Topics in Statistical Minimax Hypothesis Testing

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von MSc Maurilio Gutzeit

geb. am 20. Mai 1989 in Oranienburg

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Alexandra Carpentier

Jun.-Prof. Dr. Mathias Trabs

eingereicht am 8. Januar 2019

Verteidigung am 29. März 2019

Zusammenfassung

In dieser Dissertation präsentieren wir neue theoretische Ergebnisse im Gebiet der statistischen minimax Hypothesentests in sehr unterschiedlichen Situationen. Speziell leiten wir *nicht-asymptotische minimax "separation rates"* (Abstandsraten) für Testprobleme her, bei denen beide *Hypothesen* zusammengesetzt sind.

Im einführenden Kapitel beschreiben wir die allgemeine Situation und Idee für minimax Tests und besprechen relevante Literatur. Zudem werden für das Thema typische Ideen und Phänomene anhand eines speziellen *"signal-detection"*-Problems illustriert.

Danach gehen wir auf drei verschiedene Szenarien mit zusammengesetzten Hypothesen ein:
Zunächst wird das Testproblem untersucht, ob der Erwartungsvektor eines d-dimensionalen Gauss'schen Vektors in einer konvexen Menge $\mathcal{C} \subseteq \mathbb{R}^d$ liegt. Dieses Problem im minimax-Sinne mit kleinen *Typ-I-* und *Typ-II-Fehler-Wahrscheinlichkeiten* zu lösen, erfordert im Allgemeinen einen positiven Abstand zwischen der *Null-* und *Alternativhypothese* und wir sind am kleinsten Euklidischen Abstand interessiert, sodass ein Test mit der gewünschten Leistung existiert. Im Laufe des Kapitels entwickeln wir einen Einblick darüber, inwiefern der minimax-optimale Abstand von der Form von $\mathcal{C}$ abhängt.
Danach beschäftigen wir uns mit dem Testproblem, ob zwei zufällige Graphen auf der gleichen Verteilung beruhen. Speziell beobachten wir uiv Realisierungen zweier (unterschiedlicher) *inhomogener Erdős-Renyi-Graphen* mit Parametermatrizen P und Q auf einer gemeinsamen Knotenmenge und sind am kleinsten Abstand zwischen P und Q interessiert, sodass ein Test die Verteilungen mit kleinen Fehlerwahrscheinlichkeiten unterscheiden kann. Wir zeigen, dass die minimax-optimale Abstandsrate – und sogar die grundsätzliche Lösbarkeit des Problems – stark vom gewählten Abstandsmaß zwischen P und Q abhängt.
Das letzte Kapitel beschäftigt sich damit, den Grad der Regularität (im *Sobolev*-Sinne) einer Funktion f basierend auf einer verrauschten Beobachtung von f zu testen. Speziell nehmen wir an, dass $f \in B_t(R)$ (ein Sobolev-Ball mit Regularität $t > 0$ und Radius $R > 0$) gilt und testen, ob sogar $f \in B_s(R)$ für ein $s > t$ gilt. Nun untersuchen wir den kleinsten Abstand im L_2-Sinne zwischen $B_s(R)$ und f, sodass ein Test diesen mit kleinen Fehlerwahrscheinlichkeiten erkennen kann. Überraschenderweise hängt die minimax-optimale Abstandsrate nicht von s ab.

Abstract

In this thesis we present new theoretical results on statistical minimax hypothesis testing in very different settings. More precisely, we derive *non-asymptotic minimax separation rates* for *composite-composite testing problems*.

In the introductory chapter, we describe the general setting and idea of minimax testing and provide a literature review on the subject. Furthermore, we illustrate typical ideas and phenomena in the field through a detailed discussion of a specific *signal-detection* problem.

After that, we elaborate on three different composite-composite scenarios:
Firstly, the problem of testing if the mean of a $d-$dimensional Gaussian vector belongs to a convex set $\mathcal{C} \subseteq \mathbb{R}^d$ is considered. Solving this problem in a minimax sense with small *type-I* and *type-II error probabilities* generally requires some positive separation between the *null and alternative hypotheses* and we aim at finding the smallest separation in a Euclidean sense such that a test with the required performance exists. During this chapter we gain much insight into how the minimax-optimal separation rate depends on the shape of $\mathcal{C}$.
After that, we study the problem of testing if two random graphs have the same underlying distribution. More precisely, based on iid samples from two (different) *Erdős-Renyi-graph* distributions on a common vertex set with parameter matrices P and Q, respectively, we are interested in the smallest separation between P and Q such that a test can distinguish the distributions with small error probabilities. It turns out that the corresponding minimax-optimal separation rate – and even the feasibility of the problem – depends heavily on the chosen measure of distance between P and Q.
The last chapter deals with testing the degree of smoothness (in a *Sobolev* sense) of a function f based on a noisy observation of f. To be more precise, we assume that $f \in B_t(R)$ (Sobolev-ball with smoothness $t > 0$ and radius $R > 0$) and consider testing if indeed $f \in B_s(R)$ for some $s > t$. Now, we examine the smallest separation in an L_2-sense between $B_s(R)$ and f such that a test can detect it with small error probabilities. Surprisingly, it turns out that the corresponding minimax-optimal separation rate does not depend on s.

Acknowledgements

I have been very lucky to spend the past few years as a PhD student with people who were not only very supportive regarding my mathematical work but also great friends. I would like to take this opportunity to mention a few people I am going to miss dearly.

First of all, I want to thank my supervisor Alexandra Carpentier; she did an amazing job in every respect: On top of being a brilliant, resourceful mathematician with uncanny instincts she is an incredibly likeable person who always has a sympathetic ear. Her guidance - professional and personal - was invaluable. I have the honour to be her first student to finish their PhD and there is no doubt in my mind that many more happy, successful candidates will follow.

Next I would like to thank Sylvie Roelly and Gilles Blanchard as I would not have started to work with Alexandra without them in the first place. Also, the results of our pleasant collaboration with Gilles still fascinate and motivate me.
In Potsdam, during the first half of my PhD studies, I very much enjoyed the company of Franziska, Nicole, Oleksandr and Andrea with whom I share a number of happy memories (if you ever read this: remember Chorin, Schloss Kröchlendorff, the Spring Schools...?). I also met the excellent mathematician Debarghya Ghoshdastidar in this phase of my studies and I am truly grateful for that: Not only am I very glad about and proud of our extensive collaboration, but I also found Debarghya to be an exceptionally nice and good person; I still like to think of the time we spent together at the Spring Schools and in Amsterdam.

After spending so many years in Potsdam, following Alexandra to Magdeburg for the second half of my PhD studies was a bit strange to me. However, it turned out that I was lucky again: I immediately sensed the friendly and benevolent environment at the institute in Magdeburg. The two strongest memories from that time are clearly the GPSD in Freiburg, where I had the opportunity to spend lots of fun time with Kerstin, Christina, Alexander and Jo from Magdeburg; and the day of my defense, where many colleagues sincerely shared the thrill and joy with me, in particular Alexandra's group consisting of Anne, James, Jo and Andrea.

Finally, I would like to express my gratitude to my parents, sister and wife for their support and belief in me and to my daughter who keeps me young in the ways that matter.

Contents

Chapter 1

Preface

1.1 Minimax Statistical Hypothesis Testing

Basic Concepts

Let $(S, \mathcal{F})$ be a measure space and $(\mathbb{P}_\theta)_{\theta\in\Theta}$ a family of probability measures on $(S, \mathcal{F})$, where the set Θ of parameters need not be finite-dimensional. Suppose that X is an S-valued random element according to the statistical model

$$(S, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta}) .$$

In this very general setting, any statistical testing problem is determined by two disjoint sets $\Theta_0, \Theta' \subseteq \Theta$ and has the form

$$H_0 : \ \theta \in \Theta_0 \ \text{ vs. } \ H_1 : \ \theta \in \Theta', \tag{1.1}$$

where we call the statement H_0 the **null hypothesis** and H_1 the **alternative hypothesis**.

Let $\mathcal{E} = 2^{\{0,1\}}$. Now, any $(\mathcal{F}, \mathcal{E})$-measurable function

$$\varphi : \ S \to \{0, 1\}$$

is a statistical test for the problem (1.1). The test φ has the role of a **decision function** which should take the value 1 if the observation X is at odds with the assumption H_0 and the value 0 otherwise. That leads to the concept of **type-I-errors** and **type-II-errors** as well as their (worst case) probabilities:

	meaning	worst case probability
type-I-error	$\varphi = 1$ yet H_0 holds	$\mathbb{P}_{H_0}(\varphi = 1) := \sup_{\theta\in\Theta_0} \mathbb{P}_\theta(\varphi = 1)$
type-II-error	$\varphi = 0$ yet H_1 holds	$\mathbb{P}_{H_1}(\varphi = 0) := \sup_{\theta\in\Theta'} \mathbb{P}_\theta(\varphi = 0)$

Depending on the respective setting and if there is no risk of confusion we may drop the index θ of $\mathbb{P}_\theta$ or write another index, for instance in the context of lower

bounds (see section 1.4.1). Similarly, although $\mathbb{P}_{H_0}$ and $\mathbb{P}_{H_1}$ are generally no probability measures (if Θ_0 or Θ' is no singleton), for the sake of simplicity we will call $\mathbb{P}_{H_0}(\varphi = 1)$ and $\mathbb{P}_{H_1}(\varphi = 0)$ error probabilities on occasion.
Furthermore, expectation and variance with respect to the measure $\mathbb{P}_\theta$ are denoted as $\mathbb{E}_\theta$ and $\mathbb{V}\mathrm{ar}_\theta$, respectively.

In general, we wish to guarantee that both error probabilities are small and for fixed $\eta \in (0, 1)$, we formalise this statement as

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\varphi = 1) + \sup_{\theta \in \Theta'} \mathbb{P}_\theta(\varphi = 0) \;\leq\; \eta. \tag{1.2}$$

Clearly, in order to fulfil this requirement, we usually need to construct a capable test. However, that typically does not suffice as the testing problem also exhibits an intrinsic difficulty: Indeed, if for instance $\Theta' = \Theta \backslash \Theta_0$, we often find that no test φ can meet the condition (1.2) – essentially every result in this thesis provides an example for that phenomenon. This leads to the idea of introducing a type of gap between the hypotheses: We define a function

$$\mathrm{dist}_{\Theta_0} : \;\Theta \to [0, \infty) \cup \{\infty\}$$

which measures the distance between $\theta \in \Theta$ and Θ_0 in a way that is suitable for the respective setting. Then, for fixed $\rho > 0$, we consider the alternative hypothesis

$$\Theta' = \Theta_\rho := \{\theta \in \Theta \;;\; \mathrm{dist}_{\Theta_0}(\theta) > \rho\}$$

and write the testing problem simply as

$$H_0 : \;\theta \in \Theta_0 \;\text{ vs. }\; H_1 : \;\mathrm{dist}_{\Theta_0}(\theta) > \rho. \tag{1.3}$$

The refined requirement

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\varphi = 1) + \sup_{\theta \in \Theta_\rho} \mathbb{P}_\theta(\varphi = 0) \;\leq\; \eta \tag{1.4}$$

may then be fulfilled for large enough ρ. In this framework, we aim at finding the smallest value for ρ enabling the existence of a test φ with the property (1.4); that is,

$$\rho^*(\eta) := \inf \left\{ \rho > 0 \;;\; \exists \text{ test } \varphi : \;\sup_{\theta \in \Theta_0} \mathbb{P}_\theta(\varphi = 1) + \sup_{\theta \in \Theta_\rho} \mathbb{P}_\theta(\varphi = 0) \;\leq\; \eta \right\}.$$

This quantity, the **minimax separation distance**, can be interpreted as the intrinsic difficulty of the testing problem

$$H_0 : \;\theta \in \Theta_0 \;\text{ vs. }\; H_1 : \;\theta \in \Theta \backslash \Theta_0$$

with respect to dist_{Θ_0} and given the **total error bound $\boldsymbol{\eta}$**. A corresponding **minimax-optimal test** is a test $\varphi = \varphi^*$ which fulfils (1.4) for the choice $\rho = \rho^*(\eta)$.

Generic Tactics and Goals

In order to describe $\rho^*(\eta)$, we typically take two steps which are very different in

nature and reflect the short discussion after (1.2):
The perhaps more natural first step is constructing a test φ which has type-I-error probability at most $\frac{\eta}{2}$. In an attempt to ensure the same bound for the type-II-error probability, ρ will have to be chosen large enough. By definition, the resulting value clearly is an upper bound on $\rho^*(\eta)$.
The second step amounts to finding a small enough value for ρ such that no test can perform well in the above sense, i.e. every test violates (1.4). Such a value clearly is a lower bound on $\rho^*(\eta)$. In order to derive this kind of statement we need some more theoretical machinery based on statistical distance measures. This will be presented in section 1.4.1.
Obviously, we would like these lower and upper bounds to be as close as possible. More specifically, we aim at deriving lower and upper bounds that match with respect to the central ingredients of the testing problem – depending on the respective setting, this could be for instance the dimension of S or Θ. That is to say, multiplicative constants in η (and possibly other fixed quantities) will be secondary and we are particularly interested in the **minimax separation rates**, where such constants are suppressed. See our first result, Theorem 1.1, for a typical example of what this means.

Literature Review

Statistical hypothesis testing is clearly a classical topic in mathematical statistics. For instance, the famous Neyman-Pearson-Lemma (see [NP33]) is concerned with the case where both Θ_0 and Θ' are singletons and provides the test with smallest type-II-error probability among all tests with a prescribed type-I-error probability. The Karlin-Rubin-theorem extends this theory under certain conditions to the case of composite hypotheses if $\Theta \subseteq \mathbb{R}$, see for instance [CB02]. Our framework, minimax testing in its general form, however, is younger and not covered by this classical theory as we typically consider multi-dimensional parameters θ and composite hypotheses.
One may broadly divide the literature on minimax testing into two major groups: Works that consider signal-detection and works that consider composite-composite testing.

Signal-Detection
We speak of a signal-detection problem if Θ_0 is a singleton. A very early article on the subject by Yuri Ingster, who is considered as the topic's founding father, would be [Ing82] with Gaussian white noise; related to section 1.2 below. Ingster and his coauthors have been very productive, for instance the series of papers [Ing93] is considered to be a landmark on non-parametric signal-detection, i.e. the case of an infinite-dimensional alternative parameter space Θ'. Many articles in that direction consider cases where Θ' is determined by an ellipsoid, see e.g. [Erm91], or more generally Sobolev- or Besov balls, e.g. [Ing98], [Spo96] and [LS99] and such cases are therefore well understood. All these references have in common that they study the asymptotic regime; non-asymptotic rates – which we focus on – are rather scarce, an important reference in that direction is [Bar02]. There are also more recent works that tackle signal-detection problems, for instance [Ver12] and [CCT17] which consider the parameter's sparsity.

Composite-Composite Testing
However, much less literature is available on cases where the null hypothesis is more complex and the present thesis makes a contribution to this theory in very different settings.
There have been some developments with regards to testing certain properties of a function f in regression and white noise settings – for instance, if f is monotonous or belongs to some functional Besov ball. Relevant articles, with both asymptotic and non-asymptotic results, include [JN02, CD13, Car15, BHL05]. Such questions usually have a direct translation as discrete regression/ Gaussian sequence problems (very explicitly e.g. in [BHL05]). In chapter 4 we extend this theory and we also address its relation to the task of constructing functional confidence sets which is done for instance in [BN13].
One may view answering the question, in what way the complexity of the null hypothesis should affect the minimax separation rate, as the central task in composite-composite testing and – consequently, we tackle it throughout this thesis. A very interesting piece of work in that direction would be [GP05] due to the generality of the approach taken there: It turns out that, in a discrete Gaussian regression setting, combining signal-detection with covering the null hypothesis works well as long as the null hypothesis is small in terms of entropy (see section 2.1 below for details); however, for instance in 2.2 we see that even in a case with infinite entropy, testing can be almost trivial. Similarly, as observed in [Car15] and chapter 4, the rate might even be the same as in signal-detection, independent of the size of the composite null hypothesis.

Further Aspects Relevant in Both Cases
Some articles that consider testing problems raise the question in how far solving these problems is related to finding optimal estimators (for the quantity to be tested), or they explicitly use testing considerations to solve the estimation problem or vice versa, see for instance [CL11], [CCT17] and [CV19] in composite-composite settings. These considerations are often based on results about estimating functionals of the mean of Gaussian vectors (such as specific norms of the mean). Further relevant literature (also for the signal-detection case) in that direction would be [CL05, CCT18, DN90]. The connection between testing and estimation is clearly not trivial, as is observed e.g. in the article [LNS99] concerned with estimation of a function's L_r-norm: the compatibility of the two questions strongly depends on the choice of r. In chapters 2 and 3 (sections 2.2 and 3.3), we exhibit such effects through very explicit examples.
Another important direction in modern testing theory is multiple testing, i.e. the situation where multiple null hypotheses are tested in a joint procedure. In this field, a typical goal would be controlling the so-called false discovery rate FDR and true discovery proportion TDP, that is, the proportion of wrongly rejected null hypotheses and the proportion of correctly rejected null hypotheses (or correctly detected alternative hypotheses), respectively. Clearly, these quantities can be seen as analogues of the type-I-error probability and testing power (i.e. 1 minus type-II-error probability), respectively. We mention just a few important and interesting references for multiple testing: The FDR was firstly introduced in [BH95] and since then the topic has been further developed for instance in [BY01, FDR07, RVDW09, ACC17b]. As

we do not study such objects in this thesis we cannot claim to make a contribution here. However, note that occasionally seemingly unrelated problems naturally have a multiple testing formulation, e.g. in [CDRV18, CV19]. Also in this thesis, the tests we propose for Theorems 1.5 and 4.1 have a multiple testing flavour.

The introduction to each of the following chapters contains further relevant literature for the specific situation considered there.

1.2 Instructive Example: Signal-Detection in l_q-norm

We now devote a major part of this preface to studying a specific non-trivial signal-detection problem as it is suitable to point out some typical phenomena and technical tools in the field of minimax testing.

For fixed $d \in \mathbb{N}\backslash\{0\}$ and $\sigma > 0$, let $\Theta = \mathbb{R}^d$, $\Theta_0 = \{\mathbb{0}_d\}$ (null vector $\mathbb{0}_d \in \mathbb{R}^d$) and

$$\mathbb{P}_\theta = \mathcal{N}(\theta, \sigma\mathbb{I}_d) \quad \text{for} \quad \theta \in \Theta,$$

where $\mathcal{N}(\theta, \sigma\mathbb{I}_d)$ is the d-dimensional Gaussian distribution with expectation θ and covariance matrix $\sigma\mathbb{I}_d$ ($\mathbb{I}_d$ being the $d \times d$ identity matrix). That is, the coordinates $i \in \{1, 2, \dots, d\}$ of

$$X \overset{\mathcal{L}}{=} \theta + \sigma\epsilon$$

are independent with $\sigma\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. For $q > 0$, we choose the distance measure

$$\mathrm{dist}_{\Theta_0}(\theta) = \mathrm{dist}(\theta) = \|\theta\|_q = \left(\sum_{i=1}^{d} |\theta_i|^q\right)^{\frac{1}{q}}, \tag{1.5}$$

which is actually only a metric if $q \geq 1$. The associated testing problem in the sense of (1.3) now reads

$$H_0: \ \theta = \mathbb{0}_d \quad \text{vs.} \quad H_1: \ \|\theta\|_q > \rho \tag{1.6}$$

for fixed $\rho > 0$ and we are going to study $\rho^*(\eta) = \rho^*_q(\eta)$, $\eta \in (0, 1)$, for this problem.

Note that the dependence of $\rho^*_q(\eta)$ on d has been exposed in [IS02, Section 3.3.6] in the asymptotic regime for $d \to \infty$ and without constants in η or q, but we are not aware of a non-asymptotic treatment of the problem with explicit constants and provide it here.

The case $q \leq 2$

The case $q = 2$ is relatively easy to solve which can be attributed to the fact that the random variable $\|X\|_2^2$ is convenient to study as it is well-behaved and well-understood. This special case has already been solved in [Bar02, Propositions 1 and 2], but we provide an own analysis here since it constitutes a useful preparation for other considerations in this thesis.

Theorem 1.1. Consider the problem (1.6) with $q = 2$ and let $\eta \in (0,1)$. Then we have

$$\sigma \cdot d^{1/4} \cdot \sqrt[4]{\frac{1}{2}\ln(1+4(1-\eta)^2)} \;\leq\; \rho_q^*(\eta) \;\leq\; \sigma \cdot d^{1/4} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)},$$

that is

$$\rho_q^* \;\sim\; \sigma \cdot d^{1/4}. \tag{1.7}$$

◁

All proofs are postponed to the last section of this chapter.

Remark 1.2. As indicated earlier, we focus on the central parameters (here d and σ) and suppress constants depending on the (typically) arbitrary total error bound η and possibly other fixed parameters (here q). This step is indicated by the symbol $\sim$ for equivalence up to constants in (1.7) . The symbols $\lesssim$ and $\gtrsim$ will be used in an analogous way. ◁

Due to the geometry of the l_q-balls in $\mathbb{R}^d$, it turns out that the test and lower bound construction used for Theorem 1.1 is already sufficient to establish the minimax separation rate for $0 < q \leq 2$:

Corollary 1.3. Consider the problem (1.6) with $0 < q \leq 2$ and let $\eta \in (0,1)$. Then we have

$$\sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \cdot \sqrt[4]{\frac{1}{2}\ln(1+4(1-\eta)^2)} \;\leq\; \rho_q^*(\eta) \;\leq\; \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)},$$

that is

$$\rho_q^* \;\sim\; \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}}.$$

◁

The case $q > 2$

The upper bound for Theorem 1.1 is not difficult to obtain since, as already indicated, for any $\theta \in \mathbb{R}^d$ there are strong explicit concentration statements for $\|X\|_2^2$ available and, in particular, its expectation is known and simple.
This is no longer the case for $q > 2$: Though there are analytical formulae for $\mathbb{E}_\theta[\|X\|_q^q]$ and $\mathbb{V}\mathrm{ar}_\theta[\|X\|_q^q]$, they are not useful for our purposes as the dependence on d and q is too implicit, see [Win12]. As a consequence, the same holds for concentration inequalities on $\|X\|_q^q$. Therefore, we treat the case in a more basic, explicit manner by employing Chebyshev's and Bernstein's inequalities, where the necessary bounds on expectation and variance are given in the following auxiliary result. It is based on [IS02, Lemma 3.2] but more general. In the proof (section 1.4.4), we also fill in details and above all, provide explicit constants.

Lemma 1.4. Let $q > 2$, $\sigma > 0$, $\theta \in \mathbb{R}$ and $Y \sim \mathcal{N}(0,1)$. For $b \in (0,\infty) \cup \{\infty\}$, we define the event $\xi_b = \{|Y| \leq b\}$ and write

$$E_{b,r} := \mathbb{E}[|Y|^r \mid \xi_b] \;\text{ for }\; r > 0$$

as well as

$$l_{b,1} := \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)},\; l_{b,2} = \min\left(\frac{1}{8}, \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}(4E_{b,q})^{(2-q)/q}\right). \tag{1.8}$$

Then

$$\mathbb{E}[|\theta + \sigma Y|^q \mid \xi_b] \geq \sigma^q E_{b,q} + l_{b,1}\sigma^{q-2}\theta^2 + l_{b,2}|\theta|^q,$$
$$\mathbb{V}\mathrm{ar}[|\theta + \sigma Y|^q \mid \xi_b] \leq q^2 2^{2q-3}\left(\sigma^{2q}E_{b,2q} + \sigma^2|\theta|^{2q-2}E_{b,2}\right).$$

◁

We state and prove two different upper bounds on $\rho_q^*(\eta)$ in order to highlight another typical phenomenon in our field of work.

Theorem 1.5. Consider the problem (1.6) with $q > 2$ and let $\eta \in (0,1)$. Then there are $C_q^1, C_q^2 > 0$ such that

$$\rho_q^*(\eta) \leq \begin{cases} C_q^1 \cdot \sigma \cdot d^{\frac{1}{2q}} \cdot \sqrt{\frac{2}{\eta}}, \\ C_q^2 \cdot \sigma \cdot d^{\frac{1}{2q}} \cdot \sqrt{\ln\left(\frac{5}{\eta}\right)} \cdot \sqrt{2\ln\left(\frac{10d}{\eta}\right)}, & \text{if } d \geq \left\lceil \ln\left(\frac{10d}{\eta}\right)^q \right\rceil. \end{cases}$$

In particular, the first bound yields

$$\rho_q^* \lesssim \sigma \cdot d^{\frac{1}{2q}}.$$

◁

While the second bound displays a better dependence on the error rate η, it contains an additional ln-factor in d which renders it suboptimal for our purposes. Intuitively, in the setting of Theorem 1.5, one can interpret the bounds as two ways of paying the price for the lack of knowledge about the distribution of $\|X\|_q^q$: We must either weaken the constant with respect to η ($\frac{2}{\eta}$ rather than $\ln\left(\frac{2}{\eta}\right)$ as in Theorem 1.1) or pay additional $\ln(d)$-factors.
Throughout this thesis, we prove different bounds with either $\sqrt{\cdot}$- or $\ln(\cdot)$-dependence on η. A very interesting result in that respect is Theorem 3.8 as we eliminate "unwanted" $\ln(d)$-factors there through a so-called chaining approach.

It turns out that deriving a lower bound for the case $q > 2$ also requires new ideas: The prior we constructed for $q = 2$ under H_1 is powerful as it fully exploits the dimension d and is set precisely at l_2-distance ρ from $\mathbb{O}_d$. It also conveniently extends to $q < 2$ since its support minimises the l_2-distance (and with it statistical distance) to $\mathbb{O}_d$ among all points $\theta \in \mathbb{R}^d$ with $\|\theta\|_q = \rho$. The latter effect is reversed for $q > 2$ and it becomes more difficult to exploit the dimension while keeping the l_2-distance small. We solve this problem by introducing a mixed prior where $\|\theta\|_q \leq \rho$ is possible even under H_1; this leads to the desired rate $d^{\frac{1}{2q}}$, but as a sort of post-processing step we must exclude the case $\|\theta\|_q \leq \rho$, which requires a price in terms of the ranges of η and d:

Theorem 1.6. Consider the problem (1.6) with $q > 2, d \geq 36$ and let $\eta \in (0, \frac{1}{3})$. Then we have

$$\sigma \cdot d^{\frac{1}{2q}} \cdot \frac{\sqrt{\frac{1}{2}\ln\left(1 + \ln\left((1 + 4(\frac{1}{3} - \eta)^2)\right)\right)}}{5^{1/q}} \leq \rho_q^*(\eta),$$

that is, in conjunction with Theorem 1.5,

$$\rho_q^* \;\sim\; \sigma \cdot d^{\frac{1}{2q}}.$$

◁

We need an analogous strategy with mixed priors again later on for proving Theorem 2.6.

1.3 Outline

In the following chapters we present new results on minimax testing problems in three very different settings, where both Θ_0 and Θ_ρ are composite in each case. The chapters build upon this preface to some extent, but otherwise they are self-contained.
Firstly, chaper 2 is concerned with what can be seen as a considerable generalisation of Theorem 1.1 as we consider $\Theta_0 = \mathcal{C}$ for convex sets $\mathcal{C} \subseteq \mathbb{R}^d$. The chapter presents the results of the article [BCG18].
After that, in chapter 3 we study the problem of testing if two random graphs (inhomogeneous Erdős-Renyi model) have the same underlying distribution, where various notions of distance between the null and alternative hypotheses are considered. This chapter is closely related to the papers [GGCvL17a] and [GGCvL17b].
Finally, chapter 4 deals with testing the regularity of a function $f : [0,1] \to \mathbb{R}$ in the sense of Sobolev-norms based on a perturbed version of f, the noise being a Brownian motion. The results on this non-parametric problem are based on the article [Gut19].
Proofs are given at the end of each chapter.

1.4 Proofs

1.4.1 General Method for Obtaining Lower Bounds

The following reasoning, explained for instance in [Bar02, section 7.1], will be used throughout the thesis in order to derive lower bounds on $\rho^*(\eta)$.

Remember the general definitions from section 1.1.
Now, let ν_0 be a distribution with $S_0 := \operatorname{supp}(\nu_0) \subseteq \Theta_0$ and ν_ρ be a distribution with $S_\rho := \operatorname{supp}(\nu_\rho) \subseteq \Theta_\rho$ (priors). Furthermore, let $\mathbb{P}_{\theta\sim\nu_0}$ and $\mathbb{P}_{\theta\sim\nu_\rho}$ be the resulting probability measures when $\theta \sim \nu_0$ and $\theta \sim \nu_\rho$ respectively. Then we see that for any

test $\varphi = \mathbb{1}_A$, $A \in \mathcal{F}$, with the **total variation distance** $\|\cdot\|_{\mathrm{TV}}$ (see also [GN16]),

$$\begin{aligned}\sup_{\theta\in\Theta_0} \mathbb{P}_\theta(\varphi=1) + \sup_{\theta\in\Theta_\rho} \mathbb{P}_\theta(\varphi=0) &\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi=1) + \mathbb{P}_{\theta\sim\nu_\rho}(\varphi=0)\\ &= 1 - \big(\mathbb{P}_{\theta\sim\nu_\rho}(A) - \mathbb{P}_{\theta\sim\nu_0}(A)\big)\\ &\geq 1 - \sup_{A\in\mathcal{F}} \big|\mathbb{P}_{\theta\sim\nu_\rho}(A) - \mathbb{P}_{\theta\sim\nu_0}(A)\big|\\ &= 1 - \frac{1}{2}\|\mathbb{P}_{\theta\sim\nu_\rho} - \mathbb{P}_{\theta\sim\nu_0}\|_{\mathrm{TV}}.\\ &\geq 1 - \frac{1}{2}\left(\int \left(\frac{\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}\right)^2 \mathrm{d}\mathbb{P}_{\theta\sim\nu_0} - 1\right)^{\frac{1}{2}},\end{aligned}$$

where the last bound follows from Jensen's inequality together with a representation of the total variation distance given in [Bar02] whenever $\mathbb{P}_{\theta\sim\nu_\rho} \ll \mathbb{P}_{\theta\sim\nu_0}$.

These observations justify the following general argument:
Let $\eta \in (0,1)$. For any $\rho > 0$ such that either the total variation distance fulfils

$$\frac{1}{2}\|\mathbb{P}_{\theta\sim\nu_\rho} - \mathbb{P}_{\theta\sim\nu_0}\|_{\mathrm{TV}} < 1-\eta \tag{1.9}$$

or the **χ^2-divergence** fulfils

$$\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) := \int \left(\frac{\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}\right)^2 \mathrm{d}\mathbb{P}_{\theta\sim\nu_0} < 1 + 4(1-\eta)^2, \tag{1.10}$$

it holds for any test φ that

$$\sup_{\theta\in\Theta} \mathbb{P}_\theta(\varphi=1) + \sup_{\theta\in\Theta_\rho} \mathbb{P}_\theta(\varphi=0) \; > \; \eta.$$

Hence, for the testing problem (1.3), we have

$$\rho^*(\eta) \;\geq\; \rho.$$

1.4.2 Proof of Theorem 1.1

Lower Bound

Let $\nu_0 = \delta_{\mathbb{O}_d}$ (Dirac-δ) and ν_ρ be the uniform distribution on

$$P_h := h\cdot\{-1,1\}^d \quad \text{with} \quad h = \frac{\rho}{\sqrt{d}}.$$

The definition of h guarantees $P_h \subseteq \Theta_\rho$. For $x \in \mathbb{R}^d$, the resulting densities of $\mathbb{P}_{\theta\sim\nu_0}$ and $\mathbb{P}_{\theta\sim\nu_\rho}$ read

$$\begin{aligned}
\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}(x) &= \left(\frac{1}{2\pi\sigma^2}\right)^{d/2}\prod_{i=1}^d \exp\left(-\frac{1}{2\sigma^2}x_i^2\right);\\
\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}(x) &= \frac{1}{2^d}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{d}{2}}\sum_{v_1,\dots,v_d\in\{-1,1\}}\prod_{i=1}^d \exp\left(-\frac{1}{2\sigma^2}(x_i-h\cdot v_i)^2\right)\\
&= \frac{1}{2^d}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{d}{2}}\prod_{i=1}^d \exp\left(-\frac{1}{2\sigma^2}x_i^2-\frac{1}{2\sigma^2}h^2\right)2\cosh\left(\frac{1}{\sigma^2}hx_i\right)\\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{d}{2}}\exp\left(-\frac{d}{2\sigma^2}h^2\right)\prod_{i=1}^d \exp\left(-\frac{1}{2\sigma^2}x_i^2\right)\cosh\left(\frac{1}{\sigma^2}hx_i\right),
\end{aligned}$$

so that

$$(\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho})^2(x) = \left(\frac{1}{2\pi\sigma^2}\right)^{d}\exp\left(-\frac{d}{\sigma^2}h^2\right)\prod_{i=1}^d \exp\left(-\frac{1}{\sigma^2}x_i^2\right)\cosh^2\left(\frac{1}{\sigma^2}hx_i\right).$$

Using the fact that $\mathbb{E}[\cosh(aY)^2] = \exp(a^2\sigma^2)\cosh(a^2\sigma^2)$ for $Y \sim \mathcal{N}(0,\sigma^2)$ and Fubini's theorem we can evaluate the left hand side of (1.10):

$$\begin{aligned}
\int_{\mathbb{R}^d}\frac{(\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho})^2}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}(x) &= \exp\left(-\frac{d}{\sigma^2}h^2\right)\left(\int_{\mathbb{R}}\cosh^2\left(\frac{1}{\sigma^2}hx_i\right)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{1}{2\sigma^2}x_i^2\right)\right)^d\\
&= \exp\left(-\frac{d}{\sigma^2}h^2\right)\exp\left(-d\frac{h^2}{\sigma^2}\right)\cosh\left(\frac{h^2}{\sigma^2}\right)^d\\
&= \cosh\left(\frac{h^2}{\sigma^2}\right)^d\\
&\leq \exp\left(d\frac{h^4}{2\sigma^4}\right),
\end{aligned}$$

where the last inequality is based on the known bound $\cosh(x) \leq \exp(x^2/2)$ (compare the Taylor series). By direct elementary computation and plugging in the definition of h we now see that (1.10) is fulfilled if

$$\rho < \sigma\cdot d^{1/4}\cdot\sqrt[4]{\ln(1+4(1-\eta)^2)},$$

so we obtain the sufficiently small bound

$$\rho \leq \sigma\cdot d^{1/4}\cdot\sqrt[4]{\frac{1}{2}\ln(1+4(1-\eta)^2)}$$

claimed in Theorem 1.1.

Upper Bound

Type-I-error: finding rejection threshold

We analyse the test based on the natural test statistic for the problem at hand:

$$\varphi(X) = \mathbb{1}_{\{\|X\|_2 > \tau_{\eta/2}\}}$$

for a suitable choice of rejection threshold $\tau_{\eta/2}$. It can be chosen directly based on the following Lemma:

Lemma 1.7. For $k \in \mathbb{N}\backslash\{0\}$, $\sigma > 0$ and $\lambda \geq 0$, let $Y \sim \mathcal{N}(0, \sigma^2)$ and $Z \sim \chi^2(k, \lambda)$ (that is, a χ^2-distribution with k degrees of freedom and non-centrality parameter λ). Then for any $\delta \in (0, 1)$, we have

$$\begin{aligned} &\text{(I)} & \mathbb{P}\left(N \geq \sigma\sqrt{2\ln\left(\tfrac{1}{\delta}\right)}\right) &\leq \delta \\ &\text{(II)} & \mathbb{P}\left(Z \geq d + \lambda + 2\sqrt{(d+2\lambda)\ln\left(\tfrac{1}{\delta}\right)} + 2\ln\left(\tfrac{1}{\delta}\right)\right) &\leq \delta, \\ &\text{(III)} & \mathbb{P}\left(Z \leq d + \lambda - 2\sqrt{(d+2\lambda)\ln\left(\tfrac{1}{\delta}\right)}\right) &\leq \delta, \end{aligned} \tag{1.11}$$

◁

(Statement (I) is the classical bound based on the Chernoff strategy and proofs of (II) and (III) can be found in [Bir01].)
Since under H_0, $\frac{1}{\sigma^2}\|X\|_2^2 \sim \chi^2(d, 0)$, we can choose the rejection threshold

$$\tau_{\eta/2} := \sigma\sqrt{d + 2\sqrt{d\ln\left(\frac{2}{\eta}\right)} + 2\ln\left(\frac{2}{\eta}\right)}$$

by (II) in (1.11).

