FACULTÉ
DES SCIENCES
UNIVERSITÉ LIBRE DE BRUXELLES

Testing uniformity against rotationally symmetric alternatives on high-dimensional spheres

Thesis submitted by Christine CUTTING
in fulfilment of the requirements of the PhD Degree in Science ("Docteur en Sciences")
Academic year 2019-2020

Supervisor: Professor Davy PAINDAVEINE

Co-supervisor: Professor Thomas VERDEBOUT

ECARES and Département de Mathématique

Remerciements

Dans un premier temps, je tiens à remercier mon promoteur Davy Paindaveine, qui m'a permis d'effectuer cette thèse avec lui. Il me l'a proposé à un moment où j'étais un peu perdue dans mes études et je lui en suis infiniment reconnaissante. J'ai aussi eu la chance d'avoir Thomas Verdebout comme co-promoteur, avec lequel Davy forme une sacrée équipe. J'ai beaucoup apprécié sa personnalité très accessible et détendue. Merci, Davy, pour le cours de statistique de BA2 qui fut une révélation, et merci à tous les deux pour votre grande disponibilité et votre encadrement pendant ces six ans. De manière générale, je suis ravie de mon parcours à l'ULB et de la qualité des enseignements et des TP que j'y ai reçus.

Ensuite, je remercie mes parents qui m'ont donné les moyens de poursuivre les études que je voulais et m'ont toujours poussée vers le haut. Thank you Jo for being such an amazing aunt and for letting me confide in you. Merci bro d'être un chouette frère (mon préféré ;)).

Ces remerciements ne pourraient pas exister sans mentionner les gens merveilleux rencontrés à l'ULB qui m'ont permis de m'épanouir : Cédric, François, Mitia, Alexia, Thierry, Sylvie, Hoan-Phung, Patrick, Tania, Jérémie, Keno, Lancelot, Dimitri, Anna, Antoine, Florence, Julien, et beaucoup d'autres. Julie, je suis contente d'avoir quelqu'un comme toi sur qui je peux compter et avec qui j'ai passé autant d'années en bachelier, en colocation, comme collègues de bureau et partenaires de pâtisserie. Matthieu, je ne te remercierai jamais assez pour tout ce que tu as fait pour moi. J'ai une dette infinie de Triple Karmeliet envers toi. Nora et Clade Max, merci de partager avec moi la joie de passer des week-ends entiers en tournoi de badminton et de me rappeler que le plus important, c'est de participer ! Dibo, je ne pensais pas rencontrer quelqu'un comme toi un jour, avec qui on se comprend tellement bien. Maintenant qu'on a commencé à partir en rando ensemble tous les ans, ça doit continuer au moins pour toujours ! Dong-Yan, je suis contente que tu aies été une étudiante impertinente prête à se lier d'amitié avec ses assistants. Tu es ma plus belle rencontre amicale pendant cette thèse, ta présence, même à distance, m'est indispensable et je chéris notre relation qui doit continuer aussi longtemps que les randos avec Dibo.

Enfin, merci, Alex, de me supporter tous les jours, notamment ces derniers mois. Tu es mon pilier et mon meilleur ami avec qui je ne me lasse pas d'être confinée, et ce n'est pas peu dire. Je t'aime fort mon "homme du moment".

Contents

Notations

$\otimes$	the usual Kronecker product
$\xrightarrow{\mathscr{D}}$	convergence in distribution
$\xrightarrow{\mathrm{L}^r}$	convergence in the L^r-norm
$\mathbf{0}$	the null vector
$\mathbf{A}^-$	the Moore–Penrose generalized inverse of matrix $\mathbf{A}$
$\mathscr{B}(\Omega)$	the Borel σ-algebra on Ω
c_p	$\frac{\Gamma(p/2)}{\sqrt{\pi}\,\Gamma((p-1)/2)}$
χ^2_d	the chi-square distribution with d degrees of freedom
$\chi^2_{d,\alpha}$	the α-quantile of the chi-square distribution with d degrees of freedom
$\chi^2_d(\lambda)$	the noncentral chi-square distribution with d degrees of freedom and noncentrality parameter λ
$\mathbf{e}_\ell$	the ℓth vector of the canonical basis of $\mathbb{R}^p$
$\mathscr{F}$	the class of functions $\{f : \mathbb{R} \to \mathbb{R}^+\}$ such that f is monotone increasing, twice differentiable at 0 and $f(0) = f'(0) = 1$
$\Gamma(\cdot)$	the Euler Gamma function
$\mathbf{I}_p$	the $p \times p$ identity matrix
$\mathscr{I}_\nu(\cdot)$	the order-ν modified Bessel function of the first kind
$\mathbb{I}_\mathrm{A}$	the indicator function of condition A
iid	independent and identically distributed
$\mathbf{J}_p$	the matrix $(\mathrm{vec}\,\mathbf{I}_p)(\mathrm{vec}\,\mathbf{I}_p)'$
$\mathbf{K}_p$	the $p^2 \times p^2$ commutation matrix $\sum_{i,j=1}^{p} \left(\mathbf{e}_i\mathbf{e}_j'\right) \otimes \left(\mathbf{e}_j\mathbf{e}_i'\right)$
$\hat{\lambda}_{ni}$	the ith largest eigenvalue of the sample covariance matrix $\mathbf{S}_n$
LAN	Locally Asymptotically Normal
$\mathscr{N}(\mu, \sigma^2)$	the univariate normal distribution with mean μ and variance σ^2
$\mathscr{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$

$\mathrm{P}_0^{(n)}$	the $\mathbf{X}_{ni}$'s, $i=1,\ldots,n$ are iid and uniformly distributed on $\mathcal{S}^{p_n-1}$
$\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}$	the $\mathbf{X}_{ni}$'s, $i=1,\ldots,n$ are iid, rotationally symmetric about $\boldsymbol{\theta}_n$ on $\mathcal{S}^{p-1}(\mathbf{x})$ and $\mathbf{X}'_{ni}\boldsymbol{\theta}_n$ has cumulative distribution F_n
$\mathrm{P}_{\mathrm{F}_n}^{(n)}$	a special case of $\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}$ with $\boldsymbol{\theta}_n$ chosen as the first vector of the canonical basis of $\mathbb{R}^{p_n}$
$\mathrm{P}_{\boldsymbol{\theta}_n,\kappa_n,f}^{(n)}$	the $\mathbf{X}_{ni}$'s, $i=1,\ldots,n$ are independent and identically distributed (iid) with density $\mathbf{x}\mapsto c_{p_n,\kappa_n,f}\, f\left(\kappa_n\mathbf{x}'\boldsymbol{\theta}_n\right)\mathbb{1}_{\mathcal{S}^{p-1}}(\mathbf{x})$
$\check{\mathrm{P}}_{\boldsymbol{\theta}_n,\kappa_n,f}^{(n)}$	the $\mathbf{X}_{ni}$'s, $i=1,\ldots,n$ are independent and identically distributed (iid) with density $\mathbf{x}\mapsto c_{p_n,\kappa_n,f}\, f\left(\kappa_n(\mathbf{x}'\boldsymbol{\theta}_n)^2\right)\mathbb{1}_{\mathcal{S}^{p-1}}(\mathbf{x})$
$\phi_{\mathrm{Bing}}^{(n)}$	the Bingham test, see (3.1)
$\phi_{\mathrm{Ray}}^{(n)}$	the Rayleigh test, see (2.2)
$\Phi(\cdot)$	the cumulative distribution function of the standard normal distribution
$\Psi_p(\cdot)$	the cumulative distribution function of the chi-squared distribution with p degrees of freedom
$\Psi_p(\cdot;\lambda)$	the cumulative distribution function of the noncentral chi-squared distribution with p degrees of freedom and noncentrality parameter λ
Q_n	the Bingham test statistic, $\frac{np_n(p_n+2)}{2}\left(\mathrm{tr}\left[\mathbf{S}_n^2\right]-\frac{1}{p_n}\right)$
$\mathrm{Q}_n^{\mathrm{St}}$	the high-dimensional Bingham test statistic, $\frac{\mathrm{Q}_n-d_{p_n}}{\sqrt{2d_{p_n}}}$, where $d_{p_n}=p_n(p_n+1)/2-1$
R_n	the Rayleigh test statistic, $np_n\|\bar{\mathbf{X}}_n\|^2$
$\mathrm{R}_n^{\mathrm{St}}$	the high-dimensional Rayleigh test statistic, $\frac{\mathrm{R}_n-p_n}{\sqrt{2p_n}}$
s_a	the sign of real number a
$\mathbf{S}_n$	the covariance matrix of the observations, $n^{-1}\sum_{i=1}^n\mathbf{X}_{ni}\mathbf{X}'_{ni}$
S^{p-1}	the unit sphere in $\mathbb{R}^p$, $\left\{\mathbf{x}\in\mathbb{R}^p:\|\mathbf{x}\|=1\right\}$
$\mathrm{tr}(\mathbf{A})$	the trace of matrix $\mathbf{A}$
$\bar{\mathbf{X}}_n$	the mean vector of the observations, $n^{-1}\sum_{i=1}^n\mathbf{X}_{ni}$
$\mathrm{vec}\,\mathbf{A}$	the vector obtained by stacking the columns of matrix $\mathbf{A}$ on top of each other
z_α	the upper α-quantile of the standard normal distribution $\mathcal{N}(0,1)$

Introduction

In directional statistics, inference is based on p-variate observations lying on the unit sphere $\mathcal{S}^{p-1} := \{\mathbf{x} \in \mathbb{R}^p : \|\mathbf{x}\| = \sqrt{\mathbf{x}'\mathbf{x}} = 1\}$. This is relevant in various situations.

(i) First, the original data themselves may belong to $\mathcal{S}^{p-1}$ like the times of day at which thunderstorms occur ($p = 2$) or epicentres of earthquakes ($p = 3$) (see [Mardia and Jupp, 2000] for more examples).

(ii) Second, some fields by nature are so that only the relative magnitude of the observations is important, which leads to projecting observations onto $\mathcal{S}^{p-1}$. In shape analysis, for instance, this projection only gets rid of an overall scale factor related to the (irrelevant) object size.

(iii) Finally, even in inference problems where the full (Euclidean) observations in principle need to be considered, a common practice in nonparametric statistics is to restrict to sign procedures, that is, to procedures that are measurable with respect to the projections of the observations onto $\mathcal{S}^{p-1}$; see, e.g., [Oja, 2010] and the references therein.

While (i) is obviously restricted to small dimensions p, (ii)-(iii) nowadays increasingly involve high-dimensional data. For (ii), high-dimensional directional data were considered in [Dryden, 2005], with applications in brain shape modeling; in text mining, [Banerjee et al., 2003] and [Banerjee and Ghosh, 2004] project high-dimensional data on unit spheres to remove the bias arising due to the length of a text when performing clustering. As for (iii), the huge interest raised by high-dimensional statistics in the last decade has made it natural to consider high-dimensional *sign* tests. In particular, [Zou et al., 2014] recently considered the high-dimensional version of the [Hallin and Paindaveine, 2006] sign tests of sphericity, whereas an extension to the high-dimensional case of the location sign test from [Chaudhuri, 1992] and [Möttönen and Oja, 1995] was recently proposed in [Wang et al., 2015]. Considering (iii) in high dimensions is particularly appealing since for moderate-to-large p, sign tests show excellent (fixed-p) efficiency properties.

We consider the problem of testing uniformity on the unit sphere $\mathcal{S}^{p-1}$. This is a very classical problem in multivariate analysis that can be traced back to [Bernoulli, 1735]. It is discussed at length in strictly all textbooks in the field; see, among many others, [Fisher et al., 1987], [Ley and Verdebout, 2017], and [Mardia and Jupp, 2000]. As explained in the review paper [García-Portugués and Verdebout, 2018], the topic has recently received a lot of attention: to cite only a few contributions, [Jupp, 2008] proposed data-driven Sobolev tests, [Cuesta-Albertos et al., 2009] proposed tests based on random projections, [Sun and Lockhart, 2019] obtained Bayesian optimality properties of some tests while [García-Portugués et al., 2019] transformed some uniformity tests into tests of rotational symmetry. Possible applications of testing uniformity on high-dimensional spheres include outlier detection; see [Juan and Prieto, 2001]. Other natural applications are related with testing for sphericity in $\mathbb{R}^p$, in the spirit of (iii) above.

As p_n can diverge to infinity with n, it is necessary to consider triangular arrays of observations $\mathbf{X}_{ni}$, $i = 1,\dots,n$, $n = 1,2,\dots$ such that for any n, $\mathbf{X}_{ni} \in \mathcal{S}^{p_n-1}$ (some other works, including [Chikuse, 1991, Chikuse, 1993], actually rather consider a fixed-n large-p asymptotic scenario). Results to test uniformity on high-dimensional spheres already exist but they often impose conditions on the way p_n and n diverge to infinity. For example, [Cai et al., 2013] study the asymptotic behaviours of the pairwise angles among n uniformly distributed vectors on $\mathcal{S}^{p_n-1}$ as $n \to \infty$, while p_n is either fixed or growing to infinity with n. Let $0 \le \Theta_{ij} \le \pi$ denote the angle between the unit vectors $\overrightarrow{O\mathbf{X}}_{ni}$ and $\overrightarrow{O\mathbf{X}}_{nj}$ for all $1 \le i < j \le n$, and $\Theta_{\min} := \min\{\Theta_{ij} : 1 \le i < j \le n\}$. When $p_n \to \infty$ as $n \to \infty$, three different regimes have to be considered:

- the sub-exponential case: $\frac{\log p_n}{n} \to 0$ as $n \to \infty$. Then

$$2p_n \log\sin\Theta_{\min} + 4\log n - \log\log n \xrightarrow{\mathcal{D}} Y,$$

  as $n \to \infty$, where Y has cumulative distribution function

$$F(y) = 1 - \exp\left(-\frac{e^{y/2}}{4\sqrt{2\pi}}\right), \quad y \in \mathbb{R}.$$

- the exponential case: $\frac{\log p_n}{n} \to \beta \in (0,\infty)$ as $n \to \infty$. Then

$$2p_n \log\sin\Theta_{\min} + 4\log n - \log\log n \xrightarrow{\mathcal{D}} Y,$$

  as $n \to \infty$, where Y has cumulative distribution function

$$F(y) = 1 - \exp\left(-\sqrt{\frac{\beta}{8\pi(1-e^{-4\beta})}}\, e^{(y+8\beta)/2}\right), \quad y \in \mathbb{R}.$$

- the super-exponential case: $\frac{\log p_n}{n} \to \infty$ as $n \to \infty$. Then

$$2p_n \log\sin\Theta_{\min} + \frac{4p_n}{p_n - 1}\log n - \log p_n \xrightarrow{\mathcal{D}} Y,$$

  as $n \to \infty$, where Y has cumulative distribution function

$$F(y) = 1 - \exp\left(-\frac{e^{y/2}}{2\sqrt{2\pi}}\right), \quad y \in \mathbb{R}.$$

In each case, uniformity is rejected if the statistic based on $\Theta_{\min}$ is too small and the critical value can be found thanks to the cumulative distribution functions above. In practice, though, if only one sample is available, it is not possible to know which regime applies and therefore which test to use.

On the contrary, test statistics that work in low dimensions and that do not need to be corrected to be valid in high-dimensions are called high-dimension robust as defined by [Paindaveine and Verdebout, 2016]. This notion arises first in [Ledoit and Wolf, 2002] who showed that the Gaussian sphericity test from [John, 1972] is robust against high dimensionality. Let $\mathbf{Z}_{n1},\dots,\mathbf{Z}_{nn}$ be Gaussian observations in $\mathbb{R}^{p_n-1}$. When $p_n = p$ is fixed, the Gaussian sphericity test from [John, 1972] rejects the null when

$$\frac{np}{2} U > \chi^2_{d_p;1-\alpha},$$

where

$$\mathrm{U} := p\left(\frac{\mathrm{tr}(\mathbf{S}_n^2)}{\mathrm{tr}(\mathbf{S}_n)^2} - \frac{1}{p}\right),$$

$\mathbf{S}_n := n^{-1}\sum_{i=1}^n \mathbf{Z}_{ni}\mathbf{Z}'_{ni}$ is the covariance matrix of the observations, $\chi^2_{d,1-\alpha}$ denotes the upper α-quantile of the χ^2_d distribution and $d_p := p(p+1)/2-1$. In high dimensions, $d_{p_n} \to \infty$ as $p_n \to \infty$ and it is expected that

$$\frac{np_n\mathrm{U}/2 - d_{p_n}}{\sqrt{2d_{p_n}}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

This is indeed the case when $p_n/n \to c \in (0,\infty)$ as shown in [Ledoit and Wolf, 2002].

In an even more general framework, $p_n \to \infty$ as $n \to \infty$ without any condition on the way p_n goes to infinity with n, [Paindaveine and Verdebout, 2016] showed the high-dimension robustness of several sign test statistics thanks to a martingale theorem that will be stated in Chapter 1 and used in Chapter 4. In particular, they studied the most classical test of uniformity, the [Rayleigh, 1919] test, which rejects the null of uniformity for large values of $\mathrm{R}_n := np_n\|\bar{\mathbf{X}}_n\|^2$, where $\bar{\mathbf{X}}_n := n^{-1}\sum_{i=1}^n \mathbf{X}_{ni}$. In low dimensions ($p_n = p$ for all n), the test is based on the null asymptotic chi-square distribution with p degrees of freedom, χ^2_p, of R_n. In the high-dimensional setup ($p_n \to \infty$ as $n \to \infty$), [Paindaveine and Verdebout, 2016] showed that

$$\mathrm{R}_n^{\mathrm{St}} := \frac{\mathrm{R}_n - p_n}{\sqrt{2p_n}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Therefore, denoting by $\Phi(\cdot)$ the cumulative distribution function of the standard normal and defining $z_\alpha := \Phi^{-1}(1-\alpha)$, the high-dimensional Rayleigh test rejects the null of uniformity when

$$\mathrm{R}_n^{\mathrm{St}} > \frac{\chi^2_{p_n,1-\alpha} - p_n}{\sqrt{2p_n}} \quad (= z_\alpha + o(1))$$

has asymptotic size α under the null hypothesis, irrespective of the rate at which p_n diverges to infinity with n, and is therefore, in that sense, high-dimension robust.

In low dimensions the Rayleigh test is the score test of uniformity within the popular Fisher–Von Mises–Langevin (FvML) model. It is then natural to consider an extension of this model to study its non-null behaviour. This is why we will focus in Chapter 2 on mutually independent observations with a common rotationally symmetric density proportional to $\mathbf{x} \in \mathcal{S}^{p_n-1} \mapsto f(\kappa_n \mathbf{x}'\boldsymbol{\theta}_n)$, where $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ is a location parameter, $\kappa_n > 0$ is a concentration parameter and f is monotone strictly increasing. If κ_n is too small, the alternatives are too close to the null of uniformity and no test can be consistent. We say that the alternatives are contiguous to the null and we will see how small κ_n must be. When $\boldsymbol{\theta}_n$ is known, we will show thanks to Local Asymptotic Normality (LAN) that the Rayleigh test is not optimal even though it shows some power against contiguous alternatives in low dimensions. However, when $\boldsymbol{\theta}_n$ is unknown, it becomes optimal against general monotone alternatives in low dimensions and against FvML alternatives in high dimensions.

Despite its nice optimality properties, the Rayleigh test will detect only asymmetric deviations from uniformity. It will show no power when the common distribution of the $\mathbf{X}_{ni}$'s attributes the same probability to antipodal regions. Far from being the exception, such antipodally symmetric distributions are actually those that need be considered when practitioners are facing axial data, that is, when one does not observe genuine locations on the sphere but rather axes (a typical example of axial data relates to the directions of optical axes in quartz crystals; see, e.g., [Mardia and Jupp, 2000]). Models and inference for axial data have been considered a lot in the literature: to cite a few, [Tyler, 1987b] and more recently [Paindaveine et al., 2018] considered inference on the distribution of the spatial sign of a Gaussian vector, [Watson, 1965], [Bijral et al., 2007] and [Sra and Karp, 2013] considered inference for Watson distributions, [Dryden, 2005] obtained distributions on high-dimensional spheres while [Anderson and Stephens, 1972], [Bingham, 1974] and [Jupp, 2001] considered uniformity tests against antipodally symmetric alternatives.

In low dimensions, the textbook test of uniformity for axial data is the [Bingham, 1974] test, that rejects the null hypothesis of uniformity whenever

$$\mathrm{Q}_n = \frac{np_n(p_n+2)}{2}\left(\mathrm{tr}\left[\mathbf{S}_n^2\right] - \frac{1}{p_n}\right) > \chi^2_{d_{p_n},1-\alpha},$$

where $\mathbf{S}_n$ is the covariance matrix of the observations on $\mathcal{S}^{p_n-1}$ and d_p is defined as above. In high dimensions, Theorem 2.5 from [Paindaveine and Verdebout, 2016] implies that, under the null hypothesis of uniformity,

$$\mathrm{Q}_n^{\mathrm{St}} = \frac{\mathrm{Q}_n - d_{p_n}}{\sqrt{2d_{p_n}}} = \frac{p_n}{n}\sum_{\substack{i<j\\ i,j=1}}^{n}\left\{\left(\mathbf{X}'_{ni}\mathbf{X}_{nj}\right)^2 - \frac{1}{p_n}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

as soon as p_n diverges to infinity with n. Therefore the Bingham test is also robust to high-dimensionality. Again, this property is not enough to justify its use and we are interested in its nonnull behaviour. In this case it is natural to consider an extension of the Watson distributions, namely axial rotationally symmetric distributions. In Chapter 3 we will assume that the observations share a density proportional to $\mathbf{x} \in \mathcal{S}^{p_n-1} \mapsto f\left(\kappa_n\,(\mathbf{x}'\boldsymbol{\theta}_n)^2\right)$, where $\boldsymbol{\theta}_n$ and f are defined as above and $\kappa_n \in \mathbb{R}$ is still a concentration parameter. The contiguity rate in this model is higher than in the monotone one and suggests the problem is more complicated. When $\boldsymbol{\theta}_n$ is known, LAN yields optimal tests and enables to show that the Bingham test is not optimal when $\boldsymbol{\theta}_n$ is unknown but shows some power against contiguous alternatives in low dimensions. Consequently we turn to single-spiked tests based on the extreme eigenvalues of the covariance matrix and derive in the low-dimensional case their limiting behaviour under contiguous alternatives.

In Chapter 4 we will prove with a martingale central limit theorem that the null high-dimensional asymptotic distributions of the Rayleigh and Bingham tests, after appropriate standardisation, actually hold under broad classes of rotationally symmetric distributions. This entails that the Bingham test, which is blind to contiguous alternatives in the monotone and axial models, can detect more severe alternatives in both cases. Throughout, simulations confirm our findings.

In Chapters 2-4 proofs are to be found in the last section. To make this document as accessible as possible, theorems that require long technical proofs are followed by sketches of proofs for readers who want the general idea without having to go into too many details or need help getting their bearings. For theorems which are proven in a similar way, the outline of the proof is given after the first one.

This thesis is based on two published and one submitted articles:

- Cutting, Christine; Paindaveine, Davy; Verdebout, Thomas. Testing uniformity on high-dimensional spheres against monotone rotationally symmetric alternatives. Ann. Statist. 45 (2017), no. 3, 1024–1058. https://projecteuclid.org/euclid.aos/1497319687

- Cutting, Christine; Paindaveine, Davy; Verdebout, Thomas. On the power of axial tests of uniformity. Electronic Journal of Statistics 14, 2123-2154 (2020). https://projecteuclid.org/euclid.ejs/1589335309

- Cutting, Christine; Paindaveine, Davy; Verdebout, Thomas. Testing uniformity on high-dimensional spheres: the non-null behaviour of the Bingham test.

Chapter 1

Useful notions

Contents

In this chapter we introduce various unrelated notions that arise in subsequent chapters. We start in Section 1.1 with some distributions on the sphere: the uniform distribution and rotationally symmetric distributions. We then focus on two special cases, monotone and axial distributions. In Section 1.2 we define Local Asymptotic Normality which can be used to find optimal tests and compute asymptotic powers under suitable alternatives thanks to Le Cam's lemmas. Section 1.3 is devoted to the invariance principle and the reduction by invariance. Finally, Section 1.4 is about Billingsley's Theorem, a martingale central limit theorem that will be applied to find the non-null behaviour of the Rayleigh and Bingham tests.

1.1 Some distributions on the sphere

1.1.1 Uniform distribution

If $\mathbf{X}$ is uniformly distributed over S^{p-1}, then its density is

$$\frac{\Gamma\left(\frac{p-1}{2}\right)c_p}{2\pi^{(p-1)/2}}\mathbb{I}_{\{\mathbf{x}\in\mathrm{S}^{p-1}\}},\quad \text{with } c_p := \frac{\Gamma\left(\frac{p}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{p-1}{2}\right)},\tag{1.1}$$

where $\Gamma(\cdot)$ is the Euler Gamma function and $\mathbb{I}_{\mathrm{A}}$ is the indicator function of condition A. As $\mathbf{OX}$ has the same distribution as $\mathbf{X}$ for any orthogonal matrix $\mathbf{O}$, $\mathrm{E}[\mathbf{X}] = \mathbf{0}$ and $\mathrm{Var}[\mathbf{X}]$ must be proportional to the identity matrix $\mathbf{I}_p$. For $\|\mathbf{X}\| = 1$ almost surely,

$$\mathrm{Var}[\mathbf{X}] = \mathrm{E}\left[\mathbf{X}\mathbf{X}'\right] = \frac{1}{p}\mathbf{I}_p.\tag{1.2}$$

Besides we can prove that (see [Tyler, 1987a], page 244)

$$\mathrm{E}\left[\mathrm{vec}\left(\mathbf{X}\mathbf{X}'\right)\left(\mathrm{vec}\left(\mathbf{X}\mathbf{X}'\right)\right)'\right] = \frac{1}{p(p+2)}\left(\mathbf{I}_{p^2} + \mathbf{K}_p + \mathbf{J}_p,\right)\tag{1.3}$$

where $\mathbf{J}_p = \left(\mathrm{vec}\,\mathbf{I}_p\right)\left(\mathrm{vec}\,\mathbf{I}_p\right)'$, $\mathrm{vec}\,\mathbf{A}$ is the vector obtained by stacking the columns of matrix $\mathbf{A}$ on top of each other and $\mathbf{K}_p$ is the commutation matrix (see [Magnus and Neudecker, 2007]).

For any $\boldsymbol{\theta} \in \mathrm{S}^{p-1}$, the distribution of $\boldsymbol{\theta}'\mathbf{X}$ is symmetric about 0 and $\left(\boldsymbol{\theta}'\mathbf{X}\right)^2 \sim \mathrm{Beta}\left(\frac{1}{2}, \frac{p-1}{2}\right)$ (see, e.g., [Muirhead, 1982], Theorem 1.5.7(ii)). Therefore, $\boldsymbol{\theta}'\mathbf{X}$ has cumulative distribution function

$$\mathrm{F}_p(t) := c_p \int_{-1}^{t} (1-s^2)^{(p-3)/2}\,ds,\quad \text{for } t \in [-1,1],$$

and

$$\mathrm{E}\left[\boldsymbol{\theta}'\mathbf{X}\right] = c_p \int_{-1}^{1} s(1-s^2)^{(p-3)/2}\,ds = 0,\tag{1.4}$$

$$\mathrm{E}\left[\left(\boldsymbol{\theta}'\mathbf{X}\right)^2\right] = c_p \int_{-1}^{1} s^2(1-s^2)^{(p-3)/2}\,ds = \frac{1}{p},\tag{1.5}$$

$$\mathrm{E}\left[\left(\boldsymbol{\theta}'\mathbf{X}\right)^4\right] = c_p \int_{-1}^{1} s^4(1-s^2)^{(p-3)/2}\,ds = \frac{3}{p(p+2)}.\tag{1.6}$$

Consider triangular arrays of observations $\mathbf{X}_{ni}$, $i = 1,\ldots,n$, $n = 1,2,\ldots$; we denote by $\mathrm{P}_0^{(n)}$ the hypothesis that for any n, $\mathbf{X}_{n1}, \mathbf{X}_{n2}, \ldots, \mathbf{X}_{nn}$ are mutually independent and uniformly distributed on S^{p_n-1}.

1.1.2 Rotationally symmetric distributions

A p-dimensional vector $\mathbf{X}$ is said to be rotationally symmetric about $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ if and only if $\mathbf{OX}$ is equal in distribution to $\mathbf{X}$ for any orthogonal $p \times p$ matrix $\mathbf{O}$ satisfying $\mathbf{O}\boldsymbol{\theta} = \boldsymbol{\theta}$. Such distributions are fully characterized by the location parameter $\boldsymbol{\theta}$ and the cumulative distribution function F of $\mathbf{X}'\boldsymbol{\theta}$. Popular examples of rotationally distributions are the FvML and Watson distributions described in Sections 1.1.3 and 1.1.4.

If $\mathbf{X}$ is rotationally symmetric about $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ then we can show that (see [Saw, 1978])

$$\mathrm{E}[\mathbf{X}] = e_1 \boldsymbol{\theta}, \tag{1.7}$$

$$\mathrm{E}[\mathbf{X}\mathbf{X}'] = e_2 \boldsymbol{\theta}\boldsymbol{\theta}' + \frac{1-e_2}{p-1}\left(\mathbf{I}_p - \boldsymbol{\theta}\boldsymbol{\theta}'\right), \tag{1.8}$$

with $e_1 := \mathrm{E}[\mathbf{X}'\boldsymbol{\theta}]$ and $e_2 := \mathrm{E}\left[(\mathbf{X}'\boldsymbol{\theta})^2\right]$. Moreover, $\mathbf{X}$ can be decomposed into

$$\mathbf{X} = u\boldsymbol{\theta} + v\,\mathbf{S} \tag{1.9}$$

where $u := \mathbf{X}'\boldsymbol{\theta}$, $v := \sqrt{1-u^2}$ and

$$\mathbf{S} := \begin{cases} \frac{\mathbf{X}-u\boldsymbol{\theta}}{\|\mathbf{X}-u\boldsymbol{\theta}\|} & \text{if } \mathbf{X} \neq \boldsymbol{\theta}, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

The multivariate sign vector $\mathbf{S}$ is uniformly distributed on $\mathcal{S}^{p-1}(\boldsymbol{\theta}^\perp) := \{\mathbf{v} \in \mathcal{S}^{p-1} : \mathbf{v}'\boldsymbol{\theta} = 0\}$ and is independent of u (see (9.3.32) in [Mardia and Jupp, 2000]).

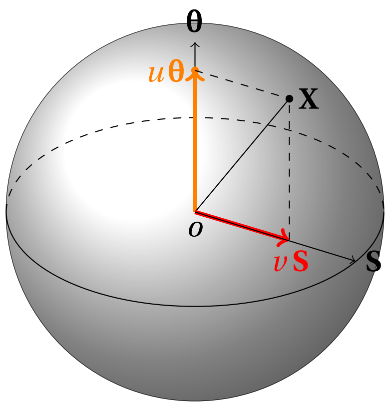


Figure 1.1 – The tangent-normal decomposition of $\mathbf{X}$, rotationally symmetric about $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$

Consider triangular arrays of observations $\mathbf{X}_{ni}$, $i = 1, \ldots, n$, $n = 1, 2, \ldots$; we denote by $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \mathrm{F}_n}$ the hypothesis that for any n, $\mathbf{X}_{n1}, \mathbf{X}_{n2}, \ldots, \mathbf{X}_{nn}$ are on $\mathcal{S}^{p_n-1}$, mutually independent and rotationally symmetric about $\boldsymbol{\theta}_n$, with cumulative distribution function F_n.

In the two following sections we will be interested in particular cases of rotationally symmetric distributions: monotone and axial ones.

1.1.3 "Monotone" rotationally symmetric distributions

We define "monotone" rotationally symmetric densities (with respect to the surface area measure on $\mathcal{S}^{p-1}$) as densities of the form

$$\mathbf{x} \mapsto \frac{\Gamma\left(\frac{p-1}{2}\right) c_{p,\kappa,f}}{2\pi^{(p-1)/2}} f\left(\kappa \mathbf{x}'\boldsymbol{\theta}\right), \qquad \mathbf{x} \in \mathcal{S}^{p-1}, \tag{1.10}$$

where

1. $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ is a location parameter;

2. $\kappa > 0$ is a concentration parameter (irrespective of f, $\kappa = 0$ corresponds to the uniform distribution and negative values of κ are discarded in this model because they only swap the roles of $\boldsymbol{\theta}$ and $-\boldsymbol{\theta}$);

3. the function f belongs to the class of functions $\mathscr{F} := \{f : \mathbb{R} \to \mathbb{R}^+ : f$ monotone increasing, twice differentiable at 0, with $f(0) = f'(0) = 1\}$;

4. the constant $c_{p,\kappa,f}$ is such that

$$c_{p,\kappa,f} \int_{-1}^{1} (1-s^2)^{(p-3)/2} f(\kappa s) ds = 1.$$

These conditions on f guarantee identifiability of $\boldsymbol{\theta}$, κ and f: clearly, the strict monotonicity of f implies that $\boldsymbol{\theta}$ is the modal location on $\mathcal{S}^{p-1}$, whereas the constraint $f'(0) = 1$ allows to identify κ and f.

This family of distributions is an extension of the Fisher–von Mises–Langevin (FvML) distributions; indeed, choosing $f(\cdot) = \exp(\cdot)$ in (1.10) provides

$$\mathbf{x} \mapsto c_{p,\kappa}^{\text{FvML}} \exp\left(\kappa \mathbf{x}'\boldsymbol{\theta}\right) \mathbb{I}_{\{\mathbf{x}\in\mathcal{S}^{p-1}\}}, \text{ with } c_{p,\kappa}^{\text{FvML}} := \frac{(\kappa/2)^{p/2-1}}{2\pi^{p/2}\mathscr{I}_{p/2-1}(\kappa)}, \tag{1.11}$$

where $\mathscr{I}_{\nu}(\cdot)$ is the order-ν modified Bessel function of the first kind; see Section 9.3.2. in [Mardia and Jupp, 2000] for more details.

Note that if $\mathbf{X}$ is a random vector with density (1.10), then $\mathbf{X}'\boldsymbol{\theta}$ has density

$$s \mapsto c_{p,\kappa,f} \left(1-s^2\right)^{(p-3)/2} f(\kappa s) \mathbb{I}_{\{|s|\le 1\}}.$$

For any $\boldsymbol{\theta}_n$, κ_n and f defined as above, we will denote as $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ the hypothesis under which the $\mathbf{X}_{ni}$'s, $i = 1, \ldots, n$, are mutually independent and share the common density

$$\mathbf{x} \mapsto \frac{\Gamma\left(\frac{p_n-1}{2}\right) c_{p_n,\kappa_n,f}}{2\pi^{(p_n-1)/2}} f\left(\kappa_n \mathbf{x}'\boldsymbol{\theta}_n\right).$$

1.1.4 "Axial" rotationally symmetric distributions

We define "axial" rotationally symmetric densities (with respect to the surface area measure on $\mathcal{S}^{p-1}$) as densities of the form

$$\mathbf{x} \mapsto \frac{\Gamma\left(\frac{p-1}{2}\right) \check{c}_{p,\kappa,f}}{2\pi^{(p-1)/2}} f\left(\kappa \left(\mathbf{x}'\boldsymbol{\theta}\right)^2\right), \qquad \mathbf{x} \in \mathcal{S}^{p-1}, \tag{1.12}$$

where

1. $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$ is a location parameter;

2. $\kappa \in \mathbb{R}$ is a concentration parameter: the larger $|\kappa|$, the more the probability mass will be concentrated: symmetrically about the poles $\pm\boldsymbol{\theta}$ for positive values of κ (bipolar case) or symmetrically about the hyperspherical equator $\mathcal{S}^{p-1}\left(\boldsymbol{\theta}^{\perp}\right)$ for negative values of κ (girdle case). The value $\kappa = 0$ (irrespective of f) corresponds to the uniform distribution over the sphere;

3. the function f belongs to the class of functions $\mathscr{F} := \left\{f : \mathbb{R} \to \mathbb{R}^+ : f \text{ monotone increasing, twice differentiable at 0, with } f(0) = f'(0) = 1\right\}$. If $\kappa \neq 0$, then f and the pair $\{\pm\boldsymbol{\theta}\}$ are identifiable, but $\boldsymbol{\theta}$ itself is not (which is natural for axial distributions). If $\kappa = 0$, the location parameter $\boldsymbol{\theta}$ is unidentifiable, even up to a sign.

4. the constant $\check{c}_{p,\kappa,f}$ is such that

$$\check{c}_{p,\kappa,f} \int_{-1}^{1} (1-s^2)^{(p-3)/2} f\left(\kappa s^2\right) ds = 1.$$

Density (1.12) is a symmetric function of $\mathbf{x}$, it attributes the same probability to antipodal regions on the sphere, hence is suitable for axial data. This family of distributions is actually an extension of the Watson distributions (the rotationally symmetric/single-spiked [Bingham, 1974] distributions) obtained by choosing $f(\cdot) = \exp(\cdot)$ in (1.12); see Section 9.4.2. in [Mardia and Jupp, 2000] for more details.

Note that if $\mathbf{X}$ is a random vector with density (1.12), then $\mathbf{X}'\boldsymbol{\theta}$ has density

$$s \mapsto \check{c}_{p,\kappa,f} \left(1-s^2\right)^{(p-3)/2} f\left(\kappa s^2\right) \mathbb{I}_{\{|s|\leq 1\}}.$$

For any $\boldsymbol{\theta}_n$, κ_n and f defined as above, we will denote as $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ the hypothesis under which the $\mathbf{X}_{ni}$'s, $i = 1,\dots,n$, are mutually independent and share the common density

$$\mathbf{x} \mapsto \frac{\Gamma\left(\frac{p_n-1}{2}\right) \check{c}_{p_n,\kappa_n,f}}{2\pi^{(p_n-1)/2}} f\left(\kappa_n \left(\mathbf{x}'\boldsymbol{\theta}_n\right)^2\right).$$

1.2 Le Cam's theory

1.2.1 Convergence of experiments

A **statistical experiment** or **statistical model** ξ is a measurable space $(\Omega, \mathscr{A})$, called the sample space, equipped with a collection of probability measures $\mathscr{P} := \left\{\mathrm{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}$. The notion of convergence for statistical models is challenging and the sequel offers a glimpse. For more details, see Sections 6.1 and 6.2 in [Liese and Miescke, 2008] or Sections 2.1 and 2.2 in [Le Cam and Yang, 2000].

Fix a decision space $\mathscr{D}$ equipped with a σ-field $\mathscr{B}_{\mathscr{D}}$, a loss function $\mathrm{L} : \mathscr{D} \times \Theta \to [0,1] : (d, \boldsymbol{\theta}) \mapsto \mathrm{L}_{\boldsymbol{\theta}}(d)$ and a mapping δ that attributes to every $x \in \Omega$ a probability measure on $(\mathscr{D}, \mathscr{B}_{\mathscr{D}})$. The average loss over all decisions d when a value $x \in \Omega$ is observed is

$$\int_{\mathscr{D}} \mathrm{L}_{\boldsymbol{\theta}}(d) \mathrm{d}\delta_x(d).$$

The risk function associated with δ is then

$$\mathrm{R}^{\delta}(\boldsymbol{\theta}) := \int_{\Omega} \left(\int_{\mathscr{D}} \mathrm{L}_{\boldsymbol{\theta}}(d) \mathrm{d}\delta_x(d)\right) \mathrm{dP}_{\boldsymbol{\theta}}(x)$$

and represents the overall average loss when x is picked according to $\mathrm{P}_{\boldsymbol{\theta}}$. Since one criterion to judge the quality of decision procedures is their risk functions, we are led to study more closely the space of risk functions available for an experiment ξ and a loss function L. Define

$$\mathscr{R}(\mathscr{P}, \mathscr{D}, \mathrm{L}) := \left\{r : \Theta \to [0,1] : \text{there exists } \delta \text{ such that } \mathrm{R}^{\delta}(\boldsymbol{\theta}) \leq r(\boldsymbol{\theta})\ \forall \boldsymbol{\theta}\right\},$$

and its pointwise closure

$$\bar{\mathscr{R}}(\mathscr{P},\mathscr{D},\mathrm{L}) := \left\{ r : \Theta \to [0,1] : r(\boldsymbol{\theta}) = \lim_{i\to\infty} r_i(\boldsymbol{\theta}) \;\forall \boldsymbol{\theta}, \text{ where } r_i \in \mathscr{R}(\mathscr{P},\mathscr{D},\mathrm{L}) \right\}.$$

Definition 1.2.1.
Let $\xi_1 := \left(\Omega_1, \mathscr{A}_1, \mathscr{P}_1 = \left\{\mathrm{P}_{1,\boldsymbol{\theta}} | \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$ and $\xi_2 := \left(\Omega_2, \mathscr{A}_2, \mathscr{P}_2 = \left\{\mathrm{P}_{2,\boldsymbol{\theta}} | \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$. The **deficiency of** ξ_1 **with respect to** ξ_2 is

$$\begin{aligned} \Delta(\xi_1,\xi_2) := \inf\Big\{ & \varepsilon \in [0,1] : \text{for every finite decision space } \mathscr{D}, \forall \mathrm{L} : \mathscr{D}\times\Theta \to [0,1] \text{ and} \\ & \forall r_2 \in \mathscr{R}(\mathscr{P}_2,\mathscr{D},\mathrm{L}), \exists r_1 \in \bar{\mathscr{R}}(\mathscr{P}_1,\mathscr{D},\mathrm{L}) \text{ such that } r_1(\boldsymbol{\theta}) \le r_2(\boldsymbol{\theta}) + \varepsilon \;\forall \boldsymbol{\theta} \in \Theta \Big\}. \end{aligned}$$

In other words, $\Delta(\xi_1,\xi_2)$ is the smallest number such that for every $\mathrm{L} : \mathscr{D}\times\Theta \to [0,1]$, the set $\bar{\mathscr{R}}(\mathscr{P}_1,\mathscr{D},\mathrm{L})$ contains $\{\mathscr{R}(\mathscr{P}_2,\mathscr{D},\mathrm{L}) + \Delta(\xi_1,\xi_2)\} := \{r_2 + \Delta(\xi_1,\xi_2) | r_2 \in \mathscr{R}(\mathscr{P}_2,\mathscr{D},\mathrm{L})\}$.

