.$$

Clearly, the CLTs under H_0 and H_1 still hold, just like in the symmetrically placed $q = 1/2$ case. \square

Bibliography

- Abbe, E. (1873). Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. *Archiv für mikroskopische Anatomie*, 9(1):413–418.
- Acuña, C. O. and Horowitz, J. (1997). A statistical approach to the resolution of point sources. *Journal of Applied Statistics*, 24(4):421–436.
- Airy, G. (1835). On the diffraction of an object-glass with circular aperture. *Transactions of the Cambridge Philosophical Society*, 5:283–291.
- Ash, E. A. and Nicholls, G. (1972). Super-resolution aperture scanning microscope. *Nature*, 237:510–512.
- Aspelmeier, T., Egner, A., and Munk, A. (2015). Modern statistical challenges in high-resolution fluorescence microscopy. *Annual Review of Statistics and Its Application*, 2(1):163–202.
- Banterle, N., Bui, K. H., Lemke, E. A., and Beck, M. (2013). Fourier ring correlation as a resolution criterion for super-resolution microscopy. *Journal of Structural Biology*, 183(3):363–367.
- Bertero, M., Boccacci, P., Desider, G., and Vicidomini, G. (2009). Image deblurring with Poisson data: from cells to galaxies. *Inverse Problems*, 25(12):123006.
- Betzig, E., Davidson, M. W., Patterson, G. H., Lippincott-Schwartz, J., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., and Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313:1642–1645.
- Billingsley, P. (1986). *Probability and Measure*. John Wiley and Sons, 2nd edition.
- Born, M. and Wolf, E. (1999). *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Cambridge University Press, 7th edition.

- Candès, E. J. and Fernandez-Granda, C. (2013). Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254.
- Candès, E. J. and Fernandez-Granda, C. (2014). Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67:906–956.
- Courjon, D. (2003). *Near-Field Microscopy and Near-Field Optics*. Imperial College Press, London.
- Cremer, C. and Masters, B. R. (2013). Resolution enhancement techniques in microscopy. *European Physical Journal H*, 38(3):281–344.
- den Dekker, A. J. and van den Bos, A. (1997). Resolution: a survey. *Journal of the Optical Society of America A*, 14(3):547–557.
- Donoho, D. L. (1992). Superresolution via sparsity constraints. *SIAM Journal on Mathematical Analysis*, 23(5):1309–1331.
- Du, C. and Kou, S. C. (2020). Statistical methodology in single-molecule experiments. *Statistical Science*, 35(1):75–91.
- Dürig, U., Pohl, D. W., and Rohner, F. (1986). Near-field optical-scanning microscopy. *Journal of Applied Physics*, 59(10):3318–3327.
- Egner, A., Geisler, C., and Siegmund, R. (2020). Basic knowledge in STED nanoscopy. In Salditt, T., Egner, A., and Luke, R., editors, *Nanoscale Photonic Imaging*, pages 3–34. Springer International Publishing.
- Egner, A., Geisler, C., Von Middendorff, C., Bock, H., Wenzel, D., Medda, R., Andresen, M., Stiel, A. C., Jakobs, S., Eggeling, C., Schönle, A., and Hell, S. W. (2007). Fluorescence nanoscopy in whole cells by asynchronous localization of photoswitching emitters. *Biophysical Journal*, 93(9):3285–3290.
- Ehrenberg, M. (2014). Super-resolved fluorescence microscopy. *Scientific Background on Nobel Prize in Chemistry*.
- Fernandez-Granda, C. (2015). Super-resolution of point sources via convex programming. *2015 IEEE 6th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, CAMSAP 2015*, pages 41–44.
- Goodman, J. W. (1985). *Statistical Optics*. Wiley-Interscience, New York.
- Grama, I. and Nussbaum, M. (2002). Asymptotic equivalence for nonparametric regression. *Mathematical Methods of Statistics*, 11(1):1–36.