Type-II-error: finding a lower confidence bound exceeding the rejection threshold
On the other hand, under H_1 we aim at finding a large enough value for ρ such that $\|\theta\|_2 \geq \rho$ implies

$$\mathbb{P}_\theta(\varphi = 0) = \mathbb{P}_\theta\left(\|X\|_2^2 \leq \sigma^2\left(d + 2\sqrt{d\ln\left(\frac{2}{\eta}\right)} + 2\ln\left(\frac{2}{\eta}\right)\right)\right) \leq \frac{\eta}{2}.$$

As $\frac{1}{\sigma^2}\|X\|_2^2 \sim \chi^2\left(d, \frac{1}{\sigma^2}\|\theta\|_2^2\right)$, equation (III) of (1.11) yields the sufficient condition

$$d + 2\sqrt{d\ln\left(\frac{2}{\eta}\right)} + 2\ln\left(\frac{2}{\eta}\right) \leq d + \frac{\|\theta\|_2^2}{\sigma^2} - 2\sqrt{\left(d + 2\frac{\|\theta\|_2^2}{\sigma^2}\right)\ln\left(\frac{2}{\eta}\right)}.$$

By direct computation, this is fulfilled if

$$\|\theta\|_2 \geq \sigma\left(\sqrt{4\sqrt{d\ln\left(\frac{2}{\eta}\right)} + 4\cdot\ln\left(\frac{2}{\eta}\right)} + \sqrt{2\cdot\ln\left(\frac{2}{\eta}\right)}\right),$$

or, very much simplified,

$$\|\theta\|_2 \geq \rho = \sigma\cdot d^{1/4}\cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)}.$$

as claimed in Theorem 1.1.

1.4.3 Proof of Corollary 1.3

Lower Bound

We use the construction from the case $q = 2$, i.e. section 1.4.2, and arrive at

$$\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) \le \exp\left(d\frac{h^4}{2\sigma^4}\right),$$

but in this more general context, $h = \rho \cdot d^{-1/q}$, so that

$$\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) \le \exp\left(d^{1-4/q}\frac{\rho^4}{2\sigma^4}\right),$$

and hence, by direct computation, (1.10) is fulfilled if

$$\rho < \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \sqrt[4]{\ln(1 + 4(1-\eta)^2)}$$

or e.g. as claimed in Corollary 1.3,

$$\rho \le \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \sqrt[4]{\frac{1}{2}\ln(1 + 4(1-\eta)^2)}$$

Upper Bound

The test φ from section 1.4.2 performs as desired for the testing problem

$$H_0 : \ \theta = \mathbb{O}_d \ \text{ vs. } \ H_1 : \ \|\theta\|_2 > \sigma \cdot d^{1/4} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)}.$$

Now, with the general equivalence property of the q-norms we have

$$\|\theta\|_q \le d^{\frac{1}{q}-\frac{1}{2}}\|\theta\|_2$$

and hence

$$\|\theta\|_q \ge \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)} \implies \|\theta\|_2 \ge \sigma \cdot d^{\frac{1}{4}} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)},$$

so that φ also solves the testing problem

$$H_0 : \ \theta = \mathbb{O}_d \ \text{ vs. } \ H_1 : \ \|\theta\|_p > \sigma \cdot d^{\frac{1}{q}-\frac{1}{4}} \cdot 6\sqrt{\ln\left(\frac{2}{\eta}\right)},$$

which concludes the argument.

1.4.4 Proof of Lemma 1.4

Bound on the expectation

We prove two lower bounds on $\mathbb{E}[|\theta + \sigma Y|^q \mid \xi_b]$ independently and then combine them.

Firstly, a simple observation is

$$\mathbb{E}[|\theta + \sigma Y|^q \mid \xi_b] \ge |\theta|^q \mathbb{P}(\theta Y > 0 \mid \xi_b) = \frac{1}{2}|\theta|^q. \tag{1.12}$$

Secondly, in a more involved step, we derive a lower bound on $\mathbb{E}[|\theta+\sigma Y|^q \mid \xi_b]$ in terms of $E_{b,q}$ and $E_{b,q-2}$. In preparation, for fixed $y \neq 0$, through Taylor expansion at $\theta_0 = 0$ we obtain that, if $\theta \neq 0$, for some $\theta' \in (-|\theta|, |\theta|)\backslash\{0\}$,

$$|\theta+\sigma y|^q = |\sigma y|^q + q|\sigma y|^{q-1}\text{sign}(y) + \frac{q(q-1)}{2}|\theta'+\sigma y|^{q-2}\theta^2. \tag{1.13}$$

Observing

$$\mathbb{E}[|\theta'+\sigma Y|^{q-2} \mid \xi_b] \geq \frac{1}{\mathbb{P}(\xi_b)}\int_{-b}^{b}|\sigma y|^{q-2}\mathbb{1}_{\{\theta' y>0\}}\,\frac{\exp(-y^2/2)}{\sqrt{2\pi}}\text{d}y = \frac{\sigma^{q-2}}{2\mathbb{P}(\xi_b)}E_{b,q-2},$$

taking the expectation in (1.13) yields

$$\mathbb{E}[|\theta+\sigma Y|^q \mid \xi_b] \geq \sigma^q E_{b,q} + \frac{q(q-1)E_{b,q-2}}{4\mathbb{P}(\xi_b)}\sigma^{q-2}\theta^2. \tag{1.14}$$

Clearly, (1.12) and (1.14) are also valid in the case $\theta = 0$. Now, as indicated above, it remains to properly combine them: If on the one hand, $|\theta|^q \geq 4\sigma^q E_{b,q}$, (1.12) tells us that

$$\mathbb{E}[|\theta+\sigma Y|^q \mid \xi_b] - \sigma^q E_{b,q} \geq \frac{|\theta|^q}{4},$$

which, together with (1.14), yields

$$\mathbb{E}[|\theta+\sigma Y|^q \mid \xi_b] \geq \sigma^q E_{b,q} + \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}\sigma^{q-2}\theta^2 + \frac{|\theta|^q}{8}. \tag{1.15}$$

If on the other hand, $|\theta|^q < 4\sigma^q E_{b,q}$, we have

$$\begin{aligned}\frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}\sigma^{q-2}\theta^2 &= \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}\sigma^{q-2}|\theta|^q|\theta|^{2-q}\\ &> \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}\sigma^{q-2}|\theta|^q(4\sigma^q E_{b,q})^{(2-q)/q}\\ &= \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}(4E_{b,q})^{(2-q)/q}|\theta|^q\end{aligned}$$

and therefore, based on (1.14),

$$\mathbb{E}[|\theta+\sigma Y|^q \mid \xi_b] \geq \sigma^q E_{b,q} + \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}\sigma^{q-2}\theta^2 + \frac{q(q-1)E_{b,q-2}}{8\mathbb{P}(\xi_b)}(4E_{b,q})^{(2-q)/q}|\theta|^q. \tag{1.16}$$

Finally, remembering (1.8), (1.15) and (1.16) yield the desired bound.

Bound on the variance

Observe that with the classical bounds

$$\begin{aligned}(|x|+|y|)^q &\leq 2^{q-1}(|x|^q+|x|^q),\\ ||x+y|^q-|y|^q| &\leq q2^{q-2}|x|(|x|^{q-1}+|y|^{q-1})\end{aligned} \tag{1.17}$$

for $x, y \in \mathbb{R}$ (see [IS02, Lemma 3.2]), we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[|\theta+\sigma Y|^q \mid \xi_\delta\right] &= \min_{x\in\mathbb{R}} \mathbb{E}\left[(|\theta+\sigma Y|^q - x)^2 \mid \xi_\delta\right] \\ &\le \mathbb{E}\left[(|\theta+\sigma Y|^q - |\theta|^q)^2 \mid \xi_\delta\right] \\ &\le \mathbb{E}\left[q^2 2^{2q-4}|\sigma Y|^2(|\sigma Y|^{q-1}+|\theta|^{q-1})^2 \mid \xi_\delta\right] \\ &\le \mathbb{E}\left[q^2 2^{2q-3}|\sigma Y|^2(|\sigma Y|^{2q-2}+|\theta|^{2q-2}) \mid \xi_\delta\right] \\ &= q^2 2^{2q-3}\left(\sigma^{2q}E_{b,2q} + \sigma^2|\theta|^{2q-2}E_{b,2}\right).\end{aligned}$$

1.4.5 Proof of Theorem 1.5

First Bound

Type-I-error: finding a suitable rejection threshold

Applying Chebyshev's inequality with the results of Lemma 1.4 for the case $b=\infty$ (i.e. ξ_b is trivial), we obtain

$$\mathbb{P}_{H_0}\left(\sum_{i=1}^d |X_i|^q \ge d\sigma^q E_{\infty,q} + \sqrt{d}\sigma^q\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,2q}}\right) \le \frac{\eta}{2}.$$

This motivates the introduction of the test $\varphi(X) := \mathbb{1}\left\{\sum_{i=1}^d |X_i|^q > \tau_{\frac{\eta}{2}}\right\}$ with rejection threshold

$$\tau_{\frac{\eta}{2}} := d\sigma^q E_{\infty,q} + \sqrt{d}\sigma^q\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,2q}}.$$

Type-II-error: finding a lower confidence bound exceeding the rejection threshold

On the other hand, we must now attempt to find a large enough ρ such that there is an appropriate $U_{\frac{\eta}{2}} = U_{\frac{\eta}{2}}(\theta) \ge \tau_{\frac{\eta}{2}}$ with $\mathbb{P}_\theta(\varphi=0) \le \mathbb{P}_\theta\left(\sum_{i=1}^d |X_i|^q \le U_{\frac{\eta}{2}}\right) \le \frac{\eta}{2}$ for any $\theta\in\Theta_\rho$. Applying Chebyshev's inequality with Lemma 1.4 again in that case suggests

$$\begin{aligned}U_{\frac{\eta}{2}} = d\sigma^q E_{\infty,q} + l_{\infty,1}\sigma^{q-2}\|\theta\|_2^2 + l_{\infty,2}\|\theta\|_q^q - \sqrt{d}\sigma^q\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,2q}} \\ - \sigma\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,q}}\|\theta\|_{2q-2}^{q-1},\end{aligned}$$

so that the remaining requirement $U_{\frac{\eta}{2}} \ge \tau_{\frac{\eta}{2}}$ reads

$$l_{\infty,1}\sigma^{q-2}\|\theta\|_2^2 + l_{\infty,2}\|\theta\|_q^q \ge 2\sqrt{d}\sigma^q\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,2q}} + \sigma\sqrt{\tfrac{2}{\eta}q^2 2^{2q-3}E_{\infty,q}}\|\theta\|_{2q-2}^{q-1}. \quad (1.18)$$

We know choose the ansatz $\|\theta\|_q = C_{\frac{\eta}{2}}\sigma d^{\frac{1}{2q}}$. Then, since the r-norms decrease in r, due to $2q-2>q$ and $\frac{q-1}{2q} \le \frac{1}{2}$,

$$\sigma\|\theta\|_{2q-2}^{q-1} \le \sigma\|\theta\|_q^{q-1} \le \sigma C_{\frac{\eta}{2}}^{q-1}\sigma^{q-1}d^{\frac{q-1}{2q}} \le C_{\frac{\eta}{2}}^{q-1}\sigma^q\sqrt{d}.$$

Then direct comparison shows that $l_{\infty,2}\|\theta\|_q^q$ exceeds the right hand side of (1.18) whenever

$$C_{\frac{\eta}{2}} \geq c_{\frac{\eta}{2}} := \max\left(\left(\sqrt{\frac{2}{\eta}}\cdot\frac{4}{l_{\infty,2}}\cdot\sqrt{q^2 2^{2q-3}E_{\infty,2q}}\right)^{1/q}, \sqrt{\frac{2}{\eta}}\cdot\frac{2}{l_{\infty,2}}\sqrt{q^2 2^{2q-3}E_{\infty,q}}\right)$$
$$\sim \sqrt{\frac{2}{\eta}}.$$

Second Bound

The second bound is based on a Bernstein-type inequality rather than Chebyshev's inequality (see for instance [Mas07, Prop. 2.9]). More specifically, we employ the following classical formulation: Let $Y_1, \dots, Y_d$ be centered independent random variables bounded in absolute value by $M > 0$. Then, for $\delta \in (0,1)$, we have

$$\mathbb{P}\left(\sum_{i=1}^d Y_i \geq \sqrt{2\ln\left(\frac{1}{\delta}\right)\sum_{i=1}^d \mathbb{V}\mathrm{ar}[Y_i]} + \frac{2}{3}M\ln\left(\frac{1}{\delta}\right)\right) \leq \delta. \tag{1.19}$$

Now, for $\delta \in (0,1)$ specified later on, we introduce the event

$$\xi_{b_\delta} = \{\forall i \in \{1,2,\dots,d\} : \ |\epsilon_i| \leq b_\delta\}, \quad b_\delta := \sqrt{2\ln\left(\frac{2d}{\delta}\right)}.$$

Applying a union bound to part (I) of (1.11) yields $\mathbb{P}(\xi_{b_\delta}) \geq 1-\delta$.

Boundedness required in Bernstein's inequality

We now study the variables $|X_i|^q$, $i \in \{1,2,\dots,d\}$, more closely: Under H_0, we clearly have

$$||X_i|^q - \mathbb{E}\left[|X_i|^q \mid \xi_{b_\delta}\right]| \leq \sigma^q b_\delta^q.$$

On the other hand, for arbitrary $\theta_i \in \mathbb{R}$, using (1.17) we see

$$|X_i|^q \leq 2^{q-1}(|\theta_i|^q + |\sigma\epsilon_i|^q),$$

so that

$$||X_i|^q - \mathbb{E}\left[|X_i|^q \mid \xi_{b_\delta}\right]| \leq 2^{q-1}(|\theta_i|^q + \sigma^q b_\delta^q) \leq 2^{q-1}(\|\theta\|_\infty^q + \sigma^q b_\delta^q).$$

Type-I-error: finding a suitable rejection threshold

Now (1.19) yields the following statements, where we use Lemma 1.4. Firstly, under the null hypothesis, we see that for $\delta \in (0,1)$,

$$\mathbb{P}_{H_0}\left(\sum_{i=1}^d |X_i|^q \geq d\sigma^q E_{b_\delta,q} + \sqrt{d}\sigma^q\sqrt{2\ln\left(\tfrac{1}{\delta}\right)q^2 2^{2q-3}E_{b,2q}} + \frac{2}{3}\sigma^q b_\delta^q \ln\left(\tfrac{1}{\delta}\right) \ \middle|\ \xi_{b_\delta}\right) \leq \delta.$$

This motivates the introduction of a preliminary test $\psi(X) := \mathbb{1}\left\{\sum_{i=1}^d |X_i|^q > \tau_\delta\right\}$ with rejection threshold

$$\tau_\delta := d\sigma^q E_{b_\delta,q} + \sqrt{d}\sigma^q\sqrt{2\ln\left(\tfrac{1}{\delta}\right)q^2 2^{2q-3}E_{b,2q}} + \frac{2}{3}\sigma^q b_\delta^q \ln\left(\tfrac{1}{\delta}\right). \tag{1.20}$$

The reason why ψ is not yet the final test φ will occur in (1.22). Moreover, the appropriate relation between δ and η will be more clear later on.

Type-II-error: finding a lower confidence bound exceeding the rejection threshold

Similar as above, we must now tune ρ in order to find $U_\delta = U_\delta(\theta) \geq \tau_\delta$ such that $\mathbb{P}_\theta(\psi = 0 \mid \xi_{b_\delta}) \leq \mathbb{P}_\theta\left(\sum_{i=1}^d |X_i|^q \leq U_\delta \mid \xi_{b_\delta}\right) \leq \delta$ for $\theta \in \Theta_\rho$. Applying (1.19) with Lemma 1.4 in that case suggests

$$\begin{aligned}U_\delta = d\sigma^q E_{b_\delta,q} + l_{b_\delta,1}\sigma^{q-2}\|\theta\|_2^2 + l_{b_\delta,2}\|\theta\|_q^q - \sqrt{d}\sigma^q\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,2q}} \\ - \sigma\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,q}}\|\theta\|_{2q-2}^{q-1}. \\ - \frac{2^q}{3}\ln\left(\tfrac{1}{\delta}\right)\|\theta\|_\infty^q - \frac{2^q}{3}\ln\left(\tfrac{1}{\delta}\right)\sigma^q b_\delta^q,\end{aligned}$$

so that the remaining requirement $U_\delta \geq \tau_\delta$ reads

$$\begin{aligned}l_{b_\delta,1}\sigma^{q-2}\|\theta\|_2^2 + l_{b_\delta,2}\|\theta\|_q^q \ \geq \ & 2\cdot\sqrt{d}\sigma^q\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,2q}} \qquad (1.21)\\ & + \sigma\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,q}}\|\theta\|_{2q-2}^{q-1} \\ & + \frac{2+2^q}{3}\ln\left(\tfrac{1}{\delta}\right)\sigma^q b_\delta^q \\ & + \frac{2^q}{3}\ln\left(\tfrac{1}{\delta}\right)\|\theta\|_\infty^q. \qquad (1.22)\end{aligned}$$

We know choose the ansatz $\|\theta\|_q = C_{\delta,d}\sigma d^{\frac{1}{2q}}$. Then again due to monotonicity, $2q-2 > q$ and $\frac{q-1}{2q} \leq \frac{1}{2}$,

$$\sigma\|\theta\|_{2q-2}^{q-1} \leq \sigma\|\theta\|_q^{q-1} \leq \sigma C_{\delta,d}^{q-1}\sigma^{q-1}d^{\frac{q-1}{2q}} \leq C_{\delta,d}^{q-1}\sigma^q\sqrt{d}.$$

Coping with $\|\theta\|_\infty^q$

The term in (1.22) requires additional steps as it cannot generally be compensated by either term on the left hand side of (1.21): We define the event

$$\Xi_\delta = \left\{\exists i \in \{1,2,\ldots,d\}: \ |X_i| > \sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}\right\}. \qquad (1.23)$$

and the assumption

$$\mathcal{A}_\delta: \ \|\theta\|_\infty \leq 2\sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}, \qquad (1.24)$$

Now, whenever $\mathcal{A}_\delta$ holds we can upper bound (1.22) as

$$\frac{2^q}{3}\ln\left(\frac{1}{\delta}\right)\|\theta\|_\infty^q \leq \sigma^q \ln\left(\frac{2d}{\delta}\right)^{q/2}\frac{2^{q/2}4^q}{3}\ln\left(\frac{1}{\delta}\right).$$

Then direct comparison shows that, as desired, $l_{b_\delta,2}\|\theta\|_q^q$ exceeds the sum on the right hand side of (1.21) through (1.22) if

$$C_{\delta,d} \geq c_{\delta,d} := \left(\frac{3}{l_{b_\delta,2}}\left(4\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,2q}} + 2\frac{2+2^q}{3}\ln\left(\tfrac{1}{\delta}\right) b_\delta^q\right)\right)^{1/q} \tag{1.25}$$

$$+ \; 4\sqrt{2}\left(\frac{\ln(\frac{1}{\delta})}{l_{b_\delta,2}}\right)^{1/q}\sqrt{\frac{\ln(2d/\delta)}{d^{1/q}}} \tag{1.26}$$

$$+ \; \frac{3}{l_{b_\delta,2}}\sqrt{2\ln\left(\tfrac{1}{\delta}\right) q^2 2^{2q-3} E_{b_\delta,q}}, \tag{1.27}$$

where we impose the restriction

$$d \geq d_0 = d_0(\delta, q) := \lceil \ln(2d/\delta)^q \rceil. \tag{1.28}$$

(It allows to control the summand (1.26)).
In summary, under the assumption (1.24) and with the restriction (1.28), whenever $\|\theta\|_q \geq C_{\delta,d}\sigma d^{\frac{1}{2q}}$, we indeed have

$$\mathbb{P}_\theta\left(\sum_{i=1}^d |X_i|^q \leq U_\delta \mid \xi_{b_\delta}\right) \leq \delta.$$

Definition of the test and conclusion
Given $\eta \in (0,1)$, let $\delta = \frac{\eta}{5}$ and take $d \geq \lceil \ln(2d/\delta)^q \rceil$. Let $T = \sum_{i=1}^d |X_i|^q$ and consider the test

$$\varphi(X) = \mathbb{1}\left\{T > \tau_\delta\right\} \vee \mathbb{1}\left\{\Xi_\delta\right\}$$

with τ_δ from (1.20) and Ξ_δ from (1.23). Firstly, under the null hypothesis, our preparation tells us that

$$\begin{aligned}\mathbb{P}_{H_0}(\varphi = 1) &\leq \mathbb{P}_{H_0}(T > \tau_\delta) + \mathbb{P}_{H_0}(\Xi_\delta)\\ &\leq \mathbb{P}_{H_0}(T > \tau_\delta \mid \xi_{b_\delta})\mathbb{P}(\xi_{b_\delta}) + \mathbb{P}_{H_0}(T > \tau_\delta \mid \xi_{b_\delta}^C)\mathbb{P}(\xi_{b_\delta}^C) + \mathbb{P}_{H_0}(\Xi_\delta)\\ &\leq \delta\cdot 1 + 1\cdot\delta + \delta = 3\delta.\end{aligned}$$

On the other hand, assume that H_1 holds, i.e. $\|\theta\|_q \geq c_{\delta,d}\sigma d^{\frac{1}{2q}}$ with $c_{\delta,d}$ from (1.25) through (1.27). We distinguish two cases: Firstly, if assumption $\mathcal{A}_\delta$ from (1.24) holds, we know that

$$\begin{aligned}\mathbb{P}_\theta(\varphi = 0) &\leq \mathbb{P}_\theta(T < \tau_\delta)\\ &= \mathbb{P}_\theta(T < \tau_\delta \mid \xi_{b_\delta})\mathbb{P}(\xi_\delta) + \mathbb{P}_\theta(T < \tau_\delta \mid \xi_\delta^C)\mathbb{P}(\xi_{b_\delta}^C)\\ &\leq \delta\cdot 1 + 1\cdot\delta = 2\delta.\end{aligned}$$

Secondly, if $\mathcal{A}_\delta$ is false, there is some $i^* \in \{1, 2, \dots, d\}$ such that $|\theta_{i^*}| > 2\sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}$. Using this index, we observe

$$\begin{aligned}
\mathbb{P}_\theta(\varphi = 0) &\leq \mathbb{P}_\theta(\Xi_\delta^C) \\
&\leq \mathbb{P}_\theta\left(|X_{i^*}| < \sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}\right) \\
&\leq \mathbb{P}_\theta\left(|\theta_{i^*}| - |\sigma\epsilon_{i^*}| < \sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}\right) \\
&\leq \mathbb{P}\left(|\sigma\epsilon_{i^*}| > \sigma\sqrt{2\ln\left(\frac{2d}{\delta}\right)}\right) \\
&\leq \frac{\delta}{d} < \delta
\end{aligned}$$

with (I) from (1.11). Therefore, as desired, with $\rho := c_{\eta/5,d}\sigma d^{\frac{1}{2q}}$ and $d \geq \lceil \ln(10d/\eta)^q \rceil$, we have

$$\mathbb{P}_{H_0}(\varphi = 1) + \sup_{\theta\in\mathbb{R}^d:\ \|\theta\|_q > \rho} \mathbb{P}_\theta(\varphi = 0) = \mathbb{P}_{H_0}(\varphi = 1) + \mathbb{P}_{H_1}(\varphi = 0) \leq \eta.$$

Discussion
We now study $c_{\delta,d}$ in order to understand the magnitude of $\rho := c_{\eta/5,d}\sigma d^{\frac{1}{2q}}$. Observe that $E_{b,p}$ is increasing in $b > 0$: Let g be the density of $\mathcal{N}(0,1)$. Then clearly

$$\frac{\mathrm{d}}{\mathrm{d}b}E_{b,q} = \frac{\mathrm{d}}{\mathrm{d}b}\left(\frac{\int_0^b |x|^q g(x)\ \mathrm{d}x}{\int_0^b g(x)\ \mathrm{d}x}\right) = \frac{|b|^q f(b)\int_0^b g(x)\ \mathrm{d}x - f(b)\int_0^b |x|^q g(x)\ \mathrm{d}x}{\left(\int_0^b g(x)\ \mathrm{d}x\right)^2}$$

exists and is positive. Since $\eta/5 \leq 1/5$, we know that $b_{\eta/5} \geq \sqrt{2\ln(10d)} \geq \sqrt{2\ln(10)}$ and hence $E_{b_{\eta/5},q-2} \geq E_{\sqrt{2\ln(10)},q-2}$. On the other hand, $E_{b_{\eta/5},q} \leq E_{\infty,q}$. Using these bounds in (1.8) leads to the absolute bound

$$\frac{1}{l_{b_{\eta/5},q}} \leq 8\max\left(1, \frac{4^{(q-2)/q}}{q(q-1)} \cdot \frac{E_{\infty,q}^{(q-2)/q}}{E_{\sqrt{\ln(10)},q-2}}\right).$$

Taking a look at the dominating factors and summands of $c_{\delta,d}$ and keeping (1.28) in mind, we see that, indeed, there is a constant C_q^2 such that

$$\rho = c_{\eta/5,d}\sigma d^{\frac{1}{2q}} \leq C_q^2 \cdot \sigma \cdot \sqrt{\ln\left(\frac{5}{\eta}\right)} \cdot \sqrt{2\ln\left(\frac{10d}{\eta}\right)} \cdot d^{\frac{1}{2q}}.$$

1.4.6 Proof of Theorem 1.6

For the null hypothesis, with $\theta = \mathbb{0}_d$, X clearly has the density function

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^d \prod_{i=1}^d \exp\left(-\frac{x_i^2}{2\sigma^2}\right), \quad x \in \mathbb{R}^d.$$

On the other hand, for the alternative hypothesis, we choose an ansatz of the form

$$\theta_1, \dots, \theta_d \overset{\text{iid}}{\sim} (1-h)\delta_0 + h\delta_z,$$

where the parameters $h, z > 0$ will be specified later in a way that ensures sufficient separation, i.e. $\|\theta\|_q$ of sufficient magnitude. The resulting density function for $x \in \mathbb{R}^d$ is given by

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}(x) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^d \prod_{i=1}^d \left((1-h)\exp\left(-\frac{x_i^2}{2\sigma^2}\right) + h\exp\left(-\frac{(x_i-z)^2}{2\sigma^2}\right)\right).$$

Statistical distance
We now have

$$\begin{aligned}
\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) &= \int_{\mathbb{R}^d} \frac{(\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho})^2}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}(x)\ \mathrm{d}x \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} \frac{\left((1-h)\exp\left(-\frac{x^2}{2\sigma^2}\right) + h\exp\left(-\frac{(x-z)^2}{2\sigma^2}\right)\right)^2}{\exp\left(-\frac{x^2}{2\sigma^2}\right)}\ \mathrm{d}x\right)^d \\
&= \left(\frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} (1-h)^2 \exp\left(-\frac{x^2}{2\sigma^2}\right)\ \mathrm{d}x\right. \\
&\quad + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} 2h(1-h)\exp\left(-\frac{(x-z)^2}{2\sigma^2}\right)\ \mathrm{d}x \\
&\quad \left. + \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\mathbb{R}} h^2 \exp\left(-\frac{(x-z)^2}{\sigma^2} + \frac{x^2}{2\sigma^2}\right)\ \mathrm{d}x\right)^d \\
&= \left((1-h)^2 + 2h(1-h) + h^2\exp\left((z/\sigma)^2\right)\right)^d \\
&= \left(1 + h^2\left(\exp((z/\sigma)^2) - 1\right)\right)^d
\end{aligned}$$

Let $\eta' \in [\frac{2}{3}, 1)$. Now, with $\ln(1+x) \le x$ for $x > -1$ we see that

$$\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) < 1 + 4(1-\eta')^2$$

is fulfilled if

$$h < \frac{1}{\sqrt{d}} \underbrace{\sqrt{\frac{\ln(1+4(1-\eta')^2)}{\exp((z/\sigma)^2) - 1}}}_{=:g}.$$

Specification of z, h, η
Motivated by the latter observation, we set $z = a\sigma$ with $a = a(\eta')$ chosen such that $g < 1$, e.g.

$$a(\eta') = \sqrt{\frac{1}{2}\ln\left(1 + \ln\left((1+4(1-\eta')^2)\right)\right)},$$

which justifies the choice $h = \frac{1}{\sqrt{d}}$.
At this point, the random number $Z = \|\theta\|_0$ has the distribution $\mathrm{Bin}(d, \frac{1}{\sqrt{d}})$ and in

particular the event $\{Z = 0\}$ has positive probability so that we have yet to ensure positive separation ρ between the priors.
To that end, we consider the event $\xi = \{Z \geq \frac{1}{5}\mathbb{E}[Z]\} = \{Z \geq \frac{1}{5}\sqrt{d}\}$. Writing Z as a sum $\sum_{i=1}^{d} B_i$ of iid variables $B_1, \dots, B_d \sim \text{Ber}(\frac{1}{\sqrt{d}})$ and assuming that $d \geq 36$, we see

$$-1 \leq \mathbb{E}[B_1] - B_1 \leq \frac{1}{6}$$

and therefore Hoeffding's inequality (see [Hoe63]) yields

$$\mathbb{P}\left(\sqrt{d} - \sum_{i=1}^{d} B_i > \frac{4}{5}\sqrt{d}\right) \leq \exp\left(-2\frac{\left(\frac{4}{5}\sqrt{d}\right)^2}{d \cdot \left(1 + \frac{1}{6}\right)^2}\right) \leq 0.4$$

and hence

$$\mathbb{P}(\xi) = 1 - \mathbb{P}\left(\sqrt{d} - \sum_{i=1}^{d} B_i > \frac{4}{5}\sqrt{d}\right) \geq 0.6.$$

Based on that, consider the conditional alternative prior distribution $\nu_\rho|\xi$ and assume that for some test φ and $\eta' > 0$ the relation

$$\mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_\rho}(\varphi = 0) > \eta'$$

holds. Then we can conclude

$$\begin{aligned}
\mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_\rho|\xi}(\varphi = 0) &= \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + 1 - \frac{\mathbb{P}_{\theta\sim\nu_\rho}(\{\varphi = 0\} \cap \xi)}{\mathbb{P}(\xi)} \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + 1 - \frac{5}{3}\mathbb{P}_{\theta\sim\nu_\rho}(\varphi = 0) \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \frac{5}{3}\mathbb{P}_{\theta\sim\nu_\rho}(\varphi = 0) - \frac{2}{3} \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_\rho}(\varphi = 0) - \frac{2}{3} \\
&> \eta' - \frac{2}{3}.
\end{aligned}$$

Therefore, if $d \geq 36$, the above construction with $h = \frac{1}{\sqrt{d}}$ and $z = a(\eta + \frac{2}{3}) \cdot \sigma$ yields a total error probability exceeding $\eta \in (0, \frac{1}{3})$ while the l_q-separation between the hypotheses (based on θ under ξ) fulfils

$$\rho \geq \|\theta\|_q \geq \left(\frac{1}{5}\sqrt{d} \cdot (a(\eta + \tfrac{2}{3})\sigma)^q\right)^{1/q} = \frac{a(\eta + \frac{2}{3})}{5^{1/q}} \cdot \sigma \cdot d^{\frac{1}{2q}}.$$

This corresponds to the lower bound claimed in the theorem.

Chapter 2

Testing Convex Hypotheses on Gaussian Means

We consider composite-composite testing problems for the expectation in the Gaussian sequence model where the null hypothesis corresponds to a closed convex subset $\mathcal{C}$ of $\mathbb{R}^d$. We adopt a minimax point of view and our primary objective is to describe the smallest Euclidean distance between the null and alternative hypotheses such that there is a test with small total error probability. In particular, we focus on the dependence of this distance on the dimension d and variance $\frac{1}{n}$ giving rise to the minimax separation rate. We provide lower and upper bounds on this rate for different smooth and non-smooth choices for $\mathcal{C}$.
This chapter is based on the article [BCG18].

2.1 Introduction

Let $d, n \in \mathbb{N}\backslash\{0\}$. Similarly as in section 1.2 we consider the d-dimensional random vector

$$X = \theta + \frac{1}{\sqrt{n}}\epsilon,$$

where ϵ is a standard Gaussian vector. Note that by construction, the variance scaling parameter n may also be interpreted as sample size since the distribution of X is precisely the distribution of the mean of n iid observations from $\mathcal{N}(\theta, \mathbb{I}_d)$.

Now, let $\Theta_0 = \mathcal{C} \subsetneq \mathbb{R}^d$ be closed, nonempty and convex. For $\theta \in \Theta = \mathbb{R}^d$ we write

$$\mathrm{dist}_{\mathcal{C}}(\theta) := \inf_{c \in \mathcal{C}} \|\theta - c\|,$$

where $\|\cdot\| := \|\cdot\|_2$ denotes Euclidean (l_2) norm, see (1.5). A corresponding open Euclidean ball with center $z \in \mathbb{R}^k$ and radius $r > 0$ is denoted $B_k(z, r)$. Moreover, we indicate vector concatenation by $[\cdot, \cdot]$, so that, for instance, $[z, 1] \in \mathbb{R}^d$ if $z \in \mathbb{R}^{d-1}$.

Testing Problem

Given $\rho > 0$, we are interested in the testing problem

$$H_0: \ \theta \in \mathcal{C} \ \text{ vs. } \ H_1: \ \mathrm{dist}_{\mathcal{C}}(\theta) > \rho. \tag{2.1}$$

Hence, in the framework of chapter 1, the problem is given by

$$\begin{cases} \Theta &= \mathbb{R}^d, \\ \mathbb{P}_\theta &= \mathcal{N}\left(\theta, \frac{1}{n}\mathbb{I}_d\right) \quad \text{for} \quad \theta \in \Theta, \\ \Theta_0 &= \mathcal{C}, \\ \Theta_\rho &= \{\theta \in \Theta \ ; \ \mathrm{dist}_{\mathcal{C}}(\theta) > \rho\}. \end{cases}$$

Thus, for $\eta \in (0, 1)$, we now aim at describing

$$\rho^*(\eta) = \rho^*_{\mathcal{C}}(\eta) = \inf\left\{\rho > 0 \ ; \ \exists \text{ test } \varphi : \ \sup_{\theta\in\mathcal{C}} \mathbb{P}_\theta(\varphi = 1) + \sup_{\theta\in\Theta_\rho} \mathbb{P}_\theta(\varphi = 0) \ \leq \ \eta\right\}.$$

In particular, similar as in section 1.2, we focus on the dependence of $\rho^*_{\mathcal{C}}(\eta)$ on the dimension d and variance scaling parameter n.

Related Questions and Literature

An instance of this problem that we have already covered is signal-detection, i.e. the case where $\mathcal{C}$ is a singleton. We can deduce from Theorem 1.1 (setting $\sigma = \frac{1}{\sqrt{n}}$) that

$$\frac{d^{1/4}}{\sqrt{n}}.$$

In its general form however, this problem is a composite-composite testing problem (i.e. neither $\mathcal{C}$ nor Θ_ρ is only a singleton). A versatile way of solving such testing problems was introduced in [GP05], where the authors combine signal-detection ideas with a covering of the null hypothesis, for deriving minimax optimal testing procedures for composite-composite testing problems, provided that the null hypothesis is not too large (i.e. that its entropy number is not too large, see Assumption (A3) in [GP05]). In this case, the authors prove that the minimax-optimal testing separation rate is the same as the signal-detection separation rate, namely $\frac{d^{1/4}}{\sqrt{n}}$. This idea can be generalised also to the case where the null hypothesis is "too large" (when Assumption (A3) in [GP05] is not satisfied); the approach then implies that an upper bound on the minimax rate of separation is the sum of the signal-detection rate and the optimal estimation rate in the null hypothesis $\mathcal{C}$ – see [BN13] for an illustration of this for a specific convex shape. Using this technique, one finds that the smaller the entropy of $\mathcal{C}$, the smaller the separation rate.
This idea has the advantage of generality, but is nevertheless sub-optimal in many simple cases. For instance, if $\mathcal{C}$ is a half-space, the minimax-optimal separation rate is $\frac{1}{\sqrt{n}}$, which is much smaller than the minimax-opimal signal-detection rate, even though a half-space has a much larger entropy (it is even infinite) and larger dimension than a single point. See Section 2.2 for an extended discussion on this case. This highlights the fact that for such a testing problem, it is in many cases not the entropy, or size, of the null hypothesis that drives the rate, but rather some other properties of the shape of $\mathcal{C}$.
In order to overcome the limitations of this approach, some other ideas were proposed. A first line of work can be found in [BHL05], where the authors consider our testing problem for separation in $\|.\|_\infty$-norm rather than $\|.\|_2$-norm. Since any

convex set can be written as a intersection of half-spaces, they rewrite the problem as a multiple testing problem. This approach is quite fruitful, but the $\|.\|_\infty$-norm results translate in a non-optimal way to $\|.\|_2$-norm in terms of the dependence on the dimension d, particularly for large d. A second main direction that was investigated was to consider testing for some *specific* convex shapes, as e.g. the cone of positive, monotone, or convex functions, see e.g. [JN02], or also balls for some metrics [LNS99, Car15]. These papers exhibit the minimax-optimal separation distance – or near optimal distance, in some cases of [JN02] and [LNS99] – for the specific convex shapes that are considered, namely cones and smoothness balls. The models considered in these works are different from our model as they consider functional estimation; also, they do not provide results for more general choices of the null hypothesis. In Sections 2.2 and 2.4, we derive results for our model and shapes related to those of these papers - namely the positive orthant and the Euclidian ball - in order to relate our work with these earlier results. Finally, a last type of results that are related to our problem is the case where the null hypothesis can be parametrised, see e.g. [CD13] where the authors consider shapes that can be parametrised by a quadratic functional. This approach and their results suggest that the smoothness of the shape of $\mathcal{C}$ has an impact on the testing rate.
In this chapter, we want to take a more general approach toward the testing problem (2.1). In section 2.2, we expose the range of possible separation rates by demonstrating that, without any further assumptions on $\mathcal{C}$, the statement

$$\frac{1}{\sqrt{n}} \lesssim \rho^*_{\mathcal{C}} \lesssim \frac{\sqrt{d}}{\sqrt{n}} \tag{2.2}$$

is sharp up to $\ln(d)$-factors. After that, in Sections 2.3 and 2.4, we investigate the potential of a geometric smoothness property of the boundary of $\mathcal{C}$. Despite its simplicity, this property takes us quite far: In particular, given any separation rate satisfying (2.2), it allows for constructing a set $\mathcal{C}$ exhibiting this rate up to $\ln(d)$-factors.