The **distance** between ξ_1 and ξ_2 is

$$\Delta^{\mathrm{dist}}(\xi_1,\xi_2) := \max\left(\Delta(\xi_1,\xi_2), \Delta(\xi_2,\xi_1)\right).$$

The equality $\Delta^{\mathrm{dist}}(\xi_1,\xi_2) = 0$ does not imply that ξ_1 and ξ_2 are the same but rather are equivalent or of the same type, like two different experiments to measure the gravity of Earth for example. If $\Delta^{\mathrm{dist}}(\xi_1,\xi_2) = \delta$, it means that any risk function you can have on one of the two experiments can be matched within δ by a risk function on the other experiment.

In this definition only loss functions L such that $0 \le \mathrm{L}_{\boldsymbol{\theta}}(d) \le 1$ are allowed. This is because multiplying the loss function by a constant $c \ge 0$ would multiply the distance by the same constant c and it could be made as large as one pleases unless the distance is zero.

We can finally give a meaning to the convergence of statistical models.

Definition 1.2.2.
Let $\xi^{(n)} := \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$ and $\xi := \left(\Omega, \mathscr{A}, \mathscr{P} := \left\{\mathrm{P}_{\boldsymbol{\theta}} | \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$. We say that $(\xi^{(n)})$ converges weakly to ξ as $n \to \infty$ if and only if, for every finite subset $\Theta_0 \subset \Theta$,

$$\Delta^{\mathrm{dist}}\left(\xi^{(n)}_{\Theta_0}, \xi_{\Theta_0}\right) \stackrel{n\to\infty}{\longrightarrow} 0.$$

The restriction to finite subsets in the definition stems from the fact that, in the finite-parameter case, equivalence of convergence in the sense of Le Cam's distance and of convergence of distributions of likelihood ratios has been proven. The infinite-parameter case is much more complex but this weak form of convergence is enough to have important consequences in the sequel.

1.2.2 Local Asymptotic Normality (LAN)

Intuitively, a sequence of statistical models is locally and asymptotically normal if it converges (in the sense of the previous section) to a Gaussian model after a suitable reparametrisation. If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is an iid sample from a distribution $\mathrm{P}_{\boldsymbol{\theta}}$ on $(\Omega, \mathscr{A})$, define

$$\xi^{(n)} := \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \middle| \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right),$$

a sequence of experiments associated with $\mathbf{X}_1, \ldots, \mathbf{X}_n$. If, for $\nu(n) \to 0$ as $n \to \infty$,

$$\xi^{(n)}_{\boldsymbol{\theta}} := \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)}_{\boldsymbol{\theta}} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\nu(n)\boldsymbol{\tau}} \middle| \boldsymbol{\tau} \in \mathbb{R}^p\right\}\right) \to \xi_{\boldsymbol{\theta}} := \left(\Omega, \mathscr{A}, \mathscr{P}_{\boldsymbol{\theta}} := \left\{\mathrm{P}_{\boldsymbol{\theta},\boldsymbol{\tau}} \middle| \boldsymbol{\tau} \in \mathbb{R}^p\right\}\right),$$

the optimal procedure in $\xi_{\boldsymbol{\theta}}$ is asymptotically optimal in $\xi^{(n)}_{\boldsymbol{\theta}}$ hence locally and asymptotically optimal in $\xi^{(n)}$.

Definition 1.2.3. A sequence of experiments $\xi^{(n)}$ is **Locally Asymptotically Normal (LAN)** if for all $\boldsymbol{\theta} \in \Theta$, there exist:

1. a sequence $\boldsymbol{\nu}^{(n)}$ of $p \times p$ full rank, non-random, symmetric and positive definite[1] matrices, called **the contiguity rate**, such that $\|\boldsymbol{\nu}^{(n)}\| \to 0$ as $n \to \infty$, where $\|\boldsymbol{\nu}^{(n)}\|$ denotes the Frobenius norm of $\boldsymbol{\nu}^{(n)}$;

2. a sequence of random p-vectors $\boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)}$ measurable with respect to $\mathbf{X}_1, \dots, \mathbf{X}_n$, called **the central sequence**;

3. a non-random $p \times p$ matrix $\boldsymbol{\Gamma}(\boldsymbol{\theta})$, called **the Fisher information matrix**;

such that for every bounded sequence of vectors $\boldsymbol{\tau}^{(n)} \in \mathbb{R}^p$, we have under $\mathrm{P}_{\boldsymbol{\theta}}^{(n)}$

$$\text{(i) } \log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} = \left(\boldsymbol{\tau}^{(n)}\right)' \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} - \frac{1}{2}\left(\boldsymbol{\tau}^{(n)}\right)' \boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}^{(n)} + o_{\mathrm{P}_{\boldsymbol{\theta}}^{(n)}}(1), \tag{1.13}$$

$$\text{(ii) } \boldsymbol{\Delta}_{\boldsymbol{\theta}}^{(n)} \xrightarrow{\mathscr{D}} \mathcal{N}_p\left(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta})\right).$$

Consider the Gaussian shift experiment

$$\xi_{\boldsymbol{\theta}} := \left(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p), \mathscr{P}_{\boldsymbol{\theta}} := \left\{\mathrm{P}_{\boldsymbol{\theta},\boldsymbol{\tau}} := \mathcal{N}_p\left(\boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\theta})\right) \,\middle|\, \boldsymbol{\tau} \in \mathbb{R}^p\right\}\right)$$

with a single observation that we denote as $\boldsymbol{\Delta}$, where $\mathscr{B}(\mathbb{R}^p)$ is the Borel σ-field on $\mathbb{R}^p$. The log-likelihood ratio in $\xi_{\boldsymbol{\theta}}$ is

$$\log \frac{d\mathrm{P}_{\boldsymbol{\theta},\boldsymbol{\tau}}(\boldsymbol{\Delta})}{d\mathrm{P}_{\boldsymbol{\theta},\mathbf{0}}(\boldsymbol{\Delta})} = \boldsymbol{\tau}'\boldsymbol{\Delta} - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}. \tag{1.14}$$

We see that the log-likelihood ratio in (1.13) behaves asymptotically like the one in (1.14). The similarity is not just a coincidence; define

$$\xi_{\boldsymbol{\theta}}^{(n)} := \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}_{\boldsymbol{\theta}}^{(n)} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}} \,\middle|\, \boldsymbol{\tau} \in \mathbb{R}^p\right\}\right).$$

It can be shown that $\xi_{\boldsymbol{\theta}}^{(n)}$ converges weakly to $\xi_{\boldsymbol{\theta}}$ (see for example Corollary 6.66 in [Liese and Miescke, 2008] or Theorem 9.4 in [van der Vaart, 1998]).

Example 1.2.4 (One-parameter exponential family)**.**

Let $\mathrm{X}_1, \dots, \mathrm{X}_n$ be iid from $(\Omega, \mathscr{A}, \mathscr{P} := \{\mathrm{P}_\theta | \theta \in \Theta_0 \subset \mathbb{R}\})$ such that $\mathscr{P}$ is absolutely continuous with respect to some σ-finite measure μ and

$$f_\theta(x) := \frac{d\mathrm{P}_\theta}{d\mu} = \exp\left(\theta \mathrm{T}(x) - \mathrm{A}(\theta)\right)$$

for all $\theta \in \Theta_0$. Note that $\mathrm{X}^{(n)} := (\mathrm{X}_1, \dots, \mathrm{X}_n)$ comes from $\left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)} := \left\{\mathrm{P}_\theta^{(n)} \,\middle|\, \theta \in \Theta_0 \subset \mathbb{R}\right\}\right)$. The set Θ_0 is the *natural parameter space* defined as

$$\left\{\theta \in \mathbb{R} \text{ such that } \int_\Omega p_\theta(x)\, d\mu(x) < \infty\right\}.$$

Let θ_0 be a point inside Θ_0. Then

$$\log \frac{d\mathrm{P}^{(n)}_{\theta_0+n^{-1/2}\tau}}{d\mathrm{P}^{(n)}_{\theta_0}} = \log \prod_{i=1}^n \frac{f_{\theta_0+n^{-1/2}\tau}(\mathrm{X}_i)}{f_{\theta_0}(\mathrm{X}_i)} = \frac{\tau}{\sqrt{n}} \sum_{i=1}^n \mathrm{T}(\mathrm{X}_i) - n\left(\mathrm{A}\left(\theta_0 + \frac{\tau}{\sqrt{n}}\right) - \mathrm{A}(\theta_0)\right). \tag{1.15}$$

[1]This condition is not necessary but is added for convenience ; see [Le Cam and Yang, 2000]

It can be shown (see for example Problem 2.16 in [Lehmann and Romano, 2005]) that

$$\begin{aligned} E_{\theta_0}[T(X_i)] &= A'(\theta_0), \\ \mathrm{Var}_{\theta_0}[T(X_i)] &= A''(\theta_0), \end{aligned}$$

so that, by a Taylor expansion, as $n \to \infty$,

$$n\left(A\left(\theta_0 + \frac{\tau}{\sqrt{n}}\right) - A(\theta_0)\right) = \sqrt{n}\tau A'(\theta_0) + \frac{1}{2}\tau^2 A''(\theta_0) + o(1).$$

The log-likelihood ratio (1.15) becomes

$$\log \frac{dP^{(n)}_{\theta_0+n^{-1/2}\tau}}{dP^{(n)}_{\theta_0}} = \tau\Delta^{(n)}_{\theta_0} - \frac{1}{2}\tau^2 A''(\theta_0) + o(1), \tag{1.16}$$

where $A''(\theta_0)$ is the Fisher information matrix and where, under θ_0, the central sequence $\Delta^{(n)}_{\theta_0}$ is such that

$$\Delta^{(n)}_{\theta_0} := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(T(X_i) - E_{\theta_0}[T(X_i)]\right) \xrightarrow{\mathscr{D}} \mathscr{N}(0, A''(\theta_0)),$$

as $n \to \infty$.

An approximate result like (1.16) can actually be obtained for much more general families like iid observations, either by setting conditions on the second derivative of the log-likelihood or under the single condition of differentiability in quadratic mean of $f^{1/2}_{\boldsymbol{\theta}}$ at $\boldsymbol{\theta}$ (see Section 7.2 of [van der Vaart, 1998] or Section 12.2 in [Lehmann and Romano, 2005]).

1.2.3 Contiguity

The concept of contiguity was introduced by [Le Cam, 1960] as an equivalent to "asymptotic absolute continuity". Contiguity implies a strong relation between the asymptotic behaviour of a sequence of statistics under laws $Q^{(n)}$ (typically an alternative hypothesis) and its asymptotic behaviour under laws $P^{(n)}$ (typically a null distribution) and enables to find the former from the latter.

Definition 1.2.5. Let $(\Omega_n, \mathscr{A}_n)$ be measurable spaces equipped with two sequences of probability distributions $P^{(n)}$ and $Q^{(n)}$.

The sequence $Q^{(n)}$ is **contiguous** to $P^{(n)}$, denoted as $Q^{(n)} \lhd P^{(n)}$, if $P^{(n)}(A_n) \to 0$ implies $Q^{(n)}(A_n) \to 0$ for every sequence of measurable sets $A_n \in \mathscr{A}_n$. If $Q^{(n)} \lhd P^{(n)}$ and $P^{(n)} \lhd Q^{(n)}$, we say that $Q^{(n)}$ and $P^{(n)}$ are **mutually contiguous**: $Q^{(n)} \rhd\lhd P^{(n)}$.

To get an idea of what contiguity is, consider the problem of testing

$$\begin{cases} \mathscr{H}^{(n)}_0 : \{P^{(n)}\} \\ \mathscr{H}^{(n)}_1 : \{Q^{(n)}\} \end{cases}$$

and suppose $P^{(n)}$ and $Q^{(n)}$ are mutually contiguous. Consider a sequence of nonrandomized tests[2] $\phi^{(n)}$. Then

$$P^{(n)}[\phi^{(n)} = 1] \to 0 \Leftrightarrow Q^{(n)}[\phi^{(n)} = 1] \to 0$$

[2]A nonrandomized test is a test statistic that can have only two values: 0 (the null hypothesis is not rejected) and 1 (the null hypothesis is rejected).

and

$$\mathrm{Q}^{(n)}[\phi^{(n)}=1]\to 1 \Leftrightarrow \mathrm{P}^{(n)}[\phi^{(n)}=1]\to 1.$$

If the size of the tests converges to zero, then the power of the tests converges to zero too. Conversely if the power of the tests converges to one, then the size of the tests converges to one too. The sequences of probability distributions $\mathrm{P}^{(n)}$ and $\mathrm{Q}^{(n)}$ are so close that no test has a power converging to one.

1.2.4 Le Cam's First Lemma

Contiguity can be proven thanks to the following result known as Le Cam's first lemma (see Lemma 6.4 in [van der Vaart, 1998]).

Lemma 1.2.6 (Le Cam's First Lemma)**.** *Let $(\Omega_n, \mathscr{A}_n)$ be measurable spaces equipped with two sequences of probability distributions $\mathrm{P}^{(n)}$ and $\mathrm{Q}^{(n)}$. Then the following statements are equivalent :*

(i) $\mathrm{Q}^{(n)} \lhd \mathrm{P}^{(n)}$.

(ii) If $d\mathrm{P}^{(n)}/d\mathrm{Q}^{(n)} \xrightarrow{\mathscr{D}} \mathrm{U}$ under $\mathrm{Q}^{(n)}$ along a subsequence, then $\mathrm{P}[\mathrm{U}>0]=1$.

(iii) If $d\mathrm{Q}^{(n)}/d\mathrm{P}^{(n)} \xrightarrow{\mathscr{D}} \mathrm{V}$ under $\mathrm{P}^{(n)}$ along a subsequence, then $\mathrm{E}[\mathrm{V}]=1$.

(iv) For any statistics $\mathrm{T}_n : \Omega_n \mapsto \mathbb{R}^k$: if $\mathrm{T}_n \to 0$ under $\mathrm{P}^{(n)}$, then $\mathrm{T}_n \to 0$ under $\mathrm{Q}^{(n)}$.

To prove mutual contiguity, one therefore needs to prove that if $d\mathrm{Q}^{(n)}/d\mathrm{P}^{(n)} \xrightarrow{\mathscr{D}} \mathrm{U}$ under $\mathrm{P}^{(n)}$ along a subsequence, then $\mathrm{P}[\mathrm{U}>0]=1$ and $\mathrm{E}[\mathrm{U}]=1$. For example, if $\mathrm{P}^{(n)}$ and $\mathrm{Q}^{(n)}$ are two probability measures on arbitrarily measurable spaces such that, under $\mathrm{P}^{(n)}$,

$$\frac{d\mathrm{Q}^{(n)}}{d\mathrm{P}^{(n)}} \xrightarrow{\mathscr{D}} e^{\mathscr{N}(\mu,\sigma^2)},$$

then $\mathrm{Q}^{(n)}$ and $\mathrm{P}^{(n)}$ are mutually contiguous if and only if $\mu=-\sigma^2/2$ (like $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)}}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ when $\xi^{(n)} = \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)} = \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \,\middle|\, \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$ is Locally Asymptotically Normal).

1.2.5 Le Cam's Third Lemma

This section allows computing the asymptotic distribution of a statistic $\mathbf{X}_n$ under $\mathrm{Q}^{(n)}$ from its asymptotic distribution under $\mathrm{P}^{(n)}$ (see Lemma 6.6 in [van der Vaart, 1998]).

Lemma 1.2.7. *Let $\mathrm{P}^{(n)}$ and $\mathrm{Q}^{(n)}$ be sequences of probability measures on probability spaces $(\Omega_n, \mathscr{A}_n)$ and let $\mathbf{X}_n : \Omega_n \mapsto \mathbb{R}^k$ be a sequence of random vectors. Suppose that $\mathrm{Q}^{(n)} \lhd \mathrm{P}^{(n)}$ and*

$$\left(\mathbf{X}_n, \frac{d\mathrm{Q}^{(n)}}{d\mathrm{P}^{(n)}}\right) \xrightarrow{\mathscr{D}} (\mathbf{X}, \mathrm{V})$$

under $\mathrm{P}^{(n)}$. Then $\mathrm{L}(\mathrm{B}) := \mathrm{E}\left[\mathbb{I}_{\{\mathbf{X}\in\mathrm{B}\}}\mathrm{V}\right]$ defines a probability measure, and $\mathbf{X}_n \xrightarrow{\mathscr{D}} \mathrm{L}$ under $\mathrm{Q}^{(n)}$.

The following lemma is a famous special case that will be useful for computing the asymptotic power of tests under sequences of experiments that are locally and asymptotically normal (see Example 6.7 in [van der Vaart, 1998]).

Lemma 1.2.8 (Le Cam's Third Lemma)**.** *If*

$$\begin{pmatrix} \mathbf{X}_n \\ \log \frac{d\mathrm{Q}^{(n)}}{d\mathrm{P}^{(n)}} \end{pmatrix} \xrightarrow{\mathscr{D}} \mathscr{N}_{k+1}\left(\begin{pmatrix} \boldsymbol{\mu} \\ -\sigma^2/2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{c} \\ \mathbf{c}' & \sigma^2 \end{pmatrix}\right)$$

under $\mathrm{P}^{(n)}$*, then* $\mathrm{Q}^{(n)} \triangleleft \mathrm{P}^{(n)}$ *and* $\mathbf{X}_n \xrightarrow{\mathscr{D}} \mathscr{N}_k\left(\boldsymbol{\mu}+\mathbf{c}, \boldsymbol{\Sigma}\right)$ *under* $\mathrm{Q}^{(n)}$.

Example 1.2.9. Consider a sequence of experiments $\xi^{(n)} := \left(\Omega^{(n)}, \mathscr{A}^{(n)}, \mathscr{P}^{(n)} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}} \middle| \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}\right)$ that is LAN, so

$$\log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}} = \boldsymbol{\tau}'\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau} + o_{\mathrm{P}^{(n)}_{\boldsymbol{\theta}}}(1),$$

where $\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} \xrightarrow{\mathscr{D}} \mathscr{N}_p\left(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta})\right)$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}}$ as $n \to \infty$. We can apply Le Cam's Third Lemma and since

$$\mathrm{Cov}_{\mathrm{P}^{(n)}_{\boldsymbol{\theta}}}\left(\log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}}}{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}}}, \boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}\right) = \mathrm{E}_{\mathrm{P}^{(n)}_{\boldsymbol{\theta}}}\left[\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}\left(\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}}\right)'\right]\boldsymbol{\tau} \to \boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau},$$

as $n \to \infty$, we obtain that

$$\boldsymbol{\Delta}^{(n)}_{\boldsymbol{\theta}} \xrightarrow{\mathscr{D}} \mathscr{N}_p\left(\boldsymbol{\Gamma}(\boldsymbol{\theta})\boldsymbol{\tau}, \boldsymbol{\Gamma}(\boldsymbol{\theta})\right)$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}+\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}}$ as $n \to \infty$.

1.3 Invariance

Let $(\Omega, \mathscr{A}, \mathscr{P})$ be a statistical model and X a sample from $\mathrm{P} \in \mathscr{P}$. Let G be a group of transformations acting on Ω. We say that $(\Omega, \mathscr{A}, \mathscr{P})$ is **invariant** under G if for all $\mathrm{P} \in \mathscr{P}$ and for all $g \in \mathrm{G}$, there is $\mathrm{Q} \in \mathscr{P}$ such that

$$\mathrm{P}^{g\mathrm{X}} = \mathrm{Q}^{\mathrm{X}},$$

where P^{X} is the measure induced by X defined by $\mathrm{P}^{\mathrm{X}}(\mathrm{B}) = \mathrm{P}\left(\mathrm{X} \in \mathrm{B}\right)$. Each transformation g induces a mapping

$$\bar{g} : \mathscr{P} \to \mathscr{P} : \mathrm{P}^{\mathrm{X}} \mapsto \mathrm{P}^{g\mathrm{X}}.$$

In the parametric case, $\mathscr{P} = \left\{\mathrm{P}_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p\right\}$, so $\bar{g} : \Theta \to \Theta : \boldsymbol{\theta} \mapsto \bar{g}\boldsymbol{\theta}$. The set of all induced transformations, $\bar{\mathrm{G}}$, is also a group.

Let $\{\mathrm{H}_0, \mathrm{H}_1\}$ be a partition of $\mathscr{P}$. We say that the testing problem

$$\begin{cases} \mathrm{H}_0 \\ \mathrm{H}_1 \end{cases} \tag{1.17}$$

is **invariant** under G if $(\Omega, \mathscr{A}, \mathrm{H}_0)$ is invariant under G. Note that if $(\Omega, \mathscr{A}, \mathscr{P})$ and (1.17) are invariant under G, then so is $(\Omega, \mathscr{A}, \mathrm{H}_1)$. According to the *invariance principle* it is then natural to restrict to **invariant tests**, that is tests ϕ such that

$$\phi\left(gx\right) = \phi(x)$$

for all $g \in \mathrm{G}$ and all $x \in \Omega$. Any invariant test can be written as a function of a **maximal invariant** test T, that is such that, for some g in G,

$$x' = gx \Leftrightarrow \mathrm{T}(x') = \mathrm{T}(x),$$

for all $x, x' \in \Omega$. Moreover, if G is a generating group (i.e. $\bar{\mathrm{G}}$ has only one orbit) then all invariant tests are distribution-free. More details on invariance can be found in Section 6.3 in [Shao, 2003] or in Chapter 6 in [Lehmann and Romano, 2005].

Now, suppose we have an invariant testing problem; we will then restrict to invariant tests that depend on the observations only through a maximal invariant statistic, $\mathrm{T} : (\Omega, \mathscr{A}) \to (\Omega_\mathrm{T}, \mathscr{A}_\mathrm{T})$ say. We may then apply a *reduction by invariance* and switch to the reduced model

$$\left(\Omega_\mathrm{T}, \mathscr{A}_\mathrm{T}, \mathrm{P}^\mathrm{T}\right) \tag{1.18}$$

where it is often easier to find an optimal test.

Example 1.3.1 (for more details, see Section 5.2 in [Liese and Miescke, 2008])**.** Suppose we want to test

$$\begin{cases} \mathrm{H}_0 : \boldsymbol{\mu} = \mathbf{0} \\ \mathrm{H}_1 : \boldsymbol{\mu} \neq \mathbf{0} \end{cases}$$

in the model

$$\mathscr{M} := \left(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p), \mathscr{P} := \left\{\mathrm{P}_{\boldsymbol{\mu}} := \mathscr{N}_p\left(\boldsymbol{\mu}, \mathbf{I}_p\right) \,\middle|\, \boldsymbol{\mu} \in \mathbb{R}^p\right\}\right).$$

This testing problem is multisided and a most powerful test does not in general exist. However, as the testing problem and the model are invariant under the orthogonal group $\mathrm{O}(p)$, we will focus on statistics invariant with respect to that group. We can show that any such statistic can be written as a function of

$$\mathrm{T}(\mathbf{x}) := \|\mathbf{x}\|^2.$$

The measure induced by T is

$$\mathrm{P}^\mathrm{T}_{\boldsymbol{\mu}}(\mathrm{A}) = \mathrm{P}_{\boldsymbol{\mu}}\left(\|\mathbf{X}\|^2 \in \mathrm{A}\right) = \mathrm{P}_{\boldsymbol{\mu}}\left(\sum_{i=1}^p \mathrm{X}_i^2 \in \mathrm{A}\right) = \mathrm{P}\left(\chi^2_p\left(\|\boldsymbol{\mu}\|^2\right) \in \mathrm{A}\right),$$

where $\chi^2_p(\lambda)$ is the non-central chi-squared distribution with p degrees of freedom and noncentrality parameter λ. The reduced model is then

$$\mathscr{M}' := \left(\mathbb{R}^+, \mathscr{B}(\mathbb{R}^+), \mathscr{P}^\mathrm{T} := \left\{\chi^2_p(\delta^2) \,\middle|\, \delta \in \mathbb{R}\right\}\right).$$

It can be shown that if T is an observation from $\mathscr{M}'$, the test that rejects H_0 when

$$\mathrm{T} > \chi^2_{p,1-\alpha},$$

where $\chi^2_{d,\alpha}$ the α-quantile of the chi-square distribution with d degrees of freedom, is uniformly most powerful among the class of level-α tests for

$$\begin{cases} \mathrm{H}_0 : \delta^2 = 0, \\ \mathrm{H}_1 : \delta^2 > 0. \end{cases}$$

Therefore, if $\mathbf{X}$ is an observation from $\mathscr{M}$, the test that rejects H_0 when

$$\|\mathbf{X}\|^2 > \chi^2_{p,1-\alpha} \tag{1.19}$$

is uniformly most powerful invariant among the class of level-α invariant tests for

$$\begin{cases} \mathrm{H}_0 : \boldsymbol{\mu} = \mathbf{0}, \\ \mathrm{H}_1 : \|\boldsymbol{\mu}\|^2 > 0. \end{cases}$$

This test has another nice optimality property. We say that a test ϕ^* is **maximin** for (1.17) within a class $\mathscr{C}$ if its minimal power is maximal in $\mathscr{C}$, i.e.

1. $\phi^* \in \mathscr{C}$,

2. for all tests $\phi \in \mathscr{C}$, $\inf_{P \in H_1} E_P[\phi^*] \geq \inf_{P \in H_1} E_P[\phi]$.

We can show that the test in (1.19) is maximin for

$$\begin{cases} H_0 : \boldsymbol{\mu} = \mathbf{0}, \\ H_{1c} : \|\boldsymbol{\mu}\|^2 > c. \end{cases}$$

in the class of level-α tests, irrespective of $c > 0$.

We conclude this section by a lemma (Lemma 2.5.1 in [Giri, 1996]) that enables us to find the likelihood function in model (1.18).

Lemma 1.3.2. *Let* G *be a group of transformations,* λ *an invariant* [3] *probability measure on* $(\Omega, \mathscr{A})$ *and* X *a random variable on* $(\Omega, \mathscr{A})$ *with density* p *with respect to* λ.

Let λ^* *be the measure induced* [4] *by a maximal invariant* $T : (\Omega, \mathscr{A}) \to (\Omega_T, \mathscr{A}_T)$ *with respect to some group of transformations* G.

The density function of the maximal invariant T *with respect to* λ^* *is given by*

$$p^*(y) = \int p(gx)\, d\mu(g),$$

where x *is any point in* Ω *for which* $T(x) = y$ *and where* μ *is an invariant probability measure on* G.

1.4 Billingsley's Theorem

We introduce in this section a theorem that will be used in Chapter 4 to prove the asymptotic normality of the Rayleigh and Bingham test statistics under general rotationally symmetric distributions. Let us first recall that if $Z_1, Z_2, \ldots$ is a sequence of random variables on a probability space $(\Omega, \mathscr{F}, P)$ and $\mathscr{F}_1, \mathscr{F}_2, \ldots$ is a sequence of σ-fields in $\mathscr{F}$, the sequence $((Z_n, \mathscr{F}_n))$ is a **martingale** if these four conditions hold:

(i) $\mathscr{F}_n \subset \mathscr{F}_{n+1}$ (i.e., the $\mathscr{F}_n$ form a filtration);

(ii) Z_n is measurable with respect to $\mathscr{F}_n$;

(iii) $E[|Z_n|] < \infty$;

(iv) $E[Z_{n+1} | \mathscr{F}_n] = Z_n$ with probability one.

Consider triangular arrays of observations Z_{ni}, $i = 1, \ldots, n$, $n = 1, 2, \ldots$; suppose that for each n, $Z_{n1}, Z_{n2}, \ldots, Z_{nn}$ is a martingale with respect to $\mathscr{F}_{n1}, \mathscr{F}_{n2}, \ldots, \mathscr{F}_{nn}$. Let us define $D_{n\ell} := Z_{n\ell} - Z_{n,\ell-1}$ and suppose the $D_{n\ell}$'s have finite second moments.

Theorem 1.4.1 (Theorem 35.12 in [Billingsley, 1995]).
Letting $\sigma^2_{n\ell} := E\left[D^2_{n\ell} | \mathscr{F}_{n,\ell-1}\right]$ *(with* $\mathscr{F}_{n0} = \{\emptyset, \Omega\}$ *for all* n*), assume that, as* $n \to \infty$,

$$\sum_{\ell=1}^{n} \sigma^2_{n\ell} = 1 + o_P(1), \tag{1.20}$$

$$\sum_{\ell=1}^{n} E\left[D^2_{n\ell} \mathbb{I}_{\{|D_{n\ell}| > \varepsilon\}}\right] \to 0, \tag{1.21}$$

for every $\varepsilon > 0$. *Then* $\sum_{\ell=1}^{n} D_{n\ell}$ *is asymptotically standard normal.*

[3] $\lambda(gA) = \lambda(A)$ for all $g \in G$, $A \in \mathscr{A}$.

[4] $\lambda^*(C) = \lambda\left(T^{-1}(C)\right)$ for all C in $\mathscr{A}_T$.

Condition (1.20) will typically be proven by showing that, as $n \to \infty$,

$$\sum_{\ell=1}^{n} \mathrm{E}\left[\sigma_{n\ell}^2\right] \to 1 \text{ and } \mathrm{Var}\left[\sum_{\ell=1}^{n} \sigma_{n\ell}^2\right] \to 0,$$

so that $\sum_{\ell=1}^{n} \sigma_{n\ell}^2 \to 1$ in the L^2-norm, and hence in probability.

Chapter 2

Testing uniformity against a semiparametric extension of the FVML distributions

Contents

2.1 Introduction

Let $\mathbf{X}_{ni}$, $i=1,\ldots,n$, $n=1,2,\ldots$, be a triangular array of observations such that, for any n, the $\mathbf{X}_{ni}$'s form a random sample on $\mathcal{S}^{p_n-1}$. Perhaps the simplest test of uniformity is the [Rayleigh, 1919] test, that rejects the null of uniformity for large values of

$$\mathrm{R}_n := np_n\|\bar{\mathbf{X}}_n\|^2,$$

where $\bar{\mathbf{X}}_n := n^{-1}\sum_{i=1}^n \mathbf{X}_{ni}$. For fixed p, $\mathrm{R}_n \xrightarrow{\mathcal{D}} \chi^2_p$. In the high-dimensional setup, Theorem 2.1 in [Paindaveine and Verdebout, 2016] implies that, under the null of uniformity,

$$\mathrm{R}_n^{\mathrm{St}} := \frac{\mathrm{R}_n - p_n}{\sqrt{2p_n}} = \frac{\sqrt{2p_n}}{n}\sum_{1\le i<j\le n}\mathbf{X}_{ni}'\mathbf{X}_{nj} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) \tag{2.1}$$

as soon as n and p_n diverge to infinity, where $\xrightarrow{\mathcal{D}}$ denotes weak convergence. The high-dimensional Rayleigh test, say $\phi^{(n)}_{\mathrm{Ray}}$, then rejects uniformity at asymptotic level α whenever

$$\mathrm{R}_n^{\mathrm{St}} > z_\alpha, \quad \text{with } z_\alpha := \Phi^{-1}(1-\alpha). \tag{2.2}$$

Remarkably, this test does not impose any condition on the way p_n goes to infinity with n, hence can be applied as soon as n and p_n are large, without bothering about their relative magnitude. It is nonetheless necessary to study its nonnull behaviour as the trivial test, that would discard the data and reject uniformity with probability α, also has asymptotic level α, yet has a power function uniformly equal to α.

In order to do so, we will focus in this chapter on specific alternatives to uniformity: monotone rotationally symmetric distributions. This family of alternatives is a natural extension of the well-known FvML distributions (see Section 1.1.3). Assume that the triangular array $\mathbf{X}_{ni}$, $i=1,\ldots,n$, $n=1,2,\ldots$, is such that, for any n, $\mathbf{X}_{n1},\mathbf{X}_{n2},\ldots,\mathbf{X}_{nn}$ are mutually independent and identically distributed. Recall that $\mathrm{P}_0^{(n)}$ is the hypothesis that they are uniformly distributed on $\mathcal{S}^{p_n-1}$ and $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ the hypothesis under which they share the common density

$$\mathbf{x}\mapsto \frac{\Gamma\left(\frac{p_n-1}{2}\right)c_{p_n,\kappa_n,f}}{2\pi^{(p_n-1)/2}} f\left(\kappa_n\mathbf{x}'\boldsymbol{\theta}_n\right).$$

The larger the concentration parameter κ_n is, the more severe the deviations from the null of uniformity are, which is obtained as κ_n goes to zero. It is then natural to wonder how small κ_n must be to make $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $\mathrm{P}_0^{(n)}$ mutually contiguous (see Section 1.2.3). This question is answered in the following theorem which covers both the low- and high-dimensional cases.

Theorem 2.1.1. *Let (p_n) be a sequence in $\{2,3,\ldots\}$, $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n\in\mathcal{S}^{p_n-1}$ for all n, (κ_n) a positive sequence such that $\kappa_n^2 = \mathrm{O}(p_n/n)$, and assume that f is twice differentiable at 0. Then, the sequence of alternative hypotheses $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and the null sequence $\mathrm{P}_0^{(n)}$ are mutually contiguous.*

Sketch of the proof. The complete proof of this theorem can be found in Section 2.5.2.

We start by proving a lemma that gives an expansion of $c_{p_n}\int_{-1}^1(1-s^2)^{(p_n-3)/2}g(\kappa_n s)\,ds$ for positive sequences κ_n such that $\kappa_n = o(\sqrt{p_n})$ as $n\to\infty$ and any $g:\mathbb{R}\to\mathbb{R}$ twice differentiable at 0.

Then we split the log-likelihood ratio $\Lambda_n := \log dP^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}/dP^{(n)}_0$ in two terms and letting $E_{n1} := E\left[\log f\left(\kappa_n \mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\right]$ and $V_n := \text{Var}\left[\log f\left(\kappa_n \mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\right]$ we expand them thanks to the lemma to get

$$\begin{aligned}\Lambda_n &= -\frac{n\kappa_n^2}{2p_n} + \sqrt{\frac{n\kappa_n^2}{p_n} + o\left(\frac{n\kappa_n^2}{p_n}\right)}\sum_{i=1}^n \frac{\log f\left(\kappa_n \mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) - E_{n1}}{\sqrt{nV_n}} + o\left(\frac{n\kappa_n^2}{p_n}\right)\\ &=: -\frac{n\kappa_n^2}{2p_n} + \sqrt{\frac{n\kappa_n^2}{p_n} + o\left(\frac{n\kappa_n^2}{p_n}\right)}\sum_{i=1}^n W_{ni} + o\left(\frac{n\kappa_n^2}{p_n}\right). \end{aligned} \tag{2.3}$$

If $\kappa_n^2 = o(p_n/n)$, then $\Lambda_n \xrightarrow{L^2} 0$ under $P^{(n)}_0$ so $\exp(\Lambda_n) \xrightarrow{\mathscr{D}} 1$ under $P^{(n)}_0$, and Le Cam's First Lemma (see Section 1.2.6) yields that $P^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $P^{(n)}_0$ are mutually contiguous.

If $\kappa_n^2 = \tau_n^2 p_n/n$, where the positive sequence (τ_n) is $O(1)$ but not $o(1)$, (2.3) can be written as

$$\Lambda_n = -\frac{\tau_n^2}{2} + \sqrt{\tau_n^2 + o(1)}\sum_{i=1}^n W_{ni} + o(1). \tag{2.4}$$

We show that $\sum_{i=1}^n W_{ni}$ satisfies the Lévy–Lindeberg condition, hence is asymptotically standard normal. Therefore for any subsequence $(\exp(\Lambda_{n_m}))$ converging in distribution, we have under $P^{(n)}_0$ that

$$\exp(\Lambda_{n_m}) \xrightarrow{\mathscr{D}} \exp(Y), \quad \text{with } Y \sim \mathcal{N}\left(-\frac{1}{2}\lim_{m\to\infty}\tau^2_{n_m}, \lim_{m\to\infty}\tau^2_{n_m}\right)$$

and mutual contiguity ensues from Le Cam's First Lemma. □

In the low-dimensional case, the usual parametric rate $\kappa_n \sim 1/\sqrt{n}$ provides contiguous alternatives, which implies that, irrespective of f, there exist no consistent tests for

$$\begin{cases}\mathscr{H}^{(n)}_0 : \left\{P^{(n)}_0\right\}\\ \mathscr{H}^{(n)}_1 : \left\{P^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\right\}\end{cases} \tag{2.5}$$

if $\kappa_n = \tau/\sqrt{n}$, $\tau > 0$. The high-dimensional case is more interesting. First, we stress that the contiguity result in Theorem 2.1.1 does not impose conditions on p_n, hence in particular applies when (a) $p_n/n \to c$ for some $c > 0$ or (b) $p_n/n \to \infty$. Interestingly, the result shows that contiguity in cases (a)-(b) can be achieved for sequences (κ_n) that do not converge to zero: a constant sequence (κ_n) ensures contiguity in case (a), whereas contiguity in case (b) may even be obtained for a sequence (κ_n) that diverges to infinity in a suitable way. In both cases, there then exist no consistent tests for (2.5), despite the fact that the sequences (κ_n) are not $o(1)$. This may be puzzling at first since such sequences are expected to lead to severe alternatives to uniformity; it actually makes sense, however, that the fast increase of the dimension p_n, despite the favourable sequences (κ_n), makes the problem difficult enough to prevent the existence of consistent tests.

In the remainder of this chapter we will start by assuming that $\boldsymbol{\theta}_n$ is known and address the testing problem

$$\begin{cases}\mathscr{H}^{(n)}_0 : \left\{P^{(n)}_0\right\}\\ \mathscr{H}^{(n)}_1 : \cup_{\kappa>0}\cup_f \left\{P^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\}.\end{cases} \tag{2.6}$$

An optimal test is given in Section 2.2 thanks to a Local Asymptotic Normality result. When $\boldsymbol{\theta}_n$ is not specified, the testing problem becomes in low dimensions

$$\begin{cases}\mathscr{H}^{(n)}_0 : \left\{P^{(n)}_0\right\}\\ \mathscr{H}^{(n)}_1 : \cup_{\boldsymbol{\theta}\in\mathcal{S}^{p-1}}\cup_{\kappa>0}\cup_f \left\{P^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}\end{cases} \tag{2.7}$$

and Section 2.3.1 shows that the Rayleigh test becomes the optimal test against general monotone rotationally symmetric distributions. In high dimensions an invariance approach yields in Section 2.3.2 that it is optimal against FvML distributions.

2.2 Optimal testing under specified modal location

When the modal location $\boldsymbol{\theta}_n$ is specified, optimal tests of uniformity can be obtained from the following Local Asymptotic Normality (LAN) result. To the best of our knowledge, this result provides the first instance of a LAN structure in high dimensions.

Theorem 2.2.1. *Let (p_n) be a sequence in $\{2,3,\ldots\}$ and $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for all n. Let $\kappa_n = \tau_n\sqrt{p_n/n}$, where the positive sequence (τ_n) is $O(1)$ but not $o(1)$, and assume that f is twice differentiable at 0. Then, as $n\to\infty$, under $\mathrm{P}_0^{(n)}$,*

$$\Lambda_n := \log\frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}}{d\mathrm{P}_0^{(n)}} = \tau_n\Delta^{(n)}_{\boldsymbol{\theta}_n} - \frac{\tau_n^2}{2} + o_{\mathrm{P}}(1), \tag{2.8}$$

where $\Delta^{(n)}_{\boldsymbol{\theta}_n} := \sqrt{np_n}\,\bar{\mathbf{X}}_n'\boldsymbol{\theta}_n$ is asymptotically standard normal.