- Harris, J. L. (1964). Resolving power and decision theory. *Journal of the Optical Society of America*, 54(5):606–611.
- Heilemann, M., Van De Linde, S., Schüttpelz, M., Kasper, R., Seefeldt, B., Mukherjee, A., Tinnefeld, P., and Sauer, M. (2008). Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. *Angewandte Chemie - International Edition*, 47(33):6172–6176.
- Heintzmann, R. and Ficz, G. (2013). Breaking the resolution limit in light microscopy. *Methods in Cell Biology*, 114(4):525–544.
- Hell, S. W. (2007). Far-field optical nanoscopy. *Science*, 316(5828):1153–1158.
- Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782.
- Helstrom, C. (1964). The detection and resolution of optical signals. *IEEE Transactions on Information Theory*, 10:275–287.
- Helstrom, C. (1965). Correction to ‘The detection and resolution of optical signals’. *IEEE Transactions on Information Theory*, 11(1):125.
- Helstrom, C. W. (1968). *Statistical Theory of Signal Detection*. Pergamon.
- Helstrom, C. W. (1973). Resolution of point sources of light as analyzed by quantum detection theory. *IEEE Transactions on Information Theory*, 19(4):389–398.
- Helstrom, C. W. (1976). *Quantum Detection and Estimation Theory*. Academic Press.
- Hess, S. T., Girirajan, T. P., and Mason, M. D. (2006). Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophysical Journal*, 91(11):4258–4272.
- Hohage, T. and Werner, F. (2016). Inverse problems with Poisson data: statistical regularization theory, applications and algorithms. *Inverse Problems*, 32(9):093001.
- Houston, W. V. (1927). A compound interferometer for fine structure work. *Physical Review*, 29(3):478–484.
- Huang, B., Bates, M., and Zhuang, X. (2009). Super-resolution fluorescence microscopy. *Annual Review of Biochemistry*, 78(1):993–1016.

- Jaffe, A. and Witten, E. (2000). Quantum Yang–Mills theory. The Millenium Prize Problems, Clay Mathematics Institute.
- Klar, T. A., Jakobs, S., Dyba, M., Egner, A., and Hell, S. W. (2000). Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8206–8210.
- Kulaitis, G., Munk, A., and Werner, F. (2020). What is resolution? A statistical minimax testing perspective on super-resolution microscopy. *Submitted*. arXiv:2005.07450.
- Le Cam, L. (1986). *Asymptotic Methods in Statistical Decision Theory*. Springer Series in Statistics. Springer, New York.
- Le Cam, L. and Yang, G. L. Y. (2000). *Asymptotics in Statistics*. Springer Series in Statistics. Springer, New York.
- Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics. Springer, New York, 3rd edition.
- Leonhardt, U. (2010). *Essential Quantum Optics: From Quantum Measurements to Black Holes*. Cambridge University Press.
- Lu, X.-M., Krovi, H., Nair, R., Guha, S., and Shapiro, J. H. (2018). Quantum-optimal detection of one-versus-two incoherent optical sources with arbitrary separation. *npj Quantum Information*, 4(1).
- Middleton, D. (1953). Statistical criteria for the detection of pulsed carriers in noise. I. *Journal of Applied Physics*, 24(4):371–378.
- Milanfar, P. and Shakouri, A. (2002). A statistical analysis of diffraction-limited imaging. In *IEEE International Conference on Image Processing*, pages 864–867.
- Minsky, M. (1961). Microscopy Apparatus. US Patent 3013467, filed 7 November 1957, granted 19 December 1961.
- Morgenshtern, V. I. and Candès, E. J. (2016). Super-resolution of positive sources: the discrete setup. *SIAM Journal on Imaging Sciences*, 9(1):412–444.
- Munk, A., Staudt, T., and Werner, F. (2020). Statistical foundations of nanoscale photonic imaging. In Salditt, T., Egner, A., and Luke, R., editors, *Nanoscale Photonic Imaging*, pages 125–143. Springer International Publishing.

- Nair, R. and Tsang, M. (2016). Far-field superresolution of thermal electromagnetic sources at the quantum limit. *Physical Review Letters*, 117(19):190801.
- Orfanidis, S. J. (2016). *Electromagnetic Waves and Antennas*. Rutgers University. Available at <http://eceweb1.rutgers.edu/~orfanidi/ewa/>.
- Oshikane, Y., Kataoka, T., Okuda, M., Hara, S., Inoue, H., and Nakano, M. (2007). Observation of nanostructure by scanning near-field optical microscope with small sphere probe. *Science and Technology of Advanced Materials*, 8(3):181–185.
- Pawley, J. B. (2006). *Handbook of Biological Confocal Microscopy*. Springer, New York, 3rd edition.
- Petrov, V. (1975). *Sums of Independent Random Variables*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 2. Folge. Springer-Verlag Berlin Heidelberg.
- Raab, M., Jusuk, I., Molle, J., Buhr, E., Bodermann, B., Bergmann, D., Bosse, H., and Tinnefeld, P. (2018). Using DNA origami nanorulers as traceable distance measurement standards and nanoscopic benchmark structures. *Scientific Reports*, 8(1):5–13.
- Ray, K. and Schmidt-Hieber, J. (2018). The Le Cam distance between density estimation, Poisson processes and Gaussian white noise. *Mathematical Statistics and Learning*, 1(2):101–170.
- Reiffen, B. and Sherman, H. (1963). An optimum demodulator for Poisson processes: photon source detectors. *IEEE Transactions on Information Theory*, 51(10):1316–1320.
- Reuss, M., Engelhardt, J., and Hell, S. W. (2010). Birefringent device converts a standard scanning microscope into a sted microscope that also maps molecular orientation. *Opt. Express*, 18(2):1049–1058.
- Rittweger, E., Han, K. Y., Irvine, S. E., Eggeling, C., and Hell, S. W. (2009). STED microscopy reveals crystal colour centres with nanometric resolution. *Nature Photonics*, 3:144–147.
- Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, 3(10):793–795.
- Schroeder, D. (1987). *Astronomical Optics*. Academic Press.