2.2 A General Guarantee and Extreme Cases

The quantity $\rho^*_{\mathcal{C}}$ clearly depends on $\mathcal{C}$.
Let us firstly examine a simple, essentially one-dimensional case, namely a half-space.

Theorem 2.1. Let $\mathcal{C} = \mathcal{C}_{\mathrm{HS}} := \mathbb{R}^{d-1} \times (-\infty, 0]$ (if $d = 1$, $\mathcal{C}_{\mathrm{HS}} = (-\infty, 0]$). Then, in the testing problem (2.1), we have

$$\sqrt{\frac{1}{2n}\ln(1 + 4(1-\eta)^2)} \ \le\ \rho^*_{\mathcal{C}_{\mathrm{HS}}}(\eta) \ \le\ \sqrt{\frac{8}{n}\ln\left(\frac{1}{\eta}\right)}$$

and therefore

$$\rho^*_{\mathcal{C}_{\mathrm{HS}}} \sim \frac{1}{\sqrt{n}}.$$

◁

Remark 2.2. As can be seen in the proof (section 2.5.2), this testing problem is essentially equivalent to the problem $\theta = 0$ vs. $\theta = \rho$ in dimension $d = 1$, so that, alternatively, the rate $\frac{1}{\sqrt{n}}$ can be obtained by analysing the optimal test in the sense of Neyman-Pearson. Furthermore, and in fact closely related to that, note that the lower bound in the previous theorem is valid for any choice of closed convex set $\mathcal{C}$ such that $\mathcal{C}$ and $\mathbb{R}^d \backslash \mathcal{C}$ are non-empty:

$$\frac{1}{\sqrt{n}} \lesssim \rho^*_{\mathcal{C}}.$$

Indeed, we find this rate by considering a fixed pair of points $(\theta_0, \theta_1) \in \mathcal{C} \times \Theta_\rho$ that minimises the distance between $\mathcal{C}$ and Θ_ρ, i.e. $\|\theta_0 - \theta_1\| = \rho$. That seems to have firstly been discussed in [Bur79]; other related (classical) literature would be for instance [Che52] and [Ing00]. ◁

Now, on the other hand, making no additional assumptions about $\mathcal{C}$, a natural choice φ for solving (2.1) is a plug-in test based on confidence balls. This gives rise to the following general upper bound:

Theorem 2.3. Let $\mathcal{C}$ be an arbitrary closed convex subset of $\mathbb{R}^d$ such that $\mathcal{C}$ and $\mathbb{R}^d \setminus \mathcal{C}$ are non-empty. Then, in the testing problem (2.1), we have

$$\rho^*_{\mathcal{C}}(\eta) \;\leq\; 2\sqrt{\frac{d}{n} + \frac{2}{n}\sqrt{d \ln\left(\frac{2}{\eta}\right)} + \frac{2}{n}\ln\left(\frac{2}{\eta}\right)}$$

and therefore

$$\rho^*_{\mathcal{C}} \;\lesssim\; \frac{\sqrt{d}}{\sqrt{n}}.$$

◁

Remark 2.4. Note that this upper bound is the rate of estimation of θ in l_2 norm in the model (2.1), see (1.11)). ◁

Remark 2.5. From Remark 2.2 and Theorem 2.3 it is clear that

$$\frac{1}{\sqrt{n}} \lesssim \rho^*_{\mathcal{C}} \;\lesssim\; \frac{\sqrt{d}}{\sqrt{n}}$$

whenever $\mathcal{C}$ is a closed convex subset of $\mathbb{R}^d$ such that $\mathcal{C}$ and $\mathbb{R}^d \backslash \mathcal{C}$ are non-empty. ◁

Given this observation, it is natural to ask if the upper bound in Theorem 2.3 is also sharp in the sense that there is a choice of $\mathcal{C}$ that requires the separation rate $\frac{\sqrt{d}}{\sqrt{n}}$, at least up to logarithmic factors. It turns out that the answer is yes when $\mathcal{C}$ is taken to be an orthant:

Theorem 2.6. Let $\mathcal{C} = \mathcal{C}_{\mathrm{O}} := (-\infty, 0]^d$, $d \geq 42$, $\eta \in (0, \frac{8}{9})$ and

$$M_\eta := \max\left(32, \left\lceil \frac{2}{1-\ln(2)} \ln(d) + 1 + \frac{2}{1-\ln(2)} \ln\left(\frac{1.8}{\frac{8}{9} - \eta}\right) \right\rceil\right).$$

Then, for the testing problem (2.1), we have

$$\rho^*_{\mathcal{C}_{\mathrm{O}}}(\eta) \;\geq\; \frac{1}{28} \frac{1}{M_\eta^{3/2}} \frac{\sqrt{d}}{\sqrt{n}}$$

and therefore, if d is large enough in the sense that $M_\eta \leq C\ln(d)$ for some $C > 0$,

$$\frac{\sqrt{d}}{\ln(d)^{3/2}\sqrt{n}} \lesssim \rho^*_{\mathcal{C}_\mathrm{O}} \lesssim \frac{\sqrt{d}}{\sqrt{n}}.$$

◁

This result heavily relies on tailoring the priors such that they have a certain number of moments in common. A related application of this approach can be found in the proof of Theorem 1 in [JN02], see also for instance [CL11].
Furthermore, in the proof of Theorem 2.6 (section 2.5.4) we face a similar situation as in the proof of Theorem 1.6 (section 1.6)) as the preliminary prior for the alternative hypothesis has positive support inside $\mathcal{C}_\mathrm{O}$. This is handled through a similar post-processing step.

2.3 A Simple Smoothness-Type Property

Clearly, the two extreme cases $\mathcal{C}_\mathrm{HS}$ and $\mathcal{C}_\mathrm{O}$ differ significantly with respect to smoothness of their boundaries. Based on this observation, in order to be able to handle $\rho^*_\mathcal{C}$ more flexibly, we propose to describe convex sets by their boundaries' degree of smoothness, where the boundary of a set $S \in \mathbb{R}^d$ is denoted by ∂S and its closure by $\overline{S} = S \cup \partial S$. To begin with, we examine the potential of the following very simple and purely geometric smoothness concept:

Definition 2.7. Let $R \geq 0$ and $S \subseteq \mathbb{R}^d$ with non-empty interior. S is called ***R*-rounded** if

$$\forall x \in \partial S \; \exists z \in S: \; x \in \overline{B_d(z,R)} \subseteq \overline{S}.$$

◁

Remark 2.8. Note that R-rounding is a stronger requirement the higher the value of R, i.e. intuitively the degree of the boundary's smoothness grows with R. In particular, a half space $\mathcal{C}_\mathrm{HS}$ is ∞-rounded, a ball $B_d(z,R)$ (with $z \in \mathbb{R}^d$, $R \in (0,\infty)$) is R-rounded and the orthant $\mathcal{C}_\mathrm{O}$ is $0-$rounded. The definition of $R-$rounding is closely related to the so-called R-rolling condition employed in [ACC17a]. In fact, R-rounding of S is equivalent to saying that $\mathbb{R}^d \backslash S$ fulfils the R-rolling condition.
Another related concept worth mentioning is the radius of curvature, though the connection is more subtle: The radius of curvature at a point $x \in \partial S$ would be the radius r of the ball B that best fits ∂S in the sense of a common tangential hyperplane of ∂S and B at x and common analytical curvature, see for instance [Cas96]. Hence, it is possible that the infimum R of these radii r with respect to $x \in \partial S$ corresponds to the parameter R in our previous definition. However, we can then still not easily guarantee that the resulting balls B of the form $B_d(z,R)$ fulfil $\overline{B_d(z,R)} \subseteq \overline{S}$ as required in Definition 2.7. ◁

Since smoothness is usually defined as a local property of a function, we provide a suggestion for how to cast the above concept in that context for a closed convex set $\mathcal{C}$: Given any $x \in \partial\mathcal{C}$, without loss of generality (w.l.o.g.) apply a rotation and translation G such that $x' = G(x) = \mathbb{0}_d$ and $\mathcal{C}' := G(\mathcal{C}) \subseteq \mathbb{R}^{d-1} \times [0,\infty)$. Now, assume that there is an $r \in (0,R]$ and a function $f : B \to [0,\infty)$, where

$B = B_{d-1}(\mathbb{O}_{d-1}, r)$, such that its graph is contained in $\partial\mathcal{C}'$ and $\mathcal{C}' \cap (B \times [0,\infty))$ is contained in the epigraph of f – see the figure below for an illustration. The following lemma states sufficient conditions for R-rounding locally at $x \in \mathcal{C}'$, i.e. at $G^{-1}(x) \in \partial\mathcal{C}$:

Lemma 2.9. In the situation described in the latter paragraph, if f is twice differentiable on B (i.e. the gradient $\nabla f(\cdot)$ and Hessian matrix $Hf(\cdot)$ exist), the following conditions are sufficient in order that the graph of f remains below $B_d([\mathbb{O}_{d-1}, R], R)$, i.e. $\mathcal{C}'$ is locally R-rounded at $\mathbb{O}_d$.

$$\begin{cases} \nabla f(\mathbb{O}_{d-1}) = \mathbb{O}_{d-1}, \\ \forall x \in B\backslash\{\mathbb{O}_{d-1}\} : \quad 0 \leq \lambda_{\min}(Hf(x)),\ \lambda_{\max}(Hf(x)) \leq \dfrac{1}{R}, \end{cases}$$

where $\lambda_{\min}(\cdot)$ and $\lambda_{\max}(\cdot)$ are the lowest and highest eigenvalues of a real symmetric matrix, respectively.

◁

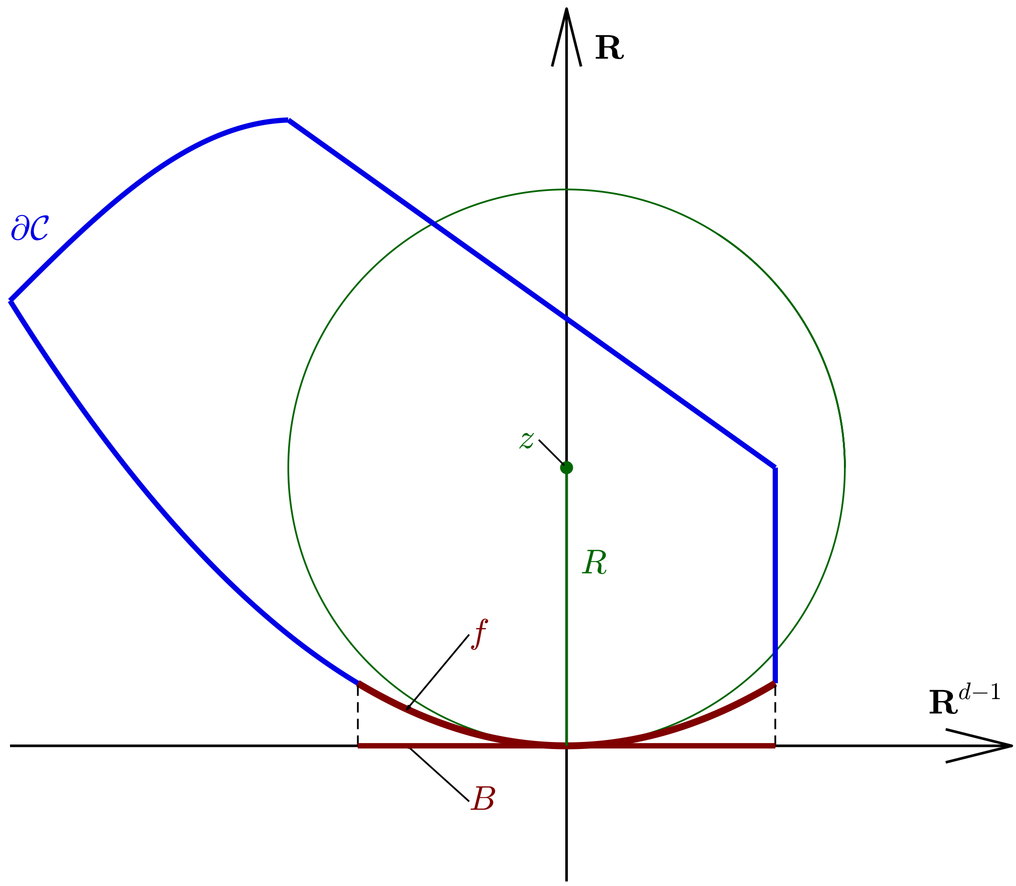

Figure 2.1: In this example $\mathcal{C}$ is only 0-rounded in the sense of Definition 2.7, but in the local sense of Lemma 2.9, there are points $x \in \partial\mathcal{C}$ with 0-rounding, ∞-rounding and "non-degenerate" rounding such as $\mathbb{O}_d$, where, however, the maximum admissible radius r of B is strictly smaller than R.

Now, let us examine how the additional assumption of $R-$rounding may affect the general upper bound of Theorem 2.3:

Theorem 2.10. If $\mathcal{C}$ is $R-$rounded for some $R \in (0, \infty)$, for the testing problem (2.1), we have

$$\rho^*_{\mathcal{C}}(\eta) \;\leq\; \sqrt{\frac{2}{n}\ln\left(\frac{8}{\eta}\right)} + \frac{d}{2nR} + \frac{2}{nR}\sqrt{d\ln\left(\frac{4}{\eta}\right)} + \frac{1}{nR}\ln\left(\frac{4}{\eta}\right) + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)}$$

and therefore, taking Theorem 2.3 into account,

$$\rho^*_{\mathcal{C}} \;\lesssim\; \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{\sqrt{d}}{\sqrt{n}}, \frac{d}{nR}\right)\right).$$

◁

The following result confirms that this upper bound can be sharp up to $\ln(d)$ factors, namely in the case where $\mathcal{C}$ is taken as an R-inflated orthant:

Theorem 2.11. Let $d \geq 43$, $\eta \in (0, \frac{8}{9})$ and

$$\mathcal{C} = \mathcal{C}_{\mathrm{IO}} = \mathcal{C}_{\mathrm{O}} + B(\mathbb{0}_d, R) = \bigcup_{z \in \mathcal{C}_{\mathrm{O}}} B(z, R),$$

where $\mathcal{C}_{\mathrm{O}} = (-\infty, 0]^d$ is the orthant from Theorem 2.6. Furthermore, let

$$M_\eta := \max\left(32, \left\lceil \frac{2}{1-\ln(2)}\ln(d-1) + 1 + \frac{2}{1-\ln(2)}\ln\left(\frac{1.8}{\frac{8}{9}-\eta}\right)\right\rceil\right).$$

Then, in the testing problem (2.1), we have with $s = \frac{\sqrt{3}}{28}\frac{1}{M_\eta^{3/2}\sqrt{n}}$

$$\rho^*_{\mathcal{C}_{\mathrm{IO}}}(\eta) \;\geq\; \frac{1}{12}\min\left(\frac{(d-1)s^2}{R}, \sqrt{3}\sqrt{d-1}\; s\right)$$

and therefore, if d is large enough in the sense that $M_\eta \leq C\ln(d-1)$ for some $C > 0$,

$$\rho^*_{\mathcal{C}_{\mathrm{IO}}} \;\gtrsim\; \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{1}{\ln(d)^3}\cdot\frac{d}{nR}, \frac{1}{\ln(d)^{3/2}}\cdot\frac{\sqrt{d}}{\sqrt{n}}\right)\right).$$

◁

2.4 Discussion

The concept of R-rounding allows for the construction of hypotheses $\mathcal{C}$ with any separation rate $\frac{1}{\sqrt{n}} \lesssim \rho^*_{\mathcal{C}} \lesssim \frac{\sqrt{d}}{\sqrt{n}}$, up to $\ln(d)$-factors. On the other hand, we must acknowledge that $R-$rounding is too weak a concept to fully describe the difficulty of testing an arbitrary $\mathcal{C}$; an examination of the natural R-rounded set, namely a ball of radius R, provides clear evidence of this drawback. The result is a direct generalisation of the known rate $\rho^*_{\mathcal{C}} \sim \frac{d^{\frac{1}{4}}}{\sqrt{n}}$ in the signal-detection setting, see [Bar02].

Theorem 2.12. Let $\eta \in (0,1)$ and $d \geq \ln(2/\eta)$. If $\mathcal{C} = \mathcal{C}_{\mathrm{B}} = B_d(z,R)$, $z \in \mathbb{R}^d$ and $R > 0$, for the testing problem (2.1), we have for $s := \frac{\sqrt{d-1}}{n}\sqrt{\ln(1+4(1-\eta)^2)}$

$$\rho^*_{\mathcal{C}_{\mathrm{B}}}(\eta) \;\geq\; \frac{s}{2\sqrt{s+R^2}} \;\gtrsim\; \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)$$

and also

$$\begin{aligned}\rho^*_{\mathcal{C}_{\mathrm{B}}}(\eta) \;&\leq\; \min\left(2\sqrt{2}\frac{d^{\frac{1}{4}}}{n^{\frac{1}{2}}}, \frac{2\sqrt{d}}{nR + 2\sqrt{n\ln\left(\frac{2}{\eta}\right)}}\right)\sqrt{\ln\left(\frac{2}{\eta}\right)} + 3\sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)}\\ &\lesssim\; \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)\right).\end{aligned}$$

Therefore,

$$\rho^*_{\mathcal{C}_{\mathrm{B}}} \;\sim\; \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)\right).$$

◁

Clearly, Theorem 2.11 does not capture this case. As a consequence, future work will be concerned with finding a stronger concept, possibly a localised version of $R-$rounding, that ideally allows for describing $\rho^*_{\mathcal{C}}$ for any choice of $\mathcal{C}$. However, we suspect this to be quite an ambitious goal.

2.5 Proofs

2.5.1 Frequently used Bounds for Expressions Containing Square Roots

We will employ the following bounds on several occasions which makes it convenient to mention them here.

Lemma. For any $a > 0, b \in \mathbb{R}$, we have

$$\frac{a}{2\sqrt{a+b^2}} \leq \sqrt{a+b^2} - b \leq \frac{a}{2b} \tag{2.3}$$

and for any $b > 0$, $a \leq b^2$ we have

$$b - \sqrt{b^2 - a} \geq \frac{a}{2b}. \tag{2.4}$$

Proof. Firstly, through Taylor expansion of $\sqrt{a+b^2} - b$ as a function in a, we see that there is a $\xi \in (0,a)$ such that

$$\sqrt{a+b^2} - b = \frac{a}{2\sqrt{\xi + b^2}}.$$

Now, with $\xi \geq 0$ and $\xi \leq a$ we obtain the upper and lower bounds in (2.3), respectively. Secondly, explicit calculation tells us that

$$b - \sqrt{b^2 - a} \geq \frac{a}{2b} \;\Leftrightarrow\; \frac{a^2}{4b^2} \geq 0,$$

which concludes the proof. □

2.5.2 Proof of Theorem 2.1

We prove independently that the order of $\rho^*_{\mathcal{C}_{\mathrm{HS}}}$ is lower and upper bounded by $\frac{1}{\sqrt{n}}$.

Lower Bound

In accordance with the framework in section 1.4.1, we verify that the bound holds in the special case $\nu_0 = \delta_{\mathbb{O}_d}$ and $\nu_\rho = \delta_{\rho \cdot e_d}$, where e_d is the last standard basis vector $e_d = [\mathbb{O}_{d-1}, 1]$. Since both the null and alternative hypotheses are simple, the corresponding density functions $\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}(x)$ and $\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}(x)$ are readily given and we obtain

$$\begin{aligned}
\int_{\mathbb{R}^d} \frac{\mathrm{d}\mathbb{P}^2_{\theta\sim\nu_\rho}}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}(x)\,\mathrm{d}x &= \sqrt{\frac{n}{2\pi}}^{\,d} \int_{\mathbb{R}^d} \exp\left(-n(x_d-\rho)^2 + \frac{n}{2}x_d^2 - \frac{n}{2}(\|x\|^2 - x_d^2)\right)\,\mathrm{d}x \\
&= \sqrt{\frac{n}{2\pi}} \int_{\mathbb{R}} \exp\left(-n(x_d-\rho)^2 + \frac{n}{2}x_d^2\right)\,\mathrm{d}x_d \\
&= \sqrt{\frac{n}{2\pi}} \exp(n\rho^2) \int_{\mathbb{R}} \exp\left(-\frac{n}{2}(x_d - 2\rho)^2\right)\,\mathrm{d}x_d \\
&= \exp(n\rho^2).
\end{aligned}$$

Therefore inequality (1.10) is satisfied for

$$\rho < \sqrt{\frac{1}{n}\ln(1 + 4(1-\eta)^2)}$$

and hence particularly for

$$\rho \leq \sqrt{\frac{1}{2n}\ln(1 + 4(1-\eta)^2)}.$$

This yields the claim.

Upper Bound

Let $\tau_{\frac{\eta}{2}} = \sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)}$ and define the test

$$\varphi(X) = \mathbb{1}_{\{X_d > \tau_{\frac{\eta}{2}}\}}.$$

Then for any $\theta \in \mathcal{C}$, we have with (I) of (1.11)

$$\mathbb{P}_\theta(\varphi = 1) \leq \mathbb{P}\left(\tfrac{1}{\sqrt{n}}\epsilon_d \geq \tau_{\frac{\eta}{2}}\right) \leq \frac{\eta}{2}.$$

On the other hand, let now $\rho = 2\tau_{\frac{\eta}{2}}$. Then for any $\theta \in \Theta_\rho$, we similarly have

$$\mathbb{P}_\theta(\varphi = 0) \leq \mathbb{P}\left(2\tau_{\frac{\eta}{2}} + \tfrac{1}{\sqrt{n}}\epsilon_d \leq \tau_{\frac{\eta}{2}}\right) \leq \mathbb{P}\left(\tfrac{1}{\sqrt{n}}\epsilon_d \leq -\tau_{\frac{\eta}{2}}\right) \leq \frac{\eta}{2}.$$

This concludes the proof and clearly $\rho \sim \frac{1}{\sqrt{n}}$.

2.5.3 Proof of Theorem 2.3

Let $\tau_{\frac{\eta}{2}} = \frac{d}{n} + \frac{2}{n}\sqrt{d \ln\left(\frac{2}{\eta}\right)} + \frac{2}{n}\ln\left(\frac{2}{\eta}\right)$ and define the test

$$\varphi(X) = \mathbb{1}_{\{\mathrm{dist}_{\mathcal{C}}(X) > \sqrt{\tau_{\frac{\eta}{2}}}\}}.$$

Then for any $\theta \in \mathcal{C}$, we have with (II) in (1.11)

$$\mathbb{P}_\theta(\varphi = 1) \leq \mathbb{P}_\theta\left(\|X - \theta\| > \sqrt{\tau_{\frac{\eta}{2}}}\right) \leq \mathbb{P}\left(\tfrac{1}{n}\|\epsilon\|^2 \geq \tau_{\frac{\eta}{2}}\right) \leq \frac{\eta}{2}.$$

On the other hand, let now $\rho = 2\sqrt{\tau_\delta}$ and $\theta \in \Theta_\rho$ arbitrary. Then similarly

$$\mathrm{dist}_{\mathcal{C}}(X) \leq \sqrt{\tau_{\frac{\eta}{2}}} \;\Rightarrow\; \|X - \theta\| \geq \sqrt{\tau_{\frac{\eta}{2}}}$$

and hence

$$\mathbb{P}_\theta(\varphi = 0) \leq \frac{\eta}{2}.$$

This concludes the proof and we see $\sqrt{\tau_{\frac{\eta}{2}}} \sim \sqrt{\frac{d}{n}}$.

2.5.4 Proof of Theorem 2.6

The arguments of this proof are related to the ones used in [JN02] and [CL11]. We decompose the proof into several steps.

Choice of priors

We make use of the following lemma used and explained in [JN02] :

Lemma. For any $M \in \mathbb{N}$ and $b > 0$, there are distributions $\widetilde{\nu}_0$ and $\widetilde{\nu}_1$ with the following properties:

$$\begin{aligned}
&\text{(I)} && \mathrm{supp}(\widetilde{\nu}_0) \subseteq [-b, 0], \;\; \mathrm{supp}(\widetilde{\nu}_1) \subseteq [-b, 0] \cup \left\{\tfrac{b}{4M^2}\right\}, \\
&\text{(II)} && \widetilde{\nu}_1\left(\left\{\tfrac{b}{4M^2}\right\}\right) \geq \tfrac{1}{2}, \\
&\text{(III)} && \forall k \in \{0, 1, \dots, M\} : \; \textstyle\int z^k \, \widetilde{\nu}_0(\mathrm{d}z) = \int z^k \, \widetilde{\nu}_1(\mathrm{d}z).
\end{aligned} \tag{2.5}$$

For now, let $\widetilde{\nu}_i$ be such distributions and $\nu_i = \widetilde{\nu}_i^{\otimes d}$, $i \in \{0, 1\}$; M, b and ρ will be specified later. Furthermore, writing $\sigma^2 = \frac{1}{n}$, let

$$\mathbb{P}_{\theta \sim \nu_i} = \left(\widetilde{\nu}_i * \mathcal{N}(0, \sigma^2)\right)^{\otimes d}, \quad i \in \{0, 1\},$$

where $*$ denotes convolution. Clearly, the corresponding density function can be written as

$$\mathrm{d}\mathbb{P}_{\theta \sim \nu_i}(x) = \prod_{j=1}^{d} \left(\mathbb{E}_{\theta_j \sim \widetilde{\nu}_i}[\phi(x_j; \theta_j, \sigma^2)]\right), \quad i \in \{0, 1\},$$

where $\phi(x;\theta,\sigma^2)$ is the density of $\mathcal{N}(\theta,\sigma^2)$. It will be convenient to examine the case $d=1$, denoted by $\widetilde{\mathbb{P}}_{\theta\sim\nu_i}$.

Controlling the total variation distance
Based on our construction, we have for $i\in\{0,1\}$ and fixed $x\in\mathbb{R}$

$$
\begin{aligned}
\mathbb{E}_{\theta\sim\widetilde{\nu}_i}[\phi(x;\theta,\sigma^2)] &= \phi(x;0,\sigma^2)\int \exp\left(\frac{2x\theta-\theta^2}{2\sigma^2}\right)\widetilde{\nu}_i(\mathrm{d}\theta) \\
&= \phi(x;0,\sigma^2)\int \sum_{k=0}^{\infty}\frac{1}{k!}\left(\frac{2x\theta-\theta^2}{2\sigma^2}\right)^k \widetilde{\nu}_i(\mathrm{d}\theta) \\
&= \phi(x;0,\sigma^2)\int \sum_{k=0}^{\infty}\frac{1}{k!(2\sigma^2)^k}(2x\theta-\theta^2)^k\ \widetilde{\nu}_i(\mathrm{d}\theta). \qquad (2.6)
\end{aligned}
$$

Let now

$$
D_k(x) := \int (2x\theta-\theta^2)^k\ \widetilde{\nu}_1(\mathrm{d}\theta) - \int (2x\theta-\theta^2)^k\ \widetilde{\nu}_0(\mathrm{d}\theta).
$$

Then (2.6) in conjunction with (2.5.III) tells us that

$$
\frac{\mathbb{E}_{\theta\sim\widetilde{\nu}_1}[\phi(x;\theta,\sigma^2)]-\mathbb{E}_{\theta\sim\widetilde{\nu}_0}[\phi(x;\theta,\sigma^2)]}{\phi(x;0,\sigma^2)} = \sum_{k=\lfloor M/2\rfloor+1}^{\infty}\frac{1}{k!(2\sigma^2)^k}D_k(x).
$$

and thus

$$
\begin{aligned}
\|\widetilde{\mathbb{P}}_{\theta\sim\nu_1}-\widetilde{\mathbb{P}}_{\theta\sim\nu_0}\|_{\mathrm{TV}} &= \int \left|\mathbb{E}_{\theta\sim\widetilde{\nu}_1}[\phi(x;\theta,\sigma^2)]-\mathbb{E}_{\theta\sim\widetilde{\nu}_0}[\phi(x;\theta,\sigma^2)]\right|\ \mathrm{d}x \\
&\le \sum_{k=\lfloor M/2\rfloor+1}^{\infty}\frac{1}{k!(2\sigma^2)^k}\left|\int D_k(x)\phi(x;0,\sigma^2)\ \mathrm{d}x\right|. \qquad (2.7)
\end{aligned}
$$

We take a moment to upper bound the individual summands: Since

$$
(2x\theta-\theta^2)^k \le 4^k|x|^k b^k + 2^k b^{2k}
$$

and, by a classical formula for Gaussian absolute moments (see [Win12]),

$$
\int |x|^k\phi(x,0,\sigma^2)\ \mathrm{d}x = \frac{\sigma^k\sqrt{2}^k}{\sqrt{\pi}}\Gamma((k+1)/2) \le \frac{\sigma^k\sqrt{2}^k}{\sqrt{\pi}}\left\lceil\frac{k}{2}\right\rceil!,
$$

we have

$$
\begin{aligned}
\left|\int D_k(x)\phi(x;0,\sigma^2)\ \mathrm{d}x\right| &\le 2\left(4^k b^k\int |x|^k\phi(x;0,\sigma^2)\ \mathrm{d}x + 2^k b^{2k}\right) \\
&\le 2\left(\frac{1}{\sqrt{\pi}}\left(4\sqrt{2}b\sigma\right)^k\left\lceil\frac{k}{2}\right\rceil! + 2^k b^{2k}\right).
\end{aligned}
$$

Now through Stirling's approximation and elementary manipulation, with $M \geq 32$ we obtain

$$\begin{aligned}\frac{\lceil \frac{k}{2} \rceil !}{k!} &\leq \frac{\mathrm{e}}{\sqrt{2\pi}} \frac{\lceil \frac{k}{2} \rceil^{\lceil \frac{k}{2} \rceil}}{k^{k+\frac{1}{2}}} \underbrace{\frac{\lceil \frac{k}{2} \rceil^{1/2}}{\mathrm{e}^{\lceil \frac{k}{2} \rceil}}}_{\leq \sqrt{3}/k} \mathrm{e}^k \\ &\leq \frac{\mathrm{e}\sqrt{3}}{\sqrt{2\pi}} \left(\sqrt{\frac{k+1}{2k^2}} \right)^{k+1} \mathrm{e}^k \\ &\leq \frac{\mathrm{e}\sqrt{3}}{5\sqrt{2\pi}} \left(\sqrt{\frac{17}{32}} \frac{1}{\sqrt{k}} \right)^{k} \\ &\leq \frac{1}{2} \left(\sqrt{\frac{17}{32}} \frac{1}{\sqrt{k}} \right)^{k}\end{aligned}$$

and

$$k! \geq \sqrt{2\pi} k^k \sqrt{k} \mathrm{e}^{-k} \geq 4\sqrt{\pi} \left(\frac{k}{\mathrm{e}} \right)^k .$$

That yields

$$\begin{aligned}\frac{1}{k!(2\sigma^2)^k} \left| \int D_k(x) \phi(x; 0, \sigma^2) \, \mathrm{d}x \right| &\leq 2 \left(\frac{1}{k!\sqrt{\pi}} \left(\frac{4}{\sqrt{2}} \frac{b}{\sigma} \right)^k \left\lceil \frac{k}{2} \right\rceil ! + \frac{1}{k!} \left(\frac{b}{\sigma} \right)^{2k} \right) \\ &\leq 2 \left(\frac{1}{2} \left(\frac{\sqrt{17}}{2} \frac{b}{\sigma\sqrt{k}} \right)^k + \frac{1}{4\sqrt{\pi}} \left(\mathrm{e} \frac{b^2}{\sigma^2 k} \right)^k \right) .\end{aligned}$$

At this point, we introduce a more explicit choice of b, namely $b = c\sqrt{M}\sigma$ with $c = \frac{2\sqrt{2}}{\sqrt{17\mathrm{e}}} \geq \frac{1}{4}$. This choice guarantees

$$\frac{1}{k!(2\sigma^2)^k} \left| \int D_k(x) \phi(x; 0, \sigma^2) \, \mathrm{d}x \right| \leq \left(1 + \frac{1}{2\sqrt{\pi}} \right) \left(\frac{\sqrt{17}}{2} \frac{b}{\sigma\sqrt{k}} \right)^k$$

and moreover, continuing (2.7),

$$\begin{aligned}\|\widetilde{\mathbb{P}}_{\theta\sim\nu_1} - \widetilde{\mathbb{P}}_{\theta\sim\nu_0}\|_{\mathrm{TV}} &\leq \left(1 + \frac{1}{2\sqrt{\pi}} \right) \sum_{k=\lfloor M/2 \rfloor + 1}^{\infty} \left(\frac{\sqrt{17}}{2} \frac{b}{\sigma\sqrt{k}} \right)^k \\ &\leq \left(1 + \frac{1}{2\sqrt{\pi}} \right) \frac{2}{\mathrm{e}-2} \left(\frac{2}{\mathrm{e}} \right)^{\lfloor M/2 \rfloor}\end{aligned}$$

and hence finally

$$\|\mathbb{P}_{\theta\sim\nu_1} - \mathbb{P}_{\theta\sim\nu_0}\|_{\mathrm{TV}} \leq d \left(1 + \frac{1}{2\sqrt{\pi}} \right) \frac{2}{\mathrm{e}-2} \left(\frac{2}{\mathrm{e}} \right)^{\lfloor M/2 \rfloor} ,$$

since the total variation distance is subadditive with respect to product measures; this is used for instance in [MR18].

Now, by direct computation, we see that for any $\eta' \in (0,1)$

$$\frac{1}{2}\|\mathbb{P}_{\theta\sim\nu_1} - \mathbb{P}_{\theta\sim\nu_0}\|_{\mathrm{TV}} < 1-\eta' \tag{2.8}$$

is fulfilled if

$$M \geq \frac{2}{1-\ln(2)}\ln(d) + \underbrace{1 + \frac{2}{1-\ln(2)}\ln\left(\frac{1.8}{1-\eta'}\right)}_{=:c_{\eta'}},$$

so we choose

$$M := \max\left(32, \left\lceil \frac{2}{1-\ln(2)}\ln(d) + c_{\eta'}\right\rceil\right).$$

Application

Note that the bound (2.8) does not formally allow for determining a lower bound on $\rho^*_{\mathcal{C}}$ yet since H_0 and H_1 are not separated in a Euclidean sense. In a final step, we will resolve this by a suitable restriction of H_1.