In other words, the sequence $\left(\left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f} : \kappa\geq 0\right\}\right)$ (where $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,0,f} := \mathrm{P}^{(n)}_0$ for any $\boldsymbol{\theta}_n$ and f) is locally asymptotically normal at $\kappa=0$ with central sequence $\Delta^{(n)}_{\boldsymbol{\theta}_n}$, Fisher information 1, and contiguity rate $\sqrt{p_n/n}$.

Sketch of the proof. The complete proof of this theorem can be found in Section 2.5.3.

The fact that $\Delta^{(n)}_{\boldsymbol{\theta}_n}$ is asymptotically standard normal follows from the central limit theorem. To prove (2.8) we rewrite (2.4) as

$$\Lambda_n = -\frac{\tau_n^2}{2} + \tau_n\sum_{i=1}^{n} W_{ni} + o_{\mathrm{P}}(1)$$

and show that $\tau_n\left[\sum_{i=1}^n W_{ni} - \Delta^{(n)}_{\boldsymbol{\theta}_n}\right]$ converges to zero in quadratic mean and therefore in probability.

□

This result, that covers both the low- and high-dimensional cases, reveals that the rate $\kappa_n \sim \sqrt{p_n/n}$ in Theorem 2.2.1 is actually the contiguity rate of the considered model (that is, more severe alternatives are not contiguous to the null of uniformity). In low dimensions, the usual parametric contiguity rate $\kappa_n \sim 1/\sqrt{n}$ is obtained. In high dimensions it is non-standard. Yet in the FvML high-dimensional case, this rate may be related to the fact that, as $p\to\infty$, one needs to consider $\kappa_p \sim \sqrt{p}$ to obtain FvML p-vectors that provide non-degenerate weak limiting results that are different from those obtained from p-vectors that are uniform over the sphere (more precisely, if $\mathbf{X}$ has density (1.11) with $\kappa = \sqrt{p}\omega$, then, as $p\to\infty$, $\sqrt{p}\boldsymbol{\theta}'\mathbf{X}$ converges weakly to the normal distribution with mean ω and variance 1; see [Watson, 1988] for more details). The contiguity rate $\kappa_n \sim \sqrt{p_n/n}$ then intuitively results from a standard $1/\sqrt{n}$-shrinkage starting from this non-trivial $\kappa_p \sim \sqrt{p}$ high-dimensional situation.

Now, assume that $\kappa_n = \tau\sqrt{p_n/n}$. By Theorem 2.2.1, as $n\to\infty$,

$$\mathrm{Cov}_{\mathrm{P}_0^{(n)}}\left[\Delta^{(n)}_{\boldsymbol{\theta}_n}, \Lambda_n\right] = \tau\mathrm{Var}_{\mathrm{P}_0^{(n)}}\left[\Delta^{(n)}_{\boldsymbol{\theta}_n}\right] = \tau$$

Therefore, by Le Cam's Third Lemma (see Lemma 1.2.8), $\Delta^{(n)}_{\boldsymbol{\theta}_n}$ is asymptotically normal with mean τ and variance one under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$. Define for some fixed $\boldsymbol{\theta}_n$ and fixed f

$$\xi^{(n)} := \left(\left(\mathcal{S}^{p_n-1}\right)^{(n)}, \mathscr{A}^{(n)}, \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f} \,\middle|\, \kappa > 0\right\}\right),$$

where $\mathscr{A}^{(n)}$ is the trace sigma-algebra of $\left(\mathcal{S}^{p_n-1}\right)^{(n)}$ in $\mathscr{B}\left(\mathbb{R}^{p_n} \times \ldots \times \mathbb{R}^{p_n}\right)$. Local asymptotic normality implies that

$$\xi_0^{(n)} := \left(\left(\mathcal{S}^{p_n-1}\right)^{(n)}, \mathscr{A}^{(n)}, \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\tau\sqrt{p_n/n},f} \,\middle|\, \tau > 0\right\}\right)$$

converges to

$$\xi := \left(\mathbb{R}, \mathscr{B}(\mathbb{R}), \left\{\mathcal{N}(\tau,1) \,\middle|\, \tau > 0\right\}\right).$$

The testing problem

$$\begin{cases} \mathscr{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\} \\ \mathscr{H}_{1,f}^{(n)} : \cup_{\kappa>0}\left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\} \end{cases} \tag{2.9}$$

can then be locally rewritten as

$$\begin{cases} \mathscr{H}_0^{(n)} : \tau = 0 \\ \mathscr{H}_0^{(n)} : \tau > 0. \end{cases}$$

Let Δ stand for an observation from ξ. From Theorem 3.4.1. in [Lehmann and Romano, 2005] we know that a uniformly most powerful test for

$$\begin{cases} \mathscr{H}_0 : \tau = 0 \\ \mathscr{H}_0 : \tau > 0. \end{cases}$$

within the class of level-α tests rejects $\mathscr{H}_0$ at level α when

$$\Delta > z_\alpha.$$

Consequently, the test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ rejecting $\mathscr{H}_0^{(n)}$ at asymptotic level α whenever

$$\Delta^{(n)}_{\boldsymbol{\theta}_n} = \sqrt{np_n}\,\bar{\mathbf{X}}_n'\boldsymbol{\theta}_n > z_\alpha \tag{2.10}$$

is locally asymptotically most powerful for (2.9). As it does not depend on the function f, it is also locally asymptotically most powerful for (2.6). Its asymptotic power under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n = \tau\sqrt{p_n/n}$, is

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\Delta^{(n)}_{\boldsymbol{\theta}_n} > z_\alpha\right] = 1 - \Phi(z_\alpha - \tau). \tag{2.11}$$

While all results of this section so far covered both the low- and high-dimensional cases, we need to treat these cases separately to investigate how the Rayleigh test compares with the optimal test $\phi^{(n)}_{\boldsymbol{\theta}_n}$.

We start with the low-dimensional case. Le Cam's Third Lemma allows to show that, under the contiguous alternatives $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n = \tau\sqrt{p/n}$,

$$\mathrm{R}_n \xrightarrow{\mathscr{D}} \chi^2_p(\tau^2) \tag{2.12}$$

as $n\to\infty$ (see proof in Appendix A). Denoting by $\Psi_\ell(\cdot;\lambda)$ the cumulative distribution function of the $\chi^2_\ell(\lambda)$ distribution, the corresponding asymptotic power of the Rayleigh test is therefore

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{R}_n > \chi^2_{p,1-\alpha}\right] = 1-\Psi_p\left(\chi^2_{p,1-\alpha};\tau^2\right), \tag{2.13}$$

which is strictly smaller than the asymptotic power in (2.11). We conclude that, in the specified-$\boldsymbol{\theta}_n$ case, the low-dimensional Rayleigh test is not locally asymptotically most powerful yet shows non-trivial asymptotic powers against contiguous alternatives.

The story is different in the high-dimensional case, as can be guessed from the following heuristic reasoning. In view of (2.12), we have that, as $n\to\infty$ under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n=\tau\sqrt{p/n}$,

$$\mathrm{R}^{\mathrm{St}}_n = \frac{\mathrm{R}_n-p}{\sqrt{2p}} \xrightarrow{\mathcal{D}} \frac{\chi^2_1(\tau^2)-1}{\sqrt{2p}} + \frac{\chi^2_{p-1}-(p-1)}{\sqrt{2p}},$$

where the two chi-square variables are independent. When both n and p are large, it is therefore expected that, under the same sequence of alternatives,

$$\mathrm{R}^{\mathrm{St}}_n \approx \mathcal{N}\left(\frac{\tau^2}{\sqrt{2p}}, 1+\frac{2\tau^2}{p}\right),$$

where $\mathrm{Z}_n\approx\mathcal{L}$ means that the distribution of Z_n is close to $\mathcal{L}$. Thus, in the high-dimensional case (where $p=p_n\to\infty$), $\mathrm{R}^{\mathrm{St}}_n$ is expected to be standard normal under these alternatives, which would imply that the high-dimensional Rayleigh test in (2.2) has asymptotic powers equal to the nominal level α.

The high-dimensional LAN result in Theorem 2.2.1 allows to confirm these heuristics. Letting $\kappa_n=\tau_n\sqrt{p_n/n}$, where τ_n is $\mathrm{O}(1)$, Theorem 2.2.1 readily yields that, as $n\to\infty$,

$$\begin{aligned}
\mathrm{Cov}_{\mathrm{P}^{(n)}_0}\left[\mathrm{R}^{\mathrm{St}}_n,\Lambda_n\right] &= \mathrm{Cov}_{\mathrm{P}^{(n)}_0}\left[\mathrm{R}^{\mathrm{St}}_n,\Delta^{(n)}_{\boldsymbol{\theta}_n}\right]\tau_n + o(1)\\
&= \frac{\sqrt{2}p_n}{n^{3/2}}\tau_n\sum_{i=1}^n\sum_{1\le k<\ell\le n}\mathrm{E}_{\mathrm{P}^{(n)}_0}\left[\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\left(\mathbf{X}'_{nk}\mathbf{X}_{n\ell}\right)\right]+o(1)\\
&= o(1),
\end{aligned}$$

so that Le Cam's Third Lemma implies that $\mathrm{R}^{\mathrm{St}}_n$ remains asymptotically standard normal under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$. This confirms that, unlike in the low-dimensional case, the high-dimensional Rayleigh test does not show any power under the contiguous alternatives from Theorem 2.2.1.

However, the high-dimensional Rayleigh test does not make use of the specified value of the modal location $\boldsymbol{\theta}_n$, hence does not primarily address the specified-$\boldsymbol{\theta}_n$ problem but rather the unspecified-$\boldsymbol{\theta}_n$ one. Therefore, the key question is whether or not the Rayleigh test is optimal for the latter problem. We answer this question in the next section.

2.3 Optimal testing under unspecified modal location

Building on the results of the previous section, two natural approaches, that may lead to an optimal test for the unspecified-$\boldsymbol{\theta}_n$ problem, are the following. The first one consists in substituting an estimator $\hat{\boldsymbol{\theta}}_n$ for $\boldsymbol{\theta}_n$ in the optimal test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ above. For the *spherical mean* $\hat{\boldsymbol{\theta}}_n=\bar{\mathbf{X}}_n/\|\bar{\mathbf{X}}_n\|$ (which is the Maximum Likelihood Estimator, MLE, for $\boldsymbol{\theta}_n$ in the FvML case), the resulting test rejects the null for large values of

$$\Delta^{(n)}_{\hat{\boldsymbol{\theta}}_n} = \sqrt{np_n}\,\bar{\mathbf{X}}'_n\hat{\boldsymbol{\theta}}_n = \sqrt{np_n}\,\|\bar{\mathbf{X}}_n\| = \mathrm{R}^{1/2}_n,$$

hence coincides with the Rayleigh test. The second approach, in the spirit of [Davies, 1977, Davies, 1987, Davies, 2002], rather consists in adopting the test statistic

$$\sup_{\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}} \Delta^{(n)}_{\boldsymbol{\theta}_n} = \sqrt{np_n}\,\|\bar{\mathbf{X}}_n\|,$$

which again leads to the Rayleigh test. These considerations suggest that the Rayleigh test may indeed be optimal for the unspecified-$\boldsymbol{\theta}_n$ problem. In this section, we investigate whether this is the case or not, both in low and high dimensions.

2.3.1 The low-dimensional case

To investigate the optimality properties of the low-dimensional Rayleigh test for the unspecified-$\boldsymbol{\theta}_n$ problem, it is helpful to adopt a new parametrisation. For fixed p and f, the model is indexed by $(\boldsymbol{\theta}, \kappa) \in \mathcal{S}^{p-1} \times \mathbb{R}^+$, where the value $\kappa = 0$ makes $\boldsymbol{\theta}$ unidentified (for fixed p, the dimension of $\boldsymbol{\theta}$ does not depend on n, so that there is no need to consider sequences $(\boldsymbol{\theta}_n)$). We then consider the alternative parametrisation in $\boldsymbol{\mu} := \kappa\boldsymbol{\theta}$, for which the fixed-$p$ result in Theorem 2.2.1 readily rewrites as follows.

Theorem 2.3.1. *Fix an integer $p \geq 2$ and let $\boldsymbol{\mu}_n = \sqrt{p/n}\,\boldsymbol{\tau}_n$ for all n, where the sequence $(\boldsymbol{\tau}_n)$ in $\mathbb{R}^p$ is $O(1)$ but not $o(1)$. Assume that f is twice differentiable at 0. For any $\boldsymbol{\mu} \in \mathbb{R}^p \setminus \{\mathbf{0}\}$, let $\mathrm{P}^{(n)}_{\boldsymbol{\mu},f} := \mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa,f}$, where $\boldsymbol{\mu} =: \kappa\boldsymbol{\theta}$, with $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$. Then, under $\mathrm{P}^{(n)}_0$,*

$$\log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\mu}_n,f}}{d\mathrm{P}^{(n)}_0} = \boldsymbol{\tau}'_n \boldsymbol{\Delta}^{(n)} - \frac{1}{2}\|\boldsymbol{\tau}_n\|^2 + o_{\mathrm{P}}(1),$$

as $n \to \infty$, where $\boldsymbol{\Delta}^{(n)} := \sqrt{np}\,\bar{\mathbf{X}}_n$ is asymptotically standard p-variate normal.

Similarly to the specified-$\boldsymbol{\theta}_n$ case, if $\boldsymbol{\mu}_n = \sqrt{p/n}\,\boldsymbol{\tau}$, by Theorem 2.3.1 and Le Cam's Third Lemma, $\boldsymbol{\Delta}^{(n)}$ is, under $\mathrm{P}^{(n)}_{\boldsymbol{\mu}_n,f}$, asymptotically normal with mean $\boldsymbol{\tau}$ and covariance matrix $\mathbf{I}_p$. Define

$$\boldsymbol{\xi}^{(n)} := \left((\mathcal{S}^{p-1})^{(n)}, \mathscr{A}^{(n)}, \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\mu},f} \,\middle|\, \boldsymbol{\mu} \in \mathbb{R}^p \right\} \right).$$

As

$$\boldsymbol{\xi}^{(n)}_{\mathbf{0}} := \left((\mathcal{S}^{p-1})^{(n)}, \mathscr{A}^{(n)}, \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\tau}\sqrt{p/n},f} \,\middle|\, \boldsymbol{\tau} \in \mathbb{R}^p \right\} \right) \to \boldsymbol{\xi} := \left(\mathbb{R}^p, \mathscr{B}(\mathbb{R}^p), \left\{ \mathcal{N}_p(\boldsymbol{\tau}, \mathbf{I}_p) \,\middle|\, \boldsymbol{\tau} \in \mathbb{R}^p \right\} \right),$$

for some fixed f, the testing problem

$$\begin{cases} \mathcal{H}^{(n)}_0 : \left\{ \mathrm{P}^{(n)}_0 \right\} \\ \mathcal{H}^{(n)}_{1,f} : \cup_{\boldsymbol{\mu} \in \mathbb{R}^p \setminus \{\mathbf{0}\}} \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\mu},f} \right\}, \end{cases} \tag{2.14}$$

which is equivalent to

$$\begin{cases} \mathcal{H}^{(n)}_0 : \left\{ \mathrm{P}^{(n)}_0 \right\} \\ \mathcal{H}^{(n)}_{1,f} : \cup_{\boldsymbol{\theta} \in \mathcal{S}^{p-1}} \cup_{\kappa > 0} \left\{ \mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa,f} \right\} \end{cases}$$

in the new parametrisation, can then be locally rewritten as

$$\begin{cases} \mathcal{H}^{(n)}_0 : \boldsymbol{\tau} = \mathbf{0} \\ \mathcal{H}^{(n)}_1 : \boldsymbol{\tau} \neq \mathbf{0}. \end{cases}$$

Let $\boldsymbol{\Delta}$ stand for an observation from $\boldsymbol{\xi}$. Example 1.3.1 implies that the test that rejects $\mathscr{H}_0$ at level α when

$$\|\boldsymbol{\Delta}\|^2 > \chi^2_{p,1-\alpha}$$

is maximin within the class of level-α tests for

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\tau} = \mathbf{0} \\ \mathscr{H}_{1c} : \|\boldsymbol{\tau}\|^2 > c \end{cases}$$

irrespective of $c > 0$, and uniformly most powerful invariant among the class of level-α invariant tests with respect to the orthogonal group $\mathrm{O}(p)$ for

$$\begin{cases} \mathscr{H}_0 : \boldsymbol{\tau} = \mathbf{0}, \\ \tilde{\mathscr{H}}_1 : \|\boldsymbol{\tau}\|^2 > 0. \end{cases} .$$

Consequently, the test rejecting $\mathscr{H}_0^{(n)}$ at asymptotic level α whenever

$$\left\|\boldsymbol{\Delta}^{(n)}\right\|^2 = np\left\|\bar{\mathbf{X}}_n\right\|^2 > \chi^2_{p,1-\alpha},$$

that is the low-dimensional Rayleigh test, is locally asymptotically maximin and locally asymptotically most powerful invariant for (2.14). As it does not depend on the function f, it is also locally asymptotically maximin and locally asymptotically most powerful invariant for (2.7). This new optimality property of the low-dimensional Rayleigh test complements the one stating that this test is locally most powerful invariant in the FvML model (see, e.g., [Chikuse, 2003], Section 6.3.5.).

The specified-$\boldsymbol{\theta}_n$ and unspecified-$\boldsymbol{\theta}_n$ testing problems are two distinct statistical problems, that, even in the low-dimensional case, provide different efficiency bounds. In low dimensions, the Rayleigh test is optimal for the unspecified-$\boldsymbol{\theta}_n$ problem, but not for the specified-$\boldsymbol{\theta}_n$ one. This thoroughly describes the optimality properties of this test in the low-dimensional case, so that we may now focus on the high-dimensional case.

2.3.2 The high-dimensional case

If p_n goes to infinity, then the dimension of the parameter $(\boldsymbol{\theta}_n, \kappa)$ increases with n, so that there cannot be a high-dimensional analogue of the LAN result in Theorem 2.3.1. We therefore rather adopt, in the present hypothesis testing context, an *invariance* approach that is close in spirit to the one used by [Moreira, 2009] in a point estimation context.

The null of uniformity and all collections of alternatives $\mathscr{P}^{(n)}_{\kappa,f} := \left\{\mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa,f} : \boldsymbol{\theta} \in \mathcal{S}^{p_n-1}\right\}$ are invariant (see Section 1.3) under the group of rotations

$$\mathscr{G}^{(n)} := \left\{g^{(n)}_{\mathbf{O}} : \mathbf{O} \in \mathrm{SO}(p_n)\right\},$$

where $g^{(n)}_{\mathbf{O}}(\mathbf{x}_1, \ldots, \mathbf{x}_n) = (\mathbf{O}\mathbf{x}_1, \ldots, \mathbf{O}\mathbf{x}_n)$ for any $(\mathbf{x}_1, \ldots, \mathbf{x}_n) \in \mathcal{S}^{p_n-1} \times \ldots \times \mathcal{S}^{p_n-1}$ (n times) and where $\mathrm{SO}(p_n)$ stands for the collection of $p_n \times p_n$ orthogonal matrices with determinant one. The problem of testing uniformity against rotationally symmetric alternatives is hence also invariant under $\mathscr{G}^{(n)}$. The invariance principle then suggests restricting to $\mathscr{G}^{(n)}$-invariant tests, that automatically are distribution-free under any $\mathscr{P}^{(n)}_{\kappa,f}$ (their distribution does not depend on $\boldsymbol{\theta}$). Indeed, $\mathscr{G}^{(n)}$ is a generating group and the set of induced transformations

$$\left\{\bar{g}_{\mathbf{O}} : \mathbf{O} \in \mathrm{SO}(p_n)\right\},$$

where $\bar{g}_{\mathbf{O}}(\boldsymbol{\theta}) = \mathbf{O}\boldsymbol{\theta}$, has only one orbit. Therefore, if $\mathbf{S}_n$ is an invariant statistic with respect to $\mathscr{G}^{(n)}$, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{S}^{p_n-1}$, there is $\bar{g}$ such that $\bar{g}_{\mathbf{O}}(\boldsymbol{\theta}_1) = \boldsymbol{\theta}_2$, so that

$$\begin{aligned}\mathrm{P}_{\boldsymbol{\theta}_2}\left(\mathbf{S}_n(\mathbf{x}_1,\ldots,\mathbf{x}_n) \in \mathrm{A}\right) = \mathrm{P}_{\bar{g}_{\mathbf{O}}(\boldsymbol{\theta}_1)}\left(\mathbf{S}_n(\mathbf{x}_1,\ldots,\mathbf{x}_n) \in \mathrm{A}\right) &= \mathrm{P}_{\boldsymbol{\theta}_1}\left(\mathbf{S}_n(\mathbf{O}\mathbf{x}_1,\ldots,\mathbf{O}\mathbf{x}_n) \in \mathrm{A}\right)\\ &= \mathrm{P}_{\boldsymbol{\theta}_1}\left(\mathbf{S}_n(\mathbf{x}_1,\ldots,\mathbf{x}_n) \in \mathrm{A}\right)\end{aligned}$$

for some $\mathrm{A} \in \mathscr{A}^{(n)}$.

As mentioned in Section 1.3, optimal invariant tests are to be determined in the image of the original model by a maximal invariant $\mathbf{T}_n$ with respect to $\mathscr{G}^{(n)}$. The likelihood (with respect to the surface area measure m_{p_n} on $\mathcal{S}^{p_n-1}$) associated with the image of $\mathscr{P}^{(n)}_{\kappa_n,f}$ by $\mathbf{T}_n$ is given by

$$\frac{d\mathrm{P}^{(n)\mathbf{T}_n}_{\kappa_n,f}}{dm_{p_n}} = \int_{\mathrm{SO}(p_n)} \prod_{i=1}^{n} \left\{ \frac{\Gamma\left(\frac{p_n-1}{2}\right) c_{p_n,\kappa_n,f}}{2\pi^{(p_n-1)/2}} f\left(\kappa_n (\mathbf{O}\mathbf{X}_{ni})'\boldsymbol{\theta}_n\right)\right\} d\mathbf{O},$$

where the integral is with respect to the Haar measure on $\mathrm{SO}(p_n)$; see Lemma 1.3.2. The resulting log-likelihood ratio to the null of uniformity is therefore

$$\begin{aligned}\Lambda^{\mathbf{T}_n}_{n,f} := \log \frac{d\mathrm{P}^{(n)\mathbf{T}_n}_{\kappa_n,f}}{d\mathrm{P}^{(n)}_0} &= \log \frac{c^n_{p_n,\kappa_n,f} \int_{\mathrm{SO}(p_n)} \prod_{i=1}^n f\left(\kappa_n \mathbf{X}'_{ni}\left(\mathbf{O}'\boldsymbol{\theta}_n\right)\right) d\mathbf{O}}{c^n_{p_n}}\\ &= \log \frac{c^n_{p_n,\kappa_n,f} \mathrm{E}\left[\prod_{i=1}^n f\left(\kappa_n \mathbf{X}'_{ni}\mathbf{U}\right) | \mathbf{X}_{n1},\ldots,\mathbf{X}_{nn}\right]}{c^n_{p_n}},\end{aligned} \tag{2.15}$$

where $\mathbf{U}$ is uniformly distributed over $\mathcal{S}^{p_n-1}$ and is independent of the $\mathbf{X}_{ni}$'s. The following theorem shows that, in the FvML case $f(\cdot) = f_{\mathrm{FvML}}(\cdot) = \exp(\cdot)$, this collection of log-likelihood ratios enjoys the LAN property.

Theorem 2.3.2. *Let (p_n) be a sequence of positive integers diverging to ∞ as $n \to \infty$ and let $\kappa_n = \tau_n p_n^{3/4}/\sqrt{n}$, where the positive sequence (τ_n) is $\mathrm{O}(1)$ but not $o(1)$.*

Then we have, under $\mathrm{P}^{(n)}_0$, that

$$\log \frac{d\mathrm{P}^{(n)\mathbf{T}_n}_{\kappa_n,\exp}}{d\mathrm{P}^{(n)}_0} = \tau_n^2 \Delta^{(n)\mathbf{T}_n} - \frac{\tau_n^4}{4} + o_{\mathrm{P}}(1), \tag{2.16}$$

as $n \to \infty$, where $\Delta^{(n)\mathbf{T}_n} := \mathrm{R}^{\mathrm{St}}_n/\sqrt{2}$ is asymptotically normal with mean zero and variance $1/2$.

Sketch of the proof. The complete proof of this theorem can be found in Section 2.5.4.

We already know from (2.1) that, under $\mathrm{P}^{(n)}_0$, $\Delta^{(n)\mathbf{T}_n}$ is asymptotically normal with mean zero and variance $1/2$. To prove (2.16), the FvML version of the log-likelihood in (2.15) is rewritten as $\Lambda^{\mathbf{T}_n}_{n,\exp} = \mathrm{L}_{n1} + \mathrm{L}_{n2}$, where

$$\begin{aligned}\mathrm{L}_{n1} &:= n \log \frac{c^{\mathrm{FvML}}_{p_n,\kappa_n}}{c_{p_n}} = -n \log \mathrm{H}_{p_n/2-1}(\kappa_n)\\ \mathrm{L}_{n2} &:= \log \mathrm{E}\left[\exp\left(\kappa_n n \bar{\mathbf{X}}'_n \mathbf{U}\right) | \bar{\mathbf{X}}_n\right] = \log \frac{c_{p_n}}{c^{\mathrm{FvML}}_{p_n, n\kappa_n \|\bar{\mathbf{X}}_n\|}} = \log \mathrm{H}_{p_n/2-1}\left(\kappa_n n \|\bar{\mathbf{X}}_n\|\right),\end{aligned}$$

with $\mathrm{H}_\nu(x) := \Gamma(\nu+1)\mathscr{I}_\nu(x)(x/2)^{-\nu}$. We then use bounds for $\log \mathrm{H}_\nu(x)$ that can be expanded if $\kappa_n = \tau_n p_n^{3/4}/\sqrt{n}$ (with $n, p_n \to \infty$ and (τ_n) bounded) and we obtain expressions of L_{n1} and L_{n2} under $\mathrm{P}_0^{(n)}$ that lead to

$$\Lambda_{n,\exp}^{\mathbf{T}_n} = \tau_n^2 \frac{\mathrm{R}_n^{\mathrm{St}}}{\sqrt{2}} - \frac{\tau_n^4 p_n}{4(p_n+2)} + o_\mathrm{P}(1).$$

□

Applying Le Cam's Third Lemma, we obtain that, as $n \to \infty$ under $\mathrm{P}_{\kappa_n,\exp}^{(n)\mathbf{T}_n}$,

$$\Delta^{(n)\mathbf{T}_n} \xrightarrow{\mathscr{D}} \mathscr{N}\left(\Gamma\tau^2, \Gamma\right)$$

with $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$ and $\Gamma = 1/2$. The model $\left\{\mathrm{P}_{\kappa,\exp}^{(n)\mathbf{T}_n} : \kappa \geq 0\right\}$ (where $\mathrm{P}_{0,\exp}^{(n)\mathbf{T}_n} := \mathrm{P}_0^{(n)}$) is thus "second-order" LAN, in the sense that the mean of the limiting Gaussian shift experiment is quadratic (rather than linear) in τ. Clearly, this does not change the form of locally asymptotically optimal tests, but only their asymptotic performances. Note that the contiguity rate $\kappa_n \sim p_n^{3/4}/\sqrt{n}$ associated with this new LAN property differs from the contiguity rate $\kappa_n \sim \sqrt{p_n/n}$ in Theorem 2.2.1.

Theorem 2.3.2 entails that the test rejecting the null of uniformity at asymptotic level α whenever

$$\Delta^{(n)\mathbf{T}_n}/\sqrt{\Gamma} = \mathrm{R}_n^{\mathrm{St}} > z_\alpha,$$

that is, the high-dimensional Rayleigh test in (2.2), is, in the FvML case, locally asymptotically most powerful invariant. This optimality result is of a high-dimensional asymptotic nature and also covers cases where κ_n does not converge to 0, hence does not follow from the aforementioned local optimality result from [Chikuse, 2003]. Le Cam's Third Lemma readily implies that as $n \to \infty$, under $\mathrm{P}_{\kappa_n,\exp}^{(n)\mathbf{T}_n}$,

$$\mathrm{R}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathscr{N}\left(\frac{\tau^2}{\sqrt{2}}, 1\right),$$

with $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$, so that the corresponding asymptotic power of the Rayleigh test is given by

$$\lim_{n\to\infty} \mathrm{P}_{\boldsymbol{\theta}_n,\kappa_n,\exp}^{(n)}\left[\mathrm{R}_n^{\mathrm{St}} > z_\alpha\right] = 1 - \Phi\left(z_\alpha - \tfrac{\tau^2}{\sqrt{2}}\right), \tag{2.17}$$

where the sequence $(\boldsymbol{\theta}_n)$ is such that $\boldsymbol{\theta}_n \in \mathrm{S}^{p_n-1}$ for all n but is otherwise arbitrary. While the Rayleigh test is blind to alternatives in $\kappa_n \sim \sqrt{p_n/n}$, it thus detects alternatives in $\kappa_n \sim p_n^{3/4}/\sqrt{n}$, which, in view of Theorem 2.3.2, is the best that can be achieved for the unspecified-$\boldsymbol{\theta}_n$ problem.

Interestingly, we might have guessed that these alternatives in $\kappa_n \sim p_n^{3/4}/\sqrt{n}$ are those that can be detected by the high-dimensional Rayleigh test. Recall indeed that heuristic arguments in Section 2.2 suggested that, under $\mathrm{P}_{\boldsymbol{\theta}_n,\kappa_n,f}^{(n)}$, with $\kappa_n = \tau\sqrt{p/n}$,

$$\mathrm{R}_n^{\mathrm{St}} \approx \mathscr{N}\left(\frac{\tau^2}{\sqrt{2p}}, 1 + \frac{2\tau^2}{p}\right)$$

for large n and p. Consequently, to obtain, in high dimensions, an asymptotic non-null distribution that differs from the limiting null (standard normal) one, we need to consider alternatives of the form $\mathrm{P}_{\boldsymbol{\theta}_n,\kappa_n,f}^{(n)}$, with $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$, under which we expect

$$\mathrm{R}_n^{\mathrm{St}} \approx \mathscr{N}\left(\frac{\tau^2}{\sqrt{2}}, 1\right)$$

for large n and p. This is fully in line with the non-null distribution and local asymptotic powers obtained from Le Cam's Third Lemma in the previous paragraph.

Provided that f is four times differentiable at 0 and that $p_n = o(n^2)$, tedious computations allow to show that a fourth-order expansion of the f-based log-likelihood ratio $\Lambda^{\mathbf{T}_n}_{n,f}$ above, still based on $\kappa_n = \tau_n p_n^{3/4}/\sqrt{n}$, exactly provides the righthand side of (2.16), with the same central sequence $\Delta^{(n)\mathbf{T}_n}$. However, turning this into a proper f-based version of Theorem 2.3.2 requires controlling the corresponding (fifth-order) remainder term, which proved to be extremely difficult. Yet we conjecture that Theorem 2.3.2 indeed extends to an arbitrary f admitting five derivatives at 0, under the aforementioned assumption that $p_n = o(n^2)$ (an assumption that is superfluous in the FvML case, since Theorem 2.3.2 allows p_n to go to infinity in an arbitrary way as a function of n). Proving this conjecture would establish that the Rayleigh test is locally asymptotically most powerful invariant under any such f, with the same asymptotic powers as in (2.17).

2.4 Simulations

In this section, we present the results of a Monte Carlo study we conducted to check the validity of our asymptotic results. We generated independent random samples of the form

$$\mathbf{X}^{(\ell)}_{i;j} \quad i = 1, \ldots, n, \quad j = 1, 2, \quad \ell = 0, 1, 2, 3, 4. \tag{2.18}$$

For $\ell = 0$, the common distribution of the $\mathbf{X}^{(\ell)}_{i;j}$'s is the uniform distribution on the unit sphere $\mathcal{S}^{p-1}$, while, for $\ell > 0$, the $\mathbf{X}^{(\ell)}_{i;j}$'s have an FvML distribution on $\mathcal{S}^{p-1}$ with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa^{(\ell)}_j$, with

$$\kappa^{(\ell)}_1 = 0.6\ell\sqrt{\frac{p}{n}} \quad \text{and} \quad \kappa^{(\ell)}_2 = 0.6\ell\,\frac{p^{3/4}}{\sqrt{n}}.$$

The case $j = 1$ relates to the contiguous alternatives (see Theorem 2.2.1), whereas $j = 2$ is associated with the alternatives under which the Rayleigh test shows non-trivial asymptotic powers in the high-dimensional setup (see Theorem 2.3.2).

For any $(n, p) \in \mathcal{C} \times \mathcal{C}$, with $\mathcal{C} := \{30, 100, 400\}$, any $j \in \{1, 2\}$, and any $\ell \in \{0, 1, 2, 3, 4\}$, we generated $M = 2{,}500$ independent random samples $\mathbf{X}^{(\ell)}_{i;j}$, $i = 1, \ldots, n$, as described above, and evaluated the rejection frequencies of

(i) the specified-$\boldsymbol{\theta}_n$ test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ in (2.10),

(ii) the high-dimensional Rayleigh test $\phi^{(n)}_{\mathrm{Ray}}$ in (2.2),

the two conducted at nominal level 5%. Rejection frequencies are plotted in Figure 2.1 as well as the corresponding asymptotic powers, obtained from (2.11), (2.17), and the fact that $\phi^{(n)}_{\boldsymbol{\theta}_n}$ is consistent against $(j = 2)$-alternatives.

Clearly, rejection frequencies match extremely well the corresponding asymptotic powers, irrespective of the tests and types of alternatives considered (the only possible exception is the test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ under $(\ell = 1, j = 2)$-alternatives; this, however, is only a consequence of the lack of continuity of the corresponding asymptotic power curves). Remarkably, this agreement is also reasonably good for moderate sample size n and dimension p. Beyond validating our asymptotic results, this Monte Carlo study therefore also shows that these results are relevant for practical values of n and p.

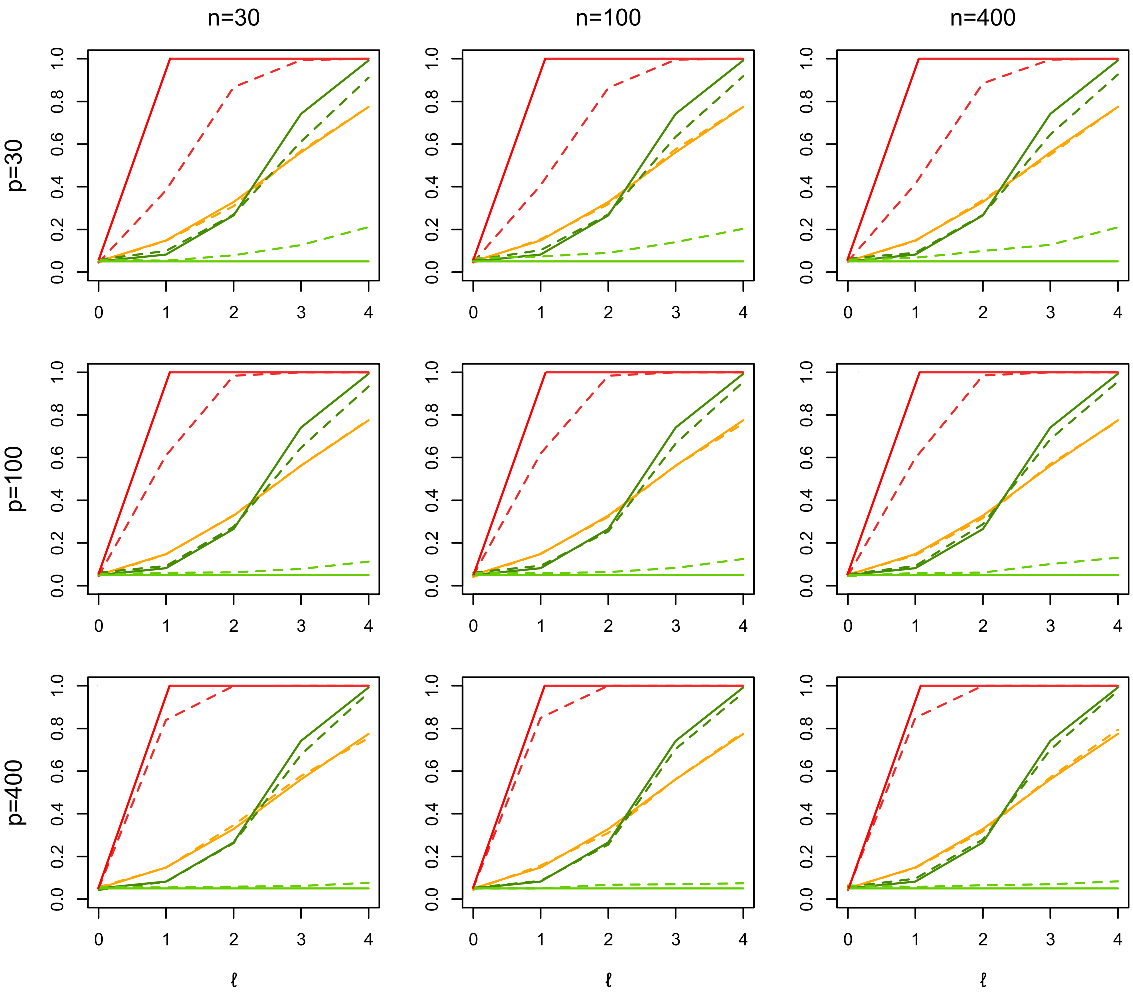


Figure 2.1 – Rejection frequencies (dashed) and asymptotic powers (solid), under the null of uniformity over the p-dimensional unit sphere ($\ell = 0$) and increasingly severe FvML alternatives ($\ell = 1, 2, 3, 4$), of the specified-$\boldsymbol{\theta}_n$ test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ in (2.10) (red/orange) and the high-dimensional Rayleigh test $\phi^{(n)}_{\text{Ray}}$ in (2.2) (light/dark green). Light colors (orange and light green) are associated with contiguous alternatives, whereas dark colors (red and dark green) correspond to the more severe alternatives under which the Rayleigh test shows non-trivial asymptotic powers in high dimensions; see Section 2.4 for details.

2.5 Proofs

2.5.1 Preliminary lemma

Lemma 2.5.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be twice differentiable at 0. Let κ_n be a positive sequence that is $o(\sqrt{p_n})$ as $n \to \infty$. Then* $\mathrm{R}_n(g) := c_{p_n} \int_{-1}^{1} (1-s^2)^{(p_n-3)/2} g(\kappa_n s)\, ds = g(0) + \frac{\kappa_n^2}{2p_n} g''(0) + o\left(\frac{\kappa_n^2}{p_n}\right)$.