- Shahram, M. (2005). *Statistical and information-theoretic analysis of resolution in imaging and array processing*. PhD, University of California, Santa Cruz.
- Shahram, M. and Milanfar, P. (2004). Imaging below the diffraction limit: a statistical analysis. *IEEE Transactions on Image Processing*, 13(5):677–689.
- Shahram, M. and Milanfar, P. (2006). Statistical and information-theoretic analysis of resolution in imaging. *IEEE Transactions on Information Theory*, 52(8):3411–3437.
- Staudt, T., Aspelmeier, T., Laitenberger, O., Geisler, C., Egner, A., and Munk, A. (2020). Statistical molecule counting in super-resolution fluorescence microscopy: towards quantitative nanoscopy. *Statistical Science*, 35(1):92–111.
- Strutt, J. W. (1879). XXXI. Investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274.
- Tenne, R., Rossman, U., Rephael, B., Israel, Y., Krupinski-Ptaszek, A., Lapkiewicz, R., Silberberg, Y., and Oron, D. (2019). Super-resolution enhancement by quantum image scanning microscopy. *Nature Photonics*, 13(2):116–122.
- Terebizh, V. (1995). Image restoration with minimum a priori information. *Uspekhi Fizicheskikh Nauk*, 165(2):143–176.
- Tham, W.-K., Ferretti, H., and Steinberg, A. M. (2017). Beating Rayleigh’s curse by imaging using phase information. *Physical Review Letters*, 118:070801.
- Thompson, R. E., Larson, D. R., and Webb, W. W. (2002). Precise nanometer localization analysis for individual fluorescent probes. *Biophysical Journal*, 82(5):2775–2783.
- Tsang, M., Nair, R., and Lu, X.-M. (2016a). Quantum information for semiclassical optics. In *Proceedings of SPIE 10029, Quantum and Nonlinear Optics IV*, page 1002903.
- Tsang, M., Nair, R., and Lu, X. M. (2016b). Quantum theory of superresolution for two incoherent optical point sources. *Physical Review X*, 6(3):031033.
- van den Bos, A. and den Dekker, A. J. (2001). Resolution reconsidered—conventional approaches and an alternative. In Hawkes, P. W., editor, *Advances in Imaging and Electron Physics*, volume 117. Elsevier.
- von Diezmann, A., Shechtman, Y., and Moerner, W. E. (2017). Three-dimensional localization of single molecules for super-resolution imaging and single-particle tracking. *Chemical Reviews*, 117(11):7244–7275.

Westphal, V. and Hell, S. W. (2005). Nanoscale resolution in the focal plane of an optical microscope. *Physical Review Letters*, 94(14):143903.

Yamagata, K., Fujiwara, A., and Gill, R. D. (2013). Quantum local asymptotic normality based on a new quantum likelihood ratio. *The Annals of Statistics*, 41(4):2197–2217.

Curriculum Vitae

Gytis Kulaitis

Hermann-Rein-Str. 2B
37075 Göttingen
Germany

gytis.kulaitis@mathematik.uni-goettingen.de

Education

- Since 2016 **PhD Student**, *Georg-August-Universität Göttingen*, Supervisors: Prof. Axel Munk and Prof. Tatyana Krivobokova.
- 2016 **M.Sc. in Theoretical and Mathematical Physics**, *LMU and TUM*, Thesis title: *Comparing Different Mathematical Definitions of 2D CFT*, Supervisor: Prof. Martin Schottenloher.
- 2013 **B.Sc. in Mathematics and Physics**, *University of St Andrews*, Thesis title: *Chaos?*, Supervisor: Dr. Mike Todd.
- 2009 **Brandos atestatas**, *Žvėryno gimnazija*, Vilnius, Lithuania.

Publications and Preprints

- [1] **G. Kulaitis**, A. Munk and F. Werner. What is resolution? A statistical minimax testing perspective on super-resolution microscopy. *Submitted*. arXiv:2005.07450.

Conference Talks

- 2019 **European Meeting of Statisticians**, Palermo, Italy.
- 2018 **Meeting of Young Mathematicians of Lithuania**, Vilnius, Lithuania.
- 2018 **International Vilnius Conference on Probability Theory and Mathematical Statistics**, Vilnius, Lithuania.