Let $Y = \sum_{i=1}^d \mathbb{1}_{\{\theta_i = u\}}$, i.e. the number of coordinates of θ taking the value $u = \frac{b}{4M^2}$. Obviously, if $\theta \sim \nu_1$, we have $Y \sim \mathrm{Bin}(d, \widetilde{\nu}_1(\{u\}))$. By property (II) of (2.5) and Hoeffding's inequality, this yields that if $d \geq 42$,

$$\mathbb{P}_{\theta\sim\nu_1}\left(Y \geq \tfrac{d}{3}\right) \geq \frac{9}{10}.$$

Now, let $\xi = \{Y \geq \frac{d}{3}\}$ and

$$H_1' : \ \theta \sim \nu_1 | \xi.$$

Assuming that for some test φ the relation

$$\mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_1}(\varphi = 0) > \eta'$$

holds, we can conclude

$$\begin{aligned}
\mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_1|\xi}(\varphi = 0) &= \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + 1 - \frac{\mathbb{P}_{\theta\sim\nu_1}(\{\varphi = 1\} \cap \xi)}{\mathbb{P}(\xi)} \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + 1 - \frac{10}{9}\mathbb{P}_{\theta\sim\nu_1}(\varphi = 1) \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \frac{10}{9}\mathbb{P}_{\theta\sim\nu_1}(\varphi = 0) - \frac{1}{9} \\
&\geq \mathbb{P}_{\theta\sim\nu_0}(\varphi = 1) + \mathbb{P}_{\theta\sim\nu_1}(\varphi = 0) - \frac{1}{9} \\
&> \eta' - \frac{1}{9}.
\end{aligned}$$

Hence, the above construction yields a total error probability exceeding $\eta = \eta' - \frac{1}{9} \in (0, \frac{8}{9})$ while the l_2-separation between the hypotheses (based on θ under ξ) fulfils

$$\rho = \frac{\sqrt{d}}{\sqrt{3}} \frac{b}{4M^2} = \frac{1}{2\mathrm{e}\sqrt{25.5}} \frac{1}{M^{3/2}} \frac{\sqrt{d}}{\sqrt{n}} \geq \frac{1}{28} \frac{1}{M^{3/2}} \frac{\sqrt{d}}{\sqrt{n}}.$$

This corresponds to the claimed lower bound.

2.5.5 Proof of Lemma 2.9

We need to ensure that on B, the graph of f remains below $\widetilde{B} = B_d([\mathbb{0}_{d-1}, R], R)$ since that corresponds to the fact that $\widetilde{B}$ is locally contained in $\mathcal{C}$, as required in Definition 2.7. This is equivalent to

$$\forall x \in B: \; 0 \leq f(x) \leq R - \sqrt{R^2 - \|x\|^2}.$$

Applying Taylor's theorem with Lagrange's remainder yields

$$\exists s \in (0,1): \; f(x) = \frac{1}{2} x^T H f(sx) x,$$

since by construction $f(\mathbb{0}_{d-1}) = 0$ and $\nabla f(\mathbb{0}_{d-1}) = \mathbb{0}_{d-1}$. Clearly, in order that $f \geq 0$ on B, it is sufficient to require $\lambda_{\min}(Hf(y)) \geq 0$ for $y \in B \backslash \{\mathbb{0}_{d-1}\}$. On the other hand, we can use a classical eigenvalue representation to obtain the desired upper bound: For some $s \in (0,1)$,

$$\begin{aligned} f(x) &= \frac{1}{2}\|x\|^2 \left(\frac{x}{\|x\|}\right)^T Hf(sx) \left(\frac{x}{\|x\|}\right) \\ &\leq \frac{1}{2}\|x\|^2 \max_{\|y\|=1} y^T Hf(sx) y \\ &= \frac{1}{2}\|x\|^2 \lambda_{\max}(Hf(sx)) \\ &\leq \frac{1}{2R}\|x\|^2 \\ &\leq R - \sqrt{R^2 - \|x\|^2} \end{aligned}$$

by assumption and (2.4).

2.5.6 Proof of Theorem 2.10

We define the test statistic

$$T(X) := \mathrm{dist}_{\mathcal{C}}(X),$$

and a corresponding test of the form $\varphi(X) = \mathbb{1}_{\{T(X) > \tau\}}$.
Let $\theta \in \mathcal{C}$. W.l.o.g. assume that $\theta' := \mathbb{0}_d \in \mathcal{C} \subseteq \mathbb{R}^{d-1} \times [0, \infty)$ and θ' minimises the distance between θ and $\partial\mathcal{C}$. Now let $z = [\mathbb{0}_{d-1}, R]$ so that by construction $\theta' \in \overline{B_d(z, R)} \subseteq \mathcal{C}$.

For $\tau > 0$, we have

$$\begin{aligned}\mathrm{dist}_{\mathcal{C}}(X) > \tau &\implies \mathrm{dist}_{B_d(z,R)}(X) > \tau \\ &\implies \left\| \tfrac{1}{\sqrt{n}}\epsilon - z \right\| - R > \tau.\end{aligned}$$

Now, writing $\epsilon_{1:(d-1)} := [\epsilon_1, \epsilon_2, \ldots, \epsilon_{d-1}]$ and using (2.3), we obtain

$$\begin{aligned}\left\| \tfrac{1}{\sqrt{n}}\epsilon - z \right\| - R &= \sqrt{\left\| \tfrac{1}{\sqrt{n}}\epsilon - z \right\|^2} - R \\ &= \sqrt{\tfrac{1}{n}\|\epsilon_{1:(d-1)}\|^2 + \left(\tfrac{1}{\sqrt{n}}\epsilon_d - R\right)^2} - R \\ &\leq \sqrt{\tfrac{1}{n}\|\epsilon_{1:(d-1)}\|^2 + \left(R + \tfrac{1}{\sqrt{n}}|\epsilon_d|\right)^2} - R \\ &\leq R + \tfrac{1}{\sqrt{n}}|\epsilon_d| + \tfrac{\|\epsilon_{1:(d-1)}\|^2}{2n(R+\frac{1}{\sqrt{n}}|\epsilon_d|)} - R \\ &\leq \tfrac{1}{\sqrt{n}}|\epsilon_d| + \tfrac{\|\epsilon_{1:(d-1)}\|^2}{2nR},\end{aligned}$$

which tells us

$$\mathbb{P}_\theta(T(X) \geq \tau) \leq \mathbb{P}\left(\tfrac{1}{\sqrt{n}}|\epsilon_d| + \tfrac{\|\epsilon_{1:(d-1)}\|^2}{2nR} \geq \tau\right).$$

Clearly, this bound holds generally in the sense

$$\sup_{\theta\in\mathcal{C}} \mathbb{P}_\theta(\varphi = 1) \leq \mathbb{P}\left(\tfrac{1}{\sqrt{n}}|\epsilon_d| + \tfrac{\|\epsilon_{1:(d-1)}\|^2}{2nR} \geq \tau\right).$$

Based on the general property

$$\mathbb{P}(A \geq \tau_1) \leq \frac{\eta}{4} \ \wedge\ \mathbb{P}(B \geq \tau_2) \leq \frac{\eta}{4} \quad\implies\quad \mathbb{P}(A + B \geq \tau_1 + \tau_2) \leq \frac{\eta}{2}$$

for random variables A and B and by using (I) and (II) of (1.11), we finally obtain the rejection threshold

$$\tau := \sqrt{\frac{2}{n}\ln\left(\frac{8}{\eta}\right)} + \frac{d}{2nR} + \frac{2}{nR}\sqrt{d\ln\left(\frac{4}{\eta}\right)} + \frac{1}{nR}\ln\left(\frac{4}{\eta}\right) \sim \max\left(\frac{1}{\sqrt{n}}, \frac{d}{nR}\right).$$

On the other hand, w.l.o.g., choose $\theta = [\mathbb{O}_{d-1}, -\rho]$. This is valid since by construction θ minimises the distance between $\mathcal{C}$ and Θ_ρ and $\mathbb{O}_d$ represents an arbitrary element of $\partial\mathcal{C}$. We have

$$\mathrm{dist}_{\mathcal{C}}(X) \leq \tau \implies X_d \geq -\tau \iff \epsilon_d \geq \sqrt{n}(\rho - \tau),$$

so that it is sufficient to ensure

$$\sup_{\theta\in\Theta_\rho} \mathbb{P}_\theta(\varphi = 0) \ \leq\ \mathbb{P}(\epsilon_d \geq \sqrt{n}(\rho - \tau)) \ \leq\ \frac{\eta}{2} \in (0, \tfrac{1}{2}),$$

which leads to the condition

$$\rho \geq \tau + \sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)} \sim \tau.$$

This concludes the proof.

2.5.7 Proof of Theorem 2.11

This is a variation on the proof of Theorem 2.6. Using the same construction and notation as previously, and taking $d \geq 3$, let now for $i \in \{0, 1\}$

$$\nu_i = \widetilde{\nu}_i^{\otimes d-1} \otimes \delta_R.$$

Since the mutual deterministic coordinate $\theta_d = R$ is irrelevant for the total variation distance between the resulting distributions $\mathbb{P}_{\theta\sim\nu_0}$ and $\mathbb{P}_{\theta\sim\nu_1}$, the bounds in Step 2 of the proof of Theorem 2.6 also hold here with $d-1$ instead of d.
The most important modification arises when calculating ρ: Now, if at least $\frac{d-1}{3}$ of the coordinates take the value $u = \frac{b}{4M^2}$, computing the Euclidean distance of θ from $\mathcal{C}$ and using (2.3) leads to

$$\begin{aligned}\rho \geq \sqrt{R^2 + \frac{d-1}{3}u^2} - R &\geq \frac{(d-1)u^2}{6\sqrt{R^2 + \frac{d-1}{3}u^2}} \\ &\geq \frac{(d-1)u^2}{6R + \frac{6}{\sqrt{3}}\sqrt{d-1}u} \\ &\geq \frac{1}{12}\min\left(\frac{(d-1)u^2}{R}, \sqrt{3}\sqrt{d-1}u\right) \\ &\sim \min\left(\frac{1}{\ln(d)^3}\cdot\frac{d}{nR}, \frac{1}{\ln(d)^{3/2}}\cdot\frac{\sqrt{d}}{\sqrt{n}}\right),\end{aligned}$$

if d is large enough in the sense that $M_\eta \leq C\ln(d-1)$ for some $C > 0$, where M_η is given in the statement of the theorem. This concludes the proof.

2.5.8 Proof of Theorem 2.12

W.l.o.g., let $z = \mathbb{O}_d$.

Lower Bound

Let $\nu_0 = \delta_{Re_d}$, giving rise to the density function

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}(x) := \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_d - R)^2\right)\prod_{i=1}^{d-1}\exp\left(-\frac{n}{2}x_i^2\right).$$

On the other hand, for a suitable $h > 0$ specified in a moment, let ν_ρ be the uniform distribution on

$$P_h := \{[h\cdot v, R] \mid v \in \{-1,1\}^{d-1}\}.$$

Since each element of P_h has Euclidean distance $\sqrt{R^2 + (d-1)h^2} - R$ from $\mathcal{C}$, which should correspond to ρ, we set $h^2 = \frac{(R+\rho)^2 - R^2}{d-1}$. This gives rise to the following density function:

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}(x) := \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_d - R)^2\right) \frac{1}{2^{d-1}} \sum_{v_1,\ldots,v_{d-1}\in\{-1,1\}} \prod_{i=1}^{d-1} \exp\left(-\frac{n}{2}(x_i - h\cdot v_i)^2\right)$$

$$= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_d - R)^2\right) \frac{1}{2^{d-1}} \prod_{i=1}^{d-1} \exp\left(-\frac{n}{2}x_i^2 - \frac{n}{2}h^2\right) 2\cosh(nhx_i)$$

$$= \left(\frac{n}{2\pi}\right)^{\frac{d}{2}} \exp\left(-\frac{n}{2}(x_d - R)^2 - (d-1)\frac{n}{2}h^2\right) \prod_{i=1}^{d-1} \exp\left(-\frac{n}{2}x_i^2\right) \cosh(nhx_i),$$

so that

$$\mathbb{P}^2_{\theta\sim\nu_\rho}(x) := \left(\frac{n}{2\pi}\right)^{d} \exp\left(-n(x_d - R)^2 - (d-1)nh^2\right) \prod_{i=1}^{d-1} \exp\left(-nx_i^2\right) \cosh^2(nhx_i).$$

Now, from the computation in section 1.4.2 and the definition of h, we can directly conclude that

$$\begin{aligned}\int_{\mathbb{R}^d} \frac{\mathrm{d}\mathbb{P}^2_{\theta\sim\nu_\rho}}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}(x)\ \mathrm{d}x &= \cosh(nh^2)^{d-1} \\ &\leq \exp\left((d-1)\frac{n^2h^4}{2}\right) \\ &= \exp\left((d-1)\frac{n^2}{2}\frac{((R+\rho)^2 - R^2)^2}{(d-1)^2}\right).\end{aligned}$$

By direct computation, this tells us that (1.10) is fulfilled if

$$\rho \leq \sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R, \text{ where } s := \sqrt{\ln(1 + 4(1-\eta)^2)}, \tag{2.9}$$

It remains to investigate (2.9) a little closer. Application of (2.3) now yields

$$\sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R \geq \frac{\frac{\sqrt{d-1}}{n}s}{2\sqrt{\frac{\sqrt{d-1}}{n}s + R^2}}$$

and distinguishing the cases $R^2 \leq \frac{\sqrt{d-1}}{n}s$ and $R^2 \geq \frac{\sqrt{d-1}}{n}s$ tells us that

$$\sqrt{\frac{\sqrt{d-1}}{n}s + R^2} - R \gtrsim \min\left(\frac{(d-1)^{\frac{1}{4}}}{n^{\frac{1}{2}}}, \frac{(d-1)^{\frac{1}{2}}}{n\cdot R}\right).$$

Upper Bound

We define the test statistic

$$T(X) := \|X\|^2 - R^2$$

and a corresponding test of the form $\varphi(X) = \mathbb{1}_{\{T(X)>\tau\}}$.

On the one hand, in order to control the type-I-error probability, take any $\theta \in \mathcal{C}$, so that $\|\theta\| \in [0, R]$. Clearly, $n\|X\|^2 \sim \chi^2_{n\|\theta\|^2}(d)$. Therefore, for $\tau' > 0$ and with the notation $Z_\lambda \sim \chi^2_\lambda(d)$, we can guarantee

$$\mathbb{P}_\theta(\varphi = 1) = \mathbb{P}(Z_{n\|\theta\|^2} > n\|\theta\|^2 + n\tau') \leq \frac{\eta}{2}$$

by setting

$$\tau' = \frac{d}{n} + 2\sqrt{\left(\frac{d}{n^2} + \frac{2}{n}\|\theta\|^2\right)\ln\left(\frac{2}{\eta}\right)} + \frac{2}{n}\ln\left(\frac{2}{\eta}\right),$$

where we use (II) from (1.11). Since $\|\theta\| \leq R$, this yields that

$$\sup_{\theta\in\mathcal{C}} \mathbb{P}_\theta(\varphi = 1) \leq \sup_{\theta\in\mathcal{C}} \mathbb{P}(Z_{n\|\theta\|^2} > nR^2 + n\tau) \leq \frac{\eta}{2}$$

for

$$\tau = \frac{d}{n} + 2\sqrt{\left(\frac{d}{n^2} + \frac{2}{n}R^2\right)\ln\left(\frac{2}{\eta}\right)} + \frac{2}{n}\ln\left(\frac{2}{\eta}\right).$$

On the other hand, in order to satisfy a prescribed level $\frac{\eta}{2}$ for the Type-II-error, take any $\theta \in \Theta_\rho$ with $\|\theta\| > R + \rho$. Then again, $n\|X\|^2 \sim \chi^2_{n\|\theta\|^2}(d)$, so that we need to ensure

$$\mathbb{P}_\theta(\varphi = 0) = \mathbb{P}(Z' \leq nR^2 + n\tau) \leq \frac{\eta}{2}, \quad \text{where } Z' \sim \chi^2_{n\|\theta\|^2}(d). \tag{2.10}$$

In this case, (III) of (1.11) yields the sufficient condition

$$d + nR^2 + 2\sqrt{(d + 2nR^2)\ln\left(\frac{2}{\eta}\right)} + 2\ln\left(\frac{2}{\eta}\right) \leq d + n\|\theta\|^2 - 2\sqrt{(d + 2n\|\theta\|^2)\ln\left(\frac{2}{\eta}\right)}.$$

The right hand side is increasing in $\|\theta\|$ if $d \geq \ln(1/\frac{\eta}{2})$, so that, similar as for the type-I-error, (2.10) holds uniformly over Θ_ρ if

$$d + nR^2 + 2\sqrt{(d + 2nR^2)\ln\left(\frac{2}{\eta}\right)} + 2\ln\left(\frac{2}{\eta}\right) \leq d + n(R+\rho)^2 - 2\sqrt{(d + 2n(R+\rho)^2)\ln\left(\frac{2}{\eta}\right)}.$$

Using $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ (for $a, b > 0$) and (2.3) respectively, we obtain two different sufficient bounds for ρ:

$$\rho \geq 2\sqrt{2}\frac{d^{\frac{1}{4}}}{\sqrt{n}}\sqrt{\ln\left(\frac{2}{\eta}\right)} + 3\sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)};$$

$$\rho \geq \frac{2\sqrt{d}}{nR + 2\sqrt{n\ln\left(\frac{2}{\eta}\right)}}\sqrt{\ln\left(\frac{2}{\eta}\right)} + 3\sqrt{\frac{2}{n}\ln\left(\frac{2}{\eta}\right)}.$$

Therefore, as claimed, the upper bound

$$\rho_{\mathcal{C}}^* \lesssim \max\left(\frac{1}{\sqrt{n}}, \min\left(\frac{d^{\frac{1}{4}}}{\sqrt{n}}, \frac{\sqrt{d}}{nR}\right)\right)$$

holds. This concludes the proof.

Chapter 3

Testing the Equality in Distribution of Two Random Graphs

In this chapter we study the problem of testing if two random graphs have the same underlying distribution. More precisely, we observe M iid samples from two inhomogeneous Erdős-Renyi-graph distributions with parameter matrices P and Q on a common set of n vertices each and make no further assumption on the two distributions. Considering a testing problem of the form

$$H_0: \; P = Q \;\text{ vs. }\; H_1: \; \text{dist}(P, Q) > \rho$$

for some $\rho > 0$ and an appropriate measure of separation, dist, we approach the task of identifying the smallest value for ρ, denoted ρ^*, enabling the existence of a test φ with small total error probability in a minimax sense.

We derive the precise dependence of ρ^* on n and M for four different natural choices of dist. In particular, it turns out that this choice is crucial as the testing problem is only feasible in two of the four cases, namely where $\text{dist}(P, Q)$ is the Frobenius or spectral norm of $P - Q$. In these cases, if $M \geq 2$, we have

$$\rho^* \sim \frac{\sqrt{n}}{\sqrt{M}}.$$

This chapter is related to the articles [GGCvL17a] and [GGCvL17b], yet most of the results are independent as they are not proved in the articles and also no simple corollaries from there.

3.1 Introduction

For $n \in \mathbb{N}\backslash\{0, 1\}$, let $V_n = \{1, 2, \ldots, n\}$ represent a fixed set of n vertices of arbitrary type – for instance, they may correspond to geographical points, brain cells or persons. These vertices are now randomly connected by edges as follows: For $i, j \in V_n$, edge (i, j) occurs with probability $P_{ij} \in [0, 1]$, where $P_{ii} = 0$. Moreover, the edges are independent and undirected, that is, we do not distinguish between (i, j) and (j, i)). Clearly, this model is fully described through the matrix

$$P = (P_{ij})_{i,j \in V_n} \in \mathcal{G}_n = \{G \in [0,1]^{n \times n} \ ; \ G = G^T \ \wedge \ \forall i \in V_n : \ G_{ii} = 0\}, \tag{3.1}$$

where G^T is the standard matrix transpose. Now, the observed random object of interest is the resulting adjacency matrix

$$A \in \mathcal{S}_n = \{S \in \{0,1\}^{n \times n} \ ; \ S = S^T \ \wedge \ \forall i \in V_n : \ S_{ii} = 0\}, \tag{3.2}$$

where A_{ij} is the indicator for the occurence of edge (i, j). By construction, this essentially corresponds to observing

$$d := \binom{n}{2}$$

independent Bernoulli random variables with respective parameters P_{ij}, $i < j$.
The type of random graph we just described is also known as an inhomogeneous Erdős-Renyi-graph ([BJR07]), hence we write

$$A \sim \mathrm{IER}(P).$$

Testing problem

Let $P, Q \in \mathcal{G}_n$ unkown and fix $M \in \mathbb{N}\backslash\{0\}$. Suppose we observe M independent realisations from each of the two models, namely

$$\begin{aligned} A^1, A^2, \dots, A^M &\overset{\text{iid}}{\sim} \mathrm{IER}(P), \\ B^1, B^2, \dots, B^M &\overset{\text{iid}}{\sim} \mathrm{IER}(Q). \end{aligned}$$

The resulting probability measure given $\theta = (P, Q)$ will be written $\mathbb{P}_\theta = \mathbb{P}_{(P,Q)}$.

Based on these observations, for $\rho > 0$ we now study the testing problem

$$H_0 : \ P = Q \ \text{ vs. } \ H_1 : \ \mathrm{dist}(\theta) = \mathrm{dist}(P, Q) > \rho, \tag{3.3}$$

where dist is an appropriate measure of distance between the two distributions – we will discuss the Kullback-Leibler-divergence as a distance measure on the level of distributions, but our main focus will be distance measures on the level of matrices, namely the zero-norm, Frobenius norm and spectral norm of $P - Q$, see the respective sections for precise definitions.
Since the null hypothesis remains the same throughout this chapter (except for the short aside for Theorem 3.7), dist does not need an index Θ_0. However, for the sake of clarity, we take account of the varying distance measure through the index of ρ^*.

In the framework of chapter 1, the testing problem at hand is given by

$$\begin{cases} \Theta &= \mathcal{G}_n^2, \\ \mathbb{P}_\theta &= \mathbb{P}_{(P,Q)} = \mathrm{IER}(P)^{\otimes M} \otimes \mathrm{IER}(Q)^{\otimes M} \quad \text{for } \ \theta = (P, Q) \in \Theta, \\ \Theta_0 &= \{(P, P) \ ; \ P \in \mathcal{G}_n\}, \\ \Theta_\rho &= \{(P, Q) \in \mathcal{G}_n^2 \ ; \ \mathrm{dist}(P, Q) > \rho\}. \end{cases}$$

Now, again, given $\eta \in (0,1)$, we aim at finding the magnitude in terms of n and M of the smallest separation distance $\rho^*(\eta)$ which enables the existence of a test φ with total error bound η in a minimax sense, i.e. of

$$\rho^*(\eta) = \inf \left\{ \rho > 0 \ ; \ \exists \text{ test } \varphi : \sup_{(P,Q)\in\Theta_0} \mathbb{P}_{(P,Q)}(\varphi = 1) + \sup_{(P,Q)\in\Theta_\rho} \mathbb{P}_{(P,Q)}(\varphi = 0) \leq \eta \right\}.$$

Related questions and literature

In this chapter we are considering a two-sample testing problem, which constitutes a significant conceptual difference to the problems discussed in the other chapters. Note that, however, since the null hypothesis from (3.3) can also be written as

$$H_0 : \ P - Q = \mathbb{0}_{d\times d},$$

the problem is closely related to signal-detection (yet Θ_0 is composite) and indeed, in order to obtain the results presented below, the basic theoretical machinery introduced in 1.4.1 and used in the previous chapters need not be extended. In particular, references like [IS02], [Bar02] and [VAC17] remain relevant.

Inference problems on random graphs have been studied extensively in both theoretical and applied contexts in the past couple of decades, which signifies the topic's relevance. For instance, see [BDER16] for an article on testing for high-dimensional geometry or [ACV14] and [VAC15] for papers on the popular problem of community detection; another signal-detection problem in a random matrix context is studied in [CN15].
However, literature on (two-sample)-testing if random graphs are equal in a certain sense is still rare. In [GGCvL17a], we consider a more general setting in which the two graphs need not be defined on a common set of vertices; clearly in that situation (3.3) is not a proper testing problem and we rather compare the graphs with respect to network statistics such as triangle counts. The articles [TAS$^+$17a] and [TAS$^+$17b] are probably the closest to the present considerations in terms of the setting since they also compare undirected, edge-independent graphs on a common vertex set. However, the underlying model of a random dot poduct graph is much more specific than a general IER-graph.
See also [GGCvL17b] for an extensive survey of related literature.

3.2 Indistinguishability Results

In this section, we will show that the testing problem is generally not feasible for certain choices of distances by providing lower bounds on $\rho^*(\eta)$ which are equal to a trivial upper bound on $\mathrm{dist}(P,Q)$.

Testing with zero-norm

For any matrix $G \in \mathbb{R}^{k\times l}$ (with $k,l \in \mathbb{N}\backslash\{0\}$), we denote the number of non-zero entries of G by $\|G\|_0$, i.e.

$$\|G\|_0 := \sum_{i=1}^{k}\sum_{j=1}^{l} \mathbb{1}_{\{G_{ij}\neq 0\}}.$$

Choosing this distance yields the following extreme case for our testing problem:

Theorem 3.1. Consider the problem (3.3) with

$$\operatorname{dist}(P,Q) = \|P-Q\|_0$$

and let $\eta \in (0,1)$. Then we have

$$\rho^*_{\text{zero}}(\eta) = n(n-1).$$

◁

We can now describe in what sense the testing problem is not feasible: Since

$$n(n-1) = \max_{P,Q\in\mathcal{G}_n} \|P-Q\|_0,$$

the result tells us that no non-trivial separation between the hypotheses can guarantee small error probabilities. In fact, from an intuitive perspective the result is not surprising since the matrices (and hence distributions) may be arbitrarily close and yet exhibit the highest possible distance in zero-norm. The proof in section 3.6.1 exploits this fact.

Testing with Kullback-Leibler-divergence

For any two probability distributions ν_0 and ν_1 on a discrete set S, the Kullback-Leibler-divergence from ν_0 to ν_1 is defined as

$$\mathrm{KL}(\nu_0||\nu_1) := \sum_{s\in S} \nu_0(\{s\}) \ln\left(\frac{\nu_0(\{s\})}{\nu_1(\{s\})}\right).$$

Based on that, we declare the symmetrised Kullback-Leibler-divergence to be

$$\begin{aligned}\mathrm{SKL}(\nu_0,\nu_1) &:= \mathrm{KL}(\nu_0||\nu_1) + \mathrm{KL}(\nu_1||\nu_0)\\ &= \sum_{s\in S} \nu_0(\{s\}) \ln\left(\frac{\nu_0(\{s\})}{\nu_1(\{s\})}\right) + \sum_{s\in S} \nu_1(\{s\}) \ln\left(\frac{\nu_1(\{s\})}{\nu_0(\{s\})}\right).\end{aligned}$$

In particular, if either $\nu_0 \ll \nu_1$ or $\nu_1 \ll \nu_0$ is false,

$$\mathrm{SKL}(\nu_0,\nu_1) = \infty.$$

Remark 3.2. The (un)symmetrised Kullback-Leibler-divergence can also be defined in more general (particularly non-discrete) contexts through

$$\mathrm{KL}(\nu_0||\nu_1) := \int_S \ln\left(\frac{\mathrm{d}\nu_0}{\mathrm{d}\nu_1}\right)\,\mathrm{d}\nu_0.$$

In fact, it is a popular measure for statistical distance along with the total variation distance and χ^2-divergence and may also be used in order to derive lower bounds as explained in section 1.4.1; the central relation enabling this is known as Pinsker's inequality, see for instance [GN16, Chapter 7]:

$$\|\nu_0-\nu_1\|_{\mathrm{TV}} \le \sqrt{2\mathrm{KL}(\nu_0||\nu_1)}.$$

Many references (including [GN16]) give a factor $\frac{1}{\sqrt{2}}$ rather than $\sqrt{2}$ on the right hand side; this is due to different conventions for the definition of the total variation distance. ◁

Now, provided that M is small, it turns out that our testing problem with the symmetrised Kullback-Leibler-divergence is also trivial in the sense of Theorem 3.1:

Theorem 3.3. Consider the problem (3.3) with

$$\mathrm{dist}(P,Q) = \mathrm{SKL}(\mathrm{IER}(P), \mathrm{IER}(Q))$$

and let $\eta \in (0,1)$. Whenever $M < \frac{\ln(1+4(1-\eta)^2)}{\ln(2)}$, we have

$$\rho^*_{\mathrm{SKL}}(\eta) = \infty.$$

◁

From the proof in section 3.6.2, it is clear that the same result holds without symmetrisation.

Remark 3.4. It is possible to obtain a very similar result when using total variation distance rather than Kullback-Leibler-divergence – see [GGCvL17b, Prop. 3.1]. ◁

3.3 Testing in Frobenius Norm

In the previous section we have exhibited some distances for problem (3.3) which do not appear to be appropriate for the task at hand. Recalling the fact that the original graph distributions are fully descibed by the matrices P and Q, as mentioned earlier, we now focus specifically on measuring the distance through classical matrix norms. It turns out that testing is possible with nontrivial separation and performance for both Frobenius and spectral norm for any $M \geq 2$.

Result

For any matrix $G \in \mathbb{R}^{k\times l}$ (with $k,l \in \mathbb{N}$), we denote its Frobenius norm by $\|G\|_{\mathrm{F}}$, i.e.

$$\|G\|_{\mathrm{F}} := \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{l} G_{ij}^2}.$$

We can state the result on Frobenius norm right away:

Theorem 3.5. Consider the problem (3.3) with

$$\mathrm{dist}(P,Q) = \|P-Q\|_{\mathrm{F}},$$

let $\eta \in (0,1)$ and take $n \geq 1 + \frac{3}{2}\ln\left(\frac{8}{\eta}\right)$. Then we have

$$\text{for } M = 1: \qquad \frac{n}{\sqrt{8}} \;\leq\; \rho^*_{\mathrm{F}}(\eta) \;\leq\; n,$$

$$\text{for } M \geq 2: \qquad \frac{\sqrt{n}}{\sqrt{M}}\sqrt[4]{\frac{1}{3}\ln(1+4(1-\eta)^2)} \;\leq\; \rho^*_{\mathrm{F}}(\eta) \;\leq\; 35\frac{\sqrt{n}}{\sqrt{M}}\sqrt{\ln\left(\frac{8}{\eta}\right)}.$$

That is,

$$\rho_{\mathrm{F}}^* \sim \begin{cases} n, & M = 1, \\ \dfrac{\sqrt{n}}{\sqrt{M}}, & M \geq 2 \end{cases}.$$

◁

The statement $\rho^* \sim n$ in the case $M = 1$ can be interpreted as a recommendation not to test for Frobenius norm separation if only one observation per model is available. Note that the phase transition between the cases $M = 1$ and $M \geq 2$ is not particularly surprising since a single observation of a Bernoulli vector cannot allow for meaningful inference about the underlying parameters (except that necessarily $P_{ij} \neq 1 - A_{ij}^1$).

Proof Strategy

While the lower bound in Theorem 3.5 is essentially a corollary of the proof of Theorem 3.1 (section 3.6.1), the upper bound is more involved and merits some explanation here: As in the previous chapters, in order to find an appropriate test statistic, we would like to use a strong estimator for the target quantity $\|P - Q\|_{\mathrm{F}}$ or, say, $\|P - Q\|_{\mathrm{F}}^2$. The naive choice would be

$$T_0 := \frac{1}{M} \sum_{m=1}^{M} \|A^m - B^m\|_{\mathrm{F}}^2.$$

However, T_0 is not an unbiased estimator of $\|P - Q\|_{\mathrm{F}}^2$ and can only lead to the very weak upper bound

$$\rho_{\mathrm{F}}^* \lesssim n,$$

see section 3.6.3 for an analysis.

Luckily, constructing an unbiased estimator here is possible without much more effort: We define two independent estimators of $P - Q$ by splitting the M samples into two groups and then multiply the results. More specifically, without loss of generality let M be even and

$$\widehat{P}^1 := \frac{2}{M} \sum_{m=1}^{M/2} A^m, \ \widehat{P}^2 := \frac{2}{M} \sum_{m=M/2+1}^{M} A^m, \ \widehat{Q}^1 := \frac{2}{M} \sum_{m=1}^{M/2} B^m, \ \widehat{Q}^2 := \frac{2}{M} \sum_{m=M/2+1}^{M} B^m.$$

Then the test statistic

$$T := \sum_{i=1}^{n} \sum_{j=1}^{n} (\widehat{P}_{ij}^1 - \widehat{Q}_{ij}^1) \cdot (\widehat{P}_{ij}^2 - \widehat{Q}_{ij}^2)$$

has the desired properties and yields the rate-optimal upper bound given in Theorem 3.5.

The central challenge of the theorem's proof (section 3.6.4) lies in finding strong concentration bounds for T. We employ the Chernoff strategy (i.e. applying Markov's inequality to $\exp(\lambda T)$ and optimizing with respect to $\lambda > 0$) and face the nontrivial problem of controlling the moment-generating function of a product of sums.

Remark 3.6. As a corollary of Theorem 3.5 and its proof in section 3.6.4 (through considering $P = \mathbb{O}_{n\times n}$), we see that the signal-detection rate for the mean of a d-dimensional Bernoulli vector is equal to $\frac{d^{1/4}}{\sqrt{M}}$. More precisely, let $d \in \mathbb{N}\backslash\{0\}$, $M \in \mathbb{N}\backslash\{0,1\}$, $P \in [0,1]^d$ and suppose that we observe independent random variables

$$A_i^m \sim \mathrm{Ber}(P_i),\ i \in \{1,2,\ldots,d\},\ m \in \{1,2,\ldots,M\}.$$

Then, in the framework of section 1 with

$$\Theta_0 = \{\mathbb{O}_d\},\ \Theta_\rho = \{P \in [0,1]^d \mid \|P\|_2 > \rho\},$$

we have

$$\rho^* \sim \frac{d^{1/4}}{\sqrt{M}}.$$

Note also the consistency with the signal-detection rate in the Gaussian sequence model from the previous section. This is essentially due to the sub-Gaussianity of the Bernoulli distribution. ◁

Problem-(In)dependent Rate and Weaker Constant

All results presented so far were problem-independent in the sense that they described the problem's difficulty in the situation where no prior information on P or Q is available. As an example and a short aside, we consider the case where a bound on $\|P+Q\|_{\mathrm{F}}$ is given. Similarly as in section 1, the upper bound in the following theorem will be based on Chebyshev's inequality which, as usual, leads to a weaker dependence on the total error bound η.

Theorem 3.7. Consider the problem (3.3) with

$$\mathrm{dist}(P,Q) = \|P-Q\|_{\mathrm{F}}$$

and where the hypotheses are restricted as

$$\begin{aligned}\Theta_{0,\Delta} &= \{(P,P)\ ;\ P \in \mathcal{G}_n\ \wedge\ \|P\|_{\mathrm{F}} \le \Delta\},\\ \Theta_{\rho,\Delta} &= \{(P,Q) \in \mathcal{G}_n^2\ ;\ \mathrm{dist}(P,Q) > \rho\ \wedge\ \|P+Q\|_{\mathrm{F}} \le 2\Delta\}\end{aligned}$$

for $\Delta \in (0, \sqrt{n(n-1)}]$. Moreover, let $\eta \in (0,1)$. Then we have

$$\frac{\sqrt{\Delta}}{\sqrt{M}}\frac{\sqrt[4]{2}\sqrt{\ln(1+4(1-\eta)^2)}}{2\sqrt{3}} \ \le\ \rho^*_{\mathrm{F},\Delta}(\eta)\ \le\ \left(\sqrt[4]{\frac{16}{\eta}} + \frac{8\sqrt{2}}{\sqrt{\eta}}\right)\sqrt{\frac{\Delta}{M}},$$

that is

$$\rho^*_{\mathrm{F},\Delta} \sim \frac{\sqrt{\Delta}}{\sqrt{M}}.$$

◁

3.4 Testing in Spectral Norm

The considerations in the previous sections were actually not specific to the matrix shape of P and Q or A^m and B^m – The zero and Frobenius norms may easily be replaced by the corresponding norms in $\mathbb{R}^d$, which we explicitly do in the proofs in order to simplify notations. In contrast to that, the last choice of distance measure we want to consider here is very sensitive to the matrices' shape, namely spectral norm.