Proof of Lemma 2.5.1. We know from (1.5) that

$$c_{p_n} \int_{-1}^{1} s^2 (1-s^2)^{(p_n-3)/2}\, ds = \frac{1}{p_n}. \tag{2.19}$$

Now, write

$$\mathrm{R}_n(g) = g(0) + c_{p_n} \int_{-1}^{1} (1-s^2)^{(p_n-3)/2} \big(g(\kappa_n s) - g(0) - \kappa_n s g'(0)\big)\, ds.$$

Letting $t = \kappa_n s$ and using the identity (2.19) then provides

$$\mathrm{R}_n(g) = g(0) + \frac{\kappa_n^2}{p_n} \int_{-\kappa_n}^{\kappa_n} h_n(t) \left(\frac{g(t) - g(0) - t g'(0)}{t^2} \right) dt,$$

where h_n is defined through

$$t \mapsto h_n(t) = \frac{(t/\kappa_n)^2 \big(1 - (t/\kappa_n)^2\big)^{(p_n-3)/2}}{\int_{-\kappa_n}^{\kappa_n} (s/\kappa_n)^2 \big(1-(s/\kappa_n)^2\big)^{(p_n-3)/2}\, ds} \mathbb{I}_{\{|t| \le \kappa_n\}}.$$

It can be checked that, since $\kappa_n = o(\sqrt{p_n})$, the h_n's form an *approximate* δ*-sequence* (see (1.157) in [Arfken et al., 2013]), in the sense that $\int_{-\infty}^{\infty} h_n(t)\, dt = 1$ for all n and $\int_{-\varepsilon}^{\varepsilon} h_n(t)\, dt \to 1$ for any $\varepsilon > 0$. Hence,

$$\mathrm{R}_n(g) = g(0) + \frac{\kappa_n^2}{p_n} \lim_{t \to 0} \left(\frac{g(t) - g(0) - t g'(0)}{t^2} \right) + o\left(\frac{\kappa_n^2}{p_n} \right)$$

which, by using L'Hôpital's rule, yields the result. □

2.5.2 Proof of Theorem 2.1.1

In this proof, all expectations and variances are taken under the null of uniformity $\mathrm{P}_0^{(n)}$ and all stochastic convergences and o_{P}'s are as $n \to \infty$ under $\mathrm{P}_0^{(n)}$. Consider then the local log-likelihood ratio

$$\begin{aligned}
\Lambda_n &:= \log \frac{d\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \kappa_n, f}}{d\mathrm{P}_0^{(n)}} = \sum_{i=1}^{n} \log \frac{c_{p_n,\kappa_n,f}\, f\left(\kappa_n \mathbf{X}_{ni}' \boldsymbol{\theta}_n\right)}{c_{p_n}} \\
&= n \left(\log \frac{c_{p_n,\kappa_n,f}}{c_{p_n}} + \mathrm{E}_{n1} \right) + \sum_{i=1}^{n} \left(\log f\left(\kappa_n \mathbf{X}_{ni}' \boldsymbol{\theta}_n\right) - \mathrm{E}_{n1} \right) \\
&=: \mathrm{L}_{n1} + \mathrm{L}_{n2}.
\end{aligned}$$

Throughout, we write $\ell_{f,k} := (\log f)^k$ and $\mathrm{E}_{nk} := \mathrm{E}\left[\ell_{f,k}(\kappa_n \mathbf{X}_{ni}' \boldsymbol{\theta}_n)\right]$ (E_{nk} actually depends on κ_n, p_n and f, but we simply write E_{nk} to avoid a heavy notation).

Lemma 2.5.1 readily yields

$$n\log\frac{c_{p_n,\kappa_n,f}}{c_{p_n}} = -n\log\left(c_{p_n}\int_{-1}^{1}(1-s^2)^{(p_n-3)/2}f(\kappa_n s)\,ds\right) = -n\log\left(1+\frac{\kappa_n^2}{2p_n}f''(0)+o\left(\frac{\kappa_n^2}{p_n}\right)\right)$$
$$= -\frac{n\kappa_n^2}{2p_n}f''(0)+o\left(\frac{n\kappa_n^2}{p_n}\right). \tag{2.20}$$

Similarly, for any positive integer k,

$$\mathrm{E}_{nk} = c_{p_n}\int_{-1}^{1}(1-s^2)^{(p_n-3)/2}\ell_{f,k}(\kappa_n s)\,ds = \frac{\kappa_n^2}{2p_n}\ell''_{f,k}(0)+o\left(\frac{\kappa_n^2}{p_n}\right). \tag{2.21}$$

Combining (2.20) and (2.21), and using the identity $\ell''_{f,1}(0) = f''(0)-1$ readily yields

$$\mathrm{L}_{n1} = \frac{n\kappa_n^2}{2p_n}\left(-f''(0)+\ell''_{f,1}(0)\right)+o\left(\frac{n\kappa_n^2}{p_n}\right) = -\frac{n\kappa_n^2}{2p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right).$$

Turning to L_{n2}, write

$$\mathrm{L}_{n2} = \sqrt{n\mathrm{V}_n}\sum_{i=1}^{n}\mathrm{W}_{ni} := \sqrt{n\mathrm{V}_n}\sum_{i=1}^{n}\frac{\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)-\mathrm{E}_{n1}}{\sqrt{n\mathrm{V}_n}},$$

where we let $\mathrm{V}_n := \mathrm{Var}\left[\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\right]$. First note that (2.21) provides

$$n\mathrm{V}_n = n\left(\mathrm{E}_{n2}-\mathrm{E}_{n1}^2\right) = \frac{n\kappa_n^2}{2p_n}\ell''_{f,2}(0)+o\left(\frac{n\kappa_n^2}{p_n}\right) = \frac{n\kappa_n^2}{p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right), \tag{2.22}$$

which leads to

$$\Lambda_n = -\frac{n\kappa_n^2}{2p_n}+\sqrt{\frac{n\kappa_n^2}{p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right)}\sum_{i=1}^{n}\mathrm{W}_{ni}+o\left(\frac{n\kappa_n^2}{p_n}\right). \tag{2.23}$$

Since the W_{ni}'s, $i=1,\ldots,n$ are mutually independent with mean zero and variance $1/n$, we obtain that

$$\mathrm{E}\left[\Lambda_n^2\right] = \mathrm{E}\left[\Lambda_n\right]^2+\mathrm{Var}\left[\Lambda_n\right] = \frac{n^2\kappa_n^4}{4p_n^2}+o\left(\frac{n^2\kappa_n^4}{p_n^2}\right)+\frac{n\kappa_n^2}{p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right). \tag{2.24}$$

If $\kappa_n^2 = o(p_n/n)$, then (2.24) implies that $\exp(\Lambda_n)\xrightarrow{\mathcal{D}}1$, so that Le Cam's First Lemma (see Section 1.2.4) yields that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $\mathrm{P}^{(n)}_0$ are mutually contiguous.

We may therefore assume that $\kappa_n^2 = \tau_n^2 p_n/n$, where the positive sequence (τ_n) is $\mathrm{O}(1)$ but not $o(1)$. In this case, (2.23) can be rewritten

$$\Lambda_n = -\frac{\tau_n^2}{2}+\sqrt{\tau_n^2+o(1)}\sum_{i=1}^{n}\mathrm{W}_{ni}+o(1).$$

Applying the Cauchy-Schwarz inequality and the Chebychev inequality, then using (2.21) and (2.22), provides that, for some positive constant C,

$$\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{W}_{ni}^2\mathbb{I}_{\{|\mathrm{W}_{ni}|>\varepsilon\}}\right] \le n\sqrt{\mathrm{E}\left[\mathrm{W}_{ni}^4\right]\mathrm{P}[|\mathrm{W}_{ni}|>\varepsilon]} \le \frac{n}{\varepsilon}\sqrt{\mathrm{E}\left[\mathrm{W}_{ni}^4\right]\mathrm{Var}[\mathrm{W}_{ni}]} = \frac{\sqrt{n}}{\varepsilon}\sqrt{\mathrm{E}\left[\mathrm{W}_{ni}^4\right]}$$
$$\le \frac{\mathrm{C}n^{1/2}\mathrm{E}_{n4}^{1/2}}{\varepsilon n\mathrm{V}_n} = \frac{\mathrm{C}\left(\frac{n\kappa_n^2\ell''_{f,4}(0)}{2p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right)\right)^{1/2}}{\varepsilon\left(\frac{n\kappa_n^2}{p_n}+o\left(\frac{n\kappa_n^2}{p_n}\right)\right)} = \frac{o(\tau_n)}{\varepsilon\left(\tau_n^2+o(\tau_n^2)\right)} = o(1),$$

where we have used the fact that $\ell''_{f,4}(0) = 0$. This shows that $\sum_{i=1}^n \mathrm{W}_{ni}$ satisfies the classical Lévy–Lindeberg condition, hence is asymptotically standard normal (as already mentioned, the W_{ni}'s, $i = 1, \ldots, n$ are mutually independent with mean zero and variance $1/n$). For any subsequence $(\exp(\Lambda_{n_m}))$ converging in distribution, we then have $\exp(\Lambda_{n_m}) \to \exp(\mathrm{Y})$, with $\mathrm{Y} \sim \mathcal{N}(-\frac{1}{2}\lim_{m\to\infty}\tau^2_{n_m}, \lim_{m\to\infty}\tau^2_{n_m})$. Mutual contiguity of $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $\mathrm{P}^{(n)}_0$ then follows from the fact that $\mathrm{P}[\exp(\mathrm{Y}) = 0] = 0$ and $\mathrm{E}[\exp(\mathrm{Y})] = 1$.

□

2.5.3 Proof of Theorem 2.2.1

As in the proof of Theorem 2.1.1, all expectations and variances in this proof are taken under the null of uniformity $\mathrm{P}^{(n)}_0$ and all stochastic convergences and o_{P}'s are as $n \to \infty$ under $\mathrm{P}^{(n)}_0$. The central limit theorem then directly establishes the asymptotic normality of $\Delta^{(n)}_{\boldsymbol{\theta}_n}$, since $\mathrm{E}\left[\Delta^{(n)}_{\boldsymbol{\theta}_n}\right] = 0$ and $\mathrm{Var}\left[\Delta^{(n)}_{\boldsymbol{\theta}_n}\right] = \frac{p_n}{n}\mathrm{Var}\left[\sum_{i=1}^n \mathbf{X}'_{ni}\boldsymbol{\theta}_n\right] = 1$.

It therefore remains to establish (2.8). Recall that, in the case where (τ_n) is $\mathrm{O}(1)$ but not $o(1)$, we have obtained in the proof of Theorem 2.1.1 that

$$\Lambda_n = -\frac{\tau_n^2}{2} + \sqrt{\tau_n^2 + o(1)}\sum_{i=1}^n \mathrm{W}_{ni} + o(1) = -\frac{\tau_n^2}{2} + \tau_n\sum_{i=1}^n \mathrm{W}_{ni} + o_{\mathrm{P}}(1),$$

where

$$\sum_{i=1}^n \mathrm{W}_{ni} = \sum_{i=1}^n \frac{\log f\left(\kappa_n \mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) - \mathrm{E}_{n1}}{\sqrt{n\mathrm{V}_n}}$$

is asymptotically standard normal. To establish the result, it is therefore sufficient to show that $\tau_n\left[\sum_{i=1}^n \mathrm{W}_{ni} - \Delta^{(n)}_{\boldsymbol{\theta}_n}\right]$ converges to zero in quadratic mean. To do so, write

$$\tau_n\left(\sum_{i=1}^n \mathrm{W}_{ni}\right) - \tau_n\Delta^{(n)}_{\boldsymbol{\theta}_n} = \frac{\tau_n}{\sqrt{n\mathrm{V}_n}}\sum_{i=1}^n\left(\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) - \mathrm{E}_{n1} - \sqrt{p_n\mathrm{V}_n}\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) =: \frac{\mathrm{M}_n}{\sqrt{n\mathrm{V}_n}}.$$

Then using $\mathrm{E}[\mathbf{X}'_{n1}\boldsymbol{\theta}_n] = 0$ and $\mathrm{E}[(\mathbf{X}'_{n1}\boldsymbol{\theta}_n)^2] = 1/p_n$, we obtain

$$\begin{aligned}
\mathrm{E}\left[\mathrm{M}_n^2\right] &= n\tau_n^2\,\mathrm{E}\left[\left(\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) - \mathrm{E}_{n1} - \sqrt{p_n\mathrm{V}_n}\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right] \\
&= n\tau_n^2\left(2\mathrm{V}_n - 2\sqrt{p_n\mathrm{V}_n}\,\mathrm{E}\left[\mathbf{X}'_{ni}\boldsymbol{\theta}_n\left(\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right) - \mathrm{E}_{n1}\right)\right]\right) \\
&= 2n\tau_n^2\mathrm{V}_n - 2\tau_n n^{3/2}\sqrt{\mathrm{V}_n}\,\mathrm{E}\left[\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\log f\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\right],
\end{aligned}$$

which, letting $g(x) := x\log f(x)$, provides

$$\mathrm{E}\left[\left(\tau_n\left(\sum_{i=1}^n \mathrm{W}_{ni}\right) - \tau_n\Delta^{(n)}_{\boldsymbol{\theta}_n}\right)^2\right] = 2\tau_n^2 - \frac{2\tau_n n}{\sqrt{n\mathrm{V}_n}}\mathrm{E}\left[g\left(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)\right]. \tag{2.25}$$

Using Lemma 2.5.1,

$$\mathrm{E}\left[g(\kappa_n\mathbf{X}'_{ni}\boldsymbol{\theta}_n)\right] = c_{p_n}\int_{-1}^{1}(1-s^2)^{(p_n-3)/2}g(\kappa_n s)\,ds = \frac{\kappa_n^2}{2p_n}g''(0) + o\left(\frac{\kappa_n^2}{p_n}\right) = \frac{\kappa_n^2}{p_n} + o\left(\frac{\kappa_n^2}{p_n}\right).$$

Plugging in (2.25) and using (2.22) then yields

$$\mathrm{E}\left[\left(\tau_n\left(\sum_{i=1}^n \mathrm{W}_{ni}\right) - \tau_n\Delta^{(n)}_{\boldsymbol{\theta}_n}\right)^2\right] = 2\tau_n^2 - \frac{2\tau_n\left(\frac{n\kappa_n^2}{p_n} + o\left(\frac{n\kappa_n^2}{p_n}\right)\right)}{\left(\frac{n\kappa_n^2}{p_n} + o\left(\frac{n\kappa_n^2}{p_n}\right)\right)^{1/2}} = o(1),$$

as was to be shown.

□

2.5.4 Proof of Theorem 2.3.2

The FvML version of the log-likelihood in (2.15) can be rewritten

$$\Lambda_{n,\exp}^{\mathbf{T}_n} = n \log \frac{c_{p_n,\kappa_n,\exp}}{c_{p_n}} + \log \mathrm{E}\left[\exp\left(\kappa_n n \bar{\mathbf{X}}_n' \mathbf{U}\right) \middle| \bar{\mathbf{X}}_n\right] =: \mathrm{L}_{n1} + \mathrm{L}_{n2}, \tag{2.26}$$

where

$$\mathrm{L}_{n1} = -n \log \frac{c_{p_n}}{c_{p_n,\kappa_n,\exp}} = -n \log \frac{\Gamma\left(p_n/2\right) \mathscr{I}_{p_n/2-1}(\kappa_n)}{(\kappa_n/2)^{p_n/2-1}} =: -n \log \mathrm{H}_{p_n/2-1}(\kappa_n),$$

(see (1.1) and (1.11) for explicit expressions of c_p and $c_{p,\kappa,\exp} = c_{p,\kappa}^{\mathrm{FvML}}$, respectively) and

$$\begin{aligned}\mathrm{L}_{n2} &= \log \mathrm{E}\left[\exp\left(n\kappa_n \|\bar{\mathbf{X}}_n\| \frac{\mathbf{U}'\bar{\mathbf{X}}_n}{\|\bar{\mathbf{X}}_n\|}\right) \middle| \bar{\mathbf{X}}_n\right] = \log\left(c_{p_n} \int_{-1}^{1} (1-s^2)^{\frac{p_n-3}{2}} \exp(n\kappa_n \|\bar{\mathbf{X}}_n\| s)\, ds\right) \\ &= \log \frac{c_{p_n}}{c^{\mathrm{FvML}}_{p_n, n\kappa_n \|\bar{\mathbf{X}}_n\|}} =: \log \mathrm{H}_{p_n/2-1}\left(\kappa_n \sqrt{n} \mathrm{T}_n\right),\end{aligned}$$

where $\mathrm{T}_n := \sqrt{n}\|\bar{\mathbf{X}}_n\|$. Now, we use the bounds

$$\mathrm{S}_{(p_n-1)/2,(p_n+1)/2}(\kappa) \leq \log \mathrm{H}_{p_n/2-1}(\kappa) \leq \mathrm{S}_{(p_n-2)/2,(p_n+2)/2}(\kappa)$$

(see (5) in [Hornik and Grün, 2014]) with

$$\begin{aligned}\mathrm{S}_{\alpha,\beta}(\kappa) &:= \sqrt{\kappa^2+\beta^2} - \alpha \log\left(\alpha + \sqrt{\kappa^2+\beta^2}\right) - \beta + \alpha \log(\alpha+\beta) \\ &= \beta\left(\sqrt{(\kappa/\beta)^2+1} - 1 + \frac{\alpha}{\beta} \log\left(\frac{\alpha/\beta+1}{\alpha/\beta + \sqrt{(\kappa/\beta)^2+1}}\right)\right).\end{aligned}$$

If $\kappa_n = \tau_n p_n^{3/4}/\sqrt{n}$, with $n, p_n \to \infty$ and (τ_n) bounded, then $\lim_{n\to\infty} \kappa_n/\beta_n = 0$ with $\beta_n \sim p_n/2$[1] so that $\mathrm{S}_{\alpha_n,\beta_n}(\kappa_n)$ can be expanded as

$$\frac{\kappa_n^2}{2(\alpha_n+\beta_n)} - \frac{\kappa_n^4}{8\beta(\alpha_n+\beta_n)^2} + \mathrm{O}\left(\frac{1}{\sqrt{p_n n^3}}\right),$$

where $\alpha_n \sim p_n/2$. It implies that

$$\mathrm{L}_{n1} = -\frac{n\kappa_n^2}{2p_n} + \frac{n\kappa_n^4}{4p_n^2(p_n+2)} + o(1) \tag{2.27}$$

under $\mathrm{P}_0^{(n)}$. Similarly, as $\mathrm{T}_n^2 - 1 = \sqrt{2/p_n}\mathrm{R}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} 0$ as $n \to \infty$ under $\mathrm{P}_0^{(n)}$ from (2.1) and Slutsky's Lemma, $\mathrm{T}_n = 1 + o_{\mathrm{P}}(1)$ and we can apply Corollary 5.1.6 from [Fuller, 1995] to get

$$\mathrm{L}_{n2} = \frac{n\kappa_n^2}{2p_n}\mathrm{T}_n^2 - \frac{n^2\kappa_n^4}{4p_n^2(p_n+2)}\mathrm{T}_n^4 + o_{\mathrm{P}}(1) \tag{2.28}$$

Plugging (2.27) and (2.28) into (2.26) and using again the fact that $\mathrm{T}_n = 1 + o_{\mathrm{P}}(1)$ entails that, as $n \to \infty$ under $\mathrm{P}_0^{(n)}$,

$$\Lambda_{n,\exp}^{\mathbf{T}_n} = \frac{n\kappa_n^2}{2p_n}(\mathrm{T}_n^2 - 1) - \frac{n^2\kappa_n^4}{4p_n^2(p_n+2)}\mathrm{T}_n^4 + o_{\mathrm{P}}(1) = \tau_n^2 \frac{\mathrm{R}_n^{\mathrm{St}}}{\sqrt{2}} - \frac{\tau_n^4 p_n}{4(p_n+2)} + o_{\mathrm{P}}(1).$$

Jointly with (2.1), this establishes the result. □

[1] Two sequences (x_n) and (y_n) are of the same order, $x_n \sim y_n$, if $\lim_{n\to\infty} \frac{x_n}{y_n} = 1$.

Chapter 3

Testing uniformity against a semiparametric extension of the Watson distributions

Contents

3.1 Introduction

In the previous chapter we were interested in testing uniformity against monotone rotationally symmetric distributions and the Rayleigh test stood out as the optimal test in the unspecified $\boldsymbol{\theta}_n$-case in low dimensions and in the FvML case in high dimensions (and presumably in general). However it will show no power when the common distribution of the $\mathbf{X}_i$'s is antipodally symmetric.

In low dimensions the most classical test of uniformity for axial data is the [Bingham, 1974] test, that rejects the null hypothesis of uniformity whenever

$$\mathrm{Q}_n = \frac{np_n(p_n+2)}{2}\left(\mathrm{tr}\,[\mathbf{S}_n^2] - \frac{1}{p_n}\right) > \chi^2_{d_{p_n},1-\alpha}, \tag{3.1}$$

where $\mathbf{S}_n := n^{-1}\sum_{i=1}^n \mathbf{X}_{ni}\mathbf{X}_{ni}'$ is the covariance matrix of the observations (using the centre of the sphere as a specified location) and $d_p := p(p+1)/2 - 1$. In the high-dimensional case, Theorem 2.5 from [Paindaveine and Verdebout, 2016] implies that, under the null hypothesis of uniformity,

$$\mathrm{Q}_n^{\mathrm{St}} = \frac{\mathrm{Q}_n - d_{p_n}}{\sqrt{2d_{p_n}}} = \frac{p_n}{n}\sum_{\substack{i<j\\ i,j=1}}^{n}\left\{\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2 - \frac{1}{p_n}\right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1), \tag{3.2}$$

as soon as p_n diverges to infinity with n. Therefore the Bingham test that rejects the null of uniformity when

$$\mathrm{Q}_n^{\mathrm{St}} > z_\alpha \tag{3.3}$$

is a natural test of uniformity in high dimensions in the context of axial data—just as the Rayleigh test is a natural test of uniformity in high dimensions in the framework of non-axial data.

In order to study the non-null behaviour of the Bingham test, we will in this chapter consider as alternatives to uniformity a family of distributions that include the classical Watson distributions: axial rotationally symmetric distributions (see Section 1.1.4). Recall that $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ is the hypothesis under which the $\mathbf{X}_{ni}$'s, $i = 1,\ldots,n$, are mutually independent and share the common density

$$\mathbf{x} \mapsto \frac{\Gamma\left(\frac{p_n-1}{2}\right)\check{c}_{p_n,\kappa_n,f}}{2\pi^{(p_n-1)/2}} f\left(\kappa_n\left(\mathbf{x}'\boldsymbol{\theta}_n\right)^2\right).$$

We first identify the sequences (κ_n) that make the corresponding sequences of alternatives contiguous to the null hypothesis of uniformity. In Section 3.2, we tackle the problem of testing uniformity under specified $\boldsymbol{\theta}_n$ and show that the resulting model is locally asymptotically normal in low- and high-dimensions. We define the resulting optimal tests of uniformity and determine their asymptotic powers under contiguous alternatives. In Section 3.3, we turn to the unspecified-$\boldsymbol{\theta}_n$ problem. In low dimensions, our LAN result naturally leads to the Bingham test that shows asymptotic power under contiguous alternatives. We show that this is not the case in high dimensions anymore. In Section 3.4, we focus on the fixed p-case and we turn our attention to tests that take into account the "single-spiked" structure of the considered alternatives, namely the tests from [Anderson and Stephens, 1972]. We derive the limiting behaviour, under sequences of contiguous alternatives, of these tests. Doing so, we obtain in particular the limiting behaviour, under local alternatives, of the extreme eigenvalues of the spatial sign covariance matrix, which is a result of independent interest; see [Dürre et al., 2016] and the references therein for a recent study of these eigenvalues. While all results above are confirmed by suitable numerical exercises in Sections 3.2–3.4, we specifically conduct, in Section 3.5, Monte Carlo simulations in order to compare the finite-sample powers of the various tests.

Our first result identifies the sequences of alternatives $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$ that are contiguous to the sequence of null hypotheses $\mathrm{P}^{(n)}_0$.

Theorem 3.1.1. *Let (p_n) be a sequence in $\{2,3,\ldots\}$, $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for all n, $f \in \mathscr{F}$ and (κ_n) a sequence in $\mathbb{R}_0 := \mathbb{R}\setminus\{0\}$ that is $\mathrm{O}\left(p_n/\sqrt{n}\right)$. Then, the sequence of alternative hypotheses $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$ and the sequence of null hypotheses $\mathrm{P}^{(n)}_0$ are mutually contiguous.*

Proof. See Section 3.7.2. □

In other words, if $\kappa_n = \mathrm{O}\left(p_n/\sqrt{n}\right)$, then no test for

$$\begin{cases} \mathscr{H}_{0n} : \left\{\mathrm{P}^{(n)}_0\right\} \\ \mathscr{H}_{1n} : \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\right\} \end{cases}$$

can be consistent. In the low-dimensional case, the contiguity rate $\kappa_n \sim 1/\sqrt{n}$ coincides with the one obtained in the non-axial case (see Theorem 2.1.1). This is not the case in high dimensions: the contiguity rate in the axial case ($p_n/\sqrt{n}$) is larger than the one in the non-axial case ($\sqrt{p_n/n}$). It is therefore more challenging to detect axial departures from uniformity than non-axial ones.

3.2 Testing uniformity under specified location

In this section, we consider the problem of testing uniformity over $\mathcal{S}^{p_n-1}$ against the class of alternatives introduced in the previous section, in a situation where the location $\boldsymbol{\theta}_n$ is specified. In other words, this corresponds to cases where it is known in which direction the possible deviation from uniformity would materialize. Depending on the exact type of alternatives we want to focus on (bipolar, girdle-type, or both), we will then consider, for $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$, the problem of testing $\mathscr{H}^{(n)}_0 : \left\{\mathrm{P}^{(n)}_0\right\}$ against

(i) $\mathscr{H}^{(n)}_1 : \cup_{\kappa>0} \cup_{f\in\mathscr{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\}$;

(ii) $\mathscr{H}^{(n)}_1 : \cup_{\kappa<0} \cup_{f\in\mathscr{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\}$;

(iii) $\mathscr{H}^{(n)}_1 : \cup_{\kappa\neq0} \cup_{f\in\mathscr{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\}$.

3.2.1 In low dimensions

As the dimension of $\boldsymbol{\theta}_n$ does not depend on n in this section, it will be written $\boldsymbol{\theta}$. Optimal testing may be based on the following Local Asymptotic Normality (LAN) result.

Theorem 3.2.1. *Fix $p \in \{2,3,\ldots\}$, $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$, and $f \in \mathscr{F}$. Let $\kappa_n = \tau_n p/\sqrt{n}$, where the real sequence (τ_n) is $\mathrm{O}(1)$. Then, letting*

$$\check{\Delta}^{(n)}_{\boldsymbol{\theta}} := \frac{p}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \left(\mathbf{X}'_{ni}\boldsymbol{\theta}\right)^2 - \frac{1}{p} \right\} \quad \text{and} \quad \Gamma_p := \frac{2(p-1)}{p+2},$$

we have that, as $n \to \infty$*, under* $\mathrm{P}_0^{(n)}$,

$$\breve{\mathbf{\Lambda}}_n = \log \frac{d\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}}{d\mathrm{P}_0^{(n)}} = \tau_n \breve{\Delta}^{(n)}_{\boldsymbol{\theta}} - \frac{\tau_n^2}{2}\Gamma_p + o_{\mathrm{P}}(1)\,, \tag{3.4}$$

where $\breve{\Delta}^{(n)}_{\boldsymbol{\theta}}$ *is asymptotically normal with mean zero and variance* Γ_p.

In other words, the sequence $\left(\left\{\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f} : \kappa \in \mathbb{R}\right\}\right)$ *(where we let* $\mathrm{P}^{(n)}_{\boldsymbol{\theta},0,f} := \mathrm{P}_0^{(n)}$*) is locally asymptotically normal at* $\kappa = 0$ *with central sequence* $\breve{\Delta}^{(n)}_{\boldsymbol{\theta}}$*, Fisher information* Γ_p*, and contiguity rate* $1/\sqrt{n}$.

Proof. See Section 3.7.3. □

This result confirms that in low dimensions the contiguity rate when testing uniformity against the considered axial alternatives is $1/\sqrt{n}$. Note also that the central sequence rewrites

$$\breve{\Delta}^{(n)}_{\boldsymbol{\theta}} = \sqrt{n}\left(p\boldsymbol{\theta}'\mathbf{S}_n\boldsymbol{\theta} - 1\right).$$

Consequently, optimal testing of uniformity for axial data will be based in low dimensions on $\mathbf{S}_n$. The optimal axial tests of uniformity in the specified location case directly result from the LAN property above: Theorem 3.2.1 entails that, for the problem of testing

$$\begin{cases} \mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\} \\ \mathcal{H}_1^{(n)} : \cup_{\kappa>0} \cup_{f\in\mathcal{F}} \left\{\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}, \end{cases} \tag{3.5}$$

the test $\phi^{(n)}_{\boldsymbol{\theta}+}$ rejecting the null hypothesis at asymptotic level α whenever

$$\mathrm{T}^{(n)}_{\boldsymbol{\theta}} := \frac{\breve{\Delta}^{(n)}_{\boldsymbol{\theta}}}{\sqrt{\Gamma_p}} > z_\alpha \tag{3.6}$$

is locally asymptotically most powerful. A routine application of Le Cam's Third Lemma shows that, under $\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$, as $n \to \infty$,

$$\mathrm{T}^{(n)}_{\boldsymbol{\theta}} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\Gamma_p^{1/2}\tau, 1\right),$$

with $\kappa_n = \tau p/\sqrt{n}$ ($\tau > 0$). Therefore, the corresponding asymptotic power of $\phi^{(n)}_{\boldsymbol{\theta}+}$ is

$$\lim_{n\to\infty} \breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}\left[\mathrm{T}^{(n)}_{\boldsymbol{\theta}} > z_\alpha\right] = 1 - \Phi\left(z_\alpha - \Gamma_p^{1/2}\tau\right). \tag{3.7}$$

Note that this asymptotic power does not converge to α as p diverges to infinity. This may be surprising at first since departures from uniformity here are of a *single-spiked* nature, that is, only materialize in a single direction out of the p directions in $\mathcal{S}^{p-1}$. The fact that this asymptotic power does not fade out for larger dimensions is actually explained by the fact that we did not consider local alternatives associated with $\kappa_n = \tau/\sqrt{n}$ but rather with $\kappa_n = \tau p/\sqrt{n}$, which properly scales local alternatives for different dimensions p (note also that the behaviour of high-dimensional Watson distributions—see Chapter 8 in [Chikuse, 2003]—intuitively explains that, at least in the Watson case, the concentration should indeed scale linearly with the dimension).

Optimal tests for the other one-sided problem and for the two-sided problem are obtained in a similar way. More precisely, for the problem of testing

$$\begin{cases} \mathscr{H}_0^{(n)} : \left\{ \mathrm{P}_0^{(n)} \right\} \\ \mathscr{H}_1^{(n)} : \cup_{\kappa<0} \cup_{f\in\mathscr{F}} \left\{ \check{\mathrm{P}}_{\boldsymbol{\theta},\kappa,f}^{(n)} \right\}, \end{cases} \tag{3.8}$$

the test $\phi_{\boldsymbol{\theta}-}^{(n)}$ rejecting the null hypothesis of uniformity at asymptotic level α whenever

$$\mathrm{T}_{\boldsymbol{\theta}}^{(n)} < -z_\alpha$$

is locally asymptotically most powerful and has asymptotic power

$$\lim_{n\to\infty} \check{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)} \left[\mathrm{T}_{\boldsymbol{\theta}}^{(n)} < -z_\alpha \right] = \Phi\left(-z_\alpha - \Gamma_p^{1/2}\tau \right)$$

under $\check{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)}$ with $\kappa_n = \tau p/\sqrt{n}$ $(\tau < 0)$.

The corresponding two-sided test, $\phi_{\boldsymbol{\theta}\pm}^{(n)}$ say, rejects the null hypothesis at asymptotic level α whenever

$$\left| \mathrm{T}_{\boldsymbol{\theta}}^{(n)} \right| > z_{\alpha/2}.$$

This test is locally asymptotically maximin for

$$\begin{cases} \mathscr{H}_0^{(n)} : \left\{ \mathrm{P}_0^{(n)} \right\} \\ \mathscr{H}_1^{(n)} : \cup_{\kappa\neq 0} \cup_{f\in\mathscr{F}} \left\{ \check{\mathrm{P}}_{\boldsymbol{\theta},\kappa,f}^{(n)} \right\}, \end{cases} \tag{3.9}$$

and has asymptotic power

$$\lim_{n\to\infty} \check{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)} \left[\left| \mathrm{T}_{\boldsymbol{\theta}}^{(n)} \right| > z_{\alpha/2} \right] = 2 - \Phi\left(z_{\alpha/2} - \Gamma_p^{1/2}\tau \right) - \Phi\left(z_{\alpha/2} + \Gamma_p^{1/2}\tau \right)$$

under $\check{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)}$ with $\kappa_n = \tau p/\sqrt{n}$ $(\tau \neq 0)$. Again, the local asymptotic powers of these tests do not fade out for larger dimensions p but rather converge to a constant larger than α.

We conducted the following Monte Carlo exercise in order to check the validity of our asymptotic results. For any combination (n, p) of sample size $n \in \{100, 1000\}$ and dimension $p \in \{3, 10\}$, we generated collections of 5000 independent random samples of size n from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa_n = \tau p/\sqrt{n}$, for $\tau = -2, -1, 0, 1, 2$. The value $\tau = 0$ corresponds to the null hypothesis of uniformity over S^{p-1}, whereas the larger the non-zero value of $|\tau|$ is, the more severe the alternative is. Kernel density estimates of the resulting values of the test statistic $\mathrm{T}_{\boldsymbol{\theta}}^{(n)}$ in (3.6) are provided in Figure 3.1, that further plots the densities of the corresponding asymptotic distributions (for the null case $\tau = 0$, histograms of the values of $\mathrm{T}_{\boldsymbol{\theta}}^{(n)}$ are also shown). Clearly, our asymptotic results are confirmed by these simulations (yet, unsurprisingly, larger dimensions require larger sample sizes for asymptotic results to materialize).

3.2.2 In high dimensions

In high dimensions we have a LAN result similar to the one in low dimensions.

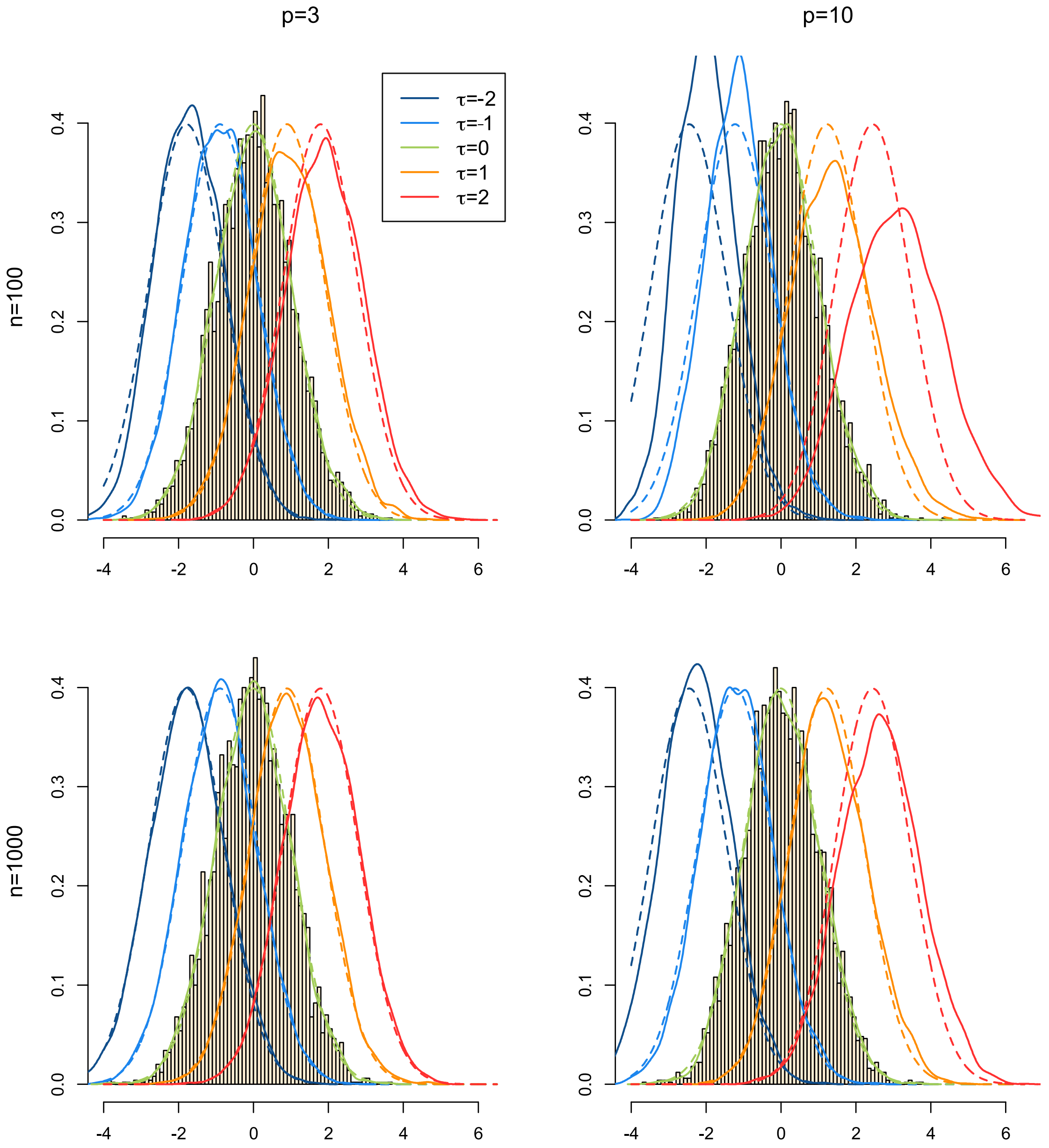


Figure 3.1 – Plots of the kernel density estimates (solid curves) of the values of the test statistic $\mathrm{T}_{\boldsymbol{\theta}}^{(n)}$ in (3.6) obtained from M = 5000 independent random samples, of size $n = 100$ (top) or 1000 (bottom), from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa = \tau p / \sqrt{n}$, with $\tau = -2, -1, 0, 1, 2$ and with $p = 3$ (left) or $p = 10$ (right); for $\tau = 0$, histograms of the values of $\mathrm{T}_{\boldsymbol{\theta}}^{(n)}$ are shown. The densities of the corresponding asymptotic $\mathcal{N}\left(\Gamma_p^{1/2}\tau, 1\right)$ distributions are also plotted (dashed curves). Throughout this chapter, kernel density estimates are obtained from the R command `density` with default parameter values.

Theorem 3.2.2. *Let (p_n) be a sequence of positive integers diverging to ∞ and $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Fix $f \in \mathcal{F}$ and let $\kappa_n = \tau_n p_n/\sqrt{n}$, where (τ_n) is $\mathrm{O}(1)$. Then, as $n \to \infty$, under $\mathrm{P}_0^{(n)}$,*

$$\breve{\boldsymbol{\Lambda}}_n = \log \frac{d\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}}{d\mathrm{P}_0^{(n)}} = \tau_n \breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n} - \tau_n^2 + o_{\mathrm{P}}(1),$$

where $\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n} := \sqrt{n}\left(p_n \boldsymbol{\theta}_n' \mathbf{S}_n \boldsymbol{\theta}_n - 1\right)$ is asymptotically normal with mean zero and variance two .

Proof. See Section 3.7.3. □

The result readily implies that Le Cam optimal tests can be found in a similar fashion as in low dimensions. For example the Le Cam optimal test of uniformity for testing

$$\begin{cases} \mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\} \\ \mathcal{H}_1^{(n)} : \cup_{\kappa \neq 0} \cup_{f \in \mathcal{F}} \left\{\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa,f}\right\}, \end{cases} \tag{3.10}$$

rejects the null hypothesis at asymptotic level α whenever

$$\left|\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n}\right| > \sqrt{2} z_{\alpha/2}.$$

Applying Le Cam's Third Lemma then entails that, under $\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ with $\kappa_n = \tau_n p_n/\sqrt{n}$ and $\tau_n \to \tau$, the central sequence $\breve{\Delta}_{\boldsymbol{\theta}_n}$ is asymptotically normal with mean 2τ and variance 2, which provides the asymptotic power

$$\lim_{n\to\infty} \breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\left|\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n}\right| > \sqrt{2} z_{\alpha/2}\right] = 2 - \Phi\left(z_{\alpha/2} - \sqrt{2}\tau\right) - \Phi\left(z_{\alpha/2} + \sqrt{2}\tau\right). \tag{3.11}$$

This optimal test can thus detect these contiguous alternatives and shows a symmetric power pattern against girdle alternatives ($\tau < 0$) and bipolar alternatives ($\tau > 0$).

These results show that optimal testing of uniformity for low-dimensional and high-dimensional axial data is based on $\mathbf{S}_n$, at least in the considered model. This is to be compared with the non-axial case investigated in Chapter 2, where optimal testing of uniformity in high dimensions is rather based on $\bar{\mathbf{X}}_n$. This will have important consequences when considering the unspecified-$\boldsymbol{\theta}$ case we turn to in the sequel.

3.3 Testing uniformity under unspecified location: the Bingham test

It is important to note that the optimal tests described in Section 3.2 are of little practical relevance because the polar axis $\pm\boldsymbol{\theta}_n$ has to be known, which is almost never the case in applications. In contrast, the Bingham test (see (3.3)) does not need such knowledge.

3.3.1 In low dimensions

We focus first on the one-sided problem of testing

$$\begin{cases} \mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\} \\ \mathcal{H}_1^{(n)} : \cup_{\boldsymbol{\theta} \in \mathcal{S}^{p-1}} \cup_{\kappa > 0} \cup_{f \in \mathcal{F}} \left\{\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}, \end{cases} \tag{3.12}$$

It is convenient to reparameterise the submodel associated with $\kappa \geq 0$ by defining $\boldsymbol{\vartheta} := \sqrt{\kappa}\boldsymbol{\theta}$. In this new parametrisation (which, unlike the original curved one, is flat), the testing problem rewrites

$$\begin{cases} \mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\} \\ \mathcal{H}_1^{(n)} : \cup_{\boldsymbol{\vartheta}\in\mathbb{R}^p\setminus\{\mathbf{0}\}} \cup_{f\in\mathcal{F}} \left\{\check{\mathrm{P}}_{\boldsymbol{\vartheta},f}^{(n)}\right\}, \end{cases}$$

that is, simply consists in testing

$$\begin{cases} \mathcal{H}_0^{(n)} : \boldsymbol{\vartheta} = \mathbf{0} \\ \mathcal{H}_1^{(n)} : \boldsymbol{\vartheta} \neq \mathbf{0}, \end{cases}$$

Theorem 3.3.1 below then describes the asymptotic behaviour of the corresponding local log-likelihood ratios.

Theorem 3.3.1. *Fix $p \in \{2,3,\ldots\}$ and $f \in \mathcal{F}$. Let $\boldsymbol{\vartheta}_n = \left(p/\sqrt{n}\right)^{1/2}\boldsymbol{\tau}_n$, where $(\boldsymbol{\tau}_n)$ is a sequence in $\mathbb{R}^p$ that is $\mathrm{O}(1)$ but not $o(1)$. Then, letting*

$$\check{\boldsymbol{\Delta}}^{(n)} := p\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right) \quad \text{and} \quad \boldsymbol{\Gamma}_p := \frac{p}{p+2}\left(\mathbf{I}_{p^2} + \mathbf{K}_p - \frac{2}{p}\mathbf{J}_p\right),$$

we have that, as $n \to \infty$, under $\mathrm{P}_0^{(n)}$,

$$\log\frac{d\check{\mathrm{P}}_{\boldsymbol{\vartheta}_n,f}^{(n)}}{d\mathrm{P}_0^{(n)}} = \left(\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right)\right)'\check{\boldsymbol{\Delta}}^{(n)} - \frac{1}{2}\left(\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right)\right)'\boldsymbol{\Gamma}_p\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right) + o_{\mathrm{P}}(1), \tag{3.13}$$

where $\check{\boldsymbol{\Delta}}^{(n)}$ is, still under $\mathrm{P}_0^{(n)}$, asymptotically normal with mean vector zero and covariance matrix $\boldsymbol{\Gamma}_p$.

Proof. See Section 3.7.4. □

Theorem 3.3.1 shows that the contiguity rate for $\boldsymbol{\vartheta}$ is $n^{-1/4}$, which corresponds to the contiguity rate $n^{-1/2}$ obtained for κ in Theorem 3.2.1 (recall that $\boldsymbol{\vartheta} = \sqrt{\kappa}\boldsymbol{\theta}$); however, as we will explain below, the limiting experiment in Theorem 3.3.1 is non-standard. A natural test of uniformity is the test rejecting the null hypothesis at asymptotic level α whenever

$$\left(\check{\boldsymbol{\Delta}}^{(n)}\right)'\boldsymbol{\Gamma}_p^-\check{\boldsymbol{\Delta}}^{(n)} = \frac{np(p+2)}{2}\left(\mathrm{tr}\left[\mathbf{S}_n^2\right] - \frac{1}{p}\right) > \chi^2_{d_p,1-\alpha}, \tag{3.14}$$

with $d_p = p(p+1)/2 - 1$. We recognize the Bingham test (see (3.1)) that will be denoted as ϕ_{Bing} in the sequel. This test, which rejects the null hypothesis when the sample variance of the eigenvalues $\hat{\lambda}_{n1},\ldots,\hat{\lambda}_{np}$ of $\mathbf{S}_n$ is too large, also addresses the problem of testing uniformity against the one-sided alternatives associated with $\kappa < 0$ or against the two-sided alternatives associated with $\kappa \neq 0$.

Local asymptotic powers of the Bingham test can be obtained from the LAN result in Theorem 3.2.1 and Le Cam's Third Lemma. We have the following result.

Proposition 3.3.2. *Fix $p \in \{2,3,\ldots\}$, $\boldsymbol{\theta} \in \mathcal{S}^{p-1}$, and $f \in \mathcal{F}$. Let $\kappa_n = \tau_n p/\sqrt{n}$, where the real sequence (τ_n) converges to τ. Then, under $\check{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)}$,*

$$\mathrm{Q}_n \xrightarrow{\mathcal{D}} \chi^2_{d_p}\left(\frac{2(p-1)\tau^2}{p+2}\right). \tag{3.15}$$

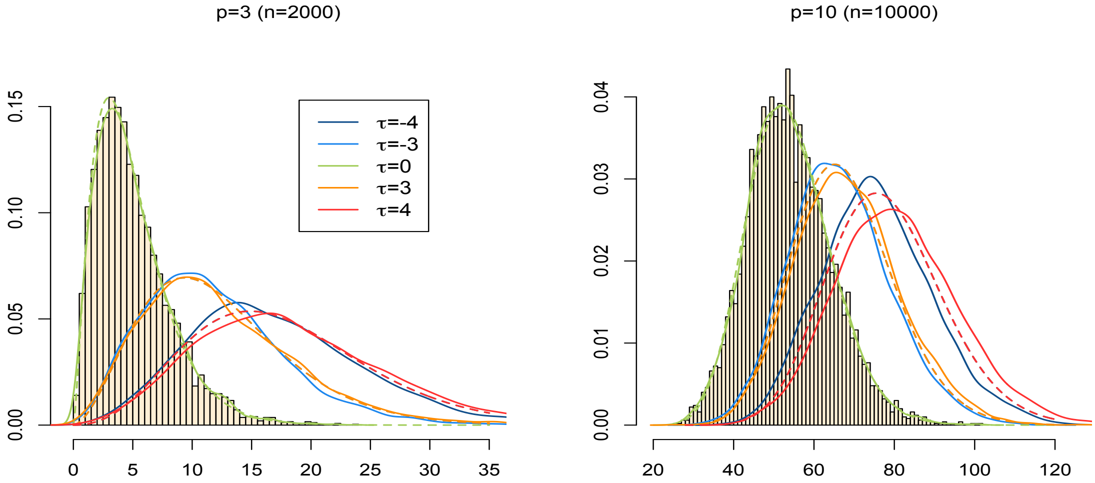


Figure 3.2 – Left: Plots of the kernel density estimates (solid curves) of the values of the Bingham test statistic Q_n in (3.14) obtained from M = 5000 independent random samples of size $n = 2000$ from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \dots, 0)' \in \mathbb{R}^p$ and concentration $\kappa = \tau p/\sqrt{n}$, with $\tau = -4, -3, 0, 3, 4$ and with $p = 3$; for $\tau = 0$, histograms of the values of Q_n are shown. The density of the corresponding asymptotic distributions in (3.15), which do not depend on the sign of τ, are also plotted (dashed curves).
Right: The corresponding results for $p = 10$ and $n = 10000$.

Under the same sequence of alternatives, the asymptotic power of the Bingham test is therefore

$$\lim_{n\to\infty} \check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}\left[Q_n > \chi^2_{d_p,1-\alpha}\right] = 1 - \Psi_{d_p}\left(\chi^2_{d_p,1-\alpha}; \frac{2(p-1)\tau^2}{p+2}\right). \tag{3.16}$$

Proof. See Section 3.7.5. □

This result in particular shows that the Bingham test is a two-sided procedure, as the asymptotic power in (3.16) exhibits a symmetric pattern with respect to girdle-type alternatives ($\tau < 0$) and bipolar alternatives ($\tau > 0$). This power, unlike the powers of the specified-$\boldsymbol{\theta}$ tests in the previous section, converges to α as p diverges to infinity, which illustrates the fact that, for larger dimensions, the Bingham test severely suffers (even asymptotically) from $\boldsymbol{\theta}$ not being specified. Note also that since the Bingham test is invariant with respect to rotations, its limiting power naturally does not depend on the location parameter $\boldsymbol{\theta}$ under the alternative.

We conducted the following simulation to check the validity of the asymptotic results of this section. In dimension $p = 3$, we generated 5000 mutually independent random samples of size $n = 2000$ from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \dots, 0)' \in \mathbb{R}^p$ and concentration $\kappa_n = \tau p/\sqrt{n}$, for $\tau = -4, -3, 0, 3, 4$. We did the same in dimension $p = 10$, with sample size $n = 10\,000$. For both dimensions p, Figure 3.2 reports kernel density estimates of the resulting values of the Bingham test statistic Q_n. They perfectly match with the corresponding asymptotic distribution in (3.15). The results also confirm the two-sided nature of the Bingham test, that, irrespective of τ_0, asymptotically behaves in the exact same way under $\tau = \pm\tau_0$.

Proposition 3.3.2 means that the Bingham test shows non-trivial asymptotic powers under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$ with $\kappa_n = \tau_n p/\sqrt{n}$ such that (τ_n) converges to τ. However, it is not Le Cam optimal for this problem. Indeed, let $(\boldsymbol{\tau}_n) \to \boldsymbol{\tau}$ in the LAN result of Theorem 3.3.1. Then, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\vartheta}_n,f}$ with $\boldsymbol{\vartheta}_n = \left(p/\sqrt{n}\right)^{1/2}\boldsymbol{\tau}_n$,

$$\check{\boldsymbol{\Delta}}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{N}_{p^2}\left(\mathbf{s}_{\boldsymbol{\tau}}, \boldsymbol{\Gamma}_p\right),$$

with $\mathbf{s}_{\boldsymbol{\tau}} = \boldsymbol{\Gamma}_p \mathrm{vec}\left(\boldsymbol{\tau}\boldsymbol{\tau}'\right)$, so that the sequence of asymptotic experiments at hand does converge to a Gaussian shift experiment $\check{\boldsymbol{\Delta}} \sim \mathcal{N}\left(\mathbf{s}_{\boldsymbol{\tau}}, \boldsymbol{\Gamma}_p\right)$ involving a *constrained* shift $\mathbf{s}_{\boldsymbol{\tau}}$. The Bingham test would be Le Cam optimal (more precisely, locally asymptotically maximin, like the low-dimensional Rayleigh test in Section 2.3.1) for an unconstrained shift $\mathbf{s} \in \mathbb{R}^{p^2}$, but is here "wasting" power against multi-spiked alternatives that are incompatible with the present single-spiked axial model. This is in line with the fact that the Bingham test, which rejects the null hypothesis of uniformity when the sample variance of the eigenvalues $\hat{\lambda}_{n1}, \ldots, \hat{\lambda}_{np}$ of $\mathbf{S}_n$ is too large, uses these eigenvalues in a permutation-invariant way. In the considered single-spiked models, it would be more natural to consider specifically $\hat{\lambda}_{n1}$ and/or $\hat{\lambda}_{np}$ to detect possible deviations from uniformity. This will be done in Section 3.4.

3.3.2 In high dimensions

It is natural to wonder whether the Bingham test also shows power against the contiguous alternatives considered in the previous section in the high-dimensional case.

Proposition 3.3.3. *Let $\left(p_n\right)$ be a sequence of positive integers diverging to ∞ and $\left(\boldsymbol{\theta}_n\right)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Let $\kappa_n = \tau_n p_n/\sqrt{n}$, with $(\tau_n) \to \tau$, and fix $f \in \mathcal{F}$. Then,* $\mathrm{Cov}\left[\mathrm{Q}_n^{\mathrm{St}}, \check{\boldsymbol{\Lambda}}_n\right] = o(1)$ *as $n \to \infty$ under $\mathrm{P}_0^{(n)}$, so that Le Cam's Third Lemma implies that $\mathrm{Q}_n^{\mathrm{St}}$ remains asymptotically standard normal under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$.*

Proof. See Section 3.7.6. □

A direct corollary is that, unlike the low-dimensional case, the Bingham test does not show power against the contiguous alternatives considered in Proposition 3.3.3. It is extremely challenging to derive a similar result to Theorem 2.3.2 with axial alternatives or even Watson distributions. We will nonetheless tackle in Chapter 4 the question of whether there exist more severe alternatives than the contiguous alternatives considered in this section that can be detected by the Bingham test.

3.4 Testing uniformity under unspecified location: single-spiked tests

A natural question is then: how to construct a test that is more powerful than the Bingham test? We will focus in this section on the low-dimensional case; the high-dimensional one is a lot trickier.

We now describe two constructions that actually lead to the same test(s). Focusing again at first on the one-sided problem involving the bipolar alternatives, we saw in Section 3.2.1 that, in the specified location case, Le Cam optimal tests of uniformity reject $\mathcal{H}_0^{(n)}: \left\{\mathrm{P}_0^{(n)}\right\}$ in favor of $\cup_{\kappa>0} \cup_{f\in\mathcal{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}$ for large values of $\check{\Delta}^{(n)}_{\boldsymbol{\theta}} = \sqrt{n}\left(p\boldsymbol{\theta}'\mathbf{S}_n\boldsymbol{\theta} - 1\right)$. In the unspecified location case, it is then natural, following [Davies, 1977, Davies, 1987, Davies, 2002], to consider the test, $\phi_+^{(n)}$ say, rejecting the null hypothesis of uniformity at asymptotic level α when

$$\mathrm{T}_+^{(n)} := \sup_{\boldsymbol{\theta}\in\mathcal{S}^{p-1}} \check{\Delta}^{(n)}_{\boldsymbol{\theta}} = \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) > c_{p,\alpha,+}, \tag{3.17}$$

where $c_{p,\alpha,+}$ is such that this test has asymptotic size α under the null hypothesis.

A similar rationale yields natural tests for the other one-sided problem and for the two-sided problem: since Le Cam optimal tests of uniformity reject $\mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\}$ in favor of $\cup_{\kappa<0} \cup_{f\in\mathcal{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}$ for large values of $-\check{\Delta}^{(n)}_{\boldsymbol{\theta}} = \sqrt{n}\left(p\boldsymbol{\theta}'\mathbf{S}_n\boldsymbol{\theta} - 1\right)$, the resulting unspecified-$\boldsymbol{\theta}$ test, $\phi_-^{(n)}$ say, will reject the null hypothesis of uniformity at asymptotic level α when

$$\mathrm{T}_-^{(n)} := \sup_{\boldsymbol{\theta}\in\mathcal{S}^{p-1}} \left(-\check{\Delta}^{(n)}_{\boldsymbol{\theta}}\right) = -\sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) > c_{p,\alpha,-}, \tag{3.18}$$

where $c_{p,\alpha,-}$ is such that this test has asymptotic size α under the null hypothesis. Finally, since Le Cam optimal tests of uniformity reject $\mathcal{H}_0^{(n)} : \left\{\mathrm{P}_0^{(n)}\right\}$ in favor of $\cup_{\kappa\neq 0} \cup_{f\in\mathcal{F}} \left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}\right\}$ for large values of $\left|\check{\Delta}^{(n)}_{\boldsymbol{\theta}}\right| = \sqrt{n}\left|p\boldsymbol{\theta}'\mathbf{S}_n\boldsymbol{\theta} - 1\right|$, the resulting unspecified-$\boldsymbol{\theta}$ test, $\phi_\pm^{(n)}$ say, will reject the null hypothesis of uniformity at asymptotic level α when

$$\mathrm{T}_\pm^{(n)} := \sup_{\boldsymbol{\theta}\in\mathcal{S}^{p-1}} \left|\check{\Delta}^{(n)}_{\boldsymbol{\theta}}\right| = \sqrt{n}\max\left\{\left|p\hat{\lambda}_{n1} - 1\right|, \left|p\hat{\lambda}_{np} - 1\right|\right\} > c_{p,\alpha,\pm}, \tag{3.19}$$

where $c_{p,\alpha,\pm}$ is still such that this test has asymptotic size α under the null hypothesis of uniformity.

Another rationale for considering the above tests is the following. For the sake of brevity, let us focus on the one-sided problem involving the bipolar alternatives, that is, the ones associated with $\kappa > 0$. A natural idea to obtain an unspecified-$\boldsymbol{\theta}$ test is to replace $\boldsymbol{\theta}$ in the corresponding optimal specified-$\boldsymbol{\theta}$ test $\phi^{(n)}_{\boldsymbol{\theta}+}$ with an estimator $\hat{\boldsymbol{\theta}}_n$. Now, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,f}$, we know from (1.7) and (1.8) that $\mathrm{E}\left[\mathbf{X}_{n1}\right] = \mathbf{0}$ and

$$\mathrm{E}\left[\mathbf{X}_{n1}\mathbf{X}_{n1}'\right] = g_f(\kappa)\,\boldsymbol{\theta}\boldsymbol{\theta}' + \frac{1 - g_f(\kappa)}{p-1}\left(\mathbf{I}_p - \boldsymbol{\theta}\boldsymbol{\theta}'\right),$$

with

$$g_f(\kappa) := \mathrm{E}_{\boldsymbol{\theta},\kappa,f}\left[\left(\mathbf{X}_{n1}'\boldsymbol{\theta}\right)^2\right] = \check{c}_{p,\kappa,f}\int_{-1}^{1}\left(1-s^2\right)^{(p-3)/2} s^2 f\left(\kappa s^2\right) ds.$$

It is easy to check that, for any $f \in \mathcal{F}$, the function $\kappa \mapsto g_f(\kappa)$ is differentiable at 0, with derivative $g_f'(0) = \mathrm{Var}_0^{(n)}\left[\left(\mathbf{X}_{n1}'\boldsymbol{\theta}\right)^2\right] > 0$, where $\mathrm{Var}_0^{(n)}$ still denotes variance under $\mathrm{P}_0^{(n)}$. Consequently, for $\kappa > 0$ small, we have $g_f(\kappa) > g_f(0) = 1/p$, so that $\boldsymbol{\theta}$ is, up to an unimportant sign (recall that only the pair $\{\pm\boldsymbol{\theta}\}$ is identifiable), the leading unit eigenvector of $\mathrm{E}\left[\mathbf{X}_{n1}\mathbf{X}_{n1}'\right]$ (for many functions f, including the Watson one $f(z) = \exp(z)$, this remains true for any $\kappa > 0$). Therefore, a moment estimator of $\boldsymbol{\theta}$ is the leading eigenvector $\hat{\boldsymbol{\theta}}_n$ of $\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^n \mathbf{X}_{ni}\mathbf{X}_{ni}'$. Note that in the Watson parametric submodel $\left\{\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa,\exp} : \kappa > 0\right\}$, this estimator $\hat{\boldsymbol{\theta}}_n$ is also the MLE of $\boldsymbol{\theta}$. The resulting test then rejects the null hypothesis of uniformity for large values of

$$\check{\Delta}^{(n)}_{\hat{\boldsymbol{\theta}}_n} := \sqrt{n}\left(p\hat{\boldsymbol{\theta}}_n'\mathbf{S}_n\hat{\boldsymbol{\theta}}_n - 1\right) = \mathrm{T}_+^{(n)},$$

hence coincides with the test $\phi_+^{(n)}$ in (3.17). A similar reasoning for the other one-sided problem leads to the test $\phi_-^{(n)}$.

The critical values in (3.17)–(3.19) above can of course be obtained from the asymptotic distribution of the corresponding test statistics under the null hypothesis. For $p = 3$, the asymptotic null distributions of $\mathrm{T}_+^{(n)}$ and $\mathrm{T}_-^{(n)}$ were obtained in [Anderson and Stephens, 1972], where the corresponding one-sided tests $\phi_+^{(n)}$ and $\phi_-^{(n)}$ were first proposed. We extend their result to the two-sided test statistic $\mathrm{T}_\pm^{(n)}$ and, more importantly, to the non-null case. The key to do so is the following result.

Theorem 3.4.1. *Fix $p \in \{2,3,\ldots\}$ and $f \in \mathscr{F}$.*
Let $\mathbf{Z}$ be a $p \times p$ random matrix such that $\mathrm{vec}\,\mathbf{Z} \sim \mathcal{N}_{p^2}\left(\mathbf{0}, \boldsymbol{\Gamma}_p\right)$, with

$$\boldsymbol{\Gamma}_p = \frac{p}{p+2}\left(\mathbf{I}_{p^2} + \mathbf{K}_p\right) - \frac{2}{p+2}\mathbf{J}_p.$$

Then,

(i) under $\mathrm{P}_0^{(n)}$,

$$\begin{pmatrix} \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) \\ \sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) \end{pmatrix} \xrightarrow{\mathscr{D}} \begin{pmatrix} \mathrm{L}_p^{\max} \\ \mathrm{L}_p^{\min} \end{pmatrix},$$

where $\mathrm{L}_{p,\max}$ (resp., $\mathrm{L}_{p,\min}$) is the largest (resp., smallest) eigenvalue of $\mathbf{Z}$;

(ii) under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$, where $\kappa_n = \tau_n p/\sqrt{n}$ involves a real sequence (τ_n) converging to τ,

$$\begin{pmatrix} \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) \\ \sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) \end{pmatrix} \xrightarrow{\mathscr{D}} \begin{pmatrix} \mathrm{L}_{p,\tau}^{\max} \\ \mathrm{L}_{p,\tau}^{\min} \end{pmatrix},$$

where $\mathrm{L}_{p,\tau}^{\max}$ (resp., $\mathrm{L}_{p,\tau}^{\min}$) is the largest (resp., smallest) eigenvalue of $\mathbf{Z}_\tau := \mathbf{Z} + \frac{2\tau}{p+2}\mathbf{W}_\tau$, with

$$\mathbf{W}_\tau := \begin{cases} \mathrm{diag}\left(p-1,-1,\ldots,-1\right) & \text{for } \tau \geq 0, \\ \mathrm{diag}\left(-1,\ldots,-1,p-1\right) & \text{for } \tau < 0. \end{cases}$$

Sketch of the proof. The complete proof of this theorem can be found in Section 3.7.7. The multivariate central limit theorem yields that under $\mathrm{P}_0^{(n)}$,

$$\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right) \xrightarrow{\mathscr{D}} \mathcal{N}_{p^2}\left(\mathbf{0}, \frac{1}{p^2}\boldsymbol{\Gamma}_p\right).$$

Theorem 3.2.1 and Le Cam's Third Lemma imply that under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$, where $\kappa_n = \tau_n p/\sqrt{n}$ is such that $\tau_n \to \tau$,

$$\sqrt{n}\,\mathrm{vec}\,(\mathbf{S}_n - \boldsymbol{\Sigma}_n) \xrightarrow{\mathscr{D}} \mathcal{N}_{p^2}\left(\mathbf{0}, \frac{1}{p^2}\boldsymbol{\Gamma}_p\right), \tag{3.20}$$

for a matrix $\boldsymbol{\Sigma}_n$ defined in the proof. Starting with $\tau \geq 0$ (the proof is similar for $\tau < 0$), the eigenvalues of $\boldsymbol{\Sigma}_n$ are $\lambda_{n1} > \lambda_{n2} = \ldots = \lambda_{np}$ and if $\boldsymbol{\Lambda}_n := \mathrm{diag}\left(\lambda_{n1}, \lambda_{n2}, \ldots, \lambda_{np}\right)$, there exists a $p \times p$ orthogonal matrix $\mathbf{G}_p$ such that $\boldsymbol{\Sigma}_n = \mathbf{G}_p\boldsymbol{\Lambda}_n\mathbf{G}_p'$. Defining

$$\xi_{ni} := \sqrt{n}p(\hat{\lambda}_{ni} - \lambda_{ni}),$$

ξ_{n1} is the largest eigenvalue of $\mathbf{Z}_n + \mathrm{diag}\,(0, -\upsilon_\tau, \ldots, -\upsilon_\tau)$, where $\mathbf{Z}_n := \sqrt{n}p\mathbf{G}_p'(\mathbf{S}_n - \boldsymbol{\Sigma}_n)\mathbf{G}_p$ and $\upsilon_\tau := 2p\tau/(p+2)$, and ξ_{np} is the smallest eigenvalue of $\mathbf{Z}_n + \mathrm{diag}\,(\upsilon_\tau, 0, \ldots, 0)$.

From (3.20), $\mathrm{vec}\,\mathbf{Z}_n \xrightarrow{\mathscr{D}} \mathrm{vec}\,\mathbf{Z}$ so that $\left(\xi_{n1}, \xi_{np}\right)' \xrightarrow{\mathscr{D}} \left(\xi_1, \xi_p\right)'$ with ξ_1 (resp. ξ_p) being the largest (resp. smallest) eigenvalue of $\mathbf{Z} + \mathrm{diag}\,(0, -\upsilon_\tau, \ldots, -\upsilon_\tau)$ (resp. $\mathbf{Z} + \mathrm{diag}\,(\upsilon_\tau, 0, \ldots, 0)$).

This implies that

$$\begin{pmatrix} \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) \\ \sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) \end{pmatrix} = \begin{pmatrix} \xi_{n1} \\ \xi_{np} \end{pmatrix} + \frac{2\tau}{p+2}\begin{pmatrix} p-1 \\ -1 \end{pmatrix} \xrightarrow{\mathscr{D}} \begin{pmatrix} \xi_1 \\ \xi_p \end{pmatrix} + \frac{2\tau}{p+2}\begin{pmatrix} p-1 \\ -1 \end{pmatrix} =: \begin{pmatrix} \eta_1 \\ \eta_p \end{pmatrix},$$

where η_1 is the largest eigenvalue of

$$\mathbf{Z} + \mathrm{diag}\,(0, -\upsilon_\tau, \ldots, -\upsilon_\tau) + \frac{2(p-1)\tau}{p+2}\mathbf{I}_p = \mathbf{Z} + \frac{2\tau}{p+2}\mathrm{diag}\left(p-1,-1,\ldots,-1\right).$$

Similarly, η_p is the smallest eigenvalue of the same matrix, $\mathbf{Z} + \frac{2\tau}{p+2}\mathrm{diag}\left(p-1,-1,\ldots,-1\right)$. □

A direct consequence of Theorem 3.4.1(i) is that simulations can be used to obtain arbitrarily precise estimates of the asymptotic critical values needed to implement the tests $\phi_+^{(n)}$, $\phi_-^{(n)}$ and $\phi_\pm^{(n)}$. For instance, the test $\phi_+^{(n)}$ will reject the null hypothesis of uniformity at asymptotic level α whenever

$$\mathrm{T}_+^{(n)} = \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) > \hat{c}_{p,\alpha,+}^{(m)},$$

where $\hat{c}_{p,\alpha,+}^{(m)}$ denotes the upper α-quantile of m independent realizations of the largest eigenvalue of $\mathbf{Z}$. Interestingly, the following corollary shows that simulations can actually be avoided in dimensions $p = 2$ and $p = 3$, as the asymptotic null distribution of $\mathrm{T}_+^{(n)}$, $\mathrm{T}_-^{(n)}$ and $\mathrm{T}_\pm^{(n)}$ can be explicitly determined for these values of p (the result for $\mathrm{T}_+^{(n)}$ and $\mathrm{T}_-^{(n)}$ in dimension $p = 3$ in (3.21) below agrees with the one from [Anderson and Stephens, 1972]).

Corollary 3.4.2.

(i) Under the null hypothesis of uniformity over S^1*, the test statistics* $\mathrm{T}_+^{(n)}$, $\mathrm{T}_-^{(n)}$*, and* $\mathrm{T}_\pm^{(n)}$ *converge weakly to* $\mathrm{L}_2^{\max}$*, where* $\mathrm{L}_2^{\max}$ *has cumulative distribution function*

$$\ell \mapsto \left(1 - \exp\left(-\ell^2\right)\right) \mathbb{I}_{\{\ell>0\}};$$

(ii) under the null hypothesis of uniformity over S^2*, the test statistics* $\mathrm{T}_+^{(n)}$ *and* $\mathrm{T}_-^{(n)}$ *converge weakly to* $\mathrm{L}_3^{\max}$*, where* $\mathrm{L}_3^{\max}$ *has cumulative distribution function*

$$\ell \mapsto \left\{\Phi\left(\sqrt{5}\ell\right) + \Phi\left(\frac{\sqrt{5}\ell}{2}\right) + 3\Phi''\left(\frac{\sqrt{5}\ell}{2}\right) - 1\right\} \mathbb{I}_{\{\ell>0\}}, \tag{3.21}$$

whereas the test statistic $\mathrm{T}_\pm^{(n)}$ *converges weakly to* $\mathrm{L}_3 := \max\left(\mathrm{L}_3^{\max}, -\mathrm{L}_3^{\min}\right)$*, where* L_3 *has cumulative distribution function*

$$\ell \mapsto \left\{2\Phi\left(\frac{\sqrt{5}\ell}{2}\right) + 6\Phi''\left(\frac{\sqrt{5}\ell}{2}\right) - 2\sqrt{3}\Phi''\left(\frac{\sqrt{5}\ell}{\sqrt{3}}\right) - 1\right\} \mathbb{I}_{\{\ell>0\}} \tag{3.22}$$

(here, Φ'' *is the second derivative of the standard normal distribution function* Φ*).*

Sketch of the proof. The complete proof of this theorem can be found in Section 3.7.8. Define

$$\boldsymbol{\ell}^{(p)} = \left(\ell_1^{(p)}, \ldots, \ell_p^{(p)}\right)',$$

where $\ell_1^{(p)} \geq \ldots \geq \ell_p^{(p)}$ are the eigenvalues of $\mathbf{Z}$ such that $\mathrm{vec}\,\mathbf{Z} \sim \mathcal{N}_{p^2}\left(\mathbf{0}, \boldsymbol{\Gamma}_p\right)$. Since $\mathrm{tr}[\mathbf{Z}] = 0$ almost surely, $\mathrm{vech}\,\mathbf{Z}$ does not admit a density and the sum of the eigenvalues of $\mathbf{Z}$ is almost surely zero and thus they do not admit a joint density over $\mathbb{R}^p$.

We consider a sequence of $p \times p$ random matrices $\mathbf{Z}_{\delta_k}$, $k = 1, 2, \ldots$, with $\delta_k > 0$ converging to zero as k goes to infinity, and such that $\mathrm{vec}\,\mathbf{Z}_\delta \sim \mathcal{N}_{p^2}\left(\mathbf{0}, \boldsymbol{\Gamma}_{p,\delta}\right)$, with

$$\boldsymbol{\Gamma}_{p,\delta} := \boldsymbol{\Gamma}_p + \frac{\delta}{p+2}\mathbf{J}_p.$$

As $\mathbf{Z}_{\delta_k} \xrightarrow{\mathcal{D}} \mathbf{Z}$ when $k \to \infty$, denoting $\ell_{1\delta}^{(p)} \geq \ldots \geq \ell_{p\delta}^{(p)}$ the eigenvalues of $\mathbf{Z}_\delta$,

$$\boldsymbol{\ell}_{\delta_k}^{(p)} = \left(\ell_{1\delta_k}^{(p)}, \ldots, \ell_{p\delta_k}^{(p)}\right)' \xrightarrow{\mathcal{D}} \boldsymbol{\ell}^{(p)} = \left(\ell_1^{(p)}, \ldots, \ell_p^{(p)}\right)'.$$

We start by computing the density of $\mathrm{vech}\,\mathbf{Z}_\delta$ for any $\delta > 0$ then that of $\mathbf{Z}_\delta$ and finally that of $\boldsymbol{\ell}_\delta^{(p)}$.

When $p = 2$, marginalisation yields the density of $\ell_{1\delta}^{(2)}$ and taking the limit, we get the one of $\ell_1^{(2)}$, which is the asymptotic distribution of $\mathrm{T}_+^{(n)}$. As $\hat{\lambda}_{n2} = 1 - \hat{\lambda}_{n1}$, $\mathrm{T}_+^{(n)} = \mathrm{T}_-^{(n)} = \mathrm{T}_\pm^{(n)}$ almost surely.

When $p = 3$, marginalisation yields the density of $\left(\ell_{1\delta}^{(3)}, \ell_{3\delta}^{(3)}\right)$ and taking the limit, we get the one of $\left(\ell_1^{(3)}, \ell_3^{(3)}\right)$. Integrating again, we obtain the density of $\ell_1^{(3)}$, which is the asymptotic distribution of $\mathrm{T}_+^{(n)}$ and $\mathrm{T}_-^{(n)}$ (see the discussion below). Finally, the asymptotic result for $\mathrm{T}_\pm^{(n)}$ stems from computing the cumulative distribution function of $\max\left(\ell_1^{(3)}, -\ell_3^{(3)}\right)$. □

Writing $\lambda_\ell(\mathbf{A})$ for the ℓth largest eigenvalue of the $p \times p$ matrix $\mathbf{A}$, Theorem 3.4.1 entails that, under the null hypothesis,

$$\mathrm{T}_+^{(n)} \xrightarrow{\mathscr{D}} \mathrm{L}_p^{\max} = \lambda_1(\mathbf{Z}) \stackrel{\mathscr{D}}{=} \lambda_1(-\mathbf{Z}) = -\lambda_p(\mathbf{Z}) = -\mathrm{L}_p^{\min} \xleftarrow{\mathscr{D}} \mathrm{T}_-^{(n)}.$$

This shows that, for any dimension p, the test statistics $\mathrm{T}_+^{(n)}$ and $\mathrm{T}_-^{(n)}$ share the same weak limit under the null hypothesis, which is confirmed in dimensions $p = 2, 3$ by Corollary 3.4.2. Maybe surprisingly, this corollary further implies that, for $p = 2$, the two-sided test statistic $\mathrm{T}_\pm^{(n)}$ has the same asymptotic null distribution as $\mathrm{T}_+^{(n)}$ and $\mathrm{T}_-^{(n)}$.

To check the validity of Theorem 3.4.1 and Corollary 3.4.2, we conducted the following numerical exercises in dimensions $p = 3$ and $p = 10$. We generated 5000 mutually independent random samples of size $n = 2000$ (for $p = 3$) and $n = 10000$ (for $p = 10$) from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa_n = \tau p/\sqrt{n}$, for $\tau = -4, -3, 0, 3, 4$. Figure 3.3 plots kernel density estimates of the resulting values of $\mathrm{T}_+^{(n)}$, $\mathrm{T}_-^{(n)}$ and $\mathrm{T}_\pm^{(n)}$, along with the densities of the corresponding asymptotic distributions; for $p = 3$ and $\tau = 0$, these densities are those associated with the distribution functions in (3.21)–(3.22), whereas, in all other cases, they are kernel density estimates obtained from 10^6 independent realizations of $\mathrm{L}_{p,\tau}^{\max}$, $-\mathrm{L}_{p,\tau}^{\min}$, and $\max\left(\mathrm{L}_{p,\tau}^{\max}, -\mathrm{L}_{p,\tau}^{\min}\right)$, respectively; see Theorem 3.4.1. Clearly, the results support our asymptotic findings. It is seen that the one-sided test $\phi_+^{(n)}$ not only shows power against the bipolar alternatives it is designed for (those associated with $\tau > 0$) but also against girdle-type ones (those associated with $\tau < 0$), which is actually desirable. The same can be said about the one-sided test $\phi_-^{(n)}$, but each of these tests, of course, shows higher powers against the alternatives it was designed for. In contrast, the two-sided test $\phi_\pm^{(n)}$ shows a symmetric power pattern for positive and negative values of τ.

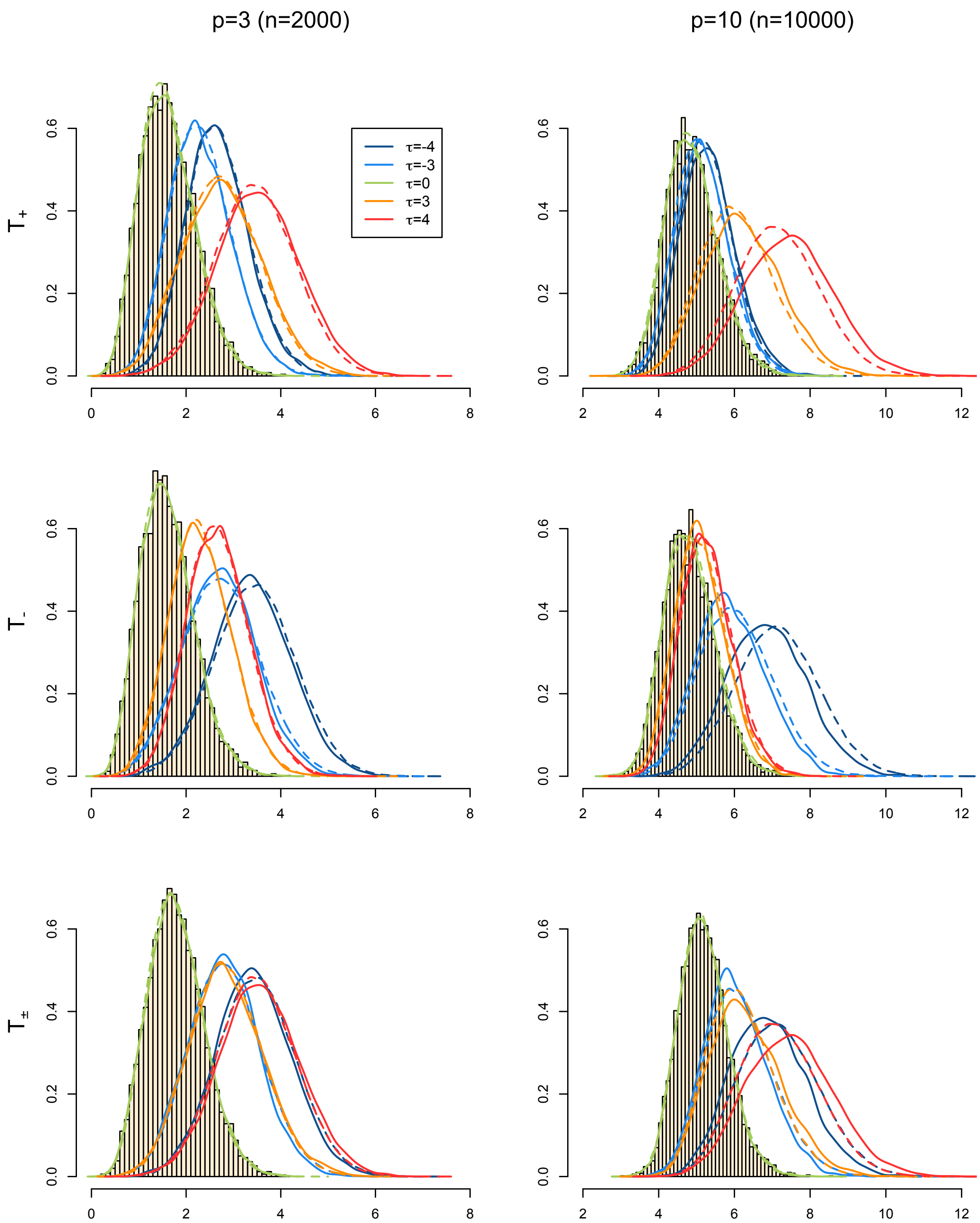


Figure 3.3 – See Section 3.4 for details.
Left: Plots of the kernel density estimates (solid curves) of (top) $T_+^{(n)} = \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right)$, (middle) $T_-^{(n)} = -\sqrt{n}\left(p\hat{\lambda}_{np} - 1\right)$ and (bottom) $T_\pm^{(n)} = \max\left(T_+^{(n)}, T_-^{(n)}\right)$, obtained from 5 000 independent random samples of size $n = 2\,000$ from the Watson distribution with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa = \tau p/\sqrt{n}$, with $\tau = -4, -3, 0, 3, 4$ and with $p = 3$; for $\tau = 0$, histograms of the corresponding test statistics are shown. The densities of the corresponding asymptotic distributions are also plotted (dashed curves).
Right: The corresponding results for $p = 10$ and $n = 10\,000$.

3.5 Finite-sample comparisons

In this section, we study the finite-sample powers of the Bingham test $\phi^{(n)}_{\mathrm{Bing}}$ and of the test $\phi^{(n)}_{+}$ (we could similarly consider the tests $\phi^{(n)}_{-}$ and $\phi^{(n)}_{\pm}$), and we compare them with those of the optimal specified-$\boldsymbol{\theta}$ test $\phi^{(n)}_{\boldsymbol{\theta}+}$. Our asymptotic results further allow us to complement these finite-sample comparisons with comparisons of the corresponding asymptotic powers.

We conducted the following Monte Carlo experiment. For any combination (n, p) of sample size $n \in \{200, 20000\}$ and dimension $p \in \{3, 10\}$, we generated collections of 2000 independent random samples of size n from the Watson distribution on $\mathcal{S}^{p-1}$ with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa_n = \tau_\ell p/\sqrt{n}$, with $\tau_\ell = 0.8\ell$, $\ell = 0, 1, \ldots, 5$. The value $\ell = 0$ corresponds to the null hypothesis of uniformity, whereas $\ell = 1, \ldots, 5$ provide increasingly severe bipolar alternatives. In each sample, we performed three tests at asymptotic level $\alpha = 5\%$, namely the specified-$\boldsymbol{\theta}$ test $\phi^{(n)}_{\boldsymbol{\theta}+}$ in (3.6), the Bingham test $\phi^{(n)}_{\mathrm{Bing}}$ in (3.14), and the test $\phi^{(n)}_{+}$ in (3.17); for $p = 3$, the asymptotic critical value for $\phi^{(n)}_{+}$ was obtained from Corollary 3.4.2(ii), whereas, for $p = 10$, an approximation of the corresponding critical value was obtained from 10 000 independent realizations of $L^{\max}_{p}$ in Theorem 3.4.1.