Result

For any matrix $G \in \mathbb{R}^{k\times k}$ (with $k \in \mathbb{N}$), we denote its spectral norm by $\|G\|_{\mathrm{S}}$, i.e.

$$\|G\|_{\mathrm{S}} := \max_{x\in\mathbb{R}^k\setminus\{\mathbb{O}_k\}} \frac{\|Gx\|_2}{\|x\|_2} = \max_{x\in\partial B_k(\mathbb{O}_k,1)} \|Gx\|_2,$$

where, as in chapter 2, $B_k(\mathbb{O}_k, 1)$ is the Euclidean ball in $\mathbb{R}^k$ of radius 1. In particular, if G is symmetric, through the min-max theorem (Courant-Fisher theory, see [Dym13]) we can also write

$$\|G\|_{\mathrm{S}} = \max_{x\in\partial B_k(\mathbb{O}_k,1)} |x^T Gx|. \tag{3.4}$$

Theorem 3.8. Consider the problem (3.3) with

$$\mathrm{dist}(P, Q) = \|P - Q\|_{\mathrm{S}}$$

and let $\eta \in (0, 1)$. Then we have

$$\frac{\sqrt{n}}{\sqrt{M}} \cdot \min\left(\frac{\sqrt[4]{\ln(1 + 4(1-\eta)^2)}}{5}, \frac{1}{10}\right) \ \le\ \rho^*_{\mathrm{S}}(\eta)$$

and

$$\rho^*_{\mathrm{S}}(\eta) \ \le\ \frac{\sqrt{n}}{\sqrt{M}}\left(2\left(2\sqrt{\ln(2)} + \ln\left(\frac{1}{4\eta}\right)\right)^2 + \sqrt{2\ln\left(\frac{4}{\eta}\right)}\right),$$

that is

$$\rho^*_{\mathrm{S}} \sim \frac{\sqrt{n}}{\sqrt{M}}.$$

◁

Proof Strategy

From a high level perspective, in order to obtain a strong lower bound we would like to construct priors ν_0 and ν_1 on P and Q respectively such that the resulting statistical distance between $(A^1, \dots, A^M)$ and $(B^1, \dots, B^M)$ is small while $\|P-Q\|_{\mathrm{S}}$ is large. By classical properties of the spectral norm, the latter requirement corresponds to $P-Q$ being of small rank. Such a construction is proposed in the proof in section 3.6.6. Interestingly, the subsequent computation of the resulting statistical distance requires a technical trick we already used for proving the upper bound in Theorem 3.5, namely the successive evaluation of expectations with respect to the individual random variables involved – see (3.12) and compare with (3.11).

Proving the upper bound also requires new ideas: Motivated by (3.4), we choose the test statistic

$$T = \max_{u\in\partial B_n(\mathbb{O}_n,1)} |u^T S_M u| \quad \text{with} \quad S_M := \frac{1}{M}\sum_{m=1}^{M}(A^m - B^m).$$

Through Hoeffding's inequality it is possible to derive a useful concentration bound for $u^T S_M u$ if u is fixed. Now, since we consider the maximum over an uncountable set, it cannot be controlled through, for instance, a simple union bound. We can however cover $\partial B_n(\mathbb{O}_n, 1)$ by finitely many balls with a prescribed radius $\rho > 0$ and consider a union bound over this covering; the resulting imprecision shrinks with decreasing ρ so that we consider a series of radii with $\rho_i \searrow 0$ – in fact we employ a chaining approach.

The curious case $M = 1$

In contrast to the case of Frobenius norm (Theorem 3.5), there is no phase transition between the cases $M = 1$ and $M \geq 2$ and the rate for $M = 1$ is not trivial as $\|P - Q\|_{\mathrm{S}} = n - 1$ is possible. This is rather surprising given the intuitive idea that only one observation cannot contain enough meaningful information in this setting. However, in fact, there is more to that now due to the sensitivity of the spectral norm to rank mentioned above: In a seemingly simple case where P and Q are far apart in each entry, the Frobenius norm may well be of order n while the spectral norm is relatively small due to a high rank of $P - Q$ or $A^1 - B^1$. More formally, remembering the general relations

$$\|G\|_{\mathrm{S}} \leq \|G\|_{\mathrm{F}} \leq \sqrt{\operatorname{rank}(G)} \cdot \|G\|_{\mathrm{S}},$$

we see for such an extreme case that

$$\|P - Q\|_{\mathrm{F}} \gtrsim n \quad \Longrightarrow \quad \|P - Q\|_{\mathrm{S}} \gtrsim \sqrt{n}.$$

3.5 Alternative Settings

Directed Graphs

It is a very natural extension of our setting to consider directed graphs, which essentially amounts to removing the symmetry assumption from (3.1) and (3.2) and observing $d' = n(n-1)$ rather than $d = \frac{n(n-1)}{2}$ random variables. Clearly, every result in this chapter which does not critically rely on the symmetry assumption can easily be extended to this case. Therefore, the only statement which does not apply to this case is the upper bound for spectral norm separation, i.e. in Theorem 3.8.

Other Problem Dependent Bounds

For any matrix $G \in \mathbb{R}^k$ $(k \in \mathbb{N}\backslash\{0\})$, we denote its maximal absolute entry as $\|G\|_{\max}$, i.e.

$$\|G\|_{\max} := \max_{i,j \in \{1,2,\ldots,k\}} |G_{ij}|.$$

In [GGCvL17b], we study the present testing problem (3.3) in generalised setting where $\mathcal{G}_n$ from (3.1) is replaced by

$$\mathcal{G}_{n,\delta} := \left\{ G \in [0,1]^{n \times n} \ ; \ G = G^T \ \wedge \ \forall i \in V_n : \ G_{ii} = 0 \ \wedge \ \|G\|_{\max} \leq \delta \right\}$$

for some fixed $\delta \in [0,1]$. This article provides bounds on the separation rate ρ^* for essentially the same choices of dist as in the present chapter; these bounds are problem-dependent in the sense that they explicitly depend on δ, comparable to Theorem 3.7 above. Clearly, the present setting corresponds to the choice $\delta = 1$ above and based on that, our Theorems 3.1 and 3.3 are corollaries of the corresponding results [GGCvL17b, Propositions 3.2 and 3.3]. However, note that our

Theorems on Frobenius norm (Theorem 3.5) and spectral norm (Theorem 3.8) cannot be directly derived from the corresponding results [GGCvL17b, Theorem 4.1, Proposition 4.4, Theorem 5.1] since they are either based on Chebyshev's inequality (rather than Chernoff bounds) or contain additional ln-factors due to different lines of arguments forced by the introduction of δ.

3.6 Proofs

3.6.1 Proof of Theorem 3.1

We firstly simplify P, Q to vectors $\widetilde{P}, \widetilde{Q} \in [0,1]^d$ by rearranging the relevant entries with $i < j$ accordingly and keep in mind that

$$\|P - Q\|_0 = 2\|\widetilde{P} - \widetilde{Q}\|_0.$$

Consider the simple prior ν_0 that sets each entry of $\widetilde{P}$ and $\widetilde{Q}$ to $\frac{1}{2}$. Now, we choose ν_ρ such that $\widetilde{P}$ is the same as before, but on the other hand $\widetilde{Q}$ consists of independent entries with distribution

$$\frac{1}{2}\delta_{\frac{1}{2}-\gamma} + \frac{1}{2}\delta_{\frac{1}{2}+\gamma}$$

(Dirac- δ) for some $\gamma \in (0, \frac{1}{2}]$. We may write the resulting distributions formally as

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0} = \left(\mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes M} \otimes \mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes M} \right)^{\otimes d}, \tag{3.5}$$

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho} = \left(\mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes M} \otimes \mathrm{Ber}\left(\frac{1}{2}\delta_{\frac{1}{2}-\gamma} + \frac{1}{2}\delta_{\frac{1}{2}+\gamma}\right)^{\otimes M} \right)^{\otimes d},$$

where we observe that in the second case (for H_1), $\rho = \|\widetilde{P} - \widetilde{Q}\|_0 = d$.

With $\Omega := \{0,1\}^M$, each of the d entries of these distributions has support Ω^2 and a mass function $f_0(\omega_1) \cdot f_0(\omega_2)$ and $f_0(\omega_1) \cdot f_\rho(\omega_2)$ respectively for $(\omega_1, \omega_2) \in \Omega^2$. As a first step, since $f_0 \equiv \frac{1}{2^M}$ note that

$$\mathrm{div}_{\chi^2}\left(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}\right) = \left(\sum_{\omega_1\in\Omega} \sum_{\omega_2\in\Omega} \frac{f_0^2(\omega_1) f_\rho^2(\omega_2)}{f_0(\omega_1) f_0(\omega_2)} \right)^d = \left(2^M \sum_{\omega\in\Omega} f_\rho^2(\omega) \right)^d. \tag{3.6}$$

Now, more explicitly, we obtain

$$\begin{aligned}
\sum_{\omega\in\Omega} f_\rho^2(\omega) &= \sum_{k=0}^{M} \binom{M}{k} \left[\frac{1}{2}\left(\frac{1}{2}-\gamma\right)^k \left(\frac{1}{2}+\gamma\right)^{M-k} + \frac{1}{2}\left(\frac{1}{2}+\gamma\right)^k \left(\frac{1}{2}-\gamma\right)^{M-k}\right]^2 \\
&= \frac{1}{4}\sum_{k=0}^{M} \binom{M}{k} \left[\left(\frac{1}{2}-\gamma\right)^2\right]^k \left[\left(\frac{1}{2}+\gamma\right)^2\right]^{M-k} \\
&\quad + \frac{1}{2}\sum_{k=0}^{M} \binom{M}{k} \left(\frac{1}{2}+\gamma\right)^M \left(\frac{1}{2}-\gamma\right)^M \\
&\quad + \frac{1}{4}\sum_{k=0}^{M} \binom{M}{k} \left[\left(\frac{1}{2}+\gamma\right)^2\right]^k \left[\left(\frac{1}{2}-\gamma\right)^2\right]^{M-k} \\
&= \frac{1}{2}\left[\left(\frac{1}{2}-\gamma\right)^2 + \left(\frac{1}{2}+\gamma\right)^2\right]^M \\
&\quad + \frac{1}{2}2^M\left(\frac{1}{4}-\gamma^2\right)^M \\
&= \frac{1}{2}\left(\frac{1}{2}+2\gamma^2\right)^M + \frac{1}{2}2^M\left(\frac{1}{4}-\gamma^2\right)^M. \qquad (3.7)
\end{aligned}$$

As a result,

$$\begin{aligned}
\operatorname{div}_{\chi^2}\left(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}\right) &= \left(\frac{1}{2}\left(1+4\gamma^2\right)^M + \frac{1}{2}\left(1-4\gamma^2\right)^M\right)^d \\
&\leq \left(\frac{1}{2}\exp(4M\gamma^2) + \frac{1}{2}\exp(-4M\gamma^2)\right)^d \\
&= \cosh(4M\gamma^2)^d \\
&\leq \exp(2dM^2\gamma^4), \qquad (3.8)
\end{aligned}$$

where we use the classical bounds

$$\forall x\in\mathbb{R}: \quad 1+x \leq \exp(x), \;\; \cosh(x) \leq \exp(x^2/2)$$

which can be established through Taylor expansion. We can now evaluate the condition (1.10): The result (3.8) tells us that

$$\operatorname{div}_{\chi^2}\left(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}\right) < 1 + 4(1-\eta)^2$$

whenever

$$\gamma < \sqrt[4]{\frac{1}{2dM^2}\ln(1+4(1-\eta)^2)}, \qquad (3.9)$$

so we may choose

$$\gamma = \frac{1}{2}\sqrt[4]{\frac{1}{2dM^2}\ln(1+4(1-\eta)^2)}.$$

Since for that choice of γ, clearly $\|P - Q\|_0 = 2\|\widetilde{P} - \widetilde{Q}\|_0 = n(n-1)$, the claim follows.

3.6.2 Proof of Theorem 3.3

Again, we consider the vectors $\widetilde{P}$ and $\widetilde{Q}$ as in section 3.6.1. As a first step, we compute the SKL between the resulting distributions

$$\nu_0 := \bigotimes_{i=1}^{d} \mathrm{Ber}(\widetilde{P}_i) \quad \text{and} \quad \nu_1 := \bigotimes_{i=1}^{d} \mathrm{Ber}(\widetilde{Q}_i).$$

We have

$$\begin{aligned}
\mathrm{KL}(\nu_0 \| \nu_1) &= \sum_{i=1}^{d} \sum_{A_i=0}^{1} \left(\prod_{k=1}^{d} \widetilde{P}_k^{A_k} (1-\widetilde{P}_k)^{1-A_k} \cdot \sum_{l=1}^{d} \ln \left(\frac{\widetilde{P}_l^{A_l}(1-\widetilde{P}_l)^{1-A_l}}{\widetilde{Q}_l^{A_l}(1-\widetilde{Q}_l)^{1-A_l}} \right) \right) \\
&= \sum_{l=1}^{d} \sum_{i=1}^{d} \sum_{A_i=0}^{1} \left(\prod_{k=1}^{d} \widetilde{P}_k^{A_k} (1-\widetilde{P}_k)^{1-A_k} \cdot \ln \left(\frac{\widetilde{P}_l^{A_l}(1-\widetilde{P}_l)^{1-A_l}}{\widetilde{Q}_l^{A_l}(1-\widetilde{Q}_l)^{1-A_l}} \right) \right) \\
&= \sum_{l=1}^{d} \sum_{A_l=0}^{1} \left(\widetilde{P}_l^{A_l} (1-\widetilde{P}_l)^{1-A_l} \cdot \ln \left(\frac{\widetilde{P}_l^{A_l}(1-\widetilde{P}_l)^{1-A_l}}{\widetilde{Q}_l^{A_l}(1-\widetilde{Q}_l)^{1-A_l}} \right) \right) \\
&= \sum_{l=1}^{d} \left((1-\widetilde{P}_l) \cdot \ln \left(\frac{1-\widetilde{P}_l}{1-\widetilde{Q}_l} \right) + \widetilde{P}_l \ln \left(\frac{\widetilde{P}_l}{\widetilde{Q}_l} \right) \right),
\end{aligned}$$

where in the second to last step we expand the product and see that for each l, any summand for $i \neq l$ only translates to a factor 1. This leads to

$$\mathrm{SKL}(\nu_0, \nu_1) = \sum_{l=1}^{d} \left((\widetilde{P}_l - \widetilde{Q}_l) \ln \left(\frac{\widetilde{P}_l(1-\widetilde{Q}_l)}{\widetilde{Q}_l(1-\widetilde{P}_l)} \right) \right),$$

Now, our choice of $\mathbb{P}_{(\widetilde{P},\widetilde{Q}) \sim \nu_0}$ of $\mathbb{P}_{(\widetilde{P},\widetilde{Q}) \sim \nu_\rho}$ is based on the fact that this divergence is infinite ($\rho = \infty$) whenever for some $l \in \{1, 2, \ldots, d\}$, $\widetilde{Q}_l = 0$. Specifically, we consider

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q}) \sim \nu_0} = \left(\mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes d} \otimes \mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes d} \right)^{\otimes M},$$

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q}) \sim \nu_\rho} = \left(\mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes d} \otimes \left[\bigotimes_{i=1}^{d} \mathrm{Ber}(V_i) \right] \right)^{\otimes M},$$

where $V \in [0,1]^d$ is uniformly distributed in

$$\mathcal{V} := \left\{ \frac{1}{2}(\mathbb{1}_{\{j \neq 1\}}, \mathbb{1}_{\{j \neq 2\}}, \ldots, \mathbb{1}_{\{j \neq d\}}) \ ; \ j \in \{1, 2, \ldots, d\} \right\},$$

i.e. only one entry of $\widetilde{Q}$ is equal to 0 and the others are equal to $\frac{1}{2}$.
For the statistical distance between these distributions, we again use the χ^2-divergence.

With $\Omega = \{0,1\}^d$, in very close analogy to the computation in (3.6), we have

$$\mathrm{div}_{\chi^2}\left(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}\right) = \left(2^d \sum_{\omega\in\Omega} f_\rho^2(\omega)\right)^M$$

with

$$f_\rho(\omega) = \mathbb{E}_V\left[\prod_{i=1}^d V_i^{\omega_i}(1-V_i)^{1-\omega_i}\right] = \frac{1}{d}\sum_{i=1}^d \frac{1}{2^{d-1}}\mathbb{1}_{\{\omega_i=0\}},$$

so that using the combinatorial structure

$$\begin{aligned}
\sum_{\omega\in\Omega} f_\rho^2(\omega) &= \frac{4}{d^2 4^d}\sum_{\omega\in\Omega}\left(\sum_{i=1}^d \mathbb{1}_{\{\omega_i=0\}}\right)^2 \\
&= \frac{4}{d^2 4^d}\cdot\sum_{k=0}^d \binom{d}{k} k^2 \\
&= \frac{4}{d^2 4^d}\cdot d\sum_{k=1}^d \binom{d-1}{k-1} k \\
&= \frac{4}{d 4^d}\cdot\sum_{k=0}^{d-1} \binom{d-1}{k}(k+1) \\
&\le \frac{4}{d4^d}\cdot d2^{d-1} \\
&= \frac{2}{2^d}
\end{aligned}$$

and finally

$$\mathrm{div}_{\chi^2}(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}) \le 2^M$$

which, plugged in (1.10), yields the claim.

Alternative Derivation

Computing the χ^2-divergence for the above construction is a worthwile exercise, but we would also like to mention another solution based on total variation distance which is also appealing as it is simpler.
For the null hypothesis, we choose the prior given in (3.5); for the alternative hypothesis, we only change the distribution of the first coordinate to

$$\mathrm{Ber}\left(\frac{1}{2}\right)^{\otimes M} \otimes \mathrm{Ber}\,(0)^{\otimes M}.$$

That is, writing $B_p := \mathrm{Ber}(p)$, $p \in [0,1]$, we consider

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0} = \left(B_{1/2}^{\otimes M}\otimes B_{1/2}^{\otimes M}\right)^{\otimes d},$$

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho} = \left(B_{1/2}^{\otimes M}\otimes B_{0}^{\otimes M}\right)\otimes\left(B_{1/2}^{\otimes M}\otimes B_{1/2}^{\otimes M}\right)^{\otimes(d-1)}.$$

Note that we are still in the regime $\rho = \infty$ as under the prior ν_ρ, $\widetilde{Q}_1 = 0$ and $\widetilde{P}_1 > 0$. Now, due to subadditivity of the total variation distance, we immediately obtain

$$\begin{aligned}\|\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho} - \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}\|_{\mathrm{TV}} &\leq \|B_0^{\otimes M} - B_{1/2}^{\otimes M}\|_{\mathrm{TV}} \\ &= 2\sup_{A\in\mathcal{A}} \left| B_0^{\otimes M}(A) - B_{1/2}^{\otimes M}(A) \right|,\end{aligned}$$

where $\mathcal{A}$ is the power set of $\{0,1\}^M$. We now upper bound this supremum by distinguishing two cases: Firstly, whenever $(0,0,\ldots,0) \in A$, we have $B_0^{\otimes M}(A) = 1$ and $1 \geq B_{1/2}^{\otimes M}(A) \geq \frac{1}{2^M}$. On the other hand, whenever $(0,0,\ldots,0) \notin A$, we have $B_0^{\otimes M}(A) = 0$ and $0 \leq B_{1/2}^{\otimes M}(A) \leq 1 - \frac{1}{2^M}$. Therefore, in any case $\left| B_0^{\otimes M}(A) - B_{1/2}^{\otimes M}(A) \right| \leq 1 - \frac{1}{2^M}$ and hence

$$\|\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho} - \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}\|_{\mathrm{TV}} \leq 2\left(1 - \frac{1}{2^M}\right).$$

Plugging this in (1.9) leads to the condition

$$M < \frac{\ln(1/\eta)}{\ln(2)}$$

which may also have been given in Theorem 3.3.

3.6.3 Analysis of the Naive Test Statistic for Frobenius Norm

Let $m \in \{1,2,\ldots,M\}$ and $i \in \{1,2,\ldots,d\}$. With the vector notation $\widetilde{A}^m, \widetilde{B}^m, \widetilde{P}, \widehat{Q}$ as defined earlier, the test statistic reads

$$T_0 := \frac{2}{M}\sum_{m=1}^{M}\sum_{i=1}^{d}(\widetilde{A}_i^m - \widetilde{B}_i^m)^2.$$

Using the fact that $\widetilde{A}_i^m$ and $\widetilde{B}_i^m$ are $\mathrm{Ber}(\widetilde{P}_i)$- and $\mathrm{Ber}(\widetilde{Q}_i)$-distributed, respectively, and the independence assumptions, we obtain

$$\mathbb{E}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^2] = \widetilde{P}_i + \widetilde{Q}_i - 2\widetilde{P}_i\widetilde{Q}_i$$

and hence

$$(\widetilde{P}_i - \widetilde{Q}_i)^2 \leq \mathbb{E}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^2] \leq \frac{1}{2},$$

The lower bound is based on $\widetilde{P}_i \geq \widetilde{P}_i^2$ and $\widetilde{Q}_i \geq \widetilde{Q}_i^2$, the upper bound can be proved by analytical maximization of $f(x,y) = x + y - 2xy$ on $[0,1]^2$. Both bounds are sharp in the sense that they can be attained for specific $\widetilde{P}_i, \widetilde{Q}_i$. Furthermore, using $(\widetilde{A}_i^m - \widetilde{B}_i^m)^4 \leq (\widetilde{A}_i^m - \widetilde{B}_i^m)^2$, we have

$$\begin{aligned}\mathbb{V}\mathrm{ar}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^2] &= \mathbb{E}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^4] - \mathbb{E}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^2]^2 \\ &\leq \mathbb{E}[(\widetilde{A}_i^m - \widetilde{B}_i^m)^2] \\ &\leq \frac{1}{2}.\end{aligned}$$

These auxiliary computations now yield the following bounds on expectation and variance of T_0: It holds that

$$\|P-Q\|_{\mathrm{F}}^2 \leq \mathbb{E}[T_0] \leq d$$

and

$$\mathbb{V}\mathrm{ar}[T_0] \leq \frac{2d}{M}.$$

Now, on the one hand, under the null hypothesis Chebyshev's inequality tells us that for $\eta \in (0,1)$,

$$\mathbb{P}_{(\widetilde{P},\widetilde{P})}\left(T_0 > d + \sqrt{\frac{4d}{M\eta}}\right) \leq \frac{\eta}{2},$$

which suggests the test

$$\varphi = \mathbb{1}_{\{T_0 > \tau_{\frac{\eta}{2}}\}} \quad \text{with} \quad \tau_{\frac{\eta}{2}} = d + \sqrt{\frac{4d}{M\eta}}.$$

On the other hand, under the alternative hypothesis, we have

$$\mathbb{P}_{(\widetilde{P},\widetilde{Q})}\left(T_0 \leq \|P-Q\|_{\mathrm{F}}^2 - \sqrt{\frac{4d}{M\eta}}\right) \leq \frac{\eta}{2}.$$

The resulting condition for a sufficiently small testing error reads

$$\|P-Q\|_{\mathrm{F}} \geq \sqrt{d + 2\sqrt{\frac{4d}{M\eta}}} \sim \sqrt{d}$$

as claimed.

3.6.4 Proof of Theorem 3.5

Lower Bound

Case $M \geq 2$

We use the same construction as in section 3.6.1 until equation (3.9). Given the distribution $\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}$, clearly we have

$$\|P-Q\|_{\mathrm{F}}^2 = 2\|\widetilde{P}-\widetilde{Q}\|_2^2 = 2d\gamma^2,$$

so that the requirement (3.9) (which comes from (1.10)) now reads

$$\frac{\|P-Q\|_{\mathrm{F}}}{\sqrt{2d}} < \sqrt[4]{\frac{1}{2dM^2}\ln(1+4(1-\eta)^2)}$$

or

$$\|P-Q\|_{\mathrm{F}} < \frac{\sqrt[4]{d}}{\sqrt{M}}\sqrt[4]{2\ln(1+4(1-\eta)^2)}.$$

We may express this in terms of n through

$$\sqrt[4]{d} = \sqrt[4]{\frac{n(n-1)}{2}} \geq \sqrt[4]{\frac{n^2}{4}} = \frac{\sqrt{n}}{\sqrt{2}}$$

and obtain the sufficiently small bound

$$\|P - Q\|_{\mathrm{F}} \leq \frac{\sqrt{n}}{\sqrt{M}} \sqrt[4]{\frac{1}{3} \ln(1 + 4(1 - \eta)^2)}$$

given in Theorem 3.5.

Case $M = 1$
We plug $M = 1$ into equation (3.7) and obtain

$$\sum_{\omega \in \Omega} f_\rho^2(\omega) = \frac{1}{2}, \quad \text{so that} \quad \mathrm{div}_{\chi^2}\left(\mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_0}, \mathbb{P}_{(\widetilde{P},\widetilde{Q})\sim\nu_\rho}\right) = 1.$$

This expression fulfils (1.10) for any η and γ, so that in particular the choice $\gamma = \frac{1}{2}$ yields

$$\|P - Q\|_{\mathrm{F}} = \sqrt{2d}\gamma = \sqrt{\frac{d}{2}} \geq \frac{n}{\sqrt{8}}.$$

for the claimed lower bound.

Upper Bound

Case $M \geq 2$

The test statistic

Again with our vector notation, we use the test statistic

$$T := \frac{8}{M^2} \sum_{i=1}^{d} \left(\sum_{m=1}^{M/2} (\widetilde{A}_i^m - \widetilde{B}_i^m) \right) \left(\sum_{m=M/2+1}^{M} (\widetilde{A}_i^m - \widetilde{B}_i^m) \right).$$

Then we have $\mathbb{E}[T] = \rho^2$. We also introduce the shifted version

$$T^* := \frac{8}{M^2} \sum_{i=1}^{d} \underbrace{\left(\sum_{m=1}^{M/2} \left[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i) \right] \right)}_{=: S_1^i} \underbrace{\left(\sum_{m=M/2+1}^{M} \left[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i) \right] \right)}_{=: S_2^i}$$

which, by direct expansion, has the decomposition

$$T^* = T + 2 \underbrace{\sum_{i=1}^{d} (\widetilde{P}_i - \widetilde{Q}_i)^2}_{= \rho^2} - \underbrace{\frac{4}{M} \sum_{i=1}^{d} (\widetilde{P}_i - \widetilde{Q}_i) \sum_{m=1}^{M} (\widetilde{A}_i^m - \widetilde{B}_i^m)}_{=: K},$$

so that, noting $\mathbb{E}[K] = 2\rho^2$,

$$T - \mathbb{E}[T] = T^* + K - \mathbb{E}[K].$$

The central part of the proof consists in studying the concentration of T^* and $K - \mathbb{E}[K]$, which enables us to control the type-I and type-II errors.

An analytical lemma

As a preparation for an explicit concentration bound on T^*, we need the following independent lemma:

Lemma 3.9. For $n \in \mathbb{N}\backslash\{0\}$ and $z, c \geq 0$, consider the sequence $(T_k)_{k\in\{1,2,\dots,n\}}$ recursively defined through

$$T_1 := z, \quad T_{k+1} = T_k + cT_k^2, \quad k \in \{1, 2, \dots, n-1\}.$$

Then, if $z \leq \frac{1}{4cn}$, for each $k \in \{1, 2, \dots, n\}$ we have

$$T_k \leq \left(1 + \frac{k}{n}\right) z \leq 2z$$

and in particular

$$\sum_{k=1}^{n} T_k^2 (n-k) \leq 4z^2 \sum_{k=1}^{n} (n-k) = 4dz^2 \leq 2n^2 z^2.$$

◁

Proof. By induction. The case $k = 1$ is clear. Now, assume that for some $k \in \{1, 2, \dots, n-1\}$, $T_k \leq \left(1 + \frac{k}{n}\right) z$. Then we have

$$\begin{aligned}
T_{k+1} &= T_k(1 + cT_k) \\
&\leq \left(1 + \frac{k}{n}\right) z \cdot \left(1 + \frac{1}{4n}\left(1 + \frac{k}{n}\right)\right) \\
&= \left[1 + \frac{k}{n} + \frac{1}{n}\left(\left(\frac{1}{2}\right)^2 + 2 \cdot \frac{1}{2} \cdot \frac{k}{2n} + \left(\frac{k}{2n}\right)^2\right)\right] z \\
&= \left[1 + \frac{k}{n} + \frac{1}{n}\left(\frac{1}{2} + \frac{k}{2n}\right)^2\right] z \\
&\leq \left(1 + \frac{k+1}{n}\right) z.
\end{aligned}$$

□

Concentration of T^*

Let $t > 0$. Then we have by the iid property across the d coordinates and the Chernoff technique

$$\mathbb{P}(T^* \geq t) = \mathbb{P}\left(\frac{8}{M^2} \sum_{i=1}^{d} S_1^i \cdot S_2^i \geq t\right) \leq \inf_{\lambda > 0} \frac{\mathbb{E}[\exp(\lambda \cdot S_1^1 \cdot S_2^1)]^d}{\exp\left(\frac{M^2}{8}\lambda t\right)}. \tag{3.10}$$

So, we ought to study $\mathbb{E}[\lambda \cdot S_1^1 \cdot S_2^1]$ a little closer: Clearly, $S_1 = S_1^1$ and $S_2 = S_2^1$ are iid. Moreover, the variables

$$X_m := (\widetilde{A}_1^m - \widetilde{B}_1^m) - (\widetilde{P}_1 - \widetilde{Q}_1)$$

are independent, centered and bounded in $[-2, 2]$, so that, by Hoeffding's lemma, their moment- generating functions (mgf) are bounded by $\exp(2(\bullet)^2)$. This means that the mgf of S_1 and S_2 is less than $\exp(M(\bullet)^2)$ and as a consequence

$$\begin{aligned}
\mathbb{E}[\lambda \cdot S_1 \cdot S_2] &= \mathbb{E}_{S_1}\left[\mathbb{E}_{S_2}\left[\exp(\lambda \cdot S_1 \cdot S_2)\right]\right] \\
&\leq \mathbb{E}_{S_1}\left[\exp(M\lambda^2 (S_1)^2)\right].
\end{aligned}$$

We now consider

$$U_k := \sum_{j=2}^{k} X_k \underbrace{\sum_{l=1}^{j-1} X_l}_{=:D_{j-1}}, \quad k \in \{2, 3, \dots, \tfrac{M}{2}\}.$$

and noting that $(S_1)^2 = \sum_{m=1}^{M/2} X_m^2 + 2U_{M/2} \leq 2M + 2U_{M/2}$, we have with $z := 2M\lambda^2$ and the notation from Lemma 3.9 with $c = 4, n = \frac{M}{2}$

$$\begin{aligned}
\mathbb{E}[\lambda \cdot S_1 \cdot S_2] &\leq \exp(2M^2\lambda^2)\mathbb{E}\left[\exp(2M\lambda^2 U_{M/2})\right] \\
&\leq \exp(zM)\mathbb{E}\left[\exp(zU_{M/2})\right] \\
&= \exp(zM)\mathbb{E}\left[\mathbb{E}_{X_{M/2}}\left[\exp\left(T_1 X_{M/2} D_{M/2-1}\right)\right]\exp(T_1 U_{M/2-1})\right] \\
&\leq \exp(zM)\mathbb{E}\left[\exp(2T_1^2 D_{M/2-1}^2 + T_1 U_{M/2-1})\right] \\
&\leq \exp(zM)\exp\left(8T_1^2\left(\tfrac{M}{2} - 1\right)\right)\mathbb{E}\left[\exp(\left(T_1 + 4T_1^2\right)U_{M/2-1})\right], \\
&= \exp(zM)\exp\left(8T_1^2\left(\tfrac{M}{2} - 1\right)\right)\mathbb{E}\left[\exp(T_2 U_{M/2-1})\right],
\end{aligned}$$

where we decompose and bound $D_{M/2-1}^2$ in analogy to $(S_1)^2$ above. We continue this process and obtain by Lemma 3.9 with the restriction $z \leq \frac{1}{4cn}$, i.e. $\lambda \leq \frac{1}{4M}$:

$$\begin{aligned}
\mathbb{E}[\lambda \cdot S_1 \cdot S_2] &\leq \exp(zM)\exp\left(8\sum_{k=1}^{M/2-2} T_k^2\left(\tfrac{M}{2} - k\right)\right)\mathbb{E}[\exp(T_{M/2-1} X_2 X_1)] \\
&\leq \exp(zM)\exp\left(8\sum_{k=1}^{M/2-1} T_k^2\left(\tfrac{M}{2} - k\right)\right) \\
&\leq \exp(2M^2\lambda^2)\exp\left(8 \cdot 2 \cdot (M/2)^2 \cdot 4 \cdot M^2 \cdot \lambda^4\right) \\
&\leq \exp(3M^2\lambda^2).
\end{aligned}$$

We are now able to carry on (3.10):

$$\begin{aligned}
\mathbb{P}(T^* \geq t) \leq \inf_{\lambda > 0} \frac{\mathbb{E}[\exp(\lambda \cdot S_1^1 \cdot S_2^1)]^d}{\exp\left(\frac{M^2}{8}\lambda t\right)} &\leq \inf_{0 < \lambda \leq \frac{1}{4M}} \frac{\exp(3M^2\lambda^2)^d}{\exp\left(\frac{M^2}{8}\lambda t\right)} \\
&= \inf_{0 < \lambda \leq \frac{1}{4M}} \exp\left(3\lambda^2 - \frac{1}{8d}\lambda t\right)^{M^2 d} \\
&= \exp\left(-3\frac{M^2}{d}\frac{t^2}{48^2}\right), \qquad (3.11)
\end{aligned}$$

where the last equality requires $t \leq \frac{12d}{M}$.
This bound also applies to $-T^*$ since the (mgf-)bounds for X_m are also valid for $-X_m$. Therefore, in summary:

$$\forall t \in \left[0, \tfrac{12d}{M}\right]: \quad \mathbb{P}(|T^*| \geq t) \leq 2\exp\left(-3\frac{M^2}{d}\frac{t^2}{48^2}\right).$$

Concentration of $\boldsymbol{K} - \mathbb{E}[\boldsymbol{K}]$

This term is only relevant under H_1, so wie assume $\rho > 0$. Firstly, write

$$K - \mathbb{E}[K] = \frac{4}{M} \sum_{i=1}^{d} \sum_{m=1}^{M} \underbrace{(\widetilde{P}_i - \widetilde{Q}_i)[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i)]}_{:=X_i^m}.$$

Now, we reuse arguments from the previous step: The X_i^m are independent, centered and bounded in $[-2|\widetilde{P}_i - \widetilde{Q}_i|, 2|\widetilde{P}_i - \widetilde{Q}_i|]$, so that their mgfs are bounded by $\exp(2(\widetilde{P}_i - \widetilde{Q}_i)^2(\bullet)^2)$. Now, for $t > 0$,

$$\begin{aligned}
\mathbb{P}(K - \mathbb{E}[K] \geq t) &= \mathbb{P}\left(\sum_{i=1}^{d} \sum_{m=1}^{M} X_i^m \geq \frac{M}{4} t\right) \\
&\leq \inf_{\lambda > 0} \frac{\prod_{i=1}^{d} \prod_{m=1}^{M} \exp(2(\widetilde{P}_i - \widetilde{Q}_i)^2 \lambda^2)}{\exp\left(\frac{M}{4} \lambda t\right)} \\
&= \inf_{\lambda > 0} \frac{\exp(M \rho^2 \lambda^2)}{\exp\left(\frac{M}{4} \lambda t\right)} \\
&= \inf_{\lambda > 0} \exp\left(\rho^2 \lambda^2 - \frac{1}{4} \lambda t\right)^M \\
&= \exp\left(-\frac{M}{\rho^2} \frac{t^2}{64}\right).
\end{aligned}$$

As above, this bound also holds for $-K + \mathbb{E}[K]$.

Computation of the upper bound

Firstly, under H_0, $T = T^*$, so that a suitable rejection threshold $\tau_{\frac{\eta}{2}}$ can be obtained by solving

$$2 \exp\left(-3 \frac{M^2}{d} \frac{t^2}{48^2}\right) \leq \frac{\eta}{2},$$

keeping in mind the restriction $\tau_{\frac{\eta}{2}} \leq \frac{12d}{M}$. By direct computation, we get the threshold

$$\tau_{\frac{\eta}{2}} = 24 \frac{n}{M} \sqrt{\frac{1}{3} \ln\left(\frac{4}{\eta}\right)},$$

being valid if $n \geq 1 + 4\sqrt{\frac{1}{3} \ln\left(\frac{4}{\eta}\right)}$. (E.g. for $\eta = 0.1$, $n \geq 6$).