Figure 3.4 shows the resulting empirical powers along with their theoretical asymptotic counterparts (for any given p and τ_ℓ, the asymptotic power of $\phi^{(n)}_{+}$ was obtained from 10 000 independent copies of the random variable $L^{\max}_{p,\tau_\ell}$ in Theorem 3.4.1). The results show that, as expected, the optimal specified-$\boldsymbol{\theta}$ test outperforms both unspecified-$\boldsymbol{\theta}$ tests. The test $\phi^{(n)}_{+}$ dominates the Bingham test $\phi^{(n)}_{\mathrm{Bing}}$ and this dominance, quite intuitively, increases with the dimension p. Clearly, rejection frequencies agree very well with our asymptotic results for large sample sizes.

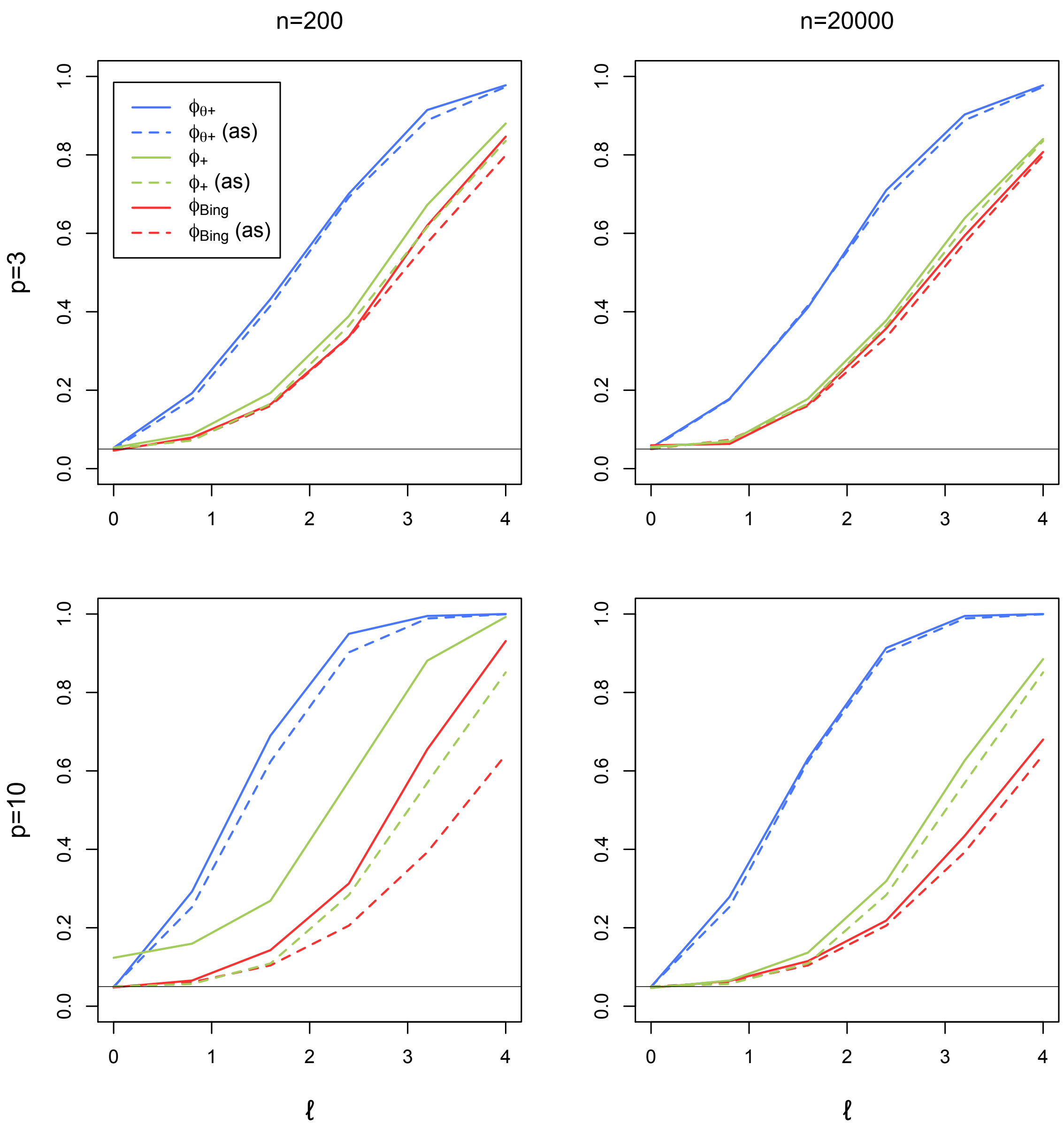


Figure 3.4 – Rejection frequencies (solid curves) of three axial tests of uniformity over S^{p-1} obtained from 2000 mutually independent random samples of size n from the Watson distribution with location $\boldsymbol{\theta} = (1,0,\ldots,0)' \in \mathbb{R}^p$ and concentration $\kappa = \tau_\ell p/\sqrt{n}$, with $\tau_0 = 0$ (null hypothesis of uniformity) and $\tau_\ell = 0.8\ell$ for $\ell = 1,2,\ldots,5$ (increasingly severe bipolar alternatives). The tests considered are the test $\phi^{(n)}_{\boldsymbol{\theta}+}$ in (3.6), the Bingham test $\phi^{(n)}_{\text{Bing}}$ in (3.14), and the test $\phi^{(n)}_{+}$ in (3.17). The corresponding asymptotic powers (dashed curves) are also provided; see Section 3.5 for details.

3.6 Applications

3.6.1 Testing for sphericity

Since the seminal paper [Ledoit and Wolf, 2002], one of the most widely considered testing problems in high-dimensional statistics is the problem of testing for sphericity. A possible approach to test for sphericity about a specified centre (without loss of generality, the origin of $\mathbb{R}^p$) is to perform a test of uniformity on the sphere $\mathcal{S}^{p-1}$ on "spatial signs", that is, on the observations projected on $\mathcal{S}^{p-1}$; see, among others, [Cai et al., 2013], where this is used in a possibly high-dimensional setup, and [Cuesta-Albertos et al., 2009], where it is argued that "*in most practical cases the violations of sphericity will arise from the non-fulfillment of uniformity on the unit sphere for projected data*". This is particularly true in the high-dimensional case, since the concentration-of-measure phenomenon there implies that information lies much more in the directions of the observations from the origin than in their distances from the origin (incidentally, note that [Juan and Prieto, 2001] also invoked the same argument to adopt a directional approach for outlier detection in high dimensions).

As shown in Chapter 2, the Rayleigh test will show power against *skewed* rotationally symmetric distributions on the sphere (skewness arises from the monotonicity of the corresponding nuisance f) and it will be blind to any non-spherical distribution in $\mathbb{R}^p$ whose projection on the sphere charges antipodal regions equally. In particular, it will show no power against elliptical alternatives, hence also against spiked alternatives (that is, against alternatives associated with scatter matrices of the form $\mathbf{\Sigma} = \sigma(\mathbf{I}_p + \lambda\boldsymbol{\beta}\boldsymbol{\beta}')$, with $\sigma, \lambda > 0$ and $\boldsymbol{\beta} \in \mathcal{S}^{p-1}$). On the contrary, the Bingham test is designed to detect elliptical or spiked alternatives, like the Gaussian sphericity test ($\phi^{(n)}_{\mathcal{N}}$, say) from [John, 1972], shown to be valid in high-dimensions in [Ledoit and Wolf, 2002].

To illustrate these antagonistic power behaviours, we performed the following simulation exercise with $n = p = 100$ involving the low-dimensional Rayleigh, $\phi^{(n)}_{\rm Ray}$, and Bingham, $\phi^{(n)}_{\rm Bing}$, tests; as mentioned in the introduction, they are HD-robust as defined in Section 2 in [Paindaveine and Verdebout, 2016], meaning that their low-dimensional versions do not need be corrected to remain valid in high dimensions. They are compared to the Gaussian sphericity test $\phi^{(n)}_{\mathcal{N}}$ in its version to test for sphericity about the origin of $\mathbb{R}^p$ and the packing/minimum angle test ($\phi^{(n)}_{\rm CFJ}$, say) from [Cai et al., 2013]. Since $n = p$ in this simulation, the asymptotic regime chosen for the latter test is the sub-exponential case.

For $\ell = 0, 1, 2, 3, 4$ and $n = p = 100$, we generated 5,000 p-dimensional independent samples $\mathbf{Z}^{(1)}_{i;\ell}$, $i = 1, \ldots, n$, and $\mathbf{Z}^{(2)}_{i;\ell}$, $i = 1, \ldots, n$, from two different alternatives to sphericity :

(i) $\mathbf{Z}^{(1)}_{i;\ell}$, $i = 1, \ldots, n$ form a random sample from the p-variate skew-normal distribution with location vector $\mathbf{0}$, scatter matrix $\mathbf{I}_p$ and skewness vector $\frac{1}{10}(\ell, \ldots, \ell)' \in \mathbb{R}^p$; see [Azzalini and Capitanio, 1999];

(ii) $\mathbf{Z}^{(2)}_{i;\ell}$, $i = 1, \ldots, n$ form a random sample from the p-variate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_p + \ell\mathbf{e}_1\mathbf{e}_1'$, with $\mathbf{e}_1 = (1, 0, \ldots, 0)' \in \mathbb{R}^p$.

For both (i)-(ii), $\ell = 0$ is associated with the null of sphericity about the origin of $\mathbb{R}^p$, whereas $\ell = 1, 2, 3, 4$ provide increasingly severe alternatives. Figure 3.5 plots the resulting empirical powers of the four tests mentioned above, all performed at nominal level 5%. Results confirm that the Rayleigh test $\phi^{(n)}_{\rm Ray}$ performs quite well under alternatives of type (i) but shows no power against alternatives of type (ii), whereas the tests $\phi^{(n)}_{\mathcal{N}}$ and $\phi^{(n)}_{\rm Bing}$ do the exact opposite. The packing test $\phi^{(n)}_{\rm CFJ}$ does not detect skew-normal alternatives and performs poorly against elliptical/spiked alternatives. In practice, thus, by comparing rejection decisions of several uniformity tests, practitioners are offered some insight on what type of deviation from sphericity they are likely to be facing.

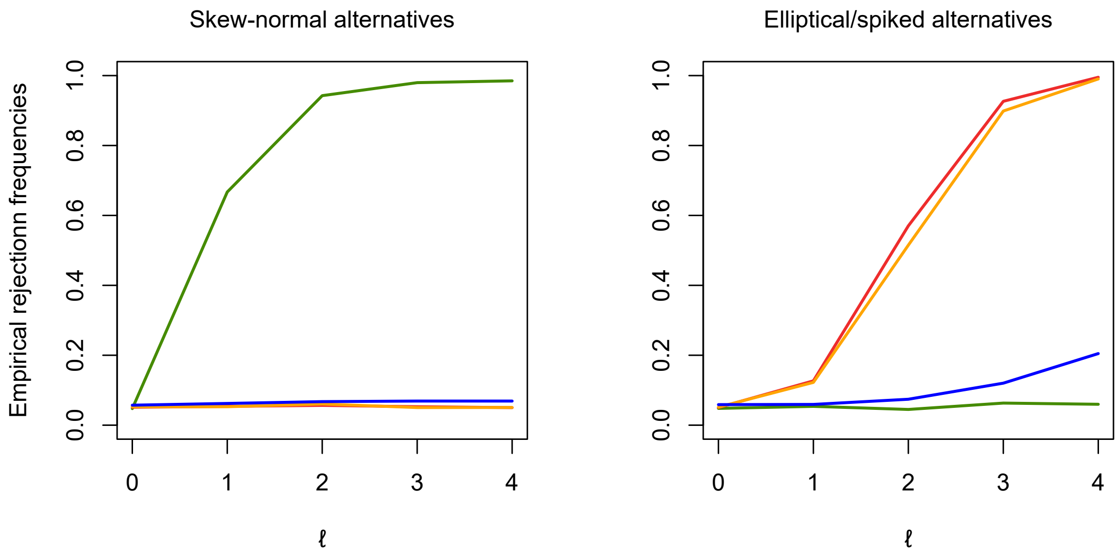


Figure 3.5 – Left: Rejection frequencies, under the null of sphericity in $\mathbb{R}^p$ ($\ell = 0$) and increasingly severe skew-normal alternatives ($\ell = 1,2,3,4$), of the Rayleigh test $\phi^{(n)}_{\text{Ray}}$ (green), the Bingham test $\phi^{(n)}_{\text{Bing}}$ (orange), the Gaussian test $\phi^{(n)}_{\mathcal{N}}$ (red) and the packing test $\phi^{(n)}_{\text{CFJ}}$ (blue)
Right: The corresponding rejection frequencies under some p-variate spiked alternatives.
In both cases, the dimension p and the sample size n are equal to 100, the nominal level is 5%, and the number of replications is 10,000; see Section 3.6.1 for details.

3.6.2 Comparison of three uniformity tests on the sphere

In order to compare the three high-dimensional tests of uniformity on the sphere at hand, for $n = 30, 100, 400$ and $p = n$, we generated 2,500 p-dimensional independent samples $\mathbf{Z}_i^{(\ell)}$, $i = 1, \ldots, n$, $\ell = 1, \ldots, 4$, from four different distributions:

(i) the components of $\mathbf{Z}_i^{(1)}$ follow independently the uniform distribution on $[-0.9, 1]$.

(ii) the components of $\mathbf{Z}_i^{(2)}$ follow the standard normal distribution with correlation 0.1;

(iii) the $\mathbf{Z}_i^{(3)}$'s have an FvML distribution on S^{p-1} with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa^{(3)} = 2p_n^{3/4}/\sqrt{n}$;

(iv) the $\mathbf{Z}_i^{(4)}$'s have a Watson distribution on S^{p-1} with location $\boldsymbol{\theta} = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and concentration $\kappa^{(4)} = p_n^{3/2}/(2\sqrt{n})$;

We then computed

$$\begin{cases} \mathbf{X}_i^{(\ell)} = \mathbf{Z}_i^{(\ell)}/\|\mathbf{Z}_i^{(\ell)}\| & \text{for } \ell = 1,2, \\ \mathbf{X}_i^{(\ell)} = \mathbf{Z}_i^{(\ell)} & \text{for } \ell = 3,4, \end{cases}$$

and applied the Rayleigh, Bingham and CFJ tests to the $\mathbf{X}_i^{(\ell)}$'s.

$p=n$	Rayleigh	Bingham	CFJ
30	0.215	0.053	0.108
100	0.998	0.094	0.103
400	1	1	0.122

(a) Case (i)

$p=n$	Rayleigh	Bingham	CFJ
30	0.062	0.722	0.213
100	0.091	1	0.979
400	0.143	1	1

(b) Case (ii)

$p=n$	Rayleigh	Bingham	CFJ
30	0.694	0.087	0.092
100	0.784	0.064	0.059
400	0.846	0.054	0.054

(c) Case (iii)

$p=n$	Rayleigh	Bingham	CFJ
30	0.069	0.894	0.231
100	0.074	1	0.508
400	0.078	1	0.921

(d) Case (iv)

Figure 3.6 – Powers of three tests of uniformity in different cases described in Section 3.6.2

When the data are not axial, like in Cases (i) and (iii), the Rayleigh test has the greatest power for all values of $n = p$ and the Bingham and CFJ test are not very efficient (even if, surprisingly, in Case (i), the power of the Bingham test leaps from 9.4% when $n = 100$ to 1 when $n = 400$). The Bingham test can detect FvML distributions but the rate chosen here, $\kappa_n \sim p_n^{3/4}/\sqrt{n}$, is below its detection threshold, $\kappa_n \sim p_n^{3/4}/n^{1/4}$). In Cases (ii) and (iv), the data are axial and as expected, it is Bingham's playground. The CFJ test fares also quite well, especially as the dimension and the number of points increase. The Rayleigh test unfortunately sees nothing.

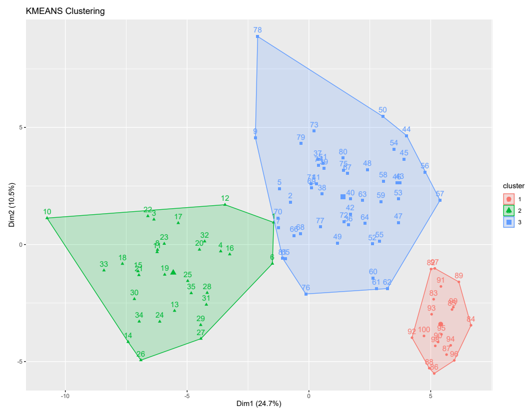
3.6.3 Yeast data set

Finally, we apply the three tests of uniformity on the sphere to a real data example. We consider the data set analyzed in [Eisen et al., 1998]: they conducted a cluster analysis of 2467 genes from a yeast, based on their expression measured at different times during several experiments (79 in total), and they identified ten clusters. We restrict to a subsample of arbitrarily 100 genes from three different clusters and we want the Rayleigh and Bingham tests to detect their presence. The data then take the form of a matrix $\mathbf{Z} = (\mathrm{Z}_{ij})$, where Z_{ij} is the jth expression value ($j = 1, \ldots, p = 79$) of the ith gene ($i = 1, \ldots, n = 100$) (even though $n > p$, the present data may be considered high-dimensional since the small value of n/p prevents relying on fixed-p asymptotic results). After imputing missing data (by replacing any missing entry in $\mathbf{Z}$ with the sample average of available measurements on the same variable), we standardize them to obtain data on $\mathcal{S}^{p-2}$. Indeed, as explained in Section 2 in [Dortet-Bernadet and Wicker, 2007], if n points in $\mathbb{R}^p$, $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$, are standardised to $\tilde{\mathbf{Z}}_1, \ldots, \tilde{\mathbf{Z}}_n$ so that

$$\sum_{j=1}^{p} \tilde{\mathrm{Z}}_{ij} = 0 \text{ and } \sum_{j=1}^{p} \tilde{\mathrm{Z}}_{ij}^2 = 1,$$

for all $i = 1, \ldots, n$, then the $\tilde{\mathbf{Z}}_i$'s lie at the intersection of a plane of dimension $p-1$ with the unit sphere of R^{p-1}. Therefore, to each sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ in $\mathbb{R}^p$ corresponds a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on $\mathcal{S}^{p-2}$. If clustering uses as a similarity index of two data points the correlation of their coordinates, this transformation does not affect this index and identifying clusters on $\mathcal{S}^{p-2}$ will prove the existence of clusters in the original data in $\mathbb{R}^p$. It can be seen in Figure 3.7: the data are projected along the two principal components and the three clusters in $\mathbb{R}^p$ also appear in $\mathcal{S}^{p-2}$.

(a) Clusters in the data in $\mathbb{R}^p$



(b) Clusters in the transformed data in $\mathcal{S}^{p-2}$

Figure 3.7

We performed the Rayleigh test $\phi^{(n)}_{\mathrm{Ray}}$, the Bingham test $\phi^{(n)}_{\mathrm{Bing}}$ and the packing test $\phi^{(n)}_{\mathrm{CFJ}}$ on the $\mathbf{X}_i$'s and they all reject uniformity with p-values so small that the software R considers them equal to 0 (the test statistics have very large absolute values, far larger than when the n points are randomly chosen among the 2467 genes).

As a comparison, if the $\mathbf{Z}_i$'s are iid normal with mean $\mathbf{0}$ and covariance matrix $\mathbf{I}_p$, the p-values of the three tests are given in Figure 3.8. They show that no test rejects uniformity on the sphere and that, as we knew, there is no cluster in the original data.

Rayleigh test	Bingham test	CFJ test
0.616	0.173	0.147

Figure 3.8 – p-values for three uniformity tests if the original data in $\mathbb{R}^p$ are multivariate standard normal

3.7 Proofs

3.7.1 Preliminary lemma

Lemma 3.7.1. *Let $g : \mathbb{R} \to \mathbb{R}$ be twice differentiable at 0. Let (p_n) be a sequence of positive integers diverging to ∞ and (κ_n) be a real sequence that is $o(p_n)$. Then,*

$$\mathrm{R}_n(g) := c_{p_n} \int_{-1}^{1} (1-s^2)^{(p_n-3)/2} g(\kappa_n s^2)\, ds = g(0) + \frac{\kappa_n}{p_n} g'(0) + \frac{3\kappa_n^2}{2p_n(p_n+2)} g''(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right)$$

as $n \to \infty$, where we let $c_p := \left(\int_{-1}^{1} (1-s^2)^{(p-3)/2}\, ds\right)^{-1}$.

Proof of Lemma 3.7.1. We know from (1.5) and (1.6) that

$$c_{p_n} \int_{-1}^{1} s^2 (1-s^2)^{(p_n-3)/2}\, ds = \frac{1}{p_n}, \tag{3.23}$$

and

$$c_{p_n} \int_{-1}^{1} s^4 (1-s^2)^{(p_n-3)/2}\, ds = \frac{3}{p_n(p_n+2)}. \tag{3.24}$$

From (3.23), we can write

$$\mathrm{R}_n(g) - g(0) - \frac{\kappa_n}{p_n} g'(0) = c_{p_n} \int_{-1}^{1} \left(1-s^2\right)^{(p_n-3)/2} \left(g\left(\kappa_n s^2\right) - g(0) - \kappa_n s^2 g'(0)\right) ds.$$

Note that we can without any loss of generality assume that (κ_n) is a sequence in $\mathbb{R}_0$, which allows us to perform the change of variables $t = |\kappa_n|^{1/2} s$. Doing so and using (3.23)–(3.24) then provides (throughout, s_{κ_n} denotes the sign of κ_n)

$$\mathrm{R}_n(g) - g(0) - \frac{\kappa_n}{p_n} g'(0) = \frac{3\kappa_n^2}{p_n(p_n+2)} \int_{-\infty}^{\infty} h_n(t) \left(\frac{g\left(s_{\kappa_n} t^2\right) - g(0) - s_{\kappa_n} t^2 g'(0)}{t^4}\right) dt,$$

or, equivalently,

$$\frac{\mathrm{R}_n(g) - g(0) - \frac{\kappa_n}{p_n} g'(0) - \frac{3\kappa_n^2}{2p_n(p_n+2)} g''(0)}{\frac{3\kappa_n^2}{p_n(p_n+2)}} = \int_{-\infty}^{\infty} h_n(t) \left(\frac{g\left(s_{\kappa_n} t^2\right) - g(0) - s_{\kappa_n} t^2 g'(0)}{t^4}\right) dt - \frac{1}{2} g''(0), \tag{3.25}$$

where h_n is defined through

$$t \mapsto h_n(t) = \frac{t^4 \left(1 - \frac{t^2}{|\kappa_n|}\right)^{(p_n-3)/2} \mathbb{I}_{\{|t| \le \sqrt{|\kappa_n|}\}}}{\int_{-\infty}^{\infty} t^4 \left(1 - \frac{t^2}{|\kappa_n|}\right)^{(p_n-3)/2} \mathbb{I}_{\{|t| \le \sqrt{|\kappa_n|}\}} \, dt}.$$

It can be checked that, since $\kappa_n = o(p_n)$, the sequence (h_n) is an *approximate* δ*-sequence* (see (1.157) in [Arfken et al., 2013]), in the sense that $\int_{-\infty}^{\infty} h_n(t)\, dt = 1$ for any n and $\int_{-\varepsilon}^{\varepsilon} h_n(t)\, dt \to 1$ for any $\varepsilon > 0$. Hence,

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} h_n(t) \left(\frac{g\left(s_{\kappa_n} t^2\right) - g(0) - s_{\kappa_n} t^2 g'(0)}{t^4}\right) dt = \lim_{t\to 0} \frac{g\left(s_{\kappa_n} t^2\right) - g(0) - s_{\kappa_n} t^2 g'(0)}{t^4},$$

which, by using L'Hôpital's rule, is equal to

$$\lim_{t\to 0} \frac{2 s_{\kappa_n} t g'\left(s_{\kappa_n} t^2\right) - 2 s_{\kappa_n} t g'(0)}{4t^3} = \frac{1}{2} \lim_{t\to 0} \frac{g'\left(s_{\kappa_n} t^2\right) - g'(0)}{s_{\kappa_n} t^2} = \frac{1}{2} g''(0)$$

Thus, (3.25) yields

$$\mathrm{R}_n(g) - g(0) - \frac{\kappa_n}{p_n} g'(0) - \frac{3\kappa_n^2}{2p_n(p_n+2)} g''(0) = o\left(\frac{\kappa_n^2}{p_n^2}\right),$$

which establishes the result. □

3.7.2 Proof of Theorem 3.1.1

In this proof, all expectations and variances are taken under the null of uniformity $\mathrm{P}_0^{(n)}$ and all stochastic convergences and o_{P}'s are as $n \to \infty$ under $\mathrm{P}_0^{(n)}$.

Consider the local log-likelihood ratio

$$
\begin{aligned}
\breve{\boldsymbol{\Lambda}}_n := \log \frac{d\breve{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}}{d\mathrm{P}^{(n)}_0} &= \sum_{i=1}^n \log \frac{\breve{c}_{p_n,\kappa_n,f}\, f\left(\kappa_n \left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right)}{c_{p_n}} \\
&= n\left(\log \frac{\breve{c}_{p_n,\kappa_n,f}}{c_{p_n}} + \breve{\mathrm{E}}_{n1}\right) + \sum_{i=1}^n \left(\log f\left(\kappa_n \left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right) - \breve{\mathrm{E}}_{n1}\right) \\
&=: \mathrm{L}_{n1} + \mathrm{L}_{n2}.
\end{aligned}
$$

Throughout, we write $\ell_{f,k} := \left(\log f\right)^k$ and $\breve{\mathrm{E}}_{nk} := \mathrm{E}\left[\ell_{f,k}\left(\kappa_n \left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right)\right]$. Lemma 3.7.1 readily yields

$$
\begin{aligned}
\log \frac{\breve{c}_{p_n,\kappa_n,f}}{c_{p_n}} &= -\log\left(c_{p_n}\int_{-1}^{1} \left(1-s^2\right)^{(p_n-3)/2} f\left(\kappa_n s^2\right) ds\right) \\
&= -\log\left(1 + \frac{\kappa_n}{p_n} + \frac{3\kappa_n^2}{2p_n\left(p_n+2\right)} f''(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right)\right) \\
&= -\frac{\kappa_n}{p_n} - \frac{3\kappa_n^2}{2p_n\left(p_n+2\right)} f''(0) + \frac{\kappa_n^2}{2{p_n}^2} + o\left(\frac{\kappa_n^2}{p_n^2}\right). \qquad (3.26)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\breve{\mathrm{E}}_{n1} &= c_{p_n}\int_{-1}^{1} \left(1-s^2\right)^{(p_n-3)/2} \ell_{f,1}\left(\kappa_n s^2\right) ds \\
&= \frac{\kappa_n}{p_n}\ell'_{f,1}(0) + \frac{3\kappa_n^2}{2p_n\left(p_n+2\right)}\ell''_{f,1}(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right) \\
&= \frac{\kappa_n}{p_n} + \frac{3\kappa_n^2}{2p_n\left(p_n+2\right)}\left(f''(0)-1\right) + o\left(\frac{\kappa_n^2}{p_n^2}\right). \qquad (3.27)
\end{aligned}
$$

Combining (3.26) and (3.27) provides

$$
\mathrm{L}_{n1} = \frac{n\kappa_n^2}{2{p_n}^2} - \frac{3n\kappa_n^2}{2p_n\left(p_n+2\right)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right) = -\frac{n\kappa_n^2\left(p_n-1\right)}{p_n^2\left(p_n+2\right)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right).
$$

Turning to L_{n2}, write

$$
\mathrm{L}_{n2} = \sqrt{n\breve{\mathrm{V}}_n}\sum_{i=1}^n \breve{\mathrm{W}}_{ni} := \sqrt{n\breve{\mathrm{V}}_n}\sum_{i=1}^n \frac{\log f\left(\kappa_n \left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right) - \breve{\mathrm{E}}_{n1}}{\sqrt{n\breve{\mathrm{V}}_n}},
$$

where we let $\breve{\mathrm{V}}_n := \mathrm{Var}\left[\log f\left(\kappa_n \left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right)\right]$. First note that, since

$$
\begin{aligned}
\breve{\mathrm{E}}_{n2} &= c_{p_n}\int_{-1}^{1} \left(1-s^2\right)^{(p_n-3)/2} \ell_{f,2}\left(\kappa_n s^2\right) ds \\
&= \frac{\kappa_n}{p_n}\ell'_{f,2}(0) + \frac{3\kappa_n^2}{2p_n\left(p_n+2\right)}\ell''_{f,2}(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right) \\
&= \frac{3\kappa_n^2}{p_n\left(p_n+2\right)} + o\left(\frac{\kappa_n^2}{p_n^2}\right), \qquad (3.28)
\end{aligned}
$$

we have

$$
n\breve{\mathrm{V}}_n = n\left(\breve{\mathrm{E}}_{n2} - \breve{\mathrm{E}}_{n1}^2\right) = \frac{2n\kappa_n^2\left(p_n-1\right)}{{p_n}^2\left(p_n+2\right)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right), \qquad (3.29)
$$

which leads to

$$\breve{\Lambda}_n = -\frac{n\kappa_n^2\left(p_n-1\right)}{p_n{}^2\left(p_n+2\right)} + \sqrt{\frac{2n\kappa_n^2\left(p_n-1\right)}{p_n{}^2\left(p_n+2\right)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right)}\sum_{i=1}^n \breve{W}_{ni} + o\left(\frac{n\kappa_n^2}{p_n^2}\right). \tag{3.30}$$

Since the $\breve{W}_{ni}$'s, $i = 1,\ldots,n$, are mutually independent with mean zero and variance $1/n$, we obtain that

$$E\left[\breve{\Lambda}_n^2\right] = E\left[\breve{\Lambda}_n\right]^2 + \text{Var}\left[\breve{\Lambda}_n\right] = \frac{n^2\kappa_n^4\left(p_n-1\right)^2}{p_n{}^4\left(p_n+2\right)^2} + o\left(\frac{n^2\kappa_n^4}{p_n^4}\right) + \frac{2n\kappa_n^2\left(p_n-1\right)}{p_n{}^2\left(p_n+2\right)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right). \tag{3.31}$$

If $\kappa_n = o\left(p_n/\sqrt{n}\right)$, then (3.31) implies that $\exp\left(\breve{\Lambda}_n\right) \xrightarrow{\mathcal{D}} 1$, so that Le Cam's First Lemma (see Section 1.2.6) yields that $\breve{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $P^{(n)}_0$ are mutually contiguous.

We may therefore assume that $\kappa_n = \tau_n p_n/\sqrt{n}$, where (τ_n) is $O(1)$ but not $o(1)$. Then, (3.30) rewrites

$$\breve{\Lambda}_n = -\frac{p_n-1}{p_n+2}\tau_n^2 + \sqrt{\frac{2\left(p_n-1\right)}{p_n+2}\tau_n^2 + o(1)}\sum_{i=1}^n \breve{W}_{ni} + o(1). \tag{3.32}$$

Applying the Cauchy–Schwarz inequality and the Chebychev inequality, then using Lemma 3.7.1 and (3.29), yields that, for some positive constant C and any $\varepsilon > 0$,

$$\begin{aligned}
\sum_{i=1}^n E\left[\breve{W}_{ni}^2 \mathbb{I}_{\{|\breve{W}_{ni}|>\varepsilon\}}\right] &\leq n\sqrt{E\left[\breve{W}_{ni}^4\right] P\left[|\breve{W}_{ni}|>\varepsilon\right]} \\
&\leq \frac{n}{\varepsilon}\sqrt{E\left[\breve{W}_{ni}^4\right]\text{Var}\left[\breve{W}_{ni}\right]} = \frac{1}{\varepsilon}\sqrt{nE\left[\breve{W}_{ni}^4\right]} \\
&\leq \frac{C\sqrt{n\breve{E}_{n4}}}{\varepsilon n\breve{V}_n} = \frac{C\left(\frac{n\kappa_n}{p_n}\ell'_{f,4}(0) + \frac{3n\kappa_n^2}{2p_n(p_n+2)}\ell''_{f,4}(0) + o\left(\frac{n\kappa_n^2}{p_n^2}\right)\right)^{1/2}}{\varepsilon\left(\frac{2n\kappa_n^2(p_n-1)}{p_n{}^2(p_n+2)} + o\left(\frac{n\kappa_n^2}{p_n^2}\right)\right)} \\
&= \frac{o(\tau_n)}{\varepsilon\left(\frac{2(p_n-1)}{p_n+2}\tau_n^2 + o\left(\tau_n^2\right)\right)} = o(1),
\end{aligned}$$

where we have used the fact that $\ell'_{f,4}(0) = \ell''_{f,4}(0) = 0$. This shows that $\sum_{i=1}^n \breve{W}_{ni}$ satisfies the classical Lévy–Lindeberg condition, hence is asymptotically standard normal (as already mentioned, the $\breve{W}_{ni}$'s, $i = 1,\ldots,n$, are mutually independent with mean zero and variance $1/n$). For any subsequence $\left(\exp\left(\breve{\Lambda}_{n_m}\right)\right)$ converging in distribution, the weak limit must then be $\exp(Y)$, with $Y \sim \mathcal{N}\left(-\eta, 2\eta\right)$, where we let

$$\eta := \begin{cases} \frac{p-1}{p+2}\lim_{m\to\infty}\tau_{n_m}^2 & \text{in the low-dimensional case}(p_n = p), \\ \lim_{m\to\infty}\tau_{n_m}^2 & \text{in the high-dimensional case } (p_n\to\infty). \end{cases}$$

Mutual contiguity of $\breve{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$ and $P^{(n)}_0$ then follows from the fact that $P[\exp(Y) = 0] = 0$ and $E\left[\exp(Y)\right] = 1$.

□

3.7.3 Proof of Theorems 3.2.1 and 3.2.2

As in the proof of Theorem 3.1.1, all expectations and variances in this proof are taken under the null of uniformity $\mathrm{P}_0^{(n)}$ and all stochastic convergences and o_P's are as $n \to \infty$ under $\mathrm{P}_0^{(n)}$.

In the low-dimensional case we will prove that Theorem 3.2.1 is actually true if the fixed location $\boldsymbol{\theta}$ is replaced with an n-dependent value $\boldsymbol{\theta}_n$. Define

$$\Gamma_p := \begin{cases} \frac{2(p-1)}{p+2} & \text{in the low-dimensional case } (p_n = p), \\ 2 & \text{in the high-dimensional case } (p_n \to \infty). \end{cases}$$

The central limit theorem directly establishes the second part of the result, since

$$\mathrm{E}\left[\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n}\right] = 0,$$
$$\mathrm{Var}\left[\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n}\right] = \frac{2\left(p_n - 1\right)}{p_n + 2}.$$

To prove the first part let $\left(p_n\right)$ be a sequence in $\{2, 3, \ldots\}$ and $\kappa_n = \tau_n p_n / \sqrt{n}$, where the real sequence (τ_n) is $\mathrm{O}(1)$. If (τ_n) is $o(1)$, then (3.30) implies that $\breve{\boldsymbol{\Lambda}}_n = o_\mathrm{P}(1)$, which proves the result in this case.

We may thus assume that (τ_n) is $\mathrm{O}(1)$ but not $o(1)$. Equation (3.32) can then be written as

$$\breve{\boldsymbol{\Lambda}}_n = -\frac{\tau_n^2}{2}\Gamma_p + |\tau_n| \Gamma_p^{1/2} \sum_{i=1}^n \breve{\mathrm{W}}_{ni} + o_\mathrm{P}(1),$$

where $\sum_{i=1}^n \breve{\mathrm{W}}_{ni}$ is asymptotically standard normal. Since $\left(\tau_n \Gamma_p^{1/2}\right)$ is $\mathrm{O}(1)$, it is therefore sufficient to show that

$$d_n := \mathrm{E}\left[\left(\breve{\Delta}^{(n)}_{\boldsymbol{\theta}_n} - \Gamma_p^{1/2} s_{\tau_n} \sum_{i=1}^n \breve{\mathrm{W}}_{ni}\right)^2\right] = o(1), \tag{3.33}$$

where s_a is the sign of real number a. To prove (3.33), define

$$\begin{aligned} \breve{\mathrm{M}}_n &:= \sqrt{n\breve{\mathrm{V}}_n}\left(\breve{\Delta}_{\boldsymbol{\theta}_n} - \Gamma_p^{1/2} s_{\tau_n} \sum_{i=1}^n \breve{\mathrm{W}}_{ni}\right) \\ &= \sum_{i=1}^n \left(p_n \sqrt{\breve{\mathrm{V}}_n}\left(\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n\right)^2 - \frac{1}{p_n}\right) - \Gamma_p^{1/2} s_{\tau_n}\left(\log f\left(\kappa_n \left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n\right)^2\right) - \breve{\mathrm{E}}_{n1}\right).\right) \end{aligned} \tag{3.34}$$

Then using (1.5) and (1.6) and noting that

$$\frac{2\left(p_n - 1\right)}{p_n + 2} = \begin{cases} \Gamma_p & \text{in the low-dimensional case } (p_n = p), \\ \Gamma_p + o(1) & \text{in the high-dimensional case } (p_n \to \infty), \end{cases}$$

we obtain

$$\begin{aligned} \mathrm{E}\left[\breve{\mathrm{M}}_n^2\right] &= n\mathrm{E}\left[\left(p_n\sqrt{\breve{\mathrm{V}}_n}\left(\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2 - \frac{1}{p_n}\right) - \Gamma_p^{1/2} s_{\tau_n}\left(\log f\left(\kappa_n\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2\right) - \breve{\mathrm{E}}_{n1}\right)\right)^2\right] \\ &= \left(2\Gamma_p + o(1)\right) n\breve{\mathrm{V}}_n - 2np_n\sqrt{\breve{\mathrm{V}}_n}\Gamma_p^{1/2} s_{\tau_n}\mathrm{E}\left[\left(\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2 - \frac{1}{p_n}\right)\left(\log f\left(\kappa_n\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2\right) - \breve{\mathrm{E}}_{n1}\right)\right] \\ &= 2\Gamma_p n\breve{\mathrm{V}}_n - 2\sqrt{\Gamma_p n\breve{\mathrm{V}}_n} s_{\tau_n}\left(\frac{\sqrt{n}p_n}{\kappa_n}\mathrm{E}\left[g\left(\kappa_n\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2\right)\right] - \sqrt{n}\breve{\mathrm{E}}_{n1}\right) + o\left(n\breve{\mathrm{V}}_n\right), \end{aligned} \tag{3.35}$$

where we let $g(x) := x\log f(x)$. Lemma 3.7.1 provides

$$\mathrm{E}\left[g\left(\kappa_n\left(\mathbf{X}_{n1}'\boldsymbol{\theta}_n\right)^2\right)\right] = c_{p_n}\int_{-1}^{1}\left(1-s^2\right)^{(p_n-3)/2} g\left(\kappa_n s^2\right)\,ds = \frac{3\kappa_n^2}{p_n\left(p_n+2\right)} + o\left(\frac{\kappa_n^2}{p_n^2}\right).$$

Using this jointly with (3.27) and (3.29), it follows from (3.34)–(3.35) that

$$\begin{aligned} d_n &= 2\Gamma_p - \frac{2\sqrt{\Gamma_p}s_{\tau_n}\left(\frac{3\sqrt{n}\kappa_n}{p_n+2}+o\left(\frac{\sqrt{n}\kappa_n}{p_n}\right)-\left(\frac{\sqrt{n}\kappa_n}{p_n}+\frac{3\sqrt{n}\kappa_n^2}{2p_n(p_n+2)}\left(f''(0)-1\right)+o\left(\frac{\sqrt{n}\kappa_n^2}{p_n^2}\right)\right)\right)}{\sqrt{\frac{2n\kappa_n^2(p_n-1)}{{p_n}^2(p_n+2)}+o\left(\frac{n\kappa_n^2}{p_n^2}\right)}} + o(1) \\ &= 2\Gamma_p - \frac{2\sqrt{\Gamma_p}s_{\tau_n}\left(\left(\Gamma_p+o(1)\right)\tau_n+o(1)\right)}{\sqrt{\left(\Gamma_p+o(1)\right)\tau_n^2+o(1)}} + o(1) \\ &= o(1), \end{aligned}$$

as was to be shown.

□

3.7.4 Proof of Theorem 3.3.1

The parameter value $\boldsymbol{\vartheta}_n = \left(p/\sqrt{n}\right)^{1/2}\boldsymbol{\tau}_n$ corresponds to $\boldsymbol{\theta}_n = \boldsymbol{\tau}_n/\|\boldsymbol{\tau}_n\|$ and $\kappa_n = p\|\boldsymbol{\tau}_n\|^2/\sqrt{n}$ in (3.4) (the fixed location $\boldsymbol{\theta}$ can be replaced with an n-dependent value $\boldsymbol{\theta}_n$ — see proof in Section 3.7.3) so that

$$\log\frac{d\check{\mathrm{P}}^{(n)}_{\boldsymbol{\vartheta}_n,f}}{d\mathrm{P}^{(n)}_0} = \log\frac{d\check{\mathrm{P}}^{(n)}_{\boldsymbol{\tau}_n/\|\boldsymbol{\tau}_n\|,p\|\boldsymbol{\tau}_n\|^2/\sqrt{n},f}}{d\mathrm{P}^{(n)}_0} = \|\boldsymbol{\tau}_n\|^2\check{\Delta}^{(n)}_{\boldsymbol{\tau}_n/\|\boldsymbol{\tau}_n\|} - \frac{\|\boldsymbol{\tau}_n\|^4}{2}\Gamma_p + o_{\mathrm{P}}(1) \tag{3.36}$$

under $\mathrm{P}^{(n)}_0$. Now, since vec(ABC) = (C′ ⊗ A)vec(B), the central sequence in Theorem 3.2.1 rewrites

$$\begin{aligned} \check{\Delta}^{(n)}_{\boldsymbol{\theta}} &= \frac{p}{\sqrt{n}}\sum_{i=1}^{n}\left\{\left(\mathbf{X}_{ni}'\boldsymbol{\theta}\right)^2-\frac{1}{p}\right\} = p\sqrt{n}\boldsymbol{\theta}'\left(\mathbf{S}_n-\frac{1}{p}\mathbf{I}_p\right)\boldsymbol{\theta} \\ &= p\sqrt{n}\left(\boldsymbol{\theta}\otimes\boldsymbol{\theta}\right)'\mathrm{vec}\left(\mathbf{S}_n-\frac{1}{p}\mathbf{I}_p\right) = \left(\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right)\right)'\check{\boldsymbol{\Delta}}^{(n)} \end{aligned} \tag{3.37}$$

and the log-likelihood ratio in (3.36) becomes

$$\log\frac{d\check{\mathrm{P}}^{(n)}_{\boldsymbol{\vartheta}_n,f}}{d\mathrm{P}^{(n)}_0} = \left(\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right)\right)'\check{\boldsymbol{\Delta}}^{(n)} - \frac{\|\boldsymbol{\tau}_n\|^4}{2}\Gamma_p + o_{\mathrm{P}}(1)$$

under $\mathrm{P}^{(n)}_0$. By using the identities $(\mathrm{vec}\,\mathbf{A})'(\mathrm{vec}\,\mathbf{B}) = \mathrm{tr}[\mathbf{A}'\mathbf{B}]$ and $\mathbf{K}_p(\mathrm{vec}\,\mathbf{A}) = \mathrm{vec}\left(\mathbf{A}'\right)$, straightforward calculations yield $\|\boldsymbol{\tau}_n\|^4\Gamma_p = \left(\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right)\right)'\boldsymbol{\Gamma}_p\mathrm{vec}\left(\boldsymbol{\tau}_n\boldsymbol{\tau}_n'\right)$, which establishes (3.13). Since the asymptotic normality result readily follows from the multivariate central limit theorem, the theorem is proved. □

3.7.5 Proof of Proposition 3.3.2

Denoting as $\mathrm{E}^{(n)}_0$ and $\mathrm{Var}^{(n)}_0$ expectation and variance under $\mathrm{P}^{(n)}_0$, one has (see (3.37))

$$\begin{aligned} \lim_{n\to\infty}\mathrm{E}^{(n)}_0\left[\check{\Delta}^{(n)}_{\boldsymbol{\theta}}\check{\boldsymbol{\Delta}}^{(n)}\right] &= \lim_{n\to\infty}\mathrm{E}^{(n)}_0\left[\check{\boldsymbol{\Delta}}^{(n)}\left(\left(\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right)\right)'\check{\boldsymbol{\Delta}}^{(n)}\right)'\right] \\ &= \lim_{n\to\infty}\mathrm{Var}^{(n)}_0\left[\check{\boldsymbol{\Delta}}^{(n)}\right]\left(\boldsymbol{\theta}\otimes\boldsymbol{\theta}\right) = \boldsymbol{\Gamma}_p\left(\boldsymbol{\theta}\otimes\boldsymbol{\theta}\right). \end{aligned}$$

Therefore, Le Cam's Third Lemma implies that, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$ with $\kappa_n = \tau p/\sqrt{n}$, $\tau \neq 0$, $\check{\boldsymbol{\Delta}}^{(n)}$ is asymptotically normal with mean vector $\tau\boldsymbol{\Gamma}_p(\boldsymbol{\theta}\otimes\boldsymbol{\theta})$ and covariance matrix $\boldsymbol{\Gamma}_p$, so that, under the same sequence of hypotheses (see Theorem 1.4.2 in [Muirhead, 1982]),

$$\mathrm{Q}_n = \left(\check{\boldsymbol{\Delta}}^{(n)}\right)'\boldsymbol{\Gamma}_p^{-}\check{\Delta}^{(n)} \xrightarrow{\mathcal{D}} \chi^2_{d_p}(\delta),$$

with

$$\delta = \tau^2(\boldsymbol{\theta}\otimes\boldsymbol{\theta})'\boldsymbol{\Gamma}_p\boldsymbol{\Gamma}_p^{-}\boldsymbol{\Gamma}_p(\boldsymbol{\theta}\otimes\boldsymbol{\theta}) = \frac{p\tau^2}{p+2}\left(2-\frac{2}{p}\right) = \frac{2(p-1)\tau^2}{p+2}.$$

The asymptotic power in (3.16) readily follows. □

3.7.6 Proof of Proposition 3.3.3

The proof of Proposition 3.3.3 requires the following lemma.

Lemma 3.7.2. *If* $\mathbf{X}_1,\ldots,\mathbf{X}_n$ *is a random sample from the uniform distribution over* $\mathcal{S}^{p-1}$, *then*

$$\mathrm{E}\left[\mathrm{tr}\,[\mathbf{S}_n^2]\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right)\right] = \mathbf{0}.$$

Proof of Lemma 3.7.2. First note that

$$\begin{aligned}\mathrm{tr}\,[\mathbf{S}_n^2]\left(\mathbf{S}_n-\frac{1}{p}\mathbf{I}_p\right) &= \frac{1}{n^3}\sum_{i,j,k=1}^{n}\mathrm{tr}\left[\mathbf{X}_i\mathbf{X}_i'\mathbf{X}_j\mathbf{X}_j'\right]\left(\mathbf{X}_k\mathbf{X}_k'-\frac{1}{p}\mathbf{I}_p\right)\\ &= \frac{1}{n^3}\sum_{i,j,k=1}^{n}\left\{(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{X}_k\mathbf{X}_k' - \frac{1}{p}(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{I}_p\right\}.\end{aligned} \quad (3.38)$$

For $i \neq j$, we have, irrespective of whether $k \in \{i,j\}$ or not,

$$\mathrm{E}\left[(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{X}_k\mathbf{X}_k'\right] = \mathrm{E}\left[\mathrm{E}\left[(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{X}_k\mathbf{X}_k'|\mathbf{X}_k\right]\right] = \frac{1}{p}\mathrm{E}[\mathbf{X}_k\mathbf{X}_k'] = \frac{1}{p^2}\mathbf{I}_p = \mathrm{E}\left[\frac{1}{p}(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{I}_p\right],$$

whereas, for $i = j$, we trivially have

$$\mathrm{E}\left[(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{X}_k\mathbf{X}_k'\right] = \mathrm{E}\left[\mathbf{X}_k\mathbf{X}_k'\right] = \frac{1}{p}\mathbf{I}_p = \mathrm{E}\left[\frac{1}{p}(\mathbf{X}_i'\mathbf{X}_j)^2\mathbf{I}_p\right].$$

The result thus follows from (3.38). □

Proof of Proposition 3.3.3. Theorem 3.2.2 implies that, as $n\to\infty$ under $\mathrm{P}_0^{(n)}$,

$$\begin{aligned}\mathrm{Cov}\,[\mathrm{Q}_n^{\mathrm{St}},\check{\boldsymbol{\Lambda}}_n] &= \frac{\tau_n}{\sqrt{2d_{p_n}}}\mathrm{E}\,[\mathrm{Q}_n\check{\Delta}_{\boldsymbol{\theta}_n}] + o(1) = \frac{np(p+2)\tau_n}{2\sqrt{2d_{p_n}}}\mathrm{E}[\mathrm{tr}\,[\mathbf{S}_n^2]\,\check{\Delta}_{\boldsymbol{\theta}_n}] + o(1)\\ &= \frac{n^{3/2}p^2(p+2)\tau_n}{2\sqrt{2d_{p_n}}}(\boldsymbol{\theta}_n\otimes\boldsymbol{\theta}_n)'\,\mathrm{E}\left[\mathrm{tr}\,[\mathbf{S}_n^2]\,\mathrm{vec}\left(\mathbf{S}_n-\frac{1}{p}\mathbf{I}_p\right)\right] + o(1),\end{aligned} \quad (3.39)$$

so that Lemma 3.7.2 yields that this covariance is $o(1)$ as $n\to\infty$. Since $\mathrm{Q}_n^{\mathrm{St}}$ is asymptotically standard normal under $\mathrm{P}_0^{(n)}$ (see (3.2)), Le Cam's Third Lemma then entails that $\mathrm{Q}_n^{\mathrm{St}}$ remains asymptotically standard normal under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$. □

3.7.7 Proof of Theorem 3.4.1

Using (1.2) and (1.3), the multivariate central limit theorem yields that under $\mathrm{P}_0^{(n)}$,

$$\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right) \xrightarrow{\mathscr{D}} \mathscr{N}_{p^2}\left(\mathbf{0}, \frac{1}{p^2}\boldsymbol{\Gamma}_p\right).$$

Now, by using (3.37), we obtain that, under $\mathrm{P}_0^{(n)}$,

$$\begin{aligned}\mathrm{E}\left[\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right)\breve{\Delta}_{\boldsymbol{\theta}}^{(n)}\right] &= np\mathrm{E}\left[\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right)\mathrm{vec}'\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right)\right]\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right)\\ &= p\left(\frac{1}{p^2}\boldsymbol{\Gamma}_p\right)\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right) = \frac{2}{p+2}\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right) - \frac{2}{p\left(p+2\right)}\mathrm{vec}\left(\mathbf{I}_p\right),\end{aligned}$$

so that Le Cam's Third Lemma shows that, under $\breve{\mathrm{P}}_{\boldsymbol{\theta},\kappa_n,f}^{(n)}$, where $\kappa_n = \tau_n p/\sqrt{n}$ is based on a sequence (τ_n) converging to τ,

$$\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \frac{1}{p}\mathbf{I}_p\right) \xrightarrow{\mathscr{D}} \mathscr{N}_{p^2}\left(\frac{2\tau}{p+2}\mathrm{vec}\left(\boldsymbol{\theta}\boldsymbol{\theta}'\right) - \frac{2\tau}{p\left(p+2\right)}\mathrm{vec}\left(\mathbf{I}_p\right), \frac{1}{p^2}\boldsymbol{\Gamma}_p\right),$$

which rewrites

$$\sqrt{n}\,\mathrm{vec}\left(\mathbf{S}_n - \boldsymbol{\Sigma}_n\right) \xrightarrow{\mathscr{D}} \mathscr{N}_{p^2}\left(\mathbf{0}, \frac{1}{p^2}\boldsymbol{\Gamma}_p\right), \tag{3.40}$$

where

$$\begin{aligned}\boldsymbol{\Sigma}_n &:= \left(\frac{1}{p} - \frac{2\tau}{\sqrt{n}p\left(p+2\right)}\right)\mathbf{I}_p + \frac{2\tau}{\sqrt{n}\left(p+2\right)}\boldsymbol{\theta}\boldsymbol{\theta}'\\ &= \left(\frac{1}{p} + \frac{2\left(p-1\right)\tau}{\sqrt{n}p\left(p+2\right)}\right)\boldsymbol{\theta}\boldsymbol{\theta}' + \left(\frac{1}{p} - \frac{2\tau}{\sqrt{n}p\left(p+2\right)}\right)\left(\mathbf{I}_p - \boldsymbol{\theta}\boldsymbol{\theta}'\right).\end{aligned}$$

We need to consider the cases (a) $\tau \geq 0$ and (b) $\tau < 0$ separately.

(a) $\boldsymbol{\Sigma}_n$ has eigenvalues

$$\lambda_{n1} = \frac{1}{p} + \frac{2\left(p-1\right)\tau}{\sqrt{n}p\left(p+2\right)} \quad \text{and} \quad \lambda_{n2} = \ldots = \lambda_{np} = \frac{1}{p} - \frac{2\tau}{\sqrt{n}p\left(p+2\right)}.$$

Starting with $\hat{\lambda}_{n1}$, we can do the decomposition

$$\sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) = \sqrt{n}p\left(\hat{\lambda}_{n1} - \lambda_{n1}\right) + \sqrt{n}\left(p\lambda_{n1} - 1\right) =: \xi_{n1} + \frac{2(p-1)\tau}{p+2}.$$

Letting $\boldsymbol{\Lambda}_n := \mathrm{diag}\left(\lambda_{n1}, \lambda_{n2}, \ldots, \lambda_{np}\right)$, there exists a $p \times p$ orthogonal matrix $\mathbf{G}_p$ such that $\boldsymbol{\Sigma}_n = \mathbf{G}_p\boldsymbol{\Lambda}_n\mathbf{G}_p'$. Clearly, ξ_{n1} is the largest eigenvalue of

$$\sqrt{n}p\left(\mathbf{S}_n - \lambda_{n1}\mathbf{I}_p\right) = \sqrt{n}p\left(\mathbf{S}_n - \boldsymbol{\Sigma}_n\right) + \sqrt{n}p\left(\boldsymbol{\Sigma}_n - \lambda_{n1}\mathbf{I}_p\right)$$

and therefore of

$$\sqrt{n}p\mathbf{G}_p'(\mathbf{S}_n - \boldsymbol{\Sigma}_n)\mathbf{G}_p + \sqrt{n}p(\boldsymbol{\Lambda}_n - \lambda_{n1}\mathbf{I}_p) =: \mathbf{Z}_n + \mathrm{diag}\left(0, -\upsilon_\tau, \ldots, -\upsilon_\tau\right),$$

where

$$\upsilon_\tau := \sqrt{n}p\left[\frac{2\left(p-1\right)\tau}{\sqrt{n}p\left(p+2\right)} + \frac{2\tau}{\sqrt{n}p\left(p+2\right)}\right] = \frac{2p\tau}{p+2}.$$

Similarly, $\xi_{np} := \sqrt{n}p\left(\hat{\lambda}_{np} - \lambda_{np}\right)$ is the smallest eigenvalue of

$$\mathbf{Z}_n + \operatorname{diag}(v_\tau, 0, \ldots, 0).$$

Note that (3.40) readily implies that $\operatorname{vec}(\mathbf{Z}_n) = \sqrt{n}p\left(\mathbf{G}_p \otimes \mathbf{G}_p\right)' \operatorname{vec}(\mathbf{S}_n - \boldsymbol{\Sigma}_n)$ converges weakly to $\operatorname{vec}(\mathbf{Z}) \sim \mathcal{N}_{p^2}\left(\mathbf{0}, \boldsymbol{\Gamma}_p\right)$. It readily follows that $\left(\xi_{n1}, \xi_{np}\right)' \xrightarrow{\mathcal{D}} \left(\xi_1, \xi_p\right)'$, where ξ_1 is the largest eigenvalue of $\mathbf{Z} + \operatorname{diag}(0, -v_\tau, \ldots, -v_\tau)$ and ξ_p is the smallest eigenvalue of $\mathbf{Z} + \operatorname{diag}(v_\tau, 0, \ldots, 0)$. This implies that

$$\begin{pmatrix} \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) \\ \sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) \end{pmatrix} = \begin{pmatrix} \xi_{n1} \\ \xi_{np} \end{pmatrix} + \frac{2\tau}{p+2} \begin{pmatrix} p-1 \\ -1 \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \xi_1 \\ \xi_p \end{pmatrix} + \frac{2\tau}{p+2} \begin{pmatrix} p-1 \\ -1 \end{pmatrix} =: \begin{pmatrix} \eta_1 \\ \eta_p \end{pmatrix}.$$

Clearly, η_1 is the largest eigenvalue of

$$\mathbf{Z} + \operatorname{diag}(0, -v_\tau, \ldots, -v_\tau) + \frac{2(p-1)\tau}{p+2}\mathbf{I}_p = \mathbf{Z} + \frac{2\tau}{p+2}\operatorname{diag}\left(p-1, -1, \ldots, -1\right)$$

whereas η_p is the smallest eigenvalue of

$$\mathbf{Z} + \operatorname{diag}(v_\tau, 0, \ldots, 0) - \frac{2\tau}{p+2}\mathbf{I}_p = \mathbf{Z} + \frac{2\tau}{p+2}\operatorname{diag}\left(p-1, -1, \ldots, -1\right).$$

This proves the result for $\tau \geq 0$.

(b) $\boldsymbol{\Sigma}_n$ has eigenvalues

$$\lambda_{n1} = \ldots = \lambda_{n,p-1} = \frac{1}{p} - \frac{2\tau}{\sqrt{n}p(p+2)} \quad \text{and} \quad \lambda_{np} = \frac{1}{p} + \frac{2(p-1)\tau}{\sqrt{n}p(p+2)}.$$

Starting with $\hat{\lambda}_{n1}$, we can decompose as above

$$\sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) = \sqrt{n}p\left(\hat{\lambda}_{n1} - \lambda_{n1}\right) + \sqrt{n}\left(p\lambda_{n1} - 1\right) =: \xi_{n1} - \frac{2\tau}{p+2}.$$

Letting $\boldsymbol{\Lambda}_n := \operatorname{diag}\left(\lambda_{n1}, \lambda_{n2}, \ldots, \lambda_{np}\right)$, there exists a $p \times p$ orthogonal matrix $\tilde{\mathbf{G}}_p$ such that $\boldsymbol{\Sigma}_n = \tilde{\mathbf{G}}_p \boldsymbol{\Lambda}_n \tilde{\mathbf{G}}_p'$. Clearly, ξ_{n1} is the largest eigenvalue of

$$\sqrt{n}p\left(\mathbf{S}_n - \lambda_{n1}\mathbf{I}_p\right) = \sqrt{n}p(\mathbf{S}_n - \boldsymbol{\Sigma}_n) + \sqrt{n}p\left(\boldsymbol{\Sigma}_n - \lambda_{n1}\mathbf{I}_p\right)$$

and therefore of

$$\sqrt{n}p\tilde{\mathbf{G}}_p'(\mathbf{S}_n - \boldsymbol{\Sigma}_n)\tilde{\mathbf{G}}_p + \sqrt{n}p(\boldsymbol{\Lambda}_n - \lambda_{n1}\mathbf{I}_p) =: \tilde{\mathbf{Z}}_n + \operatorname{diag}(0, \ldots, 0, v_\tau),$$

where v_τ remains the same as in case (a).

Similarly, $\xi_{np} := \sqrt{n}p\left(\hat{\lambda}_{np} - \lambda_{np}\right)$ is the smallest eigenvalue of

$$\tilde{\mathbf{Z}}_n + \operatorname{diag}(-v_\tau, \ldots, -v_\tau, 0).$$

As $\operatorname{vec}\left(\tilde{\mathbf{Z}}_n\right) = \sqrt{n}p\left(\tilde{\mathbf{G}}_p \otimes \tilde{\mathbf{G}}_p\right)' \operatorname{vec}(\mathbf{S}_n - \boldsymbol{\Sigma}_n)$ also converges weakly to $\operatorname{vec}(\mathbf{Z}) \sim \mathcal{N}_{p^2}\left(\mathbf{0}, \boldsymbol{\Gamma}_p\right)$, it implies that $\left(\xi_{n1}, \xi_{np}\right)' \xrightarrow{\mathcal{D}} \left(\xi_1, \xi_p\right)'$, where ξ_1 is the largest eigenvalue of $\mathbf{Z} + \operatorname{diag}(0, \ldots, 0, v_\tau)$ and ξ_p is the smallest eigenvalue of $\mathbf{Z} + \operatorname{diag}(-v_\tau, \ldots, -v_\tau, 0)$. Therefore

$$\begin{pmatrix} \sqrt{n}\left(p\hat{\lambda}_{n1} - 1\right) \\ \sqrt{n}\left(p\hat{\lambda}_{np} - 1\right) \end{pmatrix} = \begin{pmatrix} \xi_{n1} \\ \xi_{np} \end{pmatrix} + \frac{2\tau}{p+2} \begin{pmatrix} -1 \\ p-1 \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \xi_1 \\ \xi_p \end{pmatrix} + \frac{2\tau}{p+2} \begin{pmatrix} -1 \\ p-1 \end{pmatrix} =: \begin{pmatrix} \eta_1 \\ \eta_p \end{pmatrix}.$$

Clearly, η_1 is the largest eigenvalue of

$$\mathbf{Z} + \text{diag}(0, \ldots, 0, \nu_\tau) - \frac{2\tau}{p+2}\mathbf{I}_p = \mathbf{Z} + \frac{2\tau}{p+2}\text{diag}(-1, \ldots, -1, p-1)$$

whereas η_p is the smallest eigenvalue of

$$\mathbf{Z} + \text{diag}(-\nu_\tau, \ldots, -\nu_\tau, 0) + \frac{2(p-1)\tau}{p+2}\mathbf{I}_p = \mathbf{Z} + \frac{2\tau}{p+2}\text{diag}(-1, \ldots, -1, p-1).$$

This proves the result for $\tau < 0$.

3.7.8 Proof of Corollary 3.4.2

According to Theorem 3.4.1, $\mathrm{L}_p^{\max}$ is equal in distribution to the first marginal of

$$\boldsymbol{\ell}^{(p)} = \left(\ell_1^{(p)}, \ldots, \ell_p^{(p)}\right)',$$

where $\ell_1^{(p)} \geq \ldots \geq \ell_p^{(p)}$ are the eigenvalues of $\mathbf{Z} = (\mathrm{Z}_{ij})$, with $\text{vec}\,\mathbf{Z} \sim \mathcal{N}_{p^2}(\mathbf{0}, \boldsymbol{\Gamma}_p)$. Note that $\mathrm{Z}_{ij} = \mathrm{Z}_{ji}$ almost surely for any $1 \leq i < j \leq p$, so $\text{vec}\,\mathbf{Z}$ may not have a density with respect to the Lebesgue measure on $\mathbb{R}^{p^2}$, but $\text{vech}\,\mathbf{Z}$ might in principle have a density with respect to the Lebesgue measure on $\mathbb{R}^{p(p+1)/2}$. If $\text{vech}\,\mathbf{Z}$ indeed has such a density, then Theorem 13.3.1 of [Anderson, 2003] can be used to obtain the density of $\boldsymbol{\ell}^{(p)}$ from that of $\text{vech}\,\mathbf{Z}$. However, since $\text{tr}[\mathbf{Z}] = (\text{vec}\,\mathbf{I}_p)'(\text{vec}\,\mathbf{Z}) = 0$ almost surely, $\text{vech}\,\mathbf{Z}$ does not admit a density and the sum of the eigenvalues $\ell_1^{(p)}, \ldots, \ell_p^{(p)}$ of $\mathbf{Z}$ is almost surely zero (and thus they do not admit a joint density over $\mathbb{R}^p$ either).

We solve this issue by considering a sequence of $p \times p$ random matrices $\mathbf{Z}_{\delta_k}$, $k = 1, 2, \ldots$, with $\delta_k > 0$ converging to zero as k goes to infinity, and such that $\text{vec}\,\mathbf{Z}_\delta \sim \mathcal{N}_{p^2}(\mathbf{0}, \boldsymbol{\Gamma}_{p,\delta})$, with

$$\boldsymbol{\Gamma}_{p,\delta} := \frac{p}{p+2}\left(\mathbf{I}_{p^2} + \mathbf{K}_p\right) - \frac{2-\delta}{p+2}\mathbf{J}_p.$$

Of course, $\mathbf{Z}_{\delta_k}$ converges weakly to $\mathbf{Z}$, so that the continuous mapping theorem ensures that $\boldsymbol{\ell}_{\delta_k}^{(p)} = \left(\ell_{1\delta_k}^{(p)}, \ldots, \ell_{p\delta_k}^{(p)}\right)'$ (here, $\ell_{1\delta}^{(p)} \geq \ldots \geq \ell_{p\delta}^{(p)}$ are the eigenvalues of $\mathbf{Z}_\delta$) also converges weakly to $\boldsymbol{\ell}^{(p)} = \left(\ell_1^{(p)}, \ldots, \ell_p^{(p)}\right)'$. Let then $\mathbf{D}_p$ be the p-dimensional duplication matrix, that is such that $\mathbf{D}_p(\text{vech}\,\mathbf{A}) = \text{vec}\,\mathbf{A}$ for any $p \times p$ symmetric matrix $\mathbf{A}$. Write

$$\mathbf{W}_\delta := \text{vech}\,\mathbf{Z}_\delta = \mathbf{D}_p^-(\text{vec}\,\mathbf{Z}_\delta),$$

where $\mathbf{D}_p^- = \left(\mathbf{D}_p'\mathbf{D}_p\right)^{-1}\mathbf{D}_p'$ is the Moore–Penrose inverse of $\mathbf{D}_p$. By definition of $\mathbf{Z}_\delta$, the random vector $\mathbf{W}_\delta$ has density

$$\mathbf{w} \mapsto h_\delta(\mathbf{w}) = \frac{a_p}{\sqrt{\delta}}\exp\left(-\frac{1}{2}\mathbf{w}'\left(\mathbf{D}_p^-\left\{\tfrac{p}{p+2}\left(\mathbf{I}_{p^2} + \mathbf{K}_p\right) - \tfrac{2-\delta}{p+2}\mathbf{J}_p\right\}\left(\mathbf{D}_p^-\right)'\right)^{-1}\mathbf{w}\right),$$

where

$$a_p := \frac{(p+2)^{p(p+1)/4}}{2^{(p^2+3p-2)/4}(\pi p)^{p(p+1)/4}}$$

is a normalizing constant. By using the identities $\mathbf{K}_p\mathbf{D}_p = \mathbf{D}_p$ and $\mathbf{D}_p^-(\mathrm{vec}\,\mathbf{I}_p) = \mathbf{D}_p'(\mathrm{vec}\,\mathbf{I}_p) = \mathrm{vech}\,\mathbf{I}_p$, it is easy to check that

$$\left(\mathbf{D}_p^-\left\{\tfrac{p}{p+2}\left(\mathbf{I}_{p^2}+\mathbf{K}_p\right)-\tfrac{2-\delta}{p+2}\mathbf{J}_p\right\}\left(\mathbf{D}_p^-\right)'\right)^{-1} = \frac{p+2}{2p}\mathbf{D}_p'\left\{\frac{1}{2}\left(\mathbf{I}_{p^2}+\mathbf{K}_p\right)+\frac{\frac{2-\delta}{p+2}}{\frac{2p}{p+2}-\frac{p(2-\delta)}{p+2}}\mathbf{J}_p\right\}\mathbf{D}_p$$
$$= \frac{p+2}{2p}\mathbf{D}_p'\left\{\frac{1}{2}\left(\mathbf{I}_{p^2}+\mathbf{K}_p\right)+\frac{2-\delta}{\delta p}\mathbf{J}_p\right\}\mathbf{D}_p.$$

Using the identities $(\mathrm{vec}\,\mathbf{A})'(\mathrm{vec}\,\mathbf{B}) = \mathrm{tr}[\mathbf{A}'\mathbf{B}]$ and $\mathbf{K}_p(\mathrm{vec}\,\mathbf{A}) = \mathrm{vec}\left(\mathbf{A}'\right)$, the resulting density for $\mathbf{Z}_\delta$ is therefore

$$\begin{aligned}\mathbf{z}\mapsto f(\mathbf{z}) &= h_\delta(\mathrm{vech}\,\mathbf{z})\\ &= \frac{a_p}{\sqrt{\delta}}\exp\left(-\frac{p+2}{4p}(\mathrm{vech}\,\mathbf{z})'\mathbf{D}_p'\left\{\frac{1}{2}\left(\mathbf{I}_{p^2}+\mathbf{K}_p\right)+\frac{2-\delta}{\delta p}\mathbf{J}_p\right\}\mathbf{D}_p\mathrm{vech}\,\mathbf{z}\right)\\ &= \frac{a_p}{\sqrt{\delta}}\exp\left(-\frac{p+2}{4p}(\mathrm{vec}\,\mathbf{z})'\left\{\mathbf{I}_{p^2}+\frac{2-\delta}{\delta p}\mathbf{J}_p\right\}\mathrm{vec}\,\mathbf{z}\right)\\ &= \frac{a_p}{\sqrt{\delta}}\exp\left(-\frac{p+2}{4p}\left\{\mathrm{tr}\left(\mathbf{z}^2\right)+\frac{2-\delta}{\delta p}(\mathrm{tr}\,\mathbf{z})^2\right\}\right).\end{aligned}$$

Theorem 13.3.1 from [Anderson, 2003] then implies that $\boldsymbol{\ell}_\delta^{(p)} = \left(\ell_{1\delta}^{(p)},\ldots,\ell_{p\delta}^{(p)}\right)'$ has density

$$(\ell_1,\ldots,\ell_p)'\mapsto\frac{b_p}{\sqrt{\delta}}\exp\left(-\frac{p+2}{4p}\left\{\left(\sum_{j=1}^p\ell_j^2\right)+\frac{2-\delta}{\delta p}\left(\sum_{j=1}^p\ell_j\right)^2\right\}\right)\left(\prod_{1\le k<j\le p}(\ell_k-\ell_j)\right)\mathbb{I}_{\{\ell_1\ge\ldots\ge\ell_p\}},$$

with

$$b_p := \frac{(p+2)^{p(p+1)/4}}{2^{(p^2+3p-2)/4}p^{p(p+1)/4}\prod_{j=1}^p\Gamma\left(\frac{j}{2}\right)}.$$

We now turn to the particular cases (i) $p=2$ and (ii) $p=3$ considered in the statement of the corollary.

(i) We infer from above that $\boldsymbol{\ell}_\delta^{(2)} = \left(\ell_{1\delta}^{(2)},\ell_{2\delta}^{(2)}\right)'$ has density $(\ell_1,\ell_2)'\mapsto\mathscr{I}(\ell_1,\ell_2)\mathbb{I}_{\{\ell_1\ge\ell_2\}}$, with

$$\mathscr{I}(\ell_1,\ell_2) := \frac{1}{\sqrt{2\pi\delta}}(\ell_1-\ell_2)\,e^{-\frac{1}{2}\left\{\left(\ell_1^2+\ell_2^2\right)+\frac{2-\delta}{2\delta}(\ell_1+\ell_2)^2\right\}}.$$

Direct computations allow checking that $\mathscr{I}(\ell_1,\ell_2)$ is the derivative of the function

$$\ell_2\mapsto\frac{\sqrt{2\delta}}{\sqrt{\pi}(2+\delta)}e^{-\frac{1}{2}\left\{\left(\ell_1^2+\ell_2^2\right)+\frac{2-\delta}{2\delta}(\ell_1+\ell_2)^2\right\}}+\frac{2^{5/2}\ell_1}{(2+\delta)^{3/2}}e^{-\frac{2\ell_1^2}{2+\delta}}\Phi\left(\frac{(2-\delta)\ell_1+(2+\delta)\ell_2}{\sqrt{2\delta(\delta+2)}}\right).$$

It follows that $\ell_{1\delta}^{(2)}$ has density

$$\begin{aligned}\ell_1&\mapsto\int_{-\infty}^{\ell_1}\mathscr{I}(\ell_1,\ell_2)\,d\ell_2\\ &= \left[\frac{\sqrt{2\delta}}{\sqrt{\pi}(2+\delta)}e^{-\frac{1}{2}\left\{\left(\ell_1^2+\ell_2^2\right)+\frac{2-\delta}{2\delta}(\ell_1+\ell_2)^2\right\}}+\frac{2^{5/2}\ell_1}{(2+\delta)^{3/2}}e^{-\frac{2\ell_1^2}{2+\delta}}\Phi\left(\frac{(2-\delta)\ell_1+(2+\delta)\ell_2}{\sqrt{2\delta(\delta+2)}}\right)\right]_{-\infty}^{\ell_1}\\ &= \frac{\sqrt{2\delta}}{\sqrt{\pi}(2+\delta)}e^{-\frac{2\ell_1^2}{\delta}}+\frac{2^{5/2}\ell_1}{(2+\delta)^{3/2}}e^{-\frac{2\ell_1^2}{2+\delta}}\Phi\left(\frac{4\ell_1}{\sqrt{2\delta(\delta+2)}}\right).\end{aligned}$$

Taking the limit as $\delta \to 0$, we obtain that $\ell_1^{(2)} \left(\stackrel{\mathcal{D}}{=} \mathrm{L}_2^{\max} \right)$, hence $\mathrm{T}_+^{(n)}$ asymptotically, has density $\ell_1 \mapsto 2\ell_1 \exp\left(-\ell_1^2\right) \mathbb{I}_{\{\ell_1 > 0\}}$.

To prove that $\mathrm{T}_-^{(n)}$ and $\mathrm{T}_\pm^{(n)}$ have the same asymptotic null distribution as $\mathrm{T}_+^{(n)}$, note that $\mathbf{S}_n$ has trace one almost surely (a.s.) so its eigenvalues $\hat{\lambda}_{n\ell}$, $\ell = 1, \ldots, p$ sum up to one almost surely. For $p = 2$, it follows that

$$\mathrm{T}_+^{(n)} = \sqrt{n}(2\hat{\lambda}_{n1} - 1) \stackrel{a.s.}{=} -\sqrt{n}(2\hat{\lambda}_{np} - 1) = \mathrm{T}_-^{(n)}$$

so that $\mathrm{T}_+^{(n)} = \mathrm{T}_-^{(n)} = \mathrm{T}_\pm^{(n)}$ almost surely, which of course entails that the statistics of these three tests share the same weak limit, not only under the null hypothesis but under *any* sequence of hypotheses.

(ii) For $p = 3$, the density of $\boldsymbol{\ell}_\delta^{(3)} = \left(\ell_{1\delta}^{(3)}, \ell_{2\delta}^{(3)}, \ell_{3\delta}^{(3)}\right)'$ is $(\ell_1, \ell_2, \ell_3)' \mapsto \mathscr{J}(\ell_1, \ell_2, \ell_3) \mathbb{I}_{\{\ell_1 \geq \ell_2 \geq \ell_3\}}$, with

$$\mathscr{J}(\ell_1, \ell_2, \ell_3) := \frac{125}{216\pi\sqrt{\delta}} (\ell_1 - \ell_2)(\ell_1 - \ell_3)(\ell_2 - \ell_3)\, e^{-\frac{5}{12}\left\{\left(\ell_1^2 + \ell_2^2 + \ell_3^2\right) + \frac{2-\delta}{3\delta}(\ell_1 + \ell_2 + \ell_3)^2\right\}}.$$

Lengthy yet straightforward computations show that $\mathscr{J}(\ell_1, \ell_2, \ell_3)$ is the derivative of the function $\ell_3 \mapsto \mathscr{K}(\ell_1, \ell_2, \ell_3)$, with

$$\begin{aligned}
&\mathscr{K}(\ell_1, \ell_2, \ell_3) := \\
&-\frac{25\sqrt{\delta}(\ell_1 - \ell_3)}{48\pi(1+\delta)^2}(2(2\ell_1 + 2\ell_3 - \ell_2) + \delta(\ell_1 + \ell_3 - 2\ell_2))\, e^{-\frac{5}{12}\left\{\left(\ell_1^2 + \ell_2^2 + \ell_3^2\right) + \frac{2-\delta}{3\delta}(\ell_1 + \ell_2 + \ell_3)^2\right\}} \\
&-\frac{5\sqrt{10}(\ell_1 - \ell_3)}{288\sqrt{\pi}(1+\delta)^{5/2}}\left\{20\left(2\ell_1^1 + 5\ell_1\ell_3 + 2\ell_3^2\right) - 2\delta\left(5(\ell_1 - \ell_3)^2 - 18\right) - \delta^2\left(5(\ell_1 - \ell_3)^2 - 36\right)\right\} \\
&\qquad \times e^{-\frac{20\left(\ell_1^2 + \ell_1\ell_3 + \ell_3^2\right) + 5\delta(\ell_1 - \ell_3)^2}{24(1+\delta)}}\, \Phi\left(\frac{\sqrt{5}(2(\ell_1 + \ell_2 + \ell_3) - \delta(\ell_1 + \ell_3 - 2\ell_2))}{6\sqrt{\delta(1+\delta)}}\right).
\end{aligned}$$

Therefore, $\left(\ell_{1\delta}^{(3)}, \ell_{3\delta}^{(3)}\right)'$ has density

$$(\ell_1, \ell_3) \mapsto \left(\int_{\ell_3}^{\ell_1} \mathscr{J}(\ell_1, \ell_2, \ell_3)\, d\ell_2\right) \mathbb{I}_{\{\ell_1 \geq \ell_3\}} = (\mathscr{K}(\ell_1, \ell_1, \ell_3) - \mathscr{K}(\ell_1, \ell_3, \ell_3))\, \mathbb{I}_{\{\ell_1 \geq \ell_3\}},$$

with

$$\begin{aligned}
&\mathscr{K}(\ell_1, \ell_1, \ell_3) - \mathscr{K}(\ell_1, \ell_3, \ell_3) = \\
&\quad -\frac{25\sqrt{\delta}(\ell_1 - \ell_3)}{48\pi(1+\delta)^2}\{2(\ell_1 + 2\ell_3) + \delta(\ell_3 - \ell_1)\}\, e^{-\frac{5}{12}\left\{\left(2\ell_1^2 + \ell_3^2\right) + \frac{2-\delta}{3\delta}(2\ell_1 + \ell_3)^2\right\}} \\
&\quad +\frac{25\sqrt{\delta}(\ell_1 - \ell_3)}{48\pi(1+\delta)^2}\{2(2\ell_1 + \ell_3) + \delta(\ell_1 - \ell_3)\}\, e^{-\frac{5}{12}\left\{\left(\ell_1^2 + 2\ell_3^2\right) + \frac{2-\delta}{3\delta}(\ell_1 + 2\ell_3)^2\right\}} \\
&\quad -\frac{5\sqrt{10}(\ell_1 - \ell_3)}{288\sqrt{\pi}(1+\delta)^{5/2}}\, e^{-\frac{20\left(\ell_1^2 + \ell_1\ell_3 + \ell_3^2\right) + 5\delta(\ell_1 - \ell_3)^2}{24(1+\delta)}} \\
&\qquad \times \left\{20\left(2\ell_1^1 + 5\ell_1\ell_3 + 2\ell_3^2\right) - 2\delta\left(5(\ell_1 - \ell_3)^2 - 18\right) - \delta^2\left(5(\ell_1 - \ell_3)^2 - 36\right)\right\} \\
&\qquad \times \left[\Phi\left(\frac{\sqrt{5}(2(2\ell_1 + \ell_3) - \delta(\ell_3 - \ell_1))}{6\sqrt{\delta(1+\delta)}}\right) - \Phi\left(\frac{\sqrt{5}(2(\ell_1 + 2\ell_3) - \delta(\ell_1 - \ell_3))}{6\sqrt{\delta(1+\delta)}}\right)\right].
\end{aligned}$$

Taking the limit as $\delta \to 0$ shows that the density of $\left(\ell_1^{(3)}, \ell_3^{(3)}\right)'$ is

$$\mathscr{M}(\ell_1, \ell_3) := -\frac{100\sqrt{10}}{288\sqrt{\pi}} (\ell_1 - \ell_3) \left(2\ell_1^2 + 5\ell_1\ell_3 + 2\ell_3^2\right) e^{-\frac{5\left(\ell_1^2 + \ell_1\ell_3 + \ell_3^2\right)}{6}} \mathbb{I}_{\{-2\ell_1 < \ell_3 \leq -\ell_1/2\}}. \tag{3.41}$$

From Theorem 3.4.1, the density of the asymptotic null distribution of $T_+^{(n)}$ coincides with the density of $\ell_1^{(3)} \left(\stackrel{\mathscr{D}}{=} L_3^{\max}\right)$. After marginalization in (3.41), this last density is seen to be

$$\begin{aligned}
\ell_1 \mapsto & \left\{\sqrt{\frac{5}{2\pi}} e^{-\frac{5\ell_1^2}{2}} + \sqrt{\frac{5}{8\pi}} e^{-\frac{5\ell_1^2}{8}} + \frac{3}{4}\sqrt{\frac{5}{8\pi}} \left(5\ell_1^2 - 4\right) e^{-\frac{5\ell_1^2}{8}}\right\} \mathbb{I}_{\{\ell_1 \geq 0\}} \\
&= \left\{\frac{d}{d\ell_1}\Phi\left(\sqrt{5}\ell_1\right) + \frac{d}{d\ell_1}\Phi\left(\frac{\sqrt{5}\ell_1}{2}\right) + 3\frac{d}{d\ell_1}\Phi''\left(\frac{\sqrt{5}\ell_1}{2}\right)\right\} \mathbb{I}_{\{\ell_1 \geq 0\}},
\end{aligned}$$

which proves the result for $T_+^{(n)}$, hence also for $T_-^{(n)}$ (see the discussion below the corollary: $T_+^{(n)}$ and $T_-^{(n)}$ share the same null weak limit in any dimension p).