Secondly, under H_1, we assume that $\rho^2 \geq \tau_{\frac{\eta}{2}}$. We write

$$\begin{aligned}
\mathbb{P}(|T| \leq \tau_{\frac{\eta}{2}}) &\leq \mathbb{P}(|T - \mathbb{E}[T]| \geq \rho^2 - \tau_{\frac{\eta}{2}}) \\
&\leq \mathbb{P}(|T^*| + |K - \mathbb{E}[K]| \geq \rho^2 - \tau_{\frac{\eta}{2}}) \\
&\leq \mathbb{P}(|T^*| \geq \frac{\rho^2 - \tau_{\frac{\eta}{2}}}{2}) + \mathbb{P}(|K - \mathbb{E}[K]| \geq \frac{\rho^2 - \tau_{\frac{\eta}{2}}}{2}) \\
&\leq 2 \exp\left(-\frac{1}{64} \frac{M^2}{d} (\rho^2 - \tau_{\frac{\eta}{2}})^2\right) + 2 \exp\left(-\frac{M}{\rho^2} \frac{(\rho^2 - \tau)^2}{2^8}\right),
\end{aligned}$$

where we keep the condition $\frac{\rho^2-\tau_{\frac{\eta}{2}}}{2} \leq \frac{12d}{M}$ in mind. Now, we want to find $\rho > 0$ such that both terms are at most $\frac{\eta}{4}$ for. By direct computation, this yields the sufficient condition

$$\rho \geq \max\left(\sqrt{\tau_{\frac{\eta}{2}}} + \sqrt{8}\frac{\sqrt{n}}{\sqrt{M}}\sqrt{\ln\left(\frac{8}{\eta}\right)}, \sqrt{\tau_{\frac{\eta}{2}}} + \frac{16}{\sqrt{M}}\sqrt{\ln\left(\frac{8}{\eta}\right)}\right);$$

simpler (weaker) conditions are

$$\rho \geq \sqrt{\tau_{\frac{\eta}{2}}} + 16\frac{\sqrt{n}}{\sqrt{M}}\sqrt{\ln\left(\frac{8}{\eta}\right)}$$

or even

$$\rho \geq 35\frac{\sqrt{n}}{\sqrt{M}}\sqrt{\ln\left(\frac{8}{\eta}\right)},$$

being valid if $n \geq 1 + \frac{3}{2}\ln\left(\frac{8}{\eta}\right)$. (E.g. for $\eta = 0.1$, $n \geq 8$).

Finally, we observe that the right hand side is of the same order as $\sqrt{\tau_{\frac{\eta}{2}}}$. Hence, we have the desired result

$$\rho^* \lesssim \frac{\sqrt{n}}{\sqrt{M}}.$$

Case $M = 1$
Rather than proposing and analysing a test, we just observe that for any $P, Q \in \mathcal{G}_n$, by construction

$$\|P - Q\|_{\mathrm{F}} \leq \sqrt{n(n-1)} \leq n,$$

so that the claimed upper bound is in fact a trivial one.

3.6.5 Proof of Theorem 3.7

Lower Bound

Priors (simple for both hypotheses)
Let $\gamma \in (0, \frac{1}{2}]$ and $M \in \mathbb{N}$ with $M \geq 2$.
For the null hypothesis, let ν_0 take only the value $\theta_0 = (P, Q)$ such that

$$P_{12} = P_{21} = \frac{\gamma}{\sqrt{2}M} \quad \text{and} \quad P_{ij} = 0 \quad \text{otherwise.}$$

Then

$$\rho = \|P - Q\|_{\mathrm{F}} = 0 \quad \text{and} \quad \|P + Q\|_{\mathrm{F}} = 2\|P\|_{\mathrm{F}} = 2\frac{\gamma}{M},$$

so that the first Δ-restriction translates to $\gamma \leq M\Delta$.

On the other hand, let ν_ρ take only the value $\theta_1 = (P, Q)$ which differs from θ_0 above only through

$$Q_{12} = Q_{21} = \frac{\sqrt{2}\gamma}{M}.$$

Then

$$\rho = \|P - Q\|_{\mathrm{F}} = \frac{\gamma}{M} \quad \text{and} \quad \|P + Q\|_{\mathrm{F}} = 3\frac{\gamma}{M}$$

and have the stronger Δ-restriction $\gamma \leq \frac{1}{3}M\Delta$ and now we specifically choose $\gamma = \frac{1}{3}M\Delta$.

Probabilistic distance
Clearly, by construction, the probabilistic distance between $\mathbb{P}_{\theta\sim\nu_0}$ and $\mathbb{P}_{\theta\sim\nu_\rho}$ solely lies in entry $(1,2)$ of the respective matrix Q. More precisely, let $\alpha = \frac{\gamma}{\sqrt{2}}$ and

$$X \sim \mathrm{Ber}(\alpha/M)^{\otimes M}, \quad Y \sim \mathrm{Ber}(2\alpha/M)^{\otimes M}.$$

Then

$$\begin{aligned}
\mathrm{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) &= \mathrm{div}_{\chi^2}(\mathrm{Ber}(\alpha/M)^{\otimes M}, \mathrm{Ber}(2\alpha/M)^{\otimes M}) \\
&= \sum_{\omega\in\{0,1\}^M} \frac{\mathbb{P}(Y=\omega)^2}{\mathbb{P}(X=\omega)} \\
&= \sum_{k=0}^{M} \binom{M}{k} \frac{\left[\left(\frac{2\alpha}{M}\right)^k \left(1-\frac{2\alpha}{M}\right)^{M-k}\right]^2}{\left(\frac{\alpha}{M}\right)^k \left(1-\frac{\alpha}{M}\right)^{M-k}} \\
&= \sum_{k=0}^{M} \binom{M}{k} \left(\frac{4\alpha}{M}\right)^k \left(\frac{\left(1-\frac{\alpha}{M}\right)^2 - 2\left(1-\frac{\alpha}{M}\right)\frac{\alpha}{M} + \frac{\alpha^2}{M^2}}{1-\frac{\alpha}{M}}\right)^{M-k} \\
&= \sum_{k=0}^{M} \binom{M}{k} \left(\frac{4\alpha}{M}\right)^k \left(1 - \frac{\alpha}{M} - 2\frac{\alpha}{M} + \frac{\alpha^2 M}{M^2(M-\alpha)}\right)^{M-k} \\
&= \left(1 + \frac{\alpha}{M} + \frac{\alpha^2}{M(M-\alpha)}\right)^M \\
&\leq \left(1 + 2\frac{\alpha}{M-1} + \frac{\alpha^2}{(M-1)^2}\right)^M \qquad \text{(with } \alpha < 1, M \geq 2\text{)} \\
&= \left(1 + \frac{\alpha}{M-1}\right)^{2M} \\
&\leq \left(1 + \frac{2\alpha}{M}\right)^{2M}. \qquad \text{(with } M \geq 2\text{)}
\end{aligned}$$

Plugging this into (1.10) tells us that

$$\sqrt{\frac{\alpha}{M}} < \frac{1}{\sqrt{M}} \frac{\sqrt{\ln(1+4(1-\eta)^2)}}{2}$$

is sufficient for a large enough total error bound. Now, due to our prior construction, we can write the left hand side as

$$\sqrt{\frac{\alpha}{M}} = \frac{\sqrt{3}}{\sqrt[4]{2}} \frac{\|P-Q\|_{\mathrm{F}}}{\sqrt{\Delta}},$$

so that the above condition is fulfilled if e.g.

$$\|P-Q\|_{\mathrm{F}} \leq \frac{\sqrt{\Delta}}{\sqrt{M}} \frac{\sqrt[4]{2}\sqrt{\ln(1+4(1-\eta)^2)}}{2\sqrt{3}}.$$

Upper Bound

Remember

$$T - \mathbb{E}[T] = \underbrace{\frac{8}{M^2}\sum_{i=1}^{d} S_1^i S_2^i}_{=:T^*} + \underbrace{\frac{4}{M}\sum_{i=1}^{d}\sum_{m=1}^{M}(\widetilde{P}_i - \widetilde{Q}_i)[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i)]}_{=:K-\mathbb{E}[K]},$$

where

$$S_1^i = \sum_{m=1}^{M/2}\left[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i)\right], \quad S_2^i \overset{\text{iid}}{\sim} S_1^i.$$

Preparatory Computations

For $m \in \{1, 2, \ldots, M\}$ and $i \in \{1, 2, \ldots, d\}$, let

$$X_i^m = (\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i),$$
$$Y_i^m = (\widetilde{P}_i - \widetilde{Q}_i)[(\widetilde{A}_i^m - \widetilde{B}_i^m) - (\widetilde{P}_i - \widetilde{Q}_i)].$$

Then

$$\mathbb{V}\text{ar}[X_i^m] = \widetilde{P}_i(1 - \widetilde{P}_i) + \widetilde{Q}_i(1 - \widetilde{Q}_i) \leq \widetilde{P}_i + \widetilde{Q}_i$$

and

$$\begin{aligned}\mathbb{V}\text{ar}[Y_i^m] &= (\widetilde{P}_i - \widetilde{Q}_i)^2[\widetilde{P}_i(1 - \widetilde{P}_i) + \widetilde{Q}_i(1 - \widetilde{Q}_i)] \\ &\leq (\widetilde{P}_i - \widetilde{Q}_i)^2(\widetilde{P}_i + \widetilde{Q}_i).\end{aligned}$$

Hence, through independence and centrality

$$\mathbb{V}\text{ar}\left[\sum_{i=1}^{d} S_1^i S_2^i\right] = \sum_{i=1}^{d}\mathbb{V}\text{ar}[S_1^i]^2 \leq \sum_{i=1}^{d}\left(\frac{M}{2}(\widetilde{P}_i + \widetilde{Q}_i)\right)^2 = \frac{M^2}{8}\|P + Q\|_{\mathrm{F}}^2$$

and

$$\begin{aligned}\mathbb{V}\text{ar}\left[\sum_{i=1}^{d}\sum_{m=1}^{M} Y_i^m\right] = \sum_{i=1}^{d}\mathbb{E}\left[\left(\sum_{m=1}^{M} Y_i^m\right)^2\right] &= \sum_{i=1}^{d}\mathbb{V}\text{ar}\left[\sum_{m=1}^{M} Y_i^m\right] \\ &\leq \sum_{i=1}^{d} M(\widetilde{P}_i - \widetilde{Q}_i)^2(\widetilde{P}_i + \widetilde{Q}_i) \\ &\overset{(*)}{\leq} M\sqrt{\sum_{i=1}^{d}(\widetilde{P}_i - \widetilde{Q}_i)^4}\sqrt{\sum_{i=1}^{d}(\widetilde{P}_i + \widetilde{Q}_i)^2} \\ &\leq \frac{M}{2\sqrt{2}}\rho^2\|P + Q\|_{\mathrm{F}}.\end{aligned}$$

(To see $(*)$, we compare $\sum(\cdot)^4$ with ρ^4: Writing the latter, $\rho^2 \cdot \rho^2$, as a double sum $\sum_i \sum_j (\cdot)^2(\cdot)^2$, it is clear that $\sum_i (\cdot)^4$ is smaller as it corresponds to the summands with $i = j$ of the double sum.)

Concentration Statements
We can now apply Chebyshev's inequality to T^* and $K - \mathbb{E}[K]$: For any $\epsilon > 0$, we have

$$\begin{aligned}
\mathbb{P}(|T^*| \geq \rho) &\leq \frac{1}{\epsilon^2}\mathbb{V}\mathrm{ar}[T^*] \\
&\leq \frac{1}{\epsilon^2}\frac{64}{M^4} \cdot \frac{M^2}{8}\|P+Q\|_{\mathrm{F}}^2 \\
&= \frac{8}{\epsilon^2}\frac{\|P+Q\|_{\mathrm{F}}^2}{M^2} \\
&\leq \frac{8}{\epsilon^2}\frac{\Delta^2}{M^2}, \\
\mathbb{P}(|K - \mathbb{E}[K]| \geq \rho) &\leq \frac{1}{\epsilon^2}\mathbb{V}\mathrm{ar}[K] \\
&\leq \frac{1}{\epsilon^2}\frac{16}{M^2} \cdot \frac{M}{2\sqrt{2}}\rho^2\|P+Q\|_{\mathrm{F}} \\
&= \frac{8}{\sqrt{2}\epsilon^2} \cdot \frac{\rho^2\|P+Q\|_{\mathrm{F}}}{M}. \\
&\leq \frac{8}{\sqrt{2}\epsilon^2} \cdot \frac{\rho^2\Delta}{M}.
\end{aligned}$$

Computation of the Upper Bound
For $\tau_{\frac{\eta}{2}} > 0$,

$$\mathbb{P}_{H_0}(|T| > \tau_{\frac{\eta}{2}}) = \mathbb{P}_{H_0}(|T^*| > \tau_{\frac{\eta}{2}}),$$

so that, by the above concentration statement,

$$\tau_{\frac{\eta}{2}} := \frac{\Delta}{M}\sqrt{\frac{16}{\eta}}$$

is a suitable rejection threshold for the test

$$\varphi(X) = \mathbb{1}_{\{T > \tau_{\frac{\eta}{2}}\}}.$$

Now, for $\rho^2 > \tau_{\frac{\eta}{2}}$, observe

$$\begin{aligned}
\mathbb{P}_{H_1}(|T| \leq \tau_{\frac{\eta}{2}}) &\leq \mathbb{P}_{H_1}(|T - \mathbb{E}[T]| \geq \rho^2 - \tau_{\frac{\eta}{2}}) \\
&\leq \mathbb{P}_{H_1}\left(|T^*| \geq \frac{\rho^2 - \tau_{\frac{\eta}{2}}}{2}\right) + \mathbb{P}\left(|K - \mathbb{E}[K]| \geq \frac{\rho^2 - \tau_{\frac{\eta}{2}}}{2}\right).
\end{aligned}$$

By the above bounds, we see that both these terms are bounded by $\frac{\eta}{4}$ if

$$\rho^2 \geq \max\left(\tau_{\frac{\eta}{2}} + \frac{8\sqrt{2}}{\sqrt{\eta}}\frac{\Delta}{M}, \tau_{\frac{\eta}{2}} + \rho\frac{8}{\sqrt{\eta}}\sqrt{\frac{\Delta}{M}}\right),$$

which is fulfilled if

$$\rho \geq \sqrt{\tau_{\frac{\eta}{2}}} + \frac{8\sqrt{2}}{\sqrt{\eta}}\sqrt{\frac{\Delta}{M}} = \left(\sqrt[4]{\frac{16}{\eta}} + \frac{8\sqrt{2}}{\sqrt{\eta}}\right)\sqrt{\frac{\Delta}{M}}.$$

3.6.6 Proof of Theorem 3.8

Lower Bound

Construction of the prior distributions

Let $\mathbb{I}_n$ denote the $n \times n$ identity matrix and $\mathbb{1}_n$ denote the $n \times n$ matrix where all entries are 1. For the null hypothesis, define the simple prior ν_0 on $\theta = (P, Q)$ that sets $P = Q = \frac{1}{2}(\mathbb{1}_n - \mathbb{I}_n)$. The prior ν_ρ for the alternative hypothesis also sets $P = \frac{1}{2}(\mathbb{1}_n - \mathbb{I}_n)$, but now the random matrix Q is constructed as follows:
Let $\gamma \in [0, \frac{1}{2}]$, $R_0 \equiv 1$ and $R_1, R_2, \dots, R_{n-1} \overset{\text{iid}}{\sim} \mathcal{U}_{\{-1,1\}}$. Consider the column vector

$$v = \gamma \begin{pmatrix} R_0 \\ R_1 \\ \vdots \\ R_{n-1} \end{pmatrix}$$

and build the matrix $A = (v, R_1 v, R_2 v, \dots, R_{n-1} v)$. Finally, define $Q := \frac{1}{2}\mathbb{1}_n + A - (\frac{1}{2} + \gamma)\mathbb{I}_n$ - this gives rise to a uniform distribution on a family of 2^{n-1} matrices.

Spectral distance under ν_1

The previous construction for the alternative hypothesis yields

$$\rho := \|P - Q\|_{\mathrm{S}} = \|A - \gamma \mathbb{I}_n\|_{\mathrm{S}}.$$

Since A has rank 1, its spectral norm is equal to its Frobenius norm, so that $\|A\|_{\mathrm{S}} = \sqrt{n^2\gamma^2} = n\gamma$, and $\|\gamma \mathbb{I}_n\|_{\mathrm{S}} = \gamma$ is clear. Hence, by the forward and inverse triangle inequalities

$$(n-1)\gamma \leq \rho \leq (n+1)\gamma.$$

Statistical distance between the hypotheses

In this step, we derive a first tangible expression for $D := \operatorname{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho})$. In order to compute the lower bound, more considerations will be necessary afterwards. We describe the distribution of M samples based on P or Q essentially as a product of d binomially distributed random variables and write the corresponding values as $(k_{ij})_{i,j} = (k_{ij})_{(i,j)\in S}$, where $S = \{(i,j) \in \{0, 1, \dots, n-1\}^2 \mid i < j\}$ and $k_{ij} \in \{0, 1, \dots, M\}$. The resulting density functions read

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}((k_{ij})_{i,j}) = \prod_{(i,j)\in S} \binom{M}{k_{ij}} \frac{1}{2^M},$$

$$\mathrm{d}\mathbb{P}_{\theta\sim\nu_\rho}((k_{ij})_{i,j}) = \mathbb{E}_R\left[\prod_{(i,j)\in S} \binom{M}{k_{ij}} \left(\frac{1}{2} + \gamma R_i R_j\right)^{k_{ij}} \left(\frac{1}{2} - \gamma R_i R_j\right)^{M-k_{ij}}\right],$$

where $R = (R_1, R_2, \dots, R_{n-1})$. Furthermore, introducing an independent copy R' of R, we can write the square of F_1 as

$$\mathrm{d}\mathbb{P}^2_{\theta\sim\nu_\rho}((k_{ij})_{i,j}) = \mathbb{E}_{R,R'}\left[\prod_{(i,j)\in S}\binom{M}{k_{ij}}^2\left(\left(\frac{1}{2}+\gamma R_iR_j\right)\left(\frac{1}{2}+\gamma R'_iR'_j\right)\right)^{k_{ij}}\cdot\right.$$
$$\left.\left(\left(\frac{1}{2}-\gamma R_iR_j\right)\left(\frac{1}{2}-\gamma R'_iR'_j\right)\right)^{M-k_{ij}}\right].$$

Due to the equal components of ν_0 and ν_1, we now have

$$\begin{aligned}
D &= \sum_{(k_{ij})_{i,j}} \frac{\mathrm{d}\mathbb{P}^2_{\theta\sim\nu_\rho}}{\mathrm{d}\mathbb{P}_{\theta\sim\nu_0}}((k_{ij})_{i,j}) \\
&= 2^{Nd}\mathbb{E}_{R,R'}\left[\prod_{(i,j)\in S}\sum_{k_{ij}=0}^{M}\binom{M}{k_{ij}}\left(\left(\frac{1}{2}+\gamma R_iR_j\right)\left(\frac{1}{2}+\gamma R'_iR'_j\right)\right)^{k_{ij}}\cdot\right.\\
&\qquad\left.\left(\left(\frac{1}{2}-\gamma R_iR_j\right)\left(\frac{1}{2}-\gamma R'_iR'_j\right)\right)^{M-k_{ij}}\right]\\
&= 2^{Nd}\mathbb{E}_{R,R'}\left[\prod_{(i,j)\in S}\left(\frac{1}{2}+2\gamma^2R_iR'_iR_jR'_j\right)^M\right]\\
&= \mathbb{E}_R\left[\prod_{(i,j)\in S}\left(1+4\gamma^2R_iR_j\right)^M\right],
\end{aligned}$$

where in the last step we use the fact that $R_iR'_iR_jR'_j = (R_iR_j)(R'_iR'_j)$ behaves just as R_iR_j by construction. The classical bound $1 + x \le \exp(x)$ and the abbreviation $z := 4M\gamma^2$ leads to

$$D \le \mathbb{E}_R\left[\exp\left(z\sum_{(i,j)\in S}R_iR_j\right)\right],$$

We now deal with this expression in the spirit of Lemma 3.9 and Theorem 3.5 (section 3.6.4). Writing

$$U_l = \sum_{j=1}^{l} R_j \underbrace{\sum_{i=0}^{j-1} R_i}_{=:S_{j-1}}, \quad l \in \{1, 2, \dots, n-1\}$$

and using $\cosh(x) \leq \exp(\frac{x^2}{2})$ now leads to

$$\begin{aligned}
D &\leq \mathbb{E}_R\left[\exp\left(zU_{n-1}\right)\right] \\
&= \mathbb{E}_{R_1,\ldots,R_{n-2}}\left[\exp\left(zU_{n-2}\right)\mathbb{E}_{R_{n-1}}\left[\exp(zR_{n-1}S_{n-2})\right]\right] \\
&= \mathbb{E}_{R_1,\ldots,R_{n-2}}\left[\exp\left(zU_{n-2}\right)\cosh(zS_{n-2})\right]. \\
&\leq \mathbb{E}_{R_1,\ldots,R_{n-2}}\left[\exp\left(zU_{n-2}\right)\exp\left(\frac{z^2}{2}S_{n-2}^2\right)\right].
\end{aligned}$$

Since $S_{n-2}^2 = 2U_{n-2} + (n-1)$, we obtain

$$\begin{aligned}
D &\leq \exp\left(\frac{z^2}{2}(n-1)\right)\mathbb{E}_{R_1,\ldots,R_{n-2}}\left[\exp((z+z^2)U_{n-2})\right] \\
&= \exp\left(\frac{1}{2}T_1^2(n-1)\right)\mathbb{E}_{R_1,\ldots,R_{n-2}}\left[\exp(T_2U_{n-2})\right]
\end{aligned}$$

with the notation from Lemma 3.9 with $c = 1$. Under the restriction $z \leq \frac{1}{4n}$, i.e. $\gamma \leq \frac{1}{4\sqrt{Mn}}$,

$$\begin{aligned}
D &\leq \exp\left(\frac{1}{2}\sum_{k=1}^{n-2}T_k^2(n-k)\right)\mathbb{E}_{R_1}\left[\exp(T_{n-1}U_1)\right] \\
&\leq \exp\left(\frac{1}{2}\sum_{k=1}^{n-1}T_k^2(n-k)\right), \\
&\leq \exp(n^2z^2). \qquad (3.12)
\end{aligned}$$

Conclusion

Again, we want to ensure (1.10), i.e.

$$\operatorname{div}_{\chi^2}(\mathbb{P}_{\theta\sim\nu_0}, \mathbb{P}_{\theta\sim\nu_\rho}) < 1 + 4(1-\eta)^2.$$

By the previous result and ensuring $z \leq \frac{1}{4n}$, this is fulfilled if

$$z < \frac{1}{n} \cdot \min\left(\sqrt{\ln(1+4(1-\eta)^2)}, \frac{1}{4}\right).$$

Since $z = 4M\gamma^2$, this reads

$$\gamma < \frac{1}{2\sqrt{M}\sqrt{n}} \cdot \min\left(\sqrt[4]{\ln(1+4(1-\eta)^2)}, \frac{1}{2}\right).$$

Finally, with $\gamma \leq \frac{\rho}{n-1}$ and $\frac{1}{2}\sqrt{n} \leq \frac{n-1}{\sqrt{n}}$, a sufficient condition in terms of ρ is

$$\rho \leq \frac{\sqrt{n}}{5\sqrt{M}} \cdot \min\left(\sqrt[4]{\ln(1+4(1-\eta)^2)}, \frac{1}{2}\right).$$

providing the desired result

$$\rho^* \gtrsim \frac{\sqrt{n}}{\sqrt{M}}.$$

Upper Bound

We still observe $n \times n$ adjacency matrices $A^1, \ldots, A^M$ according to P and $B^1, \ldots, B^M$ according to Q. As a test statistic, we consider

$$T = \|S_M\|_{\mathrm{S}} = \max_{u \in \partial \overline{B}} |u^T S_M u|, \quad S_M := \frac{1}{M} \sum_{m=1}^{M} (A^m - B^m),$$

where $\overline{B}$ is the closed unit Euclidean ball in $\mathbb{R}^n$. Furthermore, let $d = \binom{n}{2}$.

Preparation: general concentration bound

Our first step towards finding an upper bound on ρ^* is studying the concentration of

$$f(u, v) := u^T S_M v$$

with fixed $u, v \in \overline{B}$. In order to apply Hoeffding's inequality to this quantity, we must pay some attention: There are technically $M \cdot d$ independent random variables involved; a representation isolating these variables would be

$$\begin{aligned} f(u, v) &= \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{n} \sum_{j=1}^{n} u_i \underbrace{(A_{ij}^m - B_{ij}^m)}_{Z_{ij}^m \in [-1,1]} v_j \\ &= \frac{1}{M} \sum_{m=1}^{M} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} Z_{ij}^m (u_i v_j + u_j v_i). \end{aligned}$$

Now, the sum of squared ranges of these $M \cdot d$ random variables can be bounded as follows:

$$\begin{aligned} \sum_{m=1}^{M} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} (2(|u_i v_j| + |u_j v_i|))^2 &\leq 4M \sum_{i=1}^{n} \sum_{j=1}^{n} (u_i^2 v_j^2 + 2|u_i v_i u_j v_j| + u_j^2 v_i^2) \\ &\leq 4M \left(\|u\|_2^2 \|v\|_2^2 + 2 \left(\sum_{i=1}^{n} |u_i| \cdot |v_i| \right)^2 + \|u\|_2^2 \|v\|_2^2 \right) \\ &\leq 16M \|u\|_2^2 \|v\|_2^2, \end{aligned}$$

using the Cauchy-Schwarz-inequality. In Hoeffding's inequality, this yields

$$\begin{aligned} \mathbb{P}(||f(u,v)| - |\mathbb{E}[f(u,v)]|| \geq t) &\leq \mathbb{P}(|f(u,v) - \mathbb{E}[f(u,v)]| \geq t) \\ &\leq 2 \exp\left(- \frac{2(Mt)^2}{16M \|u\|_2^2 \|v\|_2^2} \right) \\ &= 2 \exp\left(- \frac{Mt^2}{8 \|u\|_2^2 \|v\|_2^2} \right), \end{aligned} \tag{3.13}$$

or, in the form we will actually employ:

$$\mathbb{P}\left(||f(u,v)| - |\mathbb{E}[f(u,v)]|| \geq \|u\|_2 \|v\|_2 \sqrt{\frac{8}{M} \ln\left(\frac{2}{\delta}\right)} \right) \leq \delta, \quad \delta \in (0,1). \tag{3.14}$$

In particular, if $u = v \in \partial\overline{B}$, we have

$$\mathbb{P}\left(\big||u^T S_M u| - |u^T(P-Q)u|\big| \geq \sqrt{\frac{8}{M}\ln\left(\frac{2}{\delta}\right)}\right) \leq \delta, \quad \delta \in (0,1). \tag{3.15}$$

Note that all this also holds without the inner absolute values on the left hand side due to (3.13).

Type-I-error

We want to employ a chaining-approach, see [GN16]. Informally speaking, the basic idea is that a $u^* \in \partial\overline{B}$ maximising

$$|f(u,u)| = |u^T S_M u|$$

admits a representation of the form

$$u^* = \sum_{i=1}^{\infty}(u_i - u_{i-1}),\ u_0 = \mathbb{0}_n,\ u_i \in \overline{B},\ \|u_i - u_{i-1}\|_2 \leq r_i := \frac{3}{2^{2+i}},$$

where, without making it explicit in the notation, the chain $(u_i)_{i\in\mathbb{N}_0}$ is "backwards unique" in the sense that u_{i-1} depends only on u_i.

Let us for now restrict the max to the first entry of f, that is consider only

$$\max_{\|u\|_2=1} |f(u,v)|, \quad v \in \partial\overline{B} \text{ fixed.}$$

The idea mentioned above is now used for a technique of successive finite coverings of $\overline{B}$: For $i \in \mathbb{N}$, there is a set $\mathcal{B}_i \subseteq \overline{B}$ of cardinality at most $N_i := \lfloor(\frac{3}{r_i})^n\rfloor$ that provides an r_i-covering of $\overline{B}$. Now, write

$$\begin{aligned}
\max_{\|u\|_2=1} |f(u,v)| &= \max_{\substack{\|u\|_2=1,\ u_1\in\mathcal{B}_1,\\ \|u-u_1\|_2\leq\rho_1}} |f(u_1,v) + f(u-u_1,v)| \\
&\leq \max_{u_1\in\mathcal{B}_1} |f(u_1,v)| + \max_{\substack{\|u\|_2=1,\ u_1\in\mathcal{B}_1,\\ \|u-u_1\|_2\leq\rho_1}} |f(u-u_1,v)| \\
&\leq \max_{u_1\in\mathcal{B}_1} |f(u_1,v)| + \max_{u_2\in\mathcal{B}_2} |f(u_2-u_1,v)| \\
&\qquad + \max_{\substack{\|u\|_2=1,\ u_2\in\mathcal{B}_2,\\ \|u-u_2\|_2\leq\rho_2}} |f(u-u_2,v)| \\
&\ \vdots \\
&\leq \sum_{i=1}^{\infty} \max_{u_i\in\mathcal{B}_i} |f(u_i-u_{i-1},v)|.
\end{aligned}$$

The second entry, v, is now dealt with in the same manner, leading to what can be seen as double-chaining:

$$\max_{\|u\|_2} f(u,u) \leq \max_{\|u\|=\|v\|=1} f(u,v) \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \max_{u_i\in\mathcal{B}_i, v_j\in\mathcal{B}_j} |f(u_i-u_{i-1}, v_j-v_{j-1})|.$$

For fixed $i, j \in \mathbb{N}$, using the standard union bound with (3.14) and plugging in $P = Q$ and the quantities

$$\delta_{ij} = \frac{\eta/2}{2^{i+j}}, \; r_{ij} = r_i r_j, \; N_{ij} = N_i N_j,$$

we obtain

$$\mathbb{P}\left(\max_{u_i \in \mathcal{B}_i, v_j \in \mathcal{B}_j} |f(u_i - u_{i-1}, v_j - v_{j-1})| \geq \frac{9}{2^{4+i+j}} \sqrt{\frac{8}{M} \ln\left(\frac{2^{(4+i+j)(n+1)}}{8(\eta/2)} \right)} \right) \leq \delta_{ij}. \tag{3.16}$$

For the next step, we mention a small lemma.

Lemma 3.10. For $x, y \geq \exp(2)$, we have

$$\ln(xy) \leq \ln(x)\ln(y).$$

◁

Proof. By direct computation, we see that for $x, y > 0$, the bound is equivalent to

$$(\ln(x) - 1)(\ln(y) - 1) \geq 1.$$

The claim is obvious now. □

This allows a useful bound in (3.16): We write

$$\frac{2^{(4+i+j)(n+1)}}{8(\eta/2)} = \left(\frac{2^{2+i}}{(8(\eta/2))^{1/(2(n+1))}} \cdot \frac{2^{2+j}}{(8(\eta/2))^{1/(2(n+1))}} \right)^{n+1}.$$

Clearly, as long as $(\eta/2) \leq \frac{1}{8}$, the two factors above are at least $8 > \exp(2)$, so we can apply the lemma.

We are now ready for the central calculation.

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{9}{2^{4+i+j}} \sqrt{\frac{8}{M} \ln\left(\frac{2^{(4+i+j)(n+1)}}{8(\eta/2)} \right)} = \frac{9\sqrt{8}}{16\sqrt{M}} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \frac{1}{2^i} \frac{1}{2^j} \sqrt{(n+1) \ln\left(\frac{2^{4+i+j}}{(8(\eta/2))^{1/(n+1)}} \right)}$$

$$\leq \frac{9\sqrt{8}}{16} \frac{\sqrt{n+1}}{\sqrt{M}} \left(\sum_{i=1}^{\infty} \frac{1}{2^i} \sqrt{\ln\left(\frac{2^{2+i}}{(8(\eta/2))^{1/(2(n+1))}} \right)} \right)^2$$

As an auxiliary computation, using $\sum_{i=1}^{\infty} \sqrt{2+i}\, 2^{-i} \leq 2$, we have

$$\sum_{i=1}^{\infty} \frac{1}{2^i} \sqrt{\ln\left(\frac{2^{2+i}}{(8(\eta/2))^{1/(2(n+1))}} \right)} \leq \sum_{i=1}^{\infty} \frac{\sqrt{2+i}}{2^i} \sqrt{\ln(2)} + \sum_{i=1}^{\infty} \frac{1}{2^i} \ln\left(\frac{1}{4\eta} \right)$$

$$\leq 2\sqrt{\ln(2)} + \ln\left(\frac{1}{4\eta} \right).$$

This discussion particularly applies under the null hypothesis as the expectation in (3.14) vanishes. Therefore, a suitable rejection threshold for our test is given by

$$\tau_{\frac{\eta}{2}} = \frac{\sqrt{n+1}}{\sqrt{M}} \frac{9\sqrt{8}}{16} \left(2\sqrt{\ln(2)} + \ln\left(\frac{1}{4\eta} \right) \right)^2$$

with

$$\mathbb{P}_{H_0}\left(T > \tau_{\frac{\eta}{2}}\right) \leq \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\delta_{ij} = \frac{\eta}{2}.$$

Type-II-error

We assume that $\rho := \|P-Q\|_{\mathrm{S}} \geq \tau_{\frac{\eta}{2}}$. By definition, for some deterministic $u_* \in \partial\overline{B}$,

$$\rho = |u_*^T(P-Q)u_*| = |\mathbb{E}[u_*^T S_M u_*]|.$$

Using this, we note that

$$\begin{aligned} T \leq \tau_{\frac{\eta}{2}} &\iff \max_{u\in\overline{B}} f_{S_M}(u) \leq \tau_{\frac{\eta}{2}} \\ &\implies |u_*^T S_M u_*| \leq \tau_{\frac{\eta}{2}} \\ &\implies |u_*^T S_M u_* - \mathbb{E}[u_*^T S_M u_*]| \geq |\mathbb{E}[u_*^T S_M u_*]| - \tau_{\frac{\eta}{2}} = \rho - \tau_{\frac{\eta}{2}} \geq 0. \end{aligned}$$

Hence,

$$\mathbb{P}(T \leq \tau_{\frac{\eta}{2}}) \leq \mathbb{P}(|u_*^T S_M u_* - \mathbb{E}[u_*^T S_M u_*]| \geq \rho - \tau_{\frac{\eta}{2}}),$$

which, using (3.15), yields the sufficient separation distance

$$\rho \geq \tau_{\frac{\eta}{2}} + \sqrt{\frac{8}{M}\ln\left(\frac{4}{\eta}\right)},$$

which may be upper bounded by

$$\frac{\sqrt{n}}{\sqrt{M}}\left(2\left(2\sqrt{\ln(2)} + \ln\left(\frac{1}{4\eta}\right)\right)^2 + \sqrt{2\ln\left(\frac{4}{\eta}\right)}\right).$$

Chapter 4

Testing the Sobolev-Type Regularity of a Function

In this chapter we study the problem of testing if an L_2-function f belonging to a certain l_2-Sobolev-ball $B_t(R)$ of radius $R > 0$ with smoothness level $t > 0$ indeed exhibits a higher smoothness level $s > t$, that is, belongs to $B_s(R)$. We assume that only a perturbed version of f is available, where the noise is governed by a standard Brownian motion scaled by $\frac{1}{\sqrt{n}}$. More precisely, considering a testing problem of the form

$$H_0: \; f \in B_s(R) \;\; \text{vs.} \;\; H_1: \; f \in B_t(R), \;\; \inf_{h \in B_s} \|f - h\|_{L_2} > \rho$$

for some $\rho > 0$, we approach the task of identifying the smallest value for ρ, denoted ρ^*, enabling the existence of a test φ with small error probability in a minimax sense. By deriving lower and upper bounds on ρ^*, we expose its precise dependence on n:

$$\rho^* \sim n^{-\frac{t}{2t+1/2}}.$$

As a remarkable aspect of this composite-composite testing problem, it turns out that the rate does not depend on s and is equal to the rate in signal-detection, i.e. the case of a simple null hypothesis.

4.1 Introduction

Let $n \in \mathbb{N}^* = \mathbb{N}\backslash\{0\}$, f a fixed unknown element of

$$L_2 := L_2([0,1]) = \left\{ g : [0,1] \to \mathbb{R} \; ; \; \int_0^1 g(x)^2 \, \mathrm{d}\lambda(x) < \infty \right\}$$

and $(B(x))_{x \in [0,1]}$ a standard Brownian motion. Suppose we observe the Gaussian process $(Y(x))_{x \in [0,1]}$ determined by the stochastic differential equation

$$\mathrm{d}Y(x) = f(x)\mathrm{d}x + \frac{1}{\sqrt{n}}\mathrm{d}B(x), \quad x \in [0,1]. \tag{4.1}$$

The resulting probability measure given f will be written $\mathbb{P}_f$.

Testing problem

We now fix $s > t > 0$ and $R, \rho > 0$. For any $r > 0$, we denote by $B_r(R)$ the l_2-Sobolev-ball of radius R of functions on $[0,1]$ with regularity at least r – see section 4.2 for a precise definition. Based on that, let

$$\widetilde{B}_{s,t}(R,\rho) := \left\{ g \in B_t(R) \ ; \ \inf_{h \in B_s(R)} \|g-h\|_{L_2} > \rho \right\}.$$

Hence, if we interpret s and t as degrees of smoothness, $\widetilde{B}_{s,t}(R,\rho)$ is the set of functions with smoothness level at least t which are separated from the class $B_s(R)$ with stronger smoothness s by ρ in L_2-sense. Now, the testing problem of interest is

$$H_0: \ f \in B_s(R) \ \text{ vs. } \ H_1: \ f \in \widetilde{B}_{s,t}(R,\rho). \tag{4.2}$$

More specifically, given $\eta \in (0,1)$, we aim at finding the magnitude in terms of n of the smallest separation distance $\rho^*(\eta) = \rho^*(n,s,t,\eta)$ which enables the existence of a test φ of level η in a minimax sense, i.e. of

$$\rho^*(\eta) = \inf \left\{ \rho > 0 \ ; \ \exists \text{ test } \varphi : \sup_{f \in B_s(R)} \mathbb{P}_f(\varphi = 1) + \sup_{f \in \widetilde{B}_{s,t}(R,\rho)} \mathbb{P}_f(\varphi = 0) \le \eta \right\}. \tag{4.3}$$

In the framework of chapter 1, the problem reads

$$\begin{cases} \Theta &= B_t(R), \\ \mathbb{P}_\theta &= \mathcal{L}\left((Y(x))_{x\in[0,1]} \mid \theta \right), \quad \text{i.e. the law of } Y \text{ given } \theta = f \in \Theta, \\ \Theta_0 &= B_s(R), \\ \Theta_\rho &= \widetilde{B}_{s,t}(R,\rho), \end{cases}$$

with

$$\text{dist}_{\Theta_0}(\theta) = \text{dist}_{B_s(R)}(f) = \inf_{h \in B_s(R)} \|f-h\|_{L_2}, \quad \theta = f \in \Theta.$$

Related questions and literature

There are in essence two lines of work with questions or ideas closely related to the present chapter.

Firstly, considering the simpler null hypothesis $H_0: \ f \equiv 0$ puts us in the so-called signal-detection setting which has already been studied, see for instance the series of seminal papers [Ing93] as well as [IS02] or [CD13] for a more recent treatment. In that context, the order of ρ^* with respect to n is shown to be

$$n^{-\frac{t}{2t+1/2}}.$$

Secondly, another closely related task is the construction of (adaptive and honest) confidence regions for f. In [BN13], the authors study such sets in terms of L_2-separation, but rather than the observation $(Y(x))_{x\in[0,1]}$ they use a Gaussian sequence model. However, due to the equivalence of these models in the sense of Le Cam (see [LC12]), it is possible to derive from their arguments that for our problem (4.2),

$$n^{-\frac{t}{2t+1/2}} \lesssim \rho^*(\eta) \lesssim \max\left(n^{-\frac{s}{2s+1}}, n^{-\frac{t}{2t+1/2}} \right). \tag{4.4}$$

While the resulting gap in the case $s < 2t$ is not essential in the confidence region setting (see also [CL06] and [JLL03]), it is quite important from a testing perspective as it raises the question how the complexity of the null hypothesis influences the separation rate.
Now, the article [Car15] is by far the closest previous work to the present chapter. Indeed, the author studies the same problem with another choice of Sobolev-ball, namely the (r, ∞)-Sobolev-balls $B_{r,\infty}(R)$. In this context, $\rho^*(\eta)$ is proved to be of magnitude

$$n^{-\frac{t}{2t+1/2}}.$$

Note that this quantity is equal to the rate in the signal-detection case and hence in particular does not depend on s. This makes the issue of the gap in (4.4) even more interesting and, from a technical perspective, it is rather striking given that moving from a simple to the composite null hypothesis is a significant step. On top of that, there are settings where the separation rate strongly depends on the shape of the null hypothesis, see e.g. [BCG18] and [JN02] or also [CL11].
To the best of our knowledge, the case of [Car15] is the only one for which the minimax L_2-separation rate is known and our main contribution is to extend that result to the $(r.2)$-Sobolev-space. While our lower bound (Theorem 4.2 in section 4.3) is essentially a corollary of the corresponding result [Car15, Theorem 3.2], the upper bound (Theorem 4.1 in section 4.3) cannot be established through a simple application of [Car15, Theorem 3.1]. As $B_r(R) \subseteq B_{r,\infty}(R)$, this might be surprising at first sight: Indeed, the test from [Car15] would perform well in the present setting in terms of type-I-error. However, ensuring sufficient power is significantly more difficult when considering l_2-Sobolev-balls.

4.2 Setting

In this section, we describe how the relevant Sobolev balls and the observed Gaussian process will be represented throughout the chapter.

Wavelet transform and associated Sobolev ball

Throughout the chapter, we make heavy use of a wavelet decomposition of f. As is well-known, we can define a scalar product and associated norm on L_2 by

$$< g, h >:= \int_0^1 g(x)h(x) \, \mathrm{d}\lambda(x) \quad \text{with} \quad \|g\|_{L_2} := \sqrt{< g, g >}, \quad g, h \in L_2.$$

There are many orthogonal wavelet bases of L_2 with respect to $< \cdot, \cdot >$. A suitable choice for our purposes is a basis developed in [CDV93] that can be written as

$$\mathcal{W} = \bigcup_{j=2}^{\infty} \{\psi_{j,k} : \ k \in \{1, 2, \ldots, 2^j\}\},$$

i.e. it is tailored such that there are exactly 2^j basis functions at resolution $j \geq 2$. Clearly, the coefficients of $g \in L_2$ with respect to $\mathcal{W}$ are given by

$$< g, \psi_{j,k} >= \int_0^1 g(x)\psi_{j,k}(x) \, \mathrm{d}x, \quad j \geq 2, k \in \{1, 2, \ldots, 2^j\}.$$

and yield the representation

$$g = \sum_{j=2}^{\infty} \sum_{k=1}^{2^j} < g, \psi_{j,k} > \psi_{j,k}.$$

Let $r > 0$. By virtue of isometry properties discussed for instance in [Tri92] and [GN16], we may now define a functional $(r, 2)$-Sobolev-ball of radius R solely through the wavelet coefficients of its elements:

$$B_r(R) := \left\{ g \in L_2 \; ; \; \sum_{j=2}^{\infty} 4^{jr} \sum_{k=1}^{2^j} < g, \psi_{j,k} >^2 \;\; \leq \;\; R^2 \right\} \tag{4.5}$$

with associated $(r, 2)$-Sobolev-norm

$$\|g\|_{\mathcal{B}_r} := \sqrt{\sum_{j=2}^{\infty} 4^{jr} \sum_{k=1}^{2^j} < g, \psi_{j,k} >^2}, \quad g \in L_2$$

or also, as mentioned at the end of the previous section,

$$B_{r,\infty}(R) := \left\{ g \in L_2 \; ; \; \sup_{j \geq 2} \; 4^{jr} \sum_{k=1}^{2^j} < g, \psi_{j,k} >^2 \;\; \leq \;\; R^2 \right\}.$$

Discrete observation scheme based on the wavelet basis

Let

$$\mathcal{I} = \{(j, k) \in \mathbb{N}^2 \mid j \geq 2, k \leq 2^j\}.$$

Motivated by (4.5), for each $(j, k) \in \mathcal{I}$ we consider

$$a_{j,k} := < f, \psi_{j,k} >$$

so that

$$f = \sum_{j=2}^{\infty} \sum_{k=1}^{2^j} a_{j,k} \psi_{j,k}.$$

The natural corresponding estimators read

$$\widehat{a}_{j,k} := < \mathrm{d}Y, \psi_{j,k} >, \quad \widehat{f} = \sum_{j=2}^{\infty} \sum_{k=1}^{2^j} \widehat{a}_{j,k} \psi_{j,k}. \tag{4.6}$$

By construction and due to the orthonormality of $\mathcal{W}$, we know that the family $(\widehat{a}_{i,j})_{(j,k) \in \mathcal{I}}$ is independent with

$$\widehat{a}_{j,k} \sim \mathcal{N}\left(a_{j,k}, \tfrac{1}{n}\right).$$

Clearly, observing this family is equivalent to observing the original process $(Y(x))_{x \in [0,1]}$.

4.3 Main results

In this section, we state and discuss our main results, that is upper and lower bounds on $\rho^*(\eta)$. We also provide a high-level description of the strategy and ideas included in the upper bound proof, which is our main contribution.

4.3.1 Upper Bound

The test

Note that $\widehat{f}$ from (4.6) is not a useful estimator as it exhibits infinite variance. Therefore, we need to carefully impose a restriction of the form $j \leq J$ for some fixed $J \in \mathbb{N}$, $J \geq 2$. More specifically, section 4.5 is primarily concerned with obtaining an upper bound on $\rho_J^*(\eta)$ for the reduced, finite-dimensional problem

$$H_0' : \underbrace{\left\| \sum_{j=2}^{J} \sum_{k=1}^{2^j} a_{j,k} \psi_{j,k} \right\|_{\mathcal{B}_s}}_{:= S_J} \leq R \quad \text{vs.} \quad H_1' : \inf_{h \in B_s(R)} \left\| \sum_{j=2}^{J} \sum_{k=1}^{2^j} a_{j,k} \psi_{j,k} - h \right\|_{L_2} > \rho_J,$$

where ρ_J and $\rho_J^*(\eta)$ are analogous in definition and relation to their counterparts in (4.2) and (4.3).

Finding a sufficient separation distance $\rho_J \geq \rho_J^*(\eta)$ here is the central and most involved part of the chapter. Indeed, it turns out that a test based on estimating S_J^2 only cannot perform well enough under the targeted separation distance of order $n^{-t/(2t+1/2)}$. Rather than that, our test estimates $S_2^2, S_3^2, \ldots, S_J^2$ through test statistics $T_2, T_3, \ldots, T_J$ and, under the alternative H_1', relies on the smallest level j^* such that $S_{j^*}^2$ significantly exceeds R^2 (Lemmas 4.3 and 4.5 below). It essentially takes the form

$$\varphi = 1 - \prod_{j=2}^{J} \mathbb{1}_{\{T_j \leq \tau_j\}}$$

of a multi-level test with suitable thresholds τ_j (equation (4.24) below).

Finally, J must be chosen such that an appropriate trade-off between $\rho_J^*(\eta)$ and the error incurred by ignoring the resolutions beyond J is reached.

In terms of technical ingredients, these considerations are remarkable in that they solely rely on elementary computations based on the Sobolev-balls' geometry and classical properties of the χ^2-distribution. The explicit result reads as follows:

Theorem 4.1. Let $\eta \in (0,1)$. Whenever

$$\rho \geq \left(\frac{1346}{\sqrt{\eta}} + \frac{R}{1 - 2^{-t}} \right) n^{-\frac{t}{2t+1/2}},$$

there is a test φ such that

$$\sup_{f \in B_s(R)} \mathbb{P}_f(\varphi = 1) + \sup_{f \in \widetilde{B}_{s,t}(R,\rho)} \mathbb{P}_f(\varphi = 0) \leq \eta.$$

Hence,

$$\rho^*(\eta) \leq \left(\frac{1346}{\sqrt{\eta}} + \frac{R}{1 - 2^{-t}} \right) n^{-\frac{t}{2t+1/2}}, \quad \text{i.e.} \quad \rho^*(\eta) \lesssim n^{-\frac{t}{2t+1/2}}.$$

◁

4.3.2 Lower Bound

Using the same choice for J as indicated above, a lower bound on $\rho^*(\eta)$ of the same order can be derived through studying the statistical distance between specific distributions agreeing with H_0 and H_1 respectively.

Theorem 4.2. Let $\eta \in (0,1)$. There are $C_\eta > 0$ and $N_\eta \in \mathbb{N}$ such that whenever $n \geq N_\eta$ and

$$\rho \leq C_\eta n^{-\frac{t}{2t+1/2}},$$

for any test φ it holds that

$$\sup_{f \in B_s(R)} \mathbb{P}_f(\varphi = 1) + \sup_{f \in \widetilde{B}_{s,t}(R,\rho)} \mathbb{P}_f(\varphi = 0) > \eta.$$

Hence,

$$\rho^*(\eta) \geq C_\eta n^{-\frac{t}{2t+1/2}}, \quad \text{i.e.} \quad \rho^*(\eta) \gtrsim n^{-\frac{t}{2t+1/2}}.$$

In particular, one may choose

$$C_\eta := \frac{R}{2} \min\left\{1, \frac{\sqrt{\ln(1+4(1-\eta)^2)}}{2^t 16R}\right\}, \quad N_\eta := \left\lceil \left(R\frac{2^{s-t}}{C_\eta}\right)^{\frac{2t+1/2}{s-t}} \right\rceil.$$

◁

Note that, as mentioned in the introduction, Theorems 4.1 and 4.2 in conjunction reveal the minimax separation rate to be of order

$$\rho^*(\eta) \sim n^{-\frac{t}{2t+1/2}},$$

which does not depend on the size of the null hypothesis and is equal to the signal-detection rate. Indeed, in order to obtain the lower bound of Theorem 4.2, the fact that H_0 is a composite hypothesis need not be used.

4.4 Alternative settings

Before presenting the proofs of our main results, we briefly discuss their possible application in two alternative settings which might also be of interest, see also [Car15, Section 3.3] and references therein.

Heteroscedastic noise

As a generalisation of (4.1), consider the model

$$\mathrm{d}Y(x) = f(x)\mathrm{d}x + \frac{\sigma(x)}{\sqrt{n}}\mathrm{d}B(x), \quad x \in [0,1], \tag{4.7}$$

where $\sigma \in L_2$ is unknown. The proof of Theorem 4.1 relies heavily on unbiased estimators of $a_{j,k}^2$, $(j,k) \in \mathcal{I}$, and hence on knowledge of the noise coefficient, so that in this generalised version we cannot directly apply our result. However, there is a relatively simple solution under certain conditions: Suppose we have access to two independent realisations $(Y^{(1)}(x))_{x\in[0,1]}$ and $(Y^{(2)}(x))_{x\in[0,1]}$ with noise coefficient, say, $\frac{\sigma(x)}{\sqrt{n/2}}$. Then we can still consider the estimates

$$\widehat{a}^{(i)}_{j,k} \;=\; <\mathrm{d}Y^{(i)}, \psi_{j,k}> \;\sim\; \mathcal{N}\left(a_{j,k}, 2\frac{\|\sigma\cdot\psi_{j,k}\|^2_{L_2}}{n}\right), \quad i\in\{1,2\}$$

and define a new unbiased estimator for $a^2_{j,k}$ based on the simple observation

$$\mathbb{E}[\widehat{a}^{(1)}_{j,k}\cdot\widehat{a}^{(2)}_{j,k}] = a^2_{j,k}.$$

If in addition we know an upper bound on $\|\sigma\|_{L_2}$, it turns out that we can state an analogous concentration result as the one for the homoscedastic model (see Lemma 4.4 below) and obtain essentially the same result.

Regression

Another possible observation scheme for testing the smoothness of f would be collecting n iid samples $(X_i, Y_i)_{i\in\{1,2,\ldots,n\}}$ according to the model

$$Y = f(X) + \frac{\sigma(X)}{\sqrt{n}}\epsilon$$

for $\epsilon\sim\mathcal{N}(0,1)$ and X uniformly distributed on $[0,1]$. This situation is particularly interesting since, as mentioned above, it is equivalent to (4.7) in the sense of Le Cam ([LC12]) We could then arrive at the same situation as in the previous setting by considering

$$\widehat{a}^{(1)}_{j,k} = \frac{2}{n}\sum_{i=1}^{n/2} Y_i\psi_{j,k}(X_i),\quad \widehat{a}^{(2)}_{j,k} = \frac{2}{n}\sum_{i=n/2+1}^{n} Y_i\psi_{j,k}(X_i).$$

Note that if X is not uniformly distributed, $\mathbb{E}[\widehat{a}^{(i)}_{j,k}] = a_{j,k}$ is generally not true and it becomes crucial to guarantee a certain spread of the design points $(X_i)_{i\in\{1,2,\ldots,n\}}$ over $[0,1]$.

4.5 Proof of Theorem 4.1

4.5.1 General preparations

Reduction of the range of resolutions

Let us make this more clear at this point already: For $j_1, j_2 \in \mathbb{N}\cup\{\infty\}$ with $2\le j_1\le j_2$ and $g\in L_2$, define the projections

$$P^{j_2}_{j_1} g = \sum_{j=j_1}^{j_2}\sum_{k=1}^{2^j} < g, \psi_{j,k} > \psi_{j,k},\quad P_{j_1} := P^{j_1}_{j_1}.$$

Now observe that since $f\in B_t(R)$, for each $j\in\mathbb{N}$, $j\ge 2$, we have

$$\|P_j f\|_{L_2} = \frac{\|P_j f\|_{\mathcal{B}_t}}{2^{jt}} \le \frac{R}{2^{jt}}$$

and hence

$$\sum_{j=J+1}^{\infty} \|P_j f\|_{L_2} \le R \sum_{j=J+1}^{\infty} (2^{-t})^j = 2^{-tJ} \frac{2^{-t}R}{1-2^{-t}}.$$

Using the triangle inequality, this tells us that under the alternative hypothesis

$$\begin{aligned}
\rho &< \inf_{h \in B_s(R)} \|f - h\|_{L_2} \\
&\le \inf_{h \in B_s(R)} \|P_2^J f - h\|_{L_2} + \|P_{J+1}^{\infty} f\|_{L_2} \\
&\le \inf_{h \in B_s(R)} \|P_2^J f - h\|_{L_2} + \sum_{j=J+1}^{\infty} \|P_j f\|_{L_2}, \\
&\le \inf_{h \in B_s(R)} \|P_2^J f - h\|_{L_2} + 2^{-tJ} \frac{2^{-t}R}{1-2^{-t}}.
\end{aligned}$$

Accordingly, under H_1 we consider the assumption

$$\rho - 2^{-tJ} \frac{2^{-t}R}{1-2^{-t}} =: \rho_J < \inf_{h \in B_s(R)} \|P_2^J f - h\|_{L_2}$$

and firstly solve (4.3) for ρ_J in terms of the reduced range $j \in \{2, 3, \ldots, J\}$, that is, subsequently, we will primarily study the testing problem

$$H_0' : \ \|P_2^J f\|_{\mathcal{B}_s} \le R \quad \text{vs.} \quad H_1' : \ \inf_{h \in B_s(R)} \|P_2^J f - h\|_{L_2} > \rho_J.$$

Finally, ρ will be determined by choosing J such that a reasonable trade-off between the two summands,

$$\rho = \rho_J + 2^{-tJ} \frac{2^{-t}R}{1-2^{-t}}, \tag{4.8}$$

is realised.

Now, more specifically, with $a = 1346$, for $j^* \in \{2, 3, \ldots, J\} =: \mathcal{J}$, let

$$\rho_1 := 0; \quad \rho_{j^*} = a \frac{2^{(3j^*+2J)/20}}{\sqrt{n}}.$$

Under the assumption H_1' it will be technically useful to detect the level $\boldsymbol{j^*} \in \mathcal{J}$ at which $\inf_{h \in B_s(R)} \|P_2^{j^*} f - h\|_{\mathcal{B}_s}$ firstly exceeds ρ_{j^*} in the sense of Lemma 4.3 below. That leads to a multiple test across the set $\mathcal{J}$ finally given in (4.27).

Decomposition of H_1'

Lemma 4.3. Under the assumption H_1', we have

$$\exists j^* \in \mathcal{J} : \quad \begin{cases} \inf_{h \in B_s(R)} \|P_2^{j^*-1} f - h\|_{L_2} \le \rho_{j^*-1}, \\ \inf_{h \in B_s(R)} \|P_2^{j^*} f - h\|_{L_2} > \rho_{j^*}. \end{cases} \tag{4.9}$$

◁

Proof. By contradiction: Assume that (4.9) is false, i.e.

$$\forall j^* \in \mathcal{J}: \quad \underbrace{\inf_{h \in B_s(R)} \|P_2^{j^*-1} f - h\|_{L_2} > \rho_{j^*-1}}_{E_{j^*}} \;\vee\; \underbrace{\inf_{h \in B_s(R)} \|P_2^{j^*} f - h\|_{L_2} \leq \rho_{j^*}}_{F_{j^*}}.$$

Then clearly F_J is false, so that E_J is true. Equivalently, F_{J-1} is false and in turn E_{J-1} must be true. Continued application of this argument leads to the contradiction

$$\inf_{h \in B_s(R)} \|P_2^1 f - h\|_{L_2} = 0 > \rho_1.$$

□

Concentration of $\|P_2^{j^*} \widehat{f}\|^2_{\mathcal{B}_s}$

Lemma 4.4. Let $j^* \in \mathcal{J}$. Then, with

$$A_{j^*} := \frac{1}{n} \sum_{j=2}^{j^*} (2 \cdot 4^s)^j, \;\; B_{j^*} := \frac{2}{n^2} \sum_{j=2}^{j^*} (2 \cdot 4^{2s})^j, \;\; V_{j^*} = \frac{4}{n} \sum_{j=2}^{j^*} 4^{2js} \|P_j f\|^2_{L_2},$$

it holds that

$$\forall \delta \in (0,1): \quad \mathbb{P}\left(\left| \|P_2^{j^*} \widehat{f}\|^2_{\mathcal{B}_s} - A_{j^*} - \|P_2^{j^*} f\|^2_{\mathcal{B}_s} \right| \geq \sqrt{\frac{1}{\delta}\left(B_{j^*} + V_{j^*}\right)} \right) \leq \delta. \quad (4.10)$$

◁

Proof. For $j \in \mathcal{J}$, let

$$\lambda_j := n \sum_{k=1}^{2^j} a^2_{j,k} = n\|P_j f\|^2_{L_2}.$$

Then, by construction, we know that

$$n\|P_j \widehat{f}\|^2_{L_2} = \sum_{k=1}^{2^j} (\sqrt{n}\, \widehat{a}_{j,k})^2 \sim \chi^2(2^j, \lambda_j),$$

i.e. a χ^2-distribution with 2^j degrees of freedom and non-centrality parameter λ_j. Classical properties of this distribution now tell us

$$\mathbb{E}\left[\|P_j \widehat{f}\|^2_{L_2}\right] = \frac{2^j}{n} + \|P_j f\|^2_{L_2}; \quad \mathbb{V}\mathrm{ar}\left[\|P_j \widehat{f}\|^2_{L_2}\right] = 2\left(\frac{2^j}{n^2} + \frac{2}{n}\|P_j f\|^2_{L_2}\right). \quad (4.11)$$

Since

$$\|P_2^{j^*} f\|^2_{\mathcal{B}_s} = \sum_{j=2}^{j^*} 4^{js} \|P_j \widehat{f}\|^2_{L_2},$$

independence in conjunction with (4.11) yields

$$\begin{aligned}
\mathbb{E}[\|P_2^{j^*}\widehat{f}\|_{\mathcal{B}_s}^2] &= \sum_{j=2}^{j^*} 4^{js}\left(\frac{2^j}{n} + \|P_j f\|_{L_2}^2\right) \\
&= \frac{1}{n}\sum_{j=2}^{j^*}(2\cdot 4^s)^j + \|P_2^{j^*} f\|_{\mathcal{B}_s}^2 \\
&= A_{j^*} + \|P_2^{j^*} f\|_{\mathcal{B}_s}^2; \\
\mathbb{V}\mathrm{ar}[\|P_2^{j^*}\widehat{f}\|_{\mathcal{B}_s}^2] &= \sum_{j=2}^{j^*} 4^{2js}\left(2\left(\frac{2^j}{n^2} + \frac{2}{n}\|P_j f\|_{L_2}^2\right)\right) \\
&= \frac{2}{n^2}\sum_{j=2}^{j^*}(2\cdot 4^{2s})^j + \frac{4}{n}\sum_{j=2}^{j^*} 4^{2js}\|P_j f\|_{L_2}^2 \\
&= B_{j^*} + V_{j^*}.
\end{aligned}$$

We obtain the desired result directly through Chebyshev's inequality: For $\epsilon > 0$,

$$\mathbb{P}\left(\left|\|P_2^{j^*}\widehat{f}\|_{\mathcal{B}_s}^2 - A_{j^*} - \|P_2^{j^*} f\|_{\mathcal{B}_s}^2\right| \geq \epsilon\right) \leq \frac{B_{j^*} + V_{j^*}}{\epsilon^2}$$

and hence the claim. □

More specifically, observe that

$$\begin{aligned}
B_{j^*} &= \frac{2}{n^2}\sum_{j=2}^{j^*}(2\cdot 4^{2s})^j \\
&= \frac{2}{n^2}(2\cdot 4^{2s})^2\frac{(2\cdot 4^{2s})^{j^*-1} - 1}{2\cdot 4^{2s} - 1} \\
&\leq \frac{2}{n^2}\frac{2\cdot 4^{2s}}{2\cdot 4^{2s} - 1}(2\cdot 4^{2s})^{j^*} \\
&\leq \frac{4}{n^2}(2\cdot 4^{2s})^{j^*},
\end{aligned}$$

(where we use that for $x \geq 2$, $\frac{x}{x-1} \leq 2$) and hence for $\delta \in (0,1)$

$$\sqrt{\frac{B_{j^*}}{\delta}} \leq \frac{2}{\sqrt{\delta}} 4^{j^* s}\frac{2^{j^*/2}}{n}.$$

Furthermore,

$$\sqrt{\frac{V_{j^*}}{\delta}} = \frac{2}{\sqrt{\delta}}\cdot\frac{1}{\sqrt{n}}\sqrt{\sum_{j=2}^{j^*} 4^{2js}\|P_j f\|_{L_2}^2} \leq \frac{2}{\sqrt{\delta}}\cdot\frac{\sqrt{j^*-1}}{\sqrt{n}}\max_{2\leq j\leq j^*} 2^{js}\|P_j f\|_{\mathcal{B}_s}.$$

The maximum in the latter computation will play an important role in the sequel. From now on we use the abbreviation

$$M_{j^*} := \max_{2\leq j\leq j^*} 2^{js}\|P_j f\|_{\mathcal{B}_s}. \tag{4.12}$$

Plugging these bounds in (4.10) leads to

$$\mathbb{P}\left(\left|\|P_2^{j^*}\widehat{f}\|_{\mathcal{B}_s}^2 - A_{j^*} - \|P_2^{j^*}f\|_{\mathcal{B}_s}^2\right| \geq \frac{2}{\sqrt{\delta}}\cdot\frac{\sqrt{j^*-1}}{\sqrt{n}}M_{j^*} + \frac{2}{\sqrt{\delta}}4^{j^*s}\frac{2^{j^*/2}}{n}\right) \leq \delta \quad (4.13)$$

for any $\delta \in (0,1)$.

4.5.2 Preliminary Bounds on $\|P_2^{j^*}f\|_{\mathcal{B}_s}$

As a next step towards controlling the type-I and type-II errors of our test, we study $\|P_2^{j^*}f\|_{\mathcal{B}_s}$ more closely.
On the one hand, under H_0', for any $j^* \in \mathcal{J}$ we clearly have $\|P_2^{j^*}f\|_{\mathcal{B}_s} \leq R$.
On the other hand, under H_1', we require a lower bound on $\|P_2^{j^*}f\|_{\mathcal{B}_s}$. The following bound is preliminary in the sense that it requires the knowledge of an index $j^* \in \mathcal{J}$ with the property from (4.9) and the corresponding M_{j^*}. The generalisation will be considered in sections 4.5.3 and 4.5.4.

Lemma 4.5. Let $j^* \in \mathcal{J}$ be an index with the property

$$\inf_{h\in B_s(R)} \|P_2^{j^*-1}f - h\|_{L_2} \leq \rho_{j^*-1}, \quad \inf_{h\in B_s(R)} \|P_2^{j^*}f - h\|_{L_2} > \rho_{j^*}. \quad (4.14)$$

Then the following assertion holds for $A = 11$:

$$\|P_2^{j^*}f\|_{\mathcal{B}_s}^2 \geq R^2 + \frac{1}{2\cdot A^2}\rho_{j^*}M_{j^*} + \frac{1}{2\cdot A^2}4^{j^*s}\rho_{j^*}^2. \quad (4.15)$$

◁

Proof. Before giving the main arguments, we need a technical preparation and a general (i.e. only depending on j^*) lower bound on $\|P_2^{j^*}f\|_{\mathcal{B}_s}$:

1. <u>Proxy minimisation of $\inf_{h\in B_s(R)}\|P_2^{j^*}f - h\|_{L_2}$</u>
   For $\widetilde{j} \in \mathcal{J}$, write $P_{j\neq\widetilde{j}} := P_2^{j^*} - P_{\widetilde{j}}$. In the case that $\|P_{j\neq\widetilde{j}}f\|_{\mathcal{B}_s} \leq R$, we can introduce the function $\widetilde{h}$ through the wavelet coefficients

$$\begin{aligned} b_{j,k} &:= a_{j,k} && \text{for } (j,k)\in\mathcal{I}, j\neq\widetilde{j}, \\ b_{\widetilde{j},k} &:= a_{\widetilde{j},k}\cdot\frac{\sqrt{R^2 - \|P_{j\neq\widetilde{j}}f\|_{\mathcal{B}_s}^2}}{\|P_{\widetilde{j}}f\|_{\mathcal{B}_s}}, && \text{for } k\in\{1,2,\dots,2^{\widetilde{j}}\}. \end{aligned}$$

   Then $\widetilde{h} \in B_s(R)$ holds since

$$\begin{aligned} \|\widetilde{h}\|_{\mathcal{B}_s}^2 &= \sum_{j=2}^{j^*} 4^{js}\sum_{k=1}^{2^j} b_{j,k}^2 \\ &= \|P_{j\neq\widetilde{j}}f\|_{\mathcal{B}_s}^2 + \left(\frac{\sqrt{R^2 - \|P_{j\neq\widetilde{j}}f\|_{\mathcal{B}_s}^2}}{\|P_{\widetilde{j}}f\|_{\mathcal{B}_s}}\right)^2 \|P_{\widetilde{j}}f\|_{\mathcal{B}_s}^2 \\ &= R^2. \end{aligned}$$

Hence, by assumption

$$\rho_{j^*}^2 < \inf_{h \in B_s(R)} \|P_2^{j^*} f - h\|_{L_2}^2 \leq \|P_2^{j^*} f - \widetilde{h}\|_{L_2}^2 =: d^2,$$

where

$$\begin{aligned} d^2 &= \|P_2^{j^*} f - \widetilde{h}\|_{L_2}^2 \\ &= \sum_{k=1}^{2^{\widetilde{j}}} \left(1 - \frac{\sqrt{R^2 - \|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s}^2}}{\|P_{\widetilde{j}} f\|_{\mathcal{B}_s}}\right)^2 a_{\widetilde{j},k}^2 \\ &= \left(1 - \frac{\sqrt{R^2 - \|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s}^2}}{\|P_{\widetilde{j}} f\|_{\mathcal{B}_s}}\right)^2 \frac{\|P_{\widetilde{j}} f\|_{\mathcal{B}_s}^2}{4^{\widetilde{j}s}} \\ &= \left(\|P_{\widetilde{j}} f\|_{\mathcal{B}_s} - \sqrt{R^2 - \|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s}^2}\right)^2 4^{-\widetilde{j}s}. \end{aligned}$$

This tells us that if $\|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s} \leq R$,

$$\|P_{\widetilde{j}} f\|_{\mathcal{B}_s} = 2^{\widetilde{j}s} d + \sqrt{R^2 - \|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s}^2} \geq 2^{\widetilde{j}s} d, \tag{4.16}$$

$$\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 = R^2 + 2 \cdot 2^{\widetilde{j}s} d \sqrt{R^2 - \|P_{j \neq \widetilde{j}} f\|_{\mathcal{B}_s}^2} + 4^{\widetilde{j}s} d^2. \tag{4.17}$$

2. Bound in terms of $4^{j^* s} \rho_{j^*}^2$

If $\|P_2^{j^*-1} f\|_{\mathcal{B}_s} \leq R$, we can use (4.17) with $\widetilde{j} = j^*$ and $d \geq \rho_{j^*} \geq 0$ and obtain

$$\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 \geq R^2 + 4^{j^* s} \rho_{j^*}^2.$$

If $\|P_2^{j^*-1} f\|_{\mathcal{B}_s} > R$, observe that by the triangle inequality

$$\inf_{h \in B_s(R)} \|P_2^{j^*} f - h\|_{L_2} \leq \inf_{h \in B_s(R)} \|P_2^{j^*-1} f - h\|_{L_2} + \|P_{j^*} f\|_{L_2} \leq \rho_{j^*-1} + \|P_{j^*} f\|_{L_2}$$

and since

$$\rho_{j^*} - \rho_{j^*-1} \geq a \frac{2^{(3j^*+2J)/20}}{\sqrt{n}} (1 - 2^{-3/20}) \geq \frac{1}{11} \rho_{j^*} = \frac{1}{A} \rho_{j^*},$$

we obtain

$$\begin{aligned} \|P_2^{j^*} f\|_{\mathcal{B}_s}^2 &= \|P_2^{j^*-1} f\|_{\mathcal{B}_s}^2 + \|P_{j^*} f\|_{\mathcal{B}_s}^2 \\ &\geq R^2 + 4^{j^* s} (\rho_{j^*} - \rho_{j^*-1})^2 \\ &\geq R^2 + \frac{1}{A^2} 4^{j^* s} \rho_{j^*}^2. \end{aligned}$$

So, in any case,

$$\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 \geq R^2 + \frac{1}{A^2} 4^{j^* s} \rho_{j^*}^2. \tag{4.18}$$

3. Main arguments
   We are now ready to prove (4.15) effectively. To that end, fix an index

$$\overline{j} \in \underset{j \in \mathcal{J}}{\operatorname{argmax}}\ 2^{js} \|P_j f\|_{\mathcal{B}_s}.$$

Case 1: $\|P_{j \neq \overline{j}} f\|_{\mathcal{B}_s} \leq R$

In that case, we can use (4.16) and (4.17) with $\widetilde{j} = \overline{j}$ in comination with (4.18) and obtain

$$\begin{aligned}
\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 &\geq R^2 + 2^{\overline{j}s} d \sqrt{R^2 - \|P_{j \neq \overline{j}} f\|_{\mathcal{B}_s}^2} + \frac{1}{2} 4^{\overline{j}s} d^2 + \frac{1}{2 \cdot A^2} 4^{j^* s} d^2 \\
&= R^2 + 2^{\overline{j}s} d \left(\|P_{\overline{j}} f\|_{\mathcal{B}_s} - 2^{\overline{j}s} d \right) + \frac{1}{2} 4^{\overline{j}s} d^2 + \frac{1}{2 \cdot A^2} 4^{j^* s} d^2 \\
&= R^2 + d \cdot 2^{\overline{j}s} \left(\|P_{\overline{j}} f\|_{\mathcal{B}_s} - \frac{1}{2} 2^{\overline{j}s} d \right) + \frac{1}{2 \cdot A^2} 4^{j^* s} d^2 \\
&\geq R^2 + \rho_{j^*} 2^{\overline{j}s} \left(\|P_{\overline{j}} f\|_{\mathcal{B}_s} - \frac{1}{2} 2^{\overline{j}s} d \right) + \frac{1}{2 \cdot A^2} 4^{j^* s} \rho_{j^*}^2 \\
&\geq R^2 + \frac{1}{2} \rho_{j^*} M_{j^*} + \frac{1}{2 \cdot A^2} 4^{j^* s} \rho_{j^*}^2,
\end{aligned}$$

remembering (4.12).

Case 2: $\|P_{j \neq \overline{j}} f\|_{\mathcal{B}_s} > R$

That case can be handled quickly by considering two subcases:

Subcase 1: $4^{j^* s} \rho_{j^*}^2 \geq \rho_{j^*} M_{j^*}$

Observe that with (4.18)

$$\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 \geq R^2 + \frac{1}{2 \cdot A^2} 4^{j^* s} \rho_{j^*}^2 + \frac{1}{2 \cdot A^2} 4^{j^* s} \rho_{j^*}^2 \geq R^2 + \frac{1}{2 \cdot A^2} \rho_{j^*} M_{j^*} + \frac{1}{2 \cdot A^2} 4^{j^* s} \rho_{j^*}^2.$$

Subcase 2: $4^{j^* s} \rho_{j^*}^2 < \rho_{j^*} M_{j^*}$

In that case we have

$$\|P_{\overline{j}} f\|_{\mathcal{B}_s} > \frac{4^{j^* s}}{2^{\overline{j}s}} \rho_{j^*} \geq 2^{\overline{j}s} \rho_{j^*}$$

and thus

$$\begin{aligned}
\|P_2^{j^*} f\|_{\mathcal{B}_s}^2 &= \|P_{j \neq \overline{j}}\|_{\mathcal{B}_s}^2 + \|P_{\overline{j}} f\|_{\mathcal{B}_s}^2 \\
&> R^2 + 2^{\overline{j}s} \rho_{j^*} \|P_{\overline{j}} f\|_{\mathcal{B}_s} \\
&= R^2 + \rho_{j^*} M_{j^*} \\
&\geq R^2 + \frac{1}{2} \rho_{j^*} M_{j^*} + \frac{1}{2} 4^{j^* s} \rho_{j^*}^2.
\end{aligned}$$

This concludes the proof since in any case (4.15) holds.

□

4.5.3 Estimation of M_{j^*}

As a last major step before directly controlling the type-I and type-II error probabilities, we need to find an appropriate estimator for M_{j^*}.

Lemma 4.6. For $\delta \in (0,1)$ and $j^* \in \mathcal{J}$, let

$$C_\delta := \sqrt{\frac{2}{\delta}},$$

$$D_{j^*,\delta} := \frac{4^{j^*s}}{\sqrt{n}}\left(\sqrt{2}C_\delta + 2^{j^*/4}\sqrt{C_\delta}\right)$$

and define the events

$$\xi^0_{j^*,\delta} := \left\{ M_{j^*} \leq \sqrt{\max_{2\leq j\leq j^*} |Y_j| + D_{\delta,j^*}} \right\}, \tag{4.19}$$

$$\xi^1_{j^*,\delta} := \left\{ M_{j^*} \geq \sqrt{\max_{2\leq j\leq j^*} |Y_j| - D_{\delta,j^*}} \right\}. \tag{4.20}$$

Then, for any monotone decreasing sequence $(\beta_j)_{j\in\mathcal{J}}$ in $(0,1)$, the following holds:

$$\mathbb{P}(\xi^1_{j^*,\beta_{j^*}}) \geq 1 - \sum_{j=2}^{j^*} \beta_j, \quad \mathbb{P}(\xi^0_{j^*,\beta_{j^*}}) \geq 1 - \beta_{j^*}. \tag{4.21}$$

◁

Proof. Remembering (4.11), we know that for $j \in \{2,3,\dots,j^*\}$

$$Z_j := 4^{js}\|P_j\widehat{f}\|^2_{\mathcal{B}_s} = 16^{js}\|P_j\widehat{f}\|^2_{L_2}$$

has the properties

$$\mathbb{E}[Z_j] = 16^{js}\frac{2^j}{n} + 4^{js}\|P_jf\|^2_{\mathcal{B}_s},$$

$$\mathbb{V}\mathrm{ar}[Z_j] = 2\cdot 16^{2js}\left(\frac{2^j}{n^2} + \frac{2}{n}\|P_jf\|^2_{L_2}\right)$$

$$= 16^{js}\left(2\cdot 16^{js}\frac{2^j}{n^2} + \frac{4}{n}4^{js}\|P_jf\|^2_{\mathcal{B}_s}\right).$$

Now observe that for $\delta \in (0,1)$

$$\sqrt{\frac{1}{\delta}\mathbb{V}\mathrm{ar}[Z_j]} \leq \sqrt{\frac{2}{\delta}}2^{j/2}\frac{16^{js}}{n} + \frac{2}{\sqrt{\delta n}}4^{js}2^{js}\|P_jf\|_{\mathcal{B}_s}$$

$$\leq C_\delta 2^{j^*/2}\frac{16^{j^*s}}{n} + \sqrt{2}C_\delta\frac{4^{j^*s}}{\sqrt{n}}M_{j^*}$$

$$=: v_{\delta,j^*}.$$

With $Y_j = Z_j - 16^{js}\frac{2^j}{n}$, Chebyshev's inequality now tells us that

$$\mathbb{P}\left(\left|Y_j - m_j^2\right| \geq v_{\delta,j^*}\right) \leq \delta. \tag{4.22}$$

We derive two bounds from this statement by lower bounding the the left hand side in two different ways:
On the one hand, observe

$$|Y_j - m_j^2| \geq ||Y_j| - m_j^2| \geq |Y_j| - m_j^2 \geq |Y_j| - M_{j^*}^2.$$

Now, since $(\beta_j)_{j\in\mathcal{J}}$ is monotone decreasing, the sequence $(v_{\beta_j,j^*})_{j\in\mathcal{J}}$ is increasing, so that via a union bound we obtain

$$\begin{aligned}
\sum_{j=2}^{j^*} \beta_j &\geq \mathbb{P}\left(\exists j \in \{2,3,\ldots,j^*\}: \; |Y_j| \geq M_{j^*}^2 + v_{\beta_j,j^*}\right) \\
&\geq \mathbb{P}\left(\exists j \in \{2,3,\ldots,j^*\}: \; |Y_j| \geq M_{j^*}^2 + v_{\beta_{j^*},j^*}\right) \\
&= \mathbb{P}\left(\sqrt{\max_{2\leq j\leq j^*} |Y_j|} \geq \sqrt{M_{j^*}^2 + v_{\beta_{j^*},j^*}}\right).
\end{aligned}$$

With

$$\begin{aligned}
\sqrt{M_{j^*}^2 + v_{\beta_{j^*},j^*}} &= \sqrt{\left(M_{j^*} + \frac{C_{\beta_{j^*}}}{\sqrt{2}} \frac{4^{j^*s}}{\sqrt{n}}\right)^2 - \frac{C_{\beta_{j^*}}^2}{2} \frac{16^{j^*s}}{n} + C_{\beta_{j^*}} 2^{j^*/2} \frac{16^{j^*s}}{n}} \\
&\leq \sqrt{\left(M_{j^*} + \frac{C_{\beta_{j^*}}}{\sqrt{2}} \frac{4^{j^*s}}{\sqrt{n}}\right)^2 + \frac{C_{\beta_{j^*}}^2}{2} \frac{16^{j^*s}}{n} + C_{\beta_{j^*}} 2^{j^*/2} \frac{16^{j^*s}}{n}} \\
&\leq M_{j^*} + \frac{4^{j^*s}}{\sqrt{n}} \left(\sqrt{2} C_{\beta_{j^*}} + 2^{j^*/4} \sqrt{C_{\beta_{j^*}}}\right),
\end{aligned}$$

we have

$$\mathbb{P}\left(\sqrt{\max_{2\leq j\leq j^*} |Y_j|} \geq M_{j^*} + \frac{4^{j^*s}}{\sqrt{n}} \left(\sqrt{2} C_{\beta_{j^*}} + 2^{j^*/4} \sqrt{C_{\beta_{j^*}}}\right)\right) \leq \sum_{j=2}^{j^*} \beta_j$$

and hence the first claim from (4.21).
On the other hand, observe

$$|Y_j - m_j^2| \geq m_j^2 - |Y_j|$$

and consider the specific case $j = \bar{j}$ in (4.22):

$$\begin{aligned}
\beta_{j^*} &\geq \mathbb{P}\left(|Y_{\bar{j}}| \leq M_{j^*}^2 - v_{\beta_{j^*},j^*}\right) \\
&\geq \mathbb{P}\left(\max_{2\leq j\leq j^*} |Y_j| \leq M_{j^*}^2 - v_{\beta_{j^*},j^*}\right) \\
&= \mathbb{P}\left(\max_{2\leq j\leq j^*} |Y_j| + \frac{16^{j^*s}}{n} \left(\frac{C_{\beta_{j^*}}^2}{2} + 2^{j^*/2} C_{\beta_{j^*}}\right) \leq \left(M_{j^*} - \frac{C_{\beta_{j^*}}}{\sqrt{2}} \frac{4^{j^*s}}{\sqrt{n}}\right)^2\right) \\
&\geq \mathbb{P}\left(\sqrt{\max_{2\leq j\leq j^*} |Y_j|} + \frac{4^{j^*s}}{\sqrt{n}} \left(\sqrt{2} C_{\beta_{j^*}} + 2^{j^*/4} \sqrt{C_{\beta_{j^*}}}\right) \leq M_{j^*}\right),
\end{aligned}$$

which asserts the second claim from (4.21). □

4.5.4 Conclusion

We will now assemble the individual results of the previous sections to obtain the claim of Theorem 4.1. For $j \in \mathcal{J}$ we introduce

$$\rho_j = \frac{1346}{\sqrt{\eta}} \cdot \frac{2^{(3j+2J)/20}}{\sqrt{n}}, \quad \alpha_j = \eta\frac{1-2^{-1/5}}{4}2^{(j-J)/5}, \quad \beta_j = \eta\frac{1-2^{-1/2}}{2}2^{-j/2}, \qquad (4.23)$$

so that in particular

$$\sum_{j=2}^{J}\alpha_j \leq \frac{\eta}{4}, \quad \sum_{j=2}^{J}\beta_j \leq \frac{\eta}{4}$$

and $(\beta_j)_{j\in\mathcal{J}}$ is monotone decreasing.

Result for fixed index

For $j^* \in \mathcal{J}$ define

$$T_{j^*,\alpha_{j^*}} = \|P_2^{j^*}\widehat{f}\|_{\mathcal{B}_s}^2 - A_{j^*} - \frac{2}{\sqrt{\alpha_{j^*}}} \cdot \frac{\sqrt{j^*-1}}{\sqrt{n}}\sqrt{\max_{2\leq j\leq j^*}|Y_j|}.$$

Then under $H_0' \cap \xi^0_{j^*,\beta_{j^*}}$, (4.19) and (4.13) yield that with probability at least $1-\alpha_{j^*}$

$$\begin{aligned} T_{j^*,\alpha_{j^*}} &\leq \|P_2^{j^*}f\|_{\mathcal{B}_s}^2 + \frac{2}{\sqrt{\alpha_{j^*}}} \cdot \frac{\sqrt{j^*-1}}{\sqrt{n}}M_{j^*} + \frac{2}{\sqrt{\alpha_{j^*}}}4^{j^*s}\frac{2^{j^*/2}}{n} \\ &\qquad - \frac{2}{\sqrt{\alpha_{j^*}}} \cdot \frac{\sqrt{j^*-1}}{\sqrt{n}}\sqrt{\max_{2\leq j\leq j^*}|Y_j|} \\ &\leq R^2 + \frac{2}{\sqrt{\alpha_{j^*}}} \cdot \frac{\sqrt{j^*-1}}{\sqrt{n}}D_{j^*,\beta_{j^*}} + \frac{2}{\sqrt{\alpha_{j^*}}}4^{j^*s}\frac{2^{j^*/2}}{n} \end{aligned}$$

so that with

$$\tau_{j^*,\alpha_{j^*}} = R^2 + \frac{2}{\sqrt{\alpha_{j^*}}}\left(\frac{\sqrt{j^*-1}}{\sqrt{n}}D_{j^*,\beta_{j^*}} + 4^{j^*s}\frac{2^{j^*/2}}{n}\right), \qquad (4.24)$$

we obtain

$$\mathbb{P}_{H_0'}(T_{j^*,\alpha_{j^*}} > \tau_{j^*,\alpha_{j^*}} \mid \xi^0_{j^*,\beta_{j^*}}) \leq \alpha_{j^*}.$$

On the other hand, let $\boldsymbol{j^*}$ be a transition index with property (4.14). Then under $H_1' \cap \xi^1_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}}$, (4.13) and (4.15) tell us that with probability at least $1-\alpha_{\boldsymbol{j^*}}$

$$\begin{aligned}
T_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} &\geq \|P_2^{\boldsymbol{j^*}} f\|^2_{\mathcal{B}_s} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} M_{\boldsymbol{j^*}} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} 4^{\boldsymbol{j^*}s} \frac{2^{\boldsymbol{j^*}/2}}{n} \\
&\qquad - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} \sqrt{\max_{2\leq j\leq j^*} |Y_j|} \\
&\geq R^2 + \left(\frac{1}{2\cdot A^2}\rho_{\boldsymbol{j^*}} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} \right) M_{\boldsymbol{j^*}} + \frac{1}{2\cdot A^2} 4^{\boldsymbol{j^*}s} \rho^2_{\boldsymbol{j^*}} \\
&\qquad - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} 4^{\boldsymbol{j^*}s} \frac{2^{\boldsymbol{j^*}/2}}{n} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} \sqrt{\max_{2\leq j\leq j^*} |Y_j|}.
\end{aligned}$$

Provided that

$$\frac{1}{2\cdot A^2}\rho_{\boldsymbol{j^*}} \geq \frac{4}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}}, \tag{4.25}$$

using (4.20) this yields

$$T_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} \geq R^2 + \frac{1}{2\cdot A^2} 4^{\boldsymbol{j^*}s} \rho^2_{\boldsymbol{j^*}} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} D_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}} - \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} 4^{\boldsymbol{j^*}s} \frac{2^{\boldsymbol{j^*}/2}}{n}. \tag{4.26}$$

Now by explicit computation we see that the choices in (4.23) ensure (4.25) as well as

$$\frac{1}{4\cdot A^2} 4^{\boldsymbol{j^*}s} \rho^2_{\boldsymbol{j^*}} \geq \frac{4}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \cdot \frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} D_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}} \quad \text{and} \quad \frac{1}{4\cdot A^2} 4^{\boldsymbol{j^*}s} \rho^2_{\boldsymbol{j^*}} \geq \frac{4}{\sqrt{\alpha_{\boldsymbol{j^*}}}} 4^{\boldsymbol{j^*}s} \frac{2^{\boldsymbol{j^*}/2}}{n},$$

so that (4.26) can be continued as

$$T_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} \geq R^2 + \frac{2}{\sqrt{\alpha_{\boldsymbol{j^*}}}} \left(\frac{\sqrt{\boldsymbol{j^*}-1}}{\sqrt{n}} D_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}} + 4^{\boldsymbol{j^*}s} \frac{2^{\boldsymbol{j^*}/2}}{n} \right) = \tau_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}}$$

and hence, finally,

$$\mathbb{P}_{H_1'}(T_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} \leq \tau_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} \mid \xi^1_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}}) \leq \alpha_{\boldsymbol{j^*}}.$$

Generalisation to unknown $\boldsymbol{j^*}$

For our test

$$\varphi(P_2^J \widehat{f}) = 1 - \prod_{j^*=2}^{J} \mathbb{1}_{\{T_{j^*,\alpha_{j^*}} \leq \tau_{j^*,\alpha_{j^*}}\}}, \tag{4.27}$$

we can conclude with (4.21) and (4.23) that on the one hand

$$\begin{aligned}
\mathbb{P}_{H_0'}(\varphi = 1) &\leq \sum_{j^*=2}^{J} \left(\mathbb{P}_{H_0'}\left(T_{j^*,\alpha_{j^*}} > \tau_{j^*,\alpha_{j^*}} \mid \xi^0_{j^*,\beta_{j^*}} \right) + (1 - \mathbb{P}(\xi^0_{j^*,\beta_{j^*}})) \right) \\
&\leq \frac{\eta}{4} + \frac{\eta}{4} = \frac{\eta}{2}
\end{aligned} \tag{4.28}$$

and on the other hand

$$\begin{aligned}
\mathbb{P}_{H_1'}(\varphi = 0) &\leq \mathbb{P}_{H_1'}(\forall j^* \in \mathcal{J} :\ T_{j^*,\alpha} \leq \tau_{j^*,\alpha}) \\
&\leq \mathbb{P}_{H_1'}\left(T_{\boldsymbol{j^*},\alpha} \leq \tau_{\boldsymbol{j^*},\alpha_{\boldsymbol{j^*}}} \mid \xi^1_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}}\right) + \left(1 - \mathbb{P}(\xi^1_{\boldsymbol{j^*},\beta_{\boldsymbol{j^*}}})\right) \\
&\leq \alpha_{\boldsymbol{j^*}} + \sum_{j=2}^{\boldsymbol{j^*}} \beta_j \\
&\leq \frac{\eta}{4} + \frac{\eta}{4} = \frac{\eta}{2}.
\end{aligned} \tag{4.29}$$

Specification of J and conclusion

We are now ready to return to (4.8). Choose

$$J := \left\lfloor \frac{1}{2t + 1/2} \frac{\ln(n)}{\ln(2)} \right\rfloor, \tag{4.30}$$

so that

$$\frac{1}{2} n^{\frac{1}{2t+1/2}} \leq 2^J \leq n^{\frac{1}{2t+1/2}}. \tag{4.31}$$

That yields

$$2^{-Jt} \leq 2^t n^{-\frac{t}{2t+1/2}} \tag{4.32}$$

and, on the other hand,

$$\rho_J = \frac{1346}{\sqrt{\eta}} \frac{2^{J/4}}{\sqrt{n}} \leq \frac{1346}{\sqrt{\eta}} \cdot n^{-\frac{t}{2t+1/2}}.$$

Therefore, whenever we choose

$$\rho \geq \left(\frac{1346}{\sqrt{\eta}} + \frac{R}{1 - 2^{-t}} \right) n^{-\frac{t}{2t+1/2}},$$

indeed by (4.28) and (4.29)

$$\sup_{f \in B_s(R)} \mathbb{P}_f(\varphi = 1) + \sup_{f \in \widetilde{B}_{s,t}(R,\rho)} \mathbb{P}_f(\varphi = 0) \leq \frac{\eta}{2} + \frac{\eta}{2} = \eta.$$

4.6 Proof of Theorem 4.2

Priors

Since the upper bound does not depend on s and we found the index J from (4.30) to be critical, we choose the following structurally simple priors: Let ν_0 be the Dirac-δ distribution on $\{0\}$ (i.e. $f \equiv 0$) and ν_ρ be the uniform distribution on

$$\mathcal{A}_{\rho,v} := \left\{ \sum_{k=1}^{2^J} a_{J,k} \psi_{j,k} \mid a_{J,1}, a_{J,2}, \ldots, a_{J,2^J} \in \{v, -v\} \right\},$$

where $v > 0$ needs further specification: On the one hand, it is necessary to ensure that each $f \in \mathcal{A}_{\rho,v}$ fulfils $\|f\|_{\mathcal{B}_t} \leq R$ - note that for any such f, $\|f\|_{L_2} = 2^{J/2}v$, so that by construction that condition reads

$$2^{J(t+1/2)}v \leq R.$$

This motivates the choice $v := a_\eta \cdot R \cdot 2^{-J(t+1/2)}$ for some $a_\eta \in (0,1]$ specified later based on further restrictions. On the other hand, we require

$$\rho \leq \inf_{h \in B_s(R)} \|f - h\|_{L_2}. \tag{4.33}$$

Since only the level J is involved, this is in fact merely the minimum over the Euclidean ball with radius $R \cdot 2^{-Js}$ so that

$$\inf_{h \in B_s(R)} \|f - h\|_{L_2} = \max\left(0, 2^{J/2}v - R \cdot 2^{-Js}\right).$$

Now, by explicit computation we see that if

$$n \geq \left(\frac{2^{1+s-t}}{a_\eta}\right)^{\frac{2t+1/2}{s-t}},$$

with our choice of v we have

$$\max\left(0, 2^{J/2}v - R \cdot 2^{-Js}\right) \geq \frac{1}{2}2^{J/2}v = a_\eta \frac{R}{2} 2^{-Jt},$$

so that (4.33) holds if

$$\rho \leq a_\eta \frac{R}{2} 2^{-Jt}.$$

Statistical distance

Again, the central task in this proof is to compute the χ^2-divergence between $\mathbb{P}_{f\sim\nu_0}$ and $\mathbb{P}_{f\sim\nu_\rho}$. By construction, $\mathbb{P}_{f\sim\nu_0}$ corresponds to the 2^J-fold product of Gaussian distributions with mean 0 and variance $\frac{1}{n}$, so that for $x \in \mathbb{R}^{2^J}$

$$\mathrm{d}\mathbb{P}_{f\sim\nu_0}(x) = \sqrt{\frac{n}{2\pi}}^{2^J} \prod_{k=1}^{2^J} \exp\left(-\frac{n}{2}x_k^2\right).$$

On the other hand, $\mathbb{P}_{f\sim\nu_\rho}$ corresponds to a uniform mixture of 2^{2^J} products of 2^J independent Gaussians with means of the form $\pm v$ and variance $\frac{1}{n}$.

Note that, in fact, we have already considered priors of that form in section 1.4.2 and another derivation here is actually unnecessary. However, we use this opportunity to present an alternative computation with slightly different flavor.

Let $\mathcal{S} := \{1, -1\}^{2^J}$ and R be uniformly distributed on $\mathcal{S}$ (i.e. the product of 2^J

Rademacher variables). Then

$$\begin{aligned}
\mathrm{d}\mathbb{P}_{f\sim\nu_\rho}(x) &= \frac{1}{2^{2^J}}\sum_{\alpha\in\mathcal{S}}\sqrt{\frac{n}{2\pi}}^{2^J}\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}\sum_{k=1}^{2^J}(x_k-\alpha_k v)^2\right)\\
&= \sqrt{\frac{n}{2\pi}}^{2^J}\mathbb{E}_R\left[\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}(x_k-R_k v)^2\right)\right]
\end{aligned}$$

and furthermore, with an independent copy R' of R,

$$\begin{aligned}
(\mathrm{d}\mathbb{P}_{f\sim\nu_\rho}(x))^2 &= \left(\frac{n}{2\pi}\right)^{2^J}\mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}\left[(x_k-R_k v)^2+(x_k-R'_k v)^2\right]\right)\right]\\
&= \left(\frac{n}{2\pi}\right)^{2^J}\exp\left(-2^J n v^2\right)\mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J}\exp\left(-n x_k^2+n v x_k(R_k+R'_k)\right)\right].
\end{aligned}$$

The quotient we need to integrate in (1.10) therefore reads

$$\begin{aligned}
\frac{(\mathrm{d}\mathbb{P}_{f\sim\nu_\rho})^2}{\mathrm{d}\mathbb{P}_{f\sim\nu_0}}(x) &= \sqrt{\frac{n}{2\pi}}^{2^J}\exp\left(-2^J n v^2\right)\mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}x_k^2+n v x_k(R_k+R'_k)\right)\right]\\
&= \sqrt{\frac{n}{2\pi}}^{2^J}\exp\left(-2^J n v^2\right)\\
&\qquad\cdot\mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}\left(x_k-v(R_k+R'_k)\right)^2\right)\exp\left(n v^2(1+R_kR'_k)\right)\right]\\
&= \sqrt{\frac{n}{2\pi}}^{2^J}\mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J}\exp\left(n v^2 R_k R'_k\right)\prod_{k=1}^{2^J}\exp\left(-\frac{n}{2}\left(x_k-v(R_k+R'_k)\right)^2\right)\right].
\end{aligned}$$

Since the product of independent Rademacher variables is itself a Rademacher variable, we obtain

$$
\begin{aligned}
\int_{\mathbb{R}^{2^J}} \frac{(\mathrm{d}\mathbb{P}_{f\sim\nu_\rho})^2}{\mathrm{d}\mathbb{P}_{f\sim\nu_0}}(x)\,\mathrm{d}x &= \mathbb{E}_{R,R'}\left[\prod_{k=1}^{2^J} \exp\left(nv^2 R_k R'_k\right)\right] \\
&= \mathbb{E}_{R}\left[\prod_{k=1}^{2^J} \exp\left(nv^2 R_k\right)\right] \\
&= \prod_{k=1}^{2^J} \mathbb{E}_{R_k}\left[\exp\left(nv^2 R_k\right)\right] \\
&= \left(\cosh(nv^2)\right)^{2^J} \\
&\leq \exp\left(2^J \frac{n^2 v^4}{2}\right).
\end{aligned}
$$

Conclusion

Now, (1.10) holds if

$$\exp\left(2^J \frac{n^2 v^4}{2}\right) < 1 + 4(1-\eta)^2$$

which, by explicit computation, is fulfilled if

$$a_\eta \leq \frac{2^{J(t+1/4)}}{\sqrt{n}R} \sqrt[4]{\ln(1+4(1-\eta)^2)}.$$

Through (4.31) and (4.32) we find that

$$\frac{2^{J(t+1/4)}}{\sqrt{n}} \geq \frac{2^{-t}}{16}$$

and obtain the stronger condition

$$a_\eta \leq \frac{\sqrt[4]{\ln(1+4(1-\eta)^2)}}{2^t 16 R}.$$

In summary: Let

$$a_\eta = \min\left\{1, \frac{\sqrt{\ln(1+4(1-\eta)^2)}}{2^t 16 R}\right\}.$$

If

$$n \geq \left\lceil \left(\frac{2^{1+s-t}}{a_\eta}\right)^{\frac{2t+1/2}{s-t}} \right\rceil,$$

the priors ν_0 and ν_ρ meet all requirements and the lower bound

$$\rho^* \geq a_\eta \frac{R}{2} 2^{-Jt} \geq a_\eta \frac{R}{2} n^{-\frac{t}{2t+1/2}}$$

is established, where we write $C_\eta := \frac{R}{2} a_\eta$.

Bibliography

[ACC17a] Ery Arias-Castro and Alberto Rodríguez Casal. On estimating the perimeter using the alpha-shape. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 53(3):1051–1068, 2017.

[ACC17b] Ery Arias-Castro and Shiyun Chen. Distribution-free multiple testing. *Electronic Journal of Statistics*, 11(1):1983–2001, 2017.

[ACV14] Ery Arias-Castro and Nicolas Verzelen. Community detection in dense random networks. *The Annals of Statistics*, 42(3):940–969, 2014.

[Bar02] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[BCG18] Gilles Blanchard, Alexandra Carpentier, and Maurilio Gutzeit. Minimax Euclidean separation rates for testing convex hypotheses in $\mathbb{R}^d$. *Electronic Journal of Statistics*, 12(2):3713–3735, 2018.

[BDER16] Sébastien Bubeck, Jian Ding, Ronen Eldan, and Miklós Z Rácz. Testing for high-dimensional geometry in random graphs. *Random Structures & Algorithms*, 49(3):503–532, 2016.

[BH95] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, pages 289–300, 1995.

[BHL05] Yannick Baraud, Sylvie Huet, and Béatrice Laurent. Testing Convex Hypotheses on the Mean of a Gaussian Vector. Application to Testing Qualitative Hypotheses on a Regression Function. *The Annals of Statistics*, pages 214–257, 2005.

[Bir01] Lucien Birgé. An alternative point of view on Lepski's method. *Lecture Notes-Monograph Series*, pages 113–133, 2001.

[BJR07] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[BN13] Adam D. Bull and Richard Nickl. Adaptive confidence sets in L_2. *Probability Theory and Related Fields*, 156(3):889–919, 2013.

[Bur79] M Burnashev. On the minimax detection of an imperfectly known signal in a white noise background. *Theory Probab. Appl.*, 24:107–119, 1979.

[BY01] Yoav Benjamini and Daniel Yekutieli. The Control of the False Discovery Rate in Multiple Testing under Dependency. *Annals of statistics*, pages 1165–1188, 2001.

[Car15] Alexandra Carpentier. Testing the regularity of a smooth signal. *Bernoulli*, 21(1):465–488, 2015.

[Cas96] James Casey. Exploring Curvature. *Vieweg Wiesbaden*, 1996.

[CB02] George Casella and Roger L Berger. Statistical inference. *Duxbury Pacific Grove, CA*, 2, 2002.

[CCT17] Olivier Collier, Laëtitia Comminges, and Alexandre B Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics*, 45(3):923–958, 2017.

[CCT18] Olivier Collier, Laëtitia Comminges, and Alexandre B Tsybakov. On Estimation of Nonsmooth Functionals of Sparse Normal Means. *arXiv preprint 1805.10791*, 2018.

[CD13] Laëtitia Comminges and Arnak Dalalyan. Minimax testing of a composite null hypothesis defined via a quadratic functional in the model of regression. *Electronic Journal of Statistics*, 7:146–190, 2013.

[CDRV18] Alexandra Carpentier, Sylvain Delattre, Etienne Roquain, and Nicolas Verzelen. Estimating Minimum Effect with Outlier Selection. *arXiv preprint 1809.08330*, 2018.

[CDV93] Albert Cohen, Ingrid Daubechies, and Pierre Vial. Wavelets on the interval and fast wavelet transforms. *Applied and computational harmonic analysis*, 1(1):54–81, 1993.

[Che52] Herman Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23:493 – 507, 1952.

[CL05] T Tony Cai and Mark G Low. Nonquadratic Estimators of a Quadratic Functional. *The Annals of Statistics*, 33(6):2930–2956, 2005.

[CL06] T. Tony Cai and Mark G. Low. Adaptive confidence balls. *The Annals of Statistics*, 34(1):202–228, 2006.

[CL11] T. Tony Cai and Mark G. Low. Testing Composite Hypotheses, Hermite Polynomials and Optimal Estimation of a Nonsmooth Functional. *The Annals of Statistics*, 39(2):1012–1041, 2011.

[CN15] Alexandra Carpentier and Richard Nickl. On signal detection and confidence sets for low rank inference problems. *Electronic Journal of Statistics*, 9(2):2675–2688, 2015.

[CV19] Alexandra Carpentier and Nicolas Verzelen. Adaptive estimation of the sparsity in the Gaussian vector model. *The Annals of Statistics*, 47(1):93–126, 2019.

[DN90] David L Donoho and Michael Nussbaum. Minimax Quadratic Estimation of a Quadratic Functional. *J. Complexity*, 6(3):290–323, 1990.

[Dym13] Harry Dym. Linear Algebra in Action. 78, 2013.

[Erm91] Michael Sergeevich Ermakov. Minimax detection of a signal in a Gaussian white noise. *Theory of Probability & Its Applications*, 35(4):667–679, 1991.

[FDR07] Helmut Finner, Thorsten Dickhaus, and Markus Roters. Dependency and False Discovery Rate: Asymptotics. *The Annals of Statistics*, pages 1432–1455, 2007.

[GGCvL17a] Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-Sample Tests for Large Random Graphs Using Network Statistics. *Proceedings of the 2017 Conference on Learning Theory*, 65:954–977, 2017.

[GGCvL17b] Debarghya Ghoshdastidar, Maurilio Gutzeit, Alexandra Carpentier, and Ulrike von Luxburg. Two-sample Tests for Random Graphs. *currently under revision, arXiv preprint 1707.00833*, 2017.

[GN16] Evarist Giné and Richard Nickl. Mathematical foundations of infinite-dimensional statistical models. *Cambridge University Press*, 40, 2016.

[GP05] Ghislaine Gayraud and Christophe Pouet. Adaptive minimax testing in the discrete regression scheme. *Probability Theory and Related Fields*, 133(4):531–558, 2005.

[Gut19] Maurilio Gutzeit. Testing the Sobolev-Type Regularity of a Function. *currently under revision, arXiv preprint 1901.00880*, 2019.

[Hoe63] Wassily Hoeffding. Probability inequalities for Sums of Bounded Random Variables. *Journal of the American statistical association*, 58(301):13–30, 1963.

[Ing82] Yuri Izmailovich Ingster. On the minimax nonparametric detection of signals in white Gaussian noise. *Problemy Peredachi Informatsii*, 18(2):61–73, 1982.

[Ing93] Y Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist*, 2(2):85–114, 1993.

[Ing98] Yuri Ingster. ADAPTATION IN MINIMAX NONPARAMETRIC HYPOTHESIS TESTING FOR ELLIPSOIDS AND BESOV BODIES, volume Technical Report 419. 1998.

[Ing00] Yuri Ingster. ON TESTING A HYPOTHESIS WHICH IS CLOSE TO A SIMPLE HYPOTHESIS. *Theory Prob. Appl.*, 45:310–323, 2000.

[IS02] Y Ingster and I A Suslina. NON-PARAMETRIC GOODNESS-OF-FIT TESTING UNDER GAUSSIAN MODELS. *Springer*, 2002.

[JLL03] Anatoli Juditsky and Sophie Lambert-Lacroix. NONPARAMETRIC CONFIDENCE SET ESTIMATION. *Math. Methods Statist*, 12(4):410–428, 2003.

[JN02] Anatoli Juditsky and Arkadi Nemirovski. ON NONPARAMETRIC TESTS OF POSITIVITY/MONOTONICITY/CONVEXITY. *The Annals of Statistics*, pages 498–527, 2002.

[LC12] Lucien Le Cam. ASYMPTOTIC METHODS IN STATISTICAL DECISION THEORY. *Springer Science & Business Media*, 2012.

[LNS99] Oleg Lepski, Arkady Nemirovski, and Vladimir Spokoiny. ON ESTIMATION OF THE L_r NORM OF A REGRESSION FUNCTION. *Probability Theory and Related Fields*, 113(2):221–253, 1999.

[LS99] Oleg V Lepski and Vladimir G Spokoiny. MINIMAX NONPARAMETRIC HYPOTHESIS TESTING: THE CASE OF AN INHOMOGENEOUS ALTERNATIVE. *Bernoulli*, 5(2):333–358, 1999.

[Mas07] Pascal Massart. CONCENTRATION INEQUALITIES AND MODEL SELECTION. *Springer*, 2007.

[MR18] Ester Mariucci and Markus Reiß. WASSERSTEIN AND TOTAL VARIATION DISTANCE BETWEEN MARGINALS OF LÉVY PROCESSES. *Electronic Journal of Statistics*, 12(2):2482–2514, 2018.

[NP33] Jerzy Neyman and Egon Pearson. ON THE PROBLEM OF THE MOST EFFICIENT TESTS OF STATISTICAL HYPOTHESES. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289âĂŞ337, 1933.

[RVDW09] Etienne Roquain and Mark A Van De Wiel. OPTIMAL WEIGHTING FOR FALSE DISCOVERY RATE CONTROL. *Electronic Journal of Statistics*, 3:678–711, 2009.

[Spo96] Vladimir G Spokoiny. ADAPTIVE HYPOTHESIS TESTING USING WAVELETS. *The Annals of Statistics*, 24(6):2477–2498, 1996.

[TAS+17a] Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, Youngser Park, and Carey E Priebe. A SEMIPARAMETRIC TWO-SAMPLE HYPOTHESIS TESTING PROBLEM FOR RANDOM GRAPHS. *Journal of Computational and Graphical Statistics*, 26(2):344–354, 2017.

[TAS+17b] Minh Tang, Avanti Athreya, Daniel L Sussman, Vince Lyzinski, and Carey E Priebe. A nonparametric two-sample hypothesis testing problem for random graphs. *Bernoulli*, 23(3):1599–1630, 2017.

[Tri92] H Triebel. Theory of function spaces II. *Birkhäuser Basel*, 1992.

[VAC15] Nicolas Verzelen and Ery Arias-Castro. Community detection in sparse random networks. *The Annals of Applied Probability*, 25(6):3465–3510, 2015.

[VAC17] Nicolas Verzelen and Ery Arias-Castro. Detection and feature selection in sparse mixture models. *The Annals of Statistics*, 45(5):1920–1950, 2017.

[Ver12] Nicolas Verzelen. Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electronic Journal of Statistics*, 6:38–90, 2012.

[Win12] Andreas Winkelbauer. Moments and Absolute Moments of the Normal Distribution. *arXiv preprint 1209.4340*, 2012.

Ehrenerklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; verwendete fremde und eigene Quellen sind als solche kenntlich gemacht.

Ich habe insbesondere nicht wissentlich:

- Ergebnisse erfunden oder widersprüchliche Ergebnisse verschwiegen,
- statistische Verfahren absichtlich missbraucht, um Daten in ungerechtfertigter Weise zu interpretieren,
- fremde Ergebnisse oder Veröffentlichungen plagiiert oder verzerrt wiedergegeben.

Mir ist bekannt, dass Verstöße gegen das Urheberrecht Unterlassungs- und Schadenersatzansprüche des Urhebers sowie eine strafrechtliche Ahndung durch die Strafverfolgungsbehörden begründen kann.

Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form als Dissertation eingereicht und ist als Ganzes auch noch nicht veröffentlicht.

Magdeburg, 8. Januar 2019

Maurilio Gutzeit

Declaration of Honor

I hereby declare that I produced this thesis without prohibited assistance and that all sources of information that were used in producing this thesis, including my own publications, have been clearly marked and referenced.

In particular I have not wilfully:

- Fabricated data or ignored or removed undesired results.
- Misused statistical methods with the aim of drawing other conclusions than those warranted by the available data.
- Plagiarised data or publications or presented them in a distorted way.

I know that violations of copyright may lead to injunction and damage claims from the author or prosecution by the law enforcement authorities.

This work has not previously been submitted as a doctoral thesis in the same or a similar form in Germany or in any other country. It has not previously been published as a whole.

Magdeburg, 8th January, 2019

Maurilio Gutzeit