Finally, $T_\pm^{(n)}$ converges weakly to $\max\left(\ell_1^{(3)}, -\ell_3^{(3)}\right)$. Using (3.41) we get

$$\begin{aligned}
P\left[\max\left(\ell_1^{(3)}, -\ell_3^{(3)}\right) \leq z\right] &= P\left[\ell_1^{(3)} \leq z, \ell_3^{(3)} \geq -z\right] \\
&= \int_0^{z/2} \left(\int_{-2\ell_1}^{-\ell_1/2} \mathscr{M}(\ell_1, \ell_3) d\ell_3\right) d\ell_1 + \int_{z/2}^{z} \left(\int_{-z}^{-\ell_1/2} \mathscr{M}(\ell_1, \ell_3) d\ell_3\right) d\ell_1 \\
&= \left\{2\Phi\left(\frac{\sqrt{5}\ell}{2}\right) + 6\Phi''\left(\frac{\sqrt{5}}{2}\right) - 2\sqrt{3}\Phi''\left(\frac{\sqrt{5}\ell}{\sqrt{3}}\right) - 1\right\} \mathbb{I}_{\{\ell > 0\}}.
\end{aligned}$$

Chapter 4

High-dimensional behaviour of the Rayleigh and Bingham tests under general rotationally symmetric distributions

Contents

In this chapter we conduct a systematic investigation, relying on martingale central limit theorems, of the high-dimensional non-null behaviour of the Rayleigh and Bingham tests under a very broad class of rotationally symmetric alternatives. This identifies their "detection threshold", that discriminates between alternatives to which the tests will be blind and those under which they will be consistent. It will also reveal whether the Bingham test shows power against more severe alternatives than the contiguous ones in the axial model.

We start with the Rayleigh test in Section 4.1; then we study the Bingham test which, unlike the Rayleigh test, shows power under both axial and monotone alternatives. We therefore apply our general non-null results to describe thoroughly its behaviour in the semiparametric class of axial alternatives in Section 4.2.2 on the one hand and its non-null behaviour in the semiparametric class of monotone alternatives in Section 4.2.3 on the other hand. Its detection threshold and asymptotic power are compared to the Rayleigh test and Monte Carlo exercices are conducted to confirm our asymptotic findings.

4.1 Asymptotic non-null behaviour of the Rayleigh test

In this section, we derive the asymptotic distribution of the high-dimensional Rayleigh test under rotationally symmetric distributions that encompass those considered in Chapter 2. Here we do not require that the rotationally symmetric alternatives are monotone in the sense of Section 1.1.3, nor absolutely continuous with respect to the surface area measure on the unit sphere, nor that they involve a concentration parameter κ. Yet one of our objectives is to interpret the results of this section in the light of those obtained in Chapter 2.

More specifically, we consider as alternatives the sequence of hypotheses $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$ (see Section 1.1.2). Under the null of uniformity $\mathrm{P}^{(n)}_0$, the test statistic $\mathrm{R}^{\mathrm{St}}_n$ in (2.1),

$$\mathrm{R}^{\mathrm{St}}_n := \frac{np_n\|\bar{\mathbf{X}}_n\|^2 - p_n}{\sqrt{2p_n}} = \frac{\sqrt{2p_n}}{n}\sum_{1\leq i<j\leq n}\mathbf{X}'_{ni}\mathbf{X}_{nj},$$

has mean zero and variance $(n-1)/n(\to 1)$. Rotationally symmetric alternatives are expected to have an impact on the asymptotic mean and variance of $\mathrm{R}^{\mathrm{St}}_n$. This is made precise in the following result.

Proposition 4.1.1. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$*,*

$$\begin{aligned}
\mathrm{E}\left[\mathrm{R}^{\mathrm{St}}_n\right] &= \frac{(n-1)\sqrt{p_n}}{\sqrt{2}}e^2_{n1}\\
\sigma^2_n := p_n\tilde{e}^2_{n2} + 2np_ne^2_{n1}\tilde{e}_{n2} + f^2_{n2} &= \mathrm{Var}\left[\mathrm{R}^{\mathrm{St}}_n\right] + o(1),
\end{aligned}$$

as $n\to\infty$*, where the expectations*

$$\begin{aligned}
e_{n\ell} &:= \mathrm{E}\left[\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^{\ell}\right],\\
\tilde{e}_{n\ell} &:= \mathrm{E}\left[\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n - e_{n1}\right)^{\ell}\right],\\
f_{n\ell} &:= \mathrm{E}\left[\left(1-\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right)^{\ell/2}\right],
\end{aligned}$$

are taken under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$.

Proof. See Section 4.3.2. □

Under $P_0^{(n)}$, $e_{n1} = 0$ and $\tilde{e}_{n2} = e_{n2} = 1/p_n$, so that Proposition 4.1.1 is compatible with the null values of $E[R_n^{St}]$ and $\mathrm{Var}[R_n^{St}]$ provided above. Now, parallel to the null case (see (2.1)), the Rayleigh test statistic, after appropriate standardization, is also asymptotically standard normal under a broad class of rotationally symmetric alternatives. More precisely, we have the following result.

Theorem 4.1.2. *Let (p_n) be a sequence of positive integers diverging to ∞ as $n \to \infty$. Assume that the sequence $\left(P^{(n)}_{\boldsymbol{\theta}_n, F_n}\right)$ is such that, as $n \to \infty$,*

(i) $\min\left(\frac{p_n \tilde{e}^2_{n2}}{f^2_{n2}}, \frac{\tilde{e}_{n2}}{n e^2_{n1}}\right) = o(1)$;

(ii) $\frac{\tilde{e}_{n4}}{\tilde{e}^2_{n2}} = o(n)$;

(iii) $\frac{f_{n4}}{f^2_{n2}} = o(n)$;

Then, under $P^{(n)}_{\boldsymbol{\theta}_n, F_n}$ and as $n \to \infty$,

$$\frac{R_n^{St} - E[R_n^{St}]}{\sigma_n} = \frac{\sqrt{2p_n}}{n\sigma_n} \sum_{1 \le i < j \le n} \left(\mathbf{X}'_{ni}\mathbf{X}_{nj} - e^2_{n1}\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Sketch of the proof. The complete proof of this theorem can be found in Section 4.3.3.

Let $\mathscr{F}_{n\ell}$ be the σ-algebra generated by $\mathbf{X}_{n1}, \ldots, \mathbf{X}_{n\ell}$ and $E_{n\ell}$ the conditional expectation with respect to $\mathscr{F}_{n\ell}$. Define

$$Z_{n\ell} := E_{n\ell}\left[\frac{R_n^{St} - E[R_n^{St}]}{\sigma_n}\right] \text{ and } D_{n\ell} := Z_{n\ell} - Z_{n,\ell-1}.$$

We prove through tedious computations that

$$\sum_{\ell=1}^{n} \sigma^2_{n\ell} = 1 + o_P(1) \text{ and } \sum_{\ell=1}^{n} E\left[D^2_{n\ell}\mathbb{I}_{\{|D_{n\ell}|>\varepsilon\}}\right] \to 0,$$

where $\sigma^2_{n\ell} := E_{n,\ell-1}\left[D^2_{n\ell}\right]$ so that by Theorem 1.4.1,

$$\sum_{\ell=1}^{n} D_{n\ell} = Z_{nn} - Z_{n,0} = \frac{R_n^{St} - E[R_n^{St}]}{\sigma_n}$$

is asymptotically standard normal. □

This result applies under very mild assumptions, that in particular do not impose absolute continuity nor any other regularity conditions. The only structural assumptions are the conditions (i)-(iii) above, that, in the FvML case, *always* hold, that is, they hold without any constraint on the concentration κ_n nor on the way the dimension p_n goes to infinity with n. The proof of this statement can be found in Appendix B.

Theorem 4.1.2 allows to compute the asymptotic power of the Rayleigh test under appropriate sequences of alternatives. As mentioned above, the null of uniformity $\mathscr{H}_0^{(n)}$ yields $e_{n1} = 0$ and $\tilde{e}_{n2} = 1/p_n$. Here, we therefore consider "local" departures from uniformity of the form

$$\mathscr{H}_1^{(n)} : \left\{P^{(n)}_{\boldsymbol{\theta}_n, F_n} : e_{n1} = 0 + \nu_n\tau,\ \tilde{e}_{n2} = \frac{1}{p_n} + \xi_n\eta\right\}.$$

The following result provides the asymptotic power of the high-dimensional Rayleigh test in (2.2) under sequences of local alternatives that, as we will show, are intimately related to those we considered in Sections 2.2-2.3.

Theorem 4.1.3. *Let (p_n) be a sequence of positive integers diverging to ∞ as $n \to \infty$. Let the sequence $(\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \mathrm{F}_n})$ satisfy the assumptions of Theorem 4.1.2 and be such that*

$$e_{n1} = \frac{\tau}{n^{1/2} p_n^{1/4}} + o\left(\frac{1}{n^{1/2} p_n^{1/4}}\right) \qquad \text{and} \qquad \tilde{e}_{n2} = \frac{1}{p_n} + o\left(\frac{1}{p_n}\right), \tag{4.1}$$

for some $\tau \geq 0$. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \mathrm{F}_n}$, the asymptotic power of the high-dimensional Rayleigh test in (2.2) is given by $1 - \Phi\left(z_\alpha - \tau^2/\sqrt{2}\right)$.

Proof. See Section 4.3.4. □

In order to link these alternatives to those considered earlier, note that, as $n \to \infty$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \kappa_n, f}$, with $\kappa_n = \xi_n \sqrt{p_n/n}$, where the positive sequence (ξ_n) is $o(\sqrt{n})$, we have

$$\begin{aligned} e_{n1} &= \left(\frac{c_{p_n}}{c_{p_n,\kappa_n,f}}\right)^{-1} \frac{c_{p_n}}{\kappa_n} \int_{-1}^{1} \left(1 - s^2\right)^{(p_n-3)/2} \kappa_n s f(\kappa_n s)\, ds \\ &= \left(1 + \frac{\kappa_n^2}{2p_n} f''(0) + o\left(\frac{\kappa_n^2}{p_n}\right)\right)^{-1} \left(\frac{\kappa_n}{p_n} + o\left(\frac{\kappa_n}{p_n}\right)\right) \end{aligned} \tag{4.2}$$

and

$$\begin{aligned} e_{n2} &= \left(\frac{c_{p_n}}{c_{p_n,\kappa_n,f}}\right)^{-1} \frac{c_{p_n}}{\kappa_n^2} \int_{-1}^{1} \left(1 - s^2\right)^{(p_n-3)/2} (\kappa_n s)^2 f(\kappa_n s)\, ds \\ &= \left(1 + \frac{\kappa_n^2}{2p_n} f''(0) + o\left(\frac{\kappa_n^2}{p_n}\right)\right)^{-1} \left(\frac{1}{p_n} + o\left(\frac{1}{p_n}\right)\right), \end{aligned} \tag{4.3}$$

where we used four times Lemma 2.5.1. For the contiguous alternatives in Theorem 2.2.1, that is for $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \kappa_n, f}$, with $\kappa_n = \tau_n \sqrt{p_n/n}$ (where (τ_n) is bounded), (4.2)-(4.3) provide

$$e_{n1} = \frac{\tau_n}{\sqrt{n p_n}} + o\left(\frac{1}{\sqrt{n p_n}}\right) \qquad \text{and} \qquad \tilde{e}_{n2} = \frac{1}{p_n} + o\left(\frac{1}{p_n}\right). \tag{4.4}$$

Theorem 4.1.3 implies that the asymptotic power of the high-dimensional Rayleigh test under the alternatives (4.4) is equal to α, which confirms that this test is blind to the contiguous alternatives from Theorem 2.2.1 (see Section 2.2).

Now, at least if $p_n = o(n^2)$ (a constraint that is actually superfluous in the FvML case, as it can be seen by using the Amos-type bounds provided in Lemma B.0.4), the more severe alternatives $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n, \kappa_n, f}$, with $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$, from Theorem 2.3.2 translate — still in view of (4.2)-(4.3) — into those in (4.1). This shows that the asymptotic powers of the high-dimensional Rayleigh test computed in the FvML case via Le Cam's Third Lemma (see (2.17)) actually also hold away from the FvML case. Clearly, this further supports the conjecture from Section 2.3 that, under the assumption that $p_n = o(n^2)$, Theorem 2.3.2 holds for an essentially arbitrary f.

Now, we conducted the same Monte Carlo study as in Section 2.4 with data from a beta distribution to check the validity of our results for general rotationally symmetric distributions. For any $(n, p) \in \mathrm{C} \times \mathrm{C}$, with $\mathrm{C} := \{30, 100, 400\}$, any $j \in \{1, 2\}$, and any $\ell \in$

$\{0,1,2,3,4\}$, we generated $M = 2,500$ independent random samples $\mathbf{X}_{i;j}^{(\ell)}$, $i = 1,\dots,n$ of the form

$$\mathbf{X}_{i;j}^{(\ell)} \quad i = 1,\dots,n, \quad j = 1,2, \quad \ell = 0,1,2,3,4.$$

The $\mathbf{X}_{i;j}^{(\ell)}$'s are rotationally symmetric with location $\boldsymbol{\theta} = (1,0,\dots,0)' \in \mathbb{R}^p$ and are such that, for all $i = 1,\dots,n$, $\frac{1}{2}(\boldsymbol{\theta}'\mathbf{X}_{i;j}^{(\ell)} + 1)$ is beta with mean $e_{1;j}^{(\ell)}$ and variance $\tilde{e}_{2;j} = 1/p$, where we let

$$e_{1;1}^{(\ell)} = \frac{0.6\ell}{\sqrt{np}} \quad \text{and} \quad e_{1;2}^{(\ell)} = \frac{0.6\ell}{n^{1/2}p^{1/4}}$$

The value $\ell = 0$ corresponds to the null hypothesis of uniformity, while $\ell = 1,2,3,4$ provide increasingly severe alternatives. The case $j = 1$ relates to the contiguous alternatives from Theorem 2.2.1 and the corresponding (more general) alternatives in (4.4), whereas $j = 2$ is associated with the alternatives under which the Rayleigh test shows non-trivial asymptotic powers in the high-dimensional setup (see Theorem 2.3.2 and the alternatives (4.1)).

We evaluated the rejection frequencies of

(i) the specified-$\boldsymbol{\theta}_n$ test $\phi_{\boldsymbol{\theta}_n}^{(n)}$ in (2.10),

(ii) the high-dimensional Rayleigh test $\phi_{\text{Ray}}^{(n)}$ in (2.2),

both conducted at nominal level 5%. Rejection frequencies are plotted in Figure 4.1 as well as the corresponding asymptotic powers, obtained from (2.11), (2.17), and the fact that $\phi_{\boldsymbol{\theta}_n}^{(n)}$ is consistent against $(j = 2)$-alternatives.

As with FvML samples, rejection frequencies match the corresponding asymptotic powers, irrespective of the tests and types of alternatives considered.

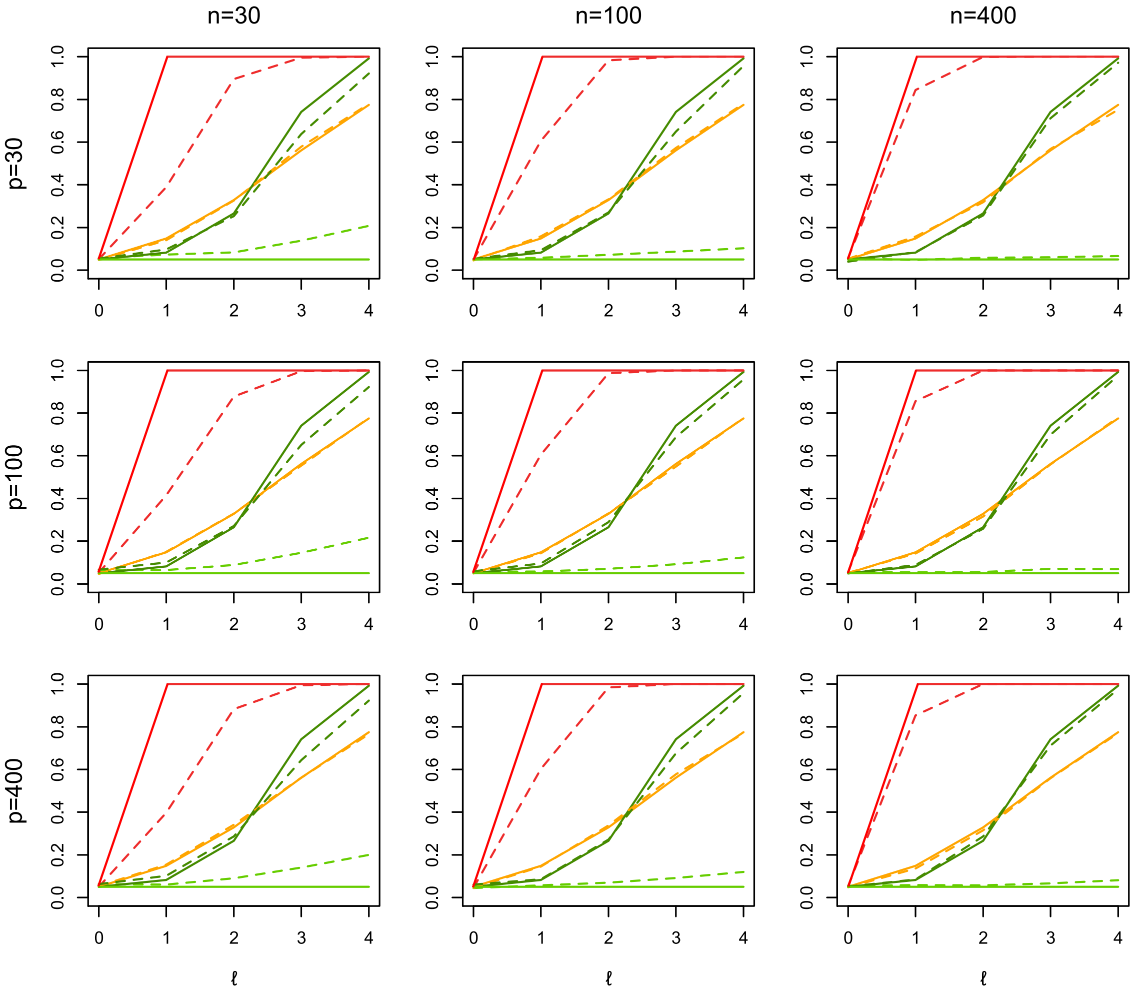


Figure 4.1 – Rejection frequencies (dashed) and asymptotic powers (solid), under the null of uniformity over the p-dimensional unit sphere ($\ell = 0$) and increasingly severe "beta" rotationally symmetric alternatives ($\ell = 1, 2, 3, 4$), of the specified-$\boldsymbol{\theta}_n$ test $\phi^{(n)}_{\boldsymbol{\theta}_n}$ in (2.10) (red/orange) and the high-dimensional Rayleigh test $\phi^{(n)}_{\mathrm{Ray}}$ in (2.2) (light/dark green). Light colors (orange and light green) are associated with contiguous alternatives, whereas dark colors (red and dark green) correspond to the more severe alternatives under which the Rayleigh test shows non-trivial asymptotic powers in high dimensions.

4.2 Asymptotic non-null behaviour of the Bingham test

4.2.1 General rotationally symmetric alternatives

We now study the non-null asymptotic behaviour of the Bingham test under general rotationally symmetric alternatives. The following result provides the expectation and variance of

$$\mathrm{Q}_n^{\mathrm{St}} = \frac{p_n}{n} \sum_{\substack{i<j \\ i,j=1}}^{n} \left\{ (\mathbf{X}_{ni}'\mathbf{X}_{nj})^2 - \frac{1}{p_n} \right\}$$

under $\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}$.

Proposition 4.2.1. *Let (p_n) be a sequence of positive integers, $(\boldsymbol{\theta}_n)$ be a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n, and (F_n) be a sequence of cumulative distribution functions on $[-1,1]$. Write*

$$e_{n\ell} := e_{n\ell}(\mathrm{F}_n) := \mathrm{E}\left[u_{n1}^{\ell}\right]$$
$$f_{n\ell} := f_{n\ell}(\mathrm{F}_n) := \mathrm{E}\left[v_{n1}^{\ell}\right]$$

for the ℓ-th moments of $u_{n1} = \mathbf{X}_{n1}'\boldsymbol{\theta}_n$ and $v_{n1} = \sqrt{1-u_{n1}^2}$ under $\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}$. Letting $g_{n2} := e_{n2} - 1/p_n$, we have under $\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}$ that

$$\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = \frac{p_n^2(n-1)}{2(p_n-1)} g_{n2}^2 \tag{4.5}$$

and

$$\begin{aligned}\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = \frac{(n-1)p_n^2}{2n} &\left\{ e_{n4}^2 + \frac{6}{p_n-1}(e_{n2}-e_{n4})^2 + \frac{3f_{n4}^2}{p_n^2-1} - \left(\frac{p_n}{p_n-1} g_{n2}^2 + \frac{1}{p_n} \right)^2 \right\} \\ &+ \frac{(n-1)(n-2)p_n^4}{n(p_n-1)^2}(e_{n4}-e_{n2}^2) g_{n2}^2. \end{aligned} \tag{4.6}$$

Proof. See Section 4.3.5. □

Under the null hypothesis of uniformity $\mathrm{P}_0^{(n)}$, the random variable $u_{n1} = \mathbf{X}_{n1}'\boldsymbol{\theta}_n$ is such that (see Section 1.1.1)

$$e_{n2} = \frac{1}{p_n} \quad \text{and} \quad e_{n4} = \frac{3}{p_n(p_n+2)}.$$

Using these values along with the identities $f_{n2} = 1 - e_{n2}$ and $f_{n4} = 1 - 2e_{n2} + e_{n4}$, Proposition 4.2.1 shows that $\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = 0$ and $\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = (n-1)(p_n-1)/(n(p_n+2))$ under $\mathrm{P}_0^{(n)}$, which is compatible with the null asymptotic normality result in (3.2). A key step in studying the non-null asymptotic behaviour of the Bingham test in high dimensions is to extend this asymptotic normality result to general rotationally symmetric alternatives.

Theorem 4.2.2. *Let (p_n) be a sequence of positive integers diverging to ∞ and $(\boldsymbol{\theta}_n)$ be a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Assume that the sequence (F_n) is such that, as $n \to \infty$,*

(a) $e_{n4} = o(1/p_n)$,

(b) $e_{n8} = o(n^{2/3}/p_n^2)$,

(c) $g_{n2} = \mathrm{O}(1/\sqrt{np_n})$.

Then,

$$\frac{\mathrm{Q}_n^{\mathrm{St}} - \mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right]}{\sqrt{\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right]}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1)$$

as $n \to \infty$ *under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$.

Proof. See proof in Section 4.3.6. □

This general asymptotic normality result extends the null one in (3.2), since Proposition 4.2.1 implies that $\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = 0$ and $\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = 1 + o(1)$ under $\mathrm{P}_0^{(n)}$. It is worth pointing out that, like the result in (3.2), Theorem 4.2.2 in principle does not impose restrictions on the rate at which p_n diverges to infinity with n (some restrictions may arise when considering some particular alternatives, though; see, e.g., Theorems 4.2.5 and 4.2.7 below). More importantly, we can now state the main theorem of this section, that describes the non-null behaviour of the Bingham test statistic under general rotationally symmetric alternatives.

Theorem 4.2.3. *Let* (p_n) *be a sequence of positive integers diverging to* ∞ *and* $(\boldsymbol{\theta}_n)$ *a sequence such that* $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ *for any* n. *Assume that the sequence* (F_n) *is such that,*

(a) $e_{n4} = o(1/p_n)$,

(b) $e_{n8} = o(n^{2/3}/p_n^2)$.

as $n \to \infty$. *Then, we have the following:*

(i) if (c) $g_{n2} = o(1/\sqrt{np_n})$, *then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$, *as* $n \to \infty$,

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1);$$

(ii) if (c) $g_{n2} = \xi_n/\sqrt{np_n}$ *with* $(\xi_n) \to \xi(\neq 0)$, *then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$, *as* $n \to \infty$,

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathscr{N}\left(\frac{\xi^2}{2}, 1\right);$$

(iii) if (c) $\sqrt{np_n}|g_{n2}| \to \infty$, *then, for any real number* M,

$$\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\mathrm{Q}_n^{\mathrm{St}} > \mathrm{M}\right] \to 1 \tag{4.7}$$

as $n \to \infty$. *Note that* (4.7) *still holds if (a)–(b) are replaced with the single condition* $e_{n4} = o(ng_{n2}^2)$, *which, in case (iii), is weaker than (a)–(b).*

Proof. See Section 4.3.7. □

Since $g_{n2} = e_{n2} - 1/p_n = 0$ under the null hypothesis of uniformity, $|g_{n2}|$ can be read as a measure of the severity of the alternatives at hand. In this context, Theorem 4.2.3 states that the Bingham test is blind to alternatives for which $g_{n2} = \xi_n/\sqrt{np_n}$ with $\xi_n \to 0$ (Part (i) of the result) and is consistent under alternatives for which $g_{n2} = \xi_n/\sqrt{np_n}$ with $|\xi_n| \to \infty$ (Part (iii) of the result). Therefore, using the terminology from [Bhattacharya, 2019], the “detection threshold" of the Bingham test in high dimensions is associated with alternatives in $g_{n2} = \xi_n/\sqrt{np_n}$, with $\xi_n \to \xi (\neq 0)$, under which the Bingham test achieves the non-trivial asymptotic power

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\mathrm{Q}_n^{\mathrm{St}} > z_\alpha\right] = 1 - \Phi\left(z_\alpha - \frac{\xi^2}{2}\right) \tag{4.8}$$

(Part (ii) of the result). We stress that the results of this section are very general: first, they do not require that the considered rotationally symmetric distributions admit a density of a specific form on the sphere, nor even that they admit a density at all. Second, they do not restrict to axial distributions, that is, they do not assume that F_n is the cumulative distribution function of a symmetric distribution over $[-1,1]$.

We conducted the following simulation exercise to check the validity of Theorem 4.2.3. For each $n \in \{100, 400, 800\}$ and each corresponding dimension $p_n = \lfloor n^a \rfloor$ with $a \in \left\{\frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}\right\}$ (leading to 12 combinations of n and p_n), we generated $\mathrm{M} = 2500$ independent random samples of size n from five rotationally symmetric distributions over $\mathcal{S}^{p_n-1}$. For each rotationally symmetric distribution, $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^{p_n}$ and the cumulative distribution function of the $\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2$'s is a Beta$(\alpha_n, \beta_n)$ distribution, where the parameters

$$\alpha_n = \frac{1}{2} - \frac{p_n^2 g_{n2}}{2p_n(g_{n2}-1)+2} \quad \text{and} \quad \beta_n = \frac{p_n - 1}{2} \tag{4.9}$$

are based on

(0) $g_{n2} = 0$

(i) $g_{n2} = 1/(np_n)$

(ii)$_a$ $g_{n2} = \xi/\sqrt{np_n}$, with $\xi = 2$

(ii)$_b$ $g_{n2} = \xi/\sqrt{np_n}$, with $\xi = 3$

(iii) $g_{n2} = 1/(n^{1/4}\sqrt{p_n})$.

As the notation suggests, α_n and β_n are such that g_{n2} in (4.9) is equal to $e_{n2} - 1/p_n$, where e_{n2} is the second moment associated with F_n, so that this quantity g_{n2} coincides with the one in Theorem 4.2.3. Case (0) yields the null hypothesis of uniformity over $\mathcal{S}^{p_n-1}$, whereas cases (i)–(iii) provide increasingly severe alternatives.

For each of the 12 combinations (n, p_n) and each of these five cases, Figure 4.2 then reports kernel density estimates (obtained from the R command `density` with default parameter values) of the resulting $\mathrm{M} = 2500$ values of the Bingham statistic $\mathrm{Q}_n^{\mathrm{st}}$ (as well as raw histograms in case (ii)$_a$). The figure also provides the densities of the corresponding asymptotic distributions in cases (0)–(ii)$_b$, that are obtained from Theorem 4.2.3(i)–(ii).

Clearly, empirical results are in perfect agreement with the theory, not only for the matching between finite-sample and asymptotic distributions in cases (0)–(ii)$_b$ but also for the consistency behaviour in case (iii) (since kernel density estimates in this case shift to infinity as expected).

4.2.2 Axial alternatives

We consider again axial alternatives to uniformity studied in Chapter 3. As Proposition 3.3.3 showed, the Bingham test is blind to alternatives of the form $\kappa_n = \tau_n p_n/\sqrt{n}$,

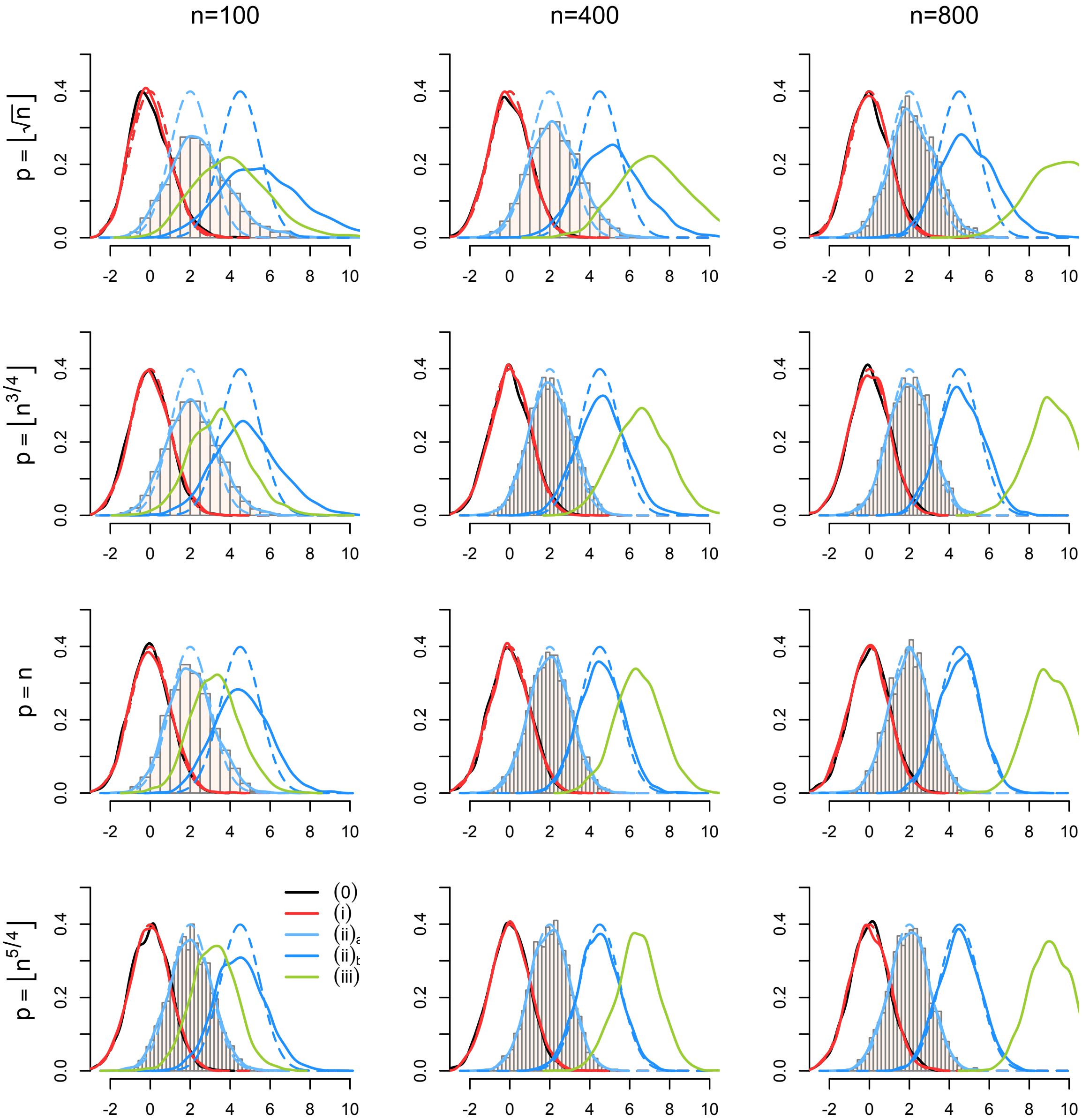


Figure 4.2 – For any $n \in \{100, 400, 800\}$ and any corresponding dimensions $p = p_n = \lfloor n^a \rfloor$ with $a \in \{\frac{1}{2}, \frac{3}{4}, 1, \frac{5}{4}\}$, kernel estimates of the density of $\mathrm{Q}_n^{\mathrm{St}}$ (solid lines) obtained from $\mathrm{M} = 2\,500$ independent samples from the rotationally symmetric distributions associated with $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^p$ and the cumulative distribution functions F_n associated with (0)–(iii) in page 85. In case (ii)$_a$, raw histograms are provided. In cases (0)–(ii)$_b$, the corresponding asymptotic densities, obtained from Theorem 4.2.3(i)–(ii), are also plotted (dashed line).

with $(\tau_n) \to \tau$. By applying the results of Section 4.2.1 we can identify the axial alternatives that can be detected by the Bingham test in high dimensions. In order to do so, we need to study the asymptotic behaviour of e_{n2}, e_{n4} and e_{n8} in the semiparametric model at hand. This is the topic of the following proposition.

Proposition 4.2.4. *Let (p_n) be a sequence of positive integers diverging to ∞, $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n, and (κ_n) a real sequence that is $o(p_n)$ as $n \to \infty$. Fix $f \in \mathcal{F}$. Then, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$,*

$$e_{n2} = \frac{1}{p_n} + \frac{2\kappa_n}{p_n^2} + o\left(\frac{\kappa_n}{p_n^2}\right) \qquad \text{and} \qquad e_{n4} = \frac{3}{p_n^2} + o\left(\frac{1}{p_n^2}\right)$$

as $n \to \infty$, so that $e_{n8} = \mathrm{O}(1/p_n^2)$ as $n \to \infty$.

Proof. See Section 4.3.8. □

The following result is then a corollary of Theorem 4.2.3.

Theorem 4.2.5. *Let (p_n) be a sequence of positive integers diverging to ∞ and $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Fix $f \in \mathcal{F}$. Then, we have the following:*

(i) if $\kappa_n = o(p_n^{3/2}/\sqrt{n})$ and $p_n = \mathrm{O}(n)$ (or more generally, if $\kappa_n = o(p_n^{3/2}/\sqrt{n})$ and $\kappa_n = o(p_n)$), then, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, as $n \to \infty$,

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1); \tag{4.10}$$

(ii) if $\kappa_n = \tau_n p_n^{3/2}/\sqrt{n}$ with $(\tau_n) \to \tau (\neq 0)$ and $p_n = o(n)$, then, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, as $n \to \infty$,

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathcal{D}} \mathcal{N}(2\tau^2,1); \tag{4.11}$$

(iii) if $\sqrt{n}|\kappa_n|/p_n^{3/2} \to \infty$ and $\kappa_n = o(p_n)$, then, for any real number M, as $n \to \infty$,

$$\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{Q}_n^{\mathrm{St}} > \mathrm{M}\right] \to 1.$$

Proof. See Section 4.3.9. □

Part (i) of this result confirms that the Bingham test is indeed blind to alternatives in $\kappa_n = o(p_n/\sqrt{n})$, irrespective of the rate at which p_n diverges to infinity with n. Under some mild assumption on this rate, this extends to alternatives in $\kappa_n = o(p_n^{3/2}/\sqrt{n})$, whereas Part (iii) of the result shows that the Bingham test is consistent under alternatives such that $\sqrt{n}|\kappa_n|/p_n^{3/2} \to \infty$. For axial alternatives, the detection threshold is thus $\kappa_n \sim p_n^{3/2}/\sqrt{n}$; the corresponding asymptotic power, under $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n = \tau_n p^{3/2}/\sqrt{n}$ and $(\tau_n) \to \tau$, is

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{Q}_n^{\mathrm{St}} > z_\alpha\right] = 1 - \Phi\left(z_\alpha - 2\tau^2\right).$$

These powers are symmetric for girdle-type alternatives ($\tau < 0$) and bipolar alternatives ($\tau > 0$), as it was already the case in low dimensions; see Proposition 3.3.2.

We conducted the following simulation. For any $n \in \{100, 400, 800\}$, we generated $\mathrm{M} = 2500$ independent random samples of size n from the Watson distribution with dimension $p_n = \lfloor\sqrt{n}\rfloor$, location $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^{p_n}$, and concentration

(i) $\kappa_n = \tau p_n/\sqrt{n}$,

(ii) $\kappa_n = \tau p_n^{3/2}/\sqrt{n}$,

(iii) $\kappa_n = \tau p_n^{7/4}/\sqrt{n}$,

in each case with $\tau = 0, 0.4, \ldots, 2$.

Figure 4.3 reports the rejection frequencies of the Bingham test as well as the corresponding asymptotic powers obtained from Theorem 4.2.5. The figure also provides the rejection frequencies and asymptotic powers of the Rayleigh test obtained from Theorem 4.1.2. Clearly, the results are in excellent agreement with the theory. In particular, the Bingham test is blind to the alternatives associated with (i) and is consistent under those in (iii). Under the threshold alternatives in (ii), this test shows rejection frequencies that are close to the corresponding asymptotic powers. The Rayleigh test is blind to all alternatives considered, which is also in line with the theory: since $e_{n1} = 0$ under any hypothesis of the form $\check{\mathrm{P}}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, Theorem 4.1.2 implies that the Rayleigh test will be blind to *all* axial alternatives, irrespective of their severity.

4.2.3 Monotone alternatives

As mentioned in Section 4.2.1, the general result in Theorem 4.2.3 is not restricted to axial distributions. It can thus also be used to investigate the non-null behaviour of the Bingham test under the monotone alternatives considered in Chapter 2.

It easily follows from Theorem 2.2.1 and Le Cam's Third Lemma that the Bingham test statistic $\mathrm{Q}_n^{\mathrm{St}}$ remains asymptotically standard normal under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n = \tau_n\sqrt{p_n/n}$ and $(\tau_n) \to \tau$. Consequently, the Bingham test is blind to such contiguous alternatives. It is then natural to wonder whether or not this test can detect more severe monotone alternatives. In order to apply our general result in Theorem 4.2.3, we need to study the asymptotic behaviour of the quantities $e_{n\ell}$, $\ell = 2,4,8$, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$.

Proposition 4.2.6. *Let (p_n) be a sequence of positive integers diverging to ∞, $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n, and (κ_n) a nonnegative real sequence that is $o(\sqrt{p_n})$ as $n \to \infty$. Fix $f \in \mathscr{F}$. Then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$,*

$$e_{n2} = \frac{1}{p_n} + \frac{\kappa_n^2}{p_n^2} f''(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right) \qquad \text{and} \qquad e_{n4} = \frac{3}{p_n^2} + o\left(\frac{1}{p_n^2}\right)$$

as $n \to \infty$, so that $e_{n8} = \mathrm{O}(1/p_n^2)$ as $n \to \infty$.

Proof. See Section 4.3.10. □

The following theorem then results from Theorem 4.2.3.

Theorem 4.2.7. *Let (p_n) be a sequence of positive integers diverging to ∞ and $(\boldsymbol{\theta}_n)$ a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Fix $f \in \mathscr{F}$. Then, we have the following:*

(i) if $\kappa_n = o(p_n^{3/4}/n^{1/4})$ and $p_n = \mathrm{O}(n)$ (or more generally, if $\kappa_n = o(p_n^{3/4}/n^{1/4})$ and $\kappa_n = o(\sqrt{p_n})$), then, under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, as $n \to \infty$,

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathscr{N}(0,1); \tag{4.12}$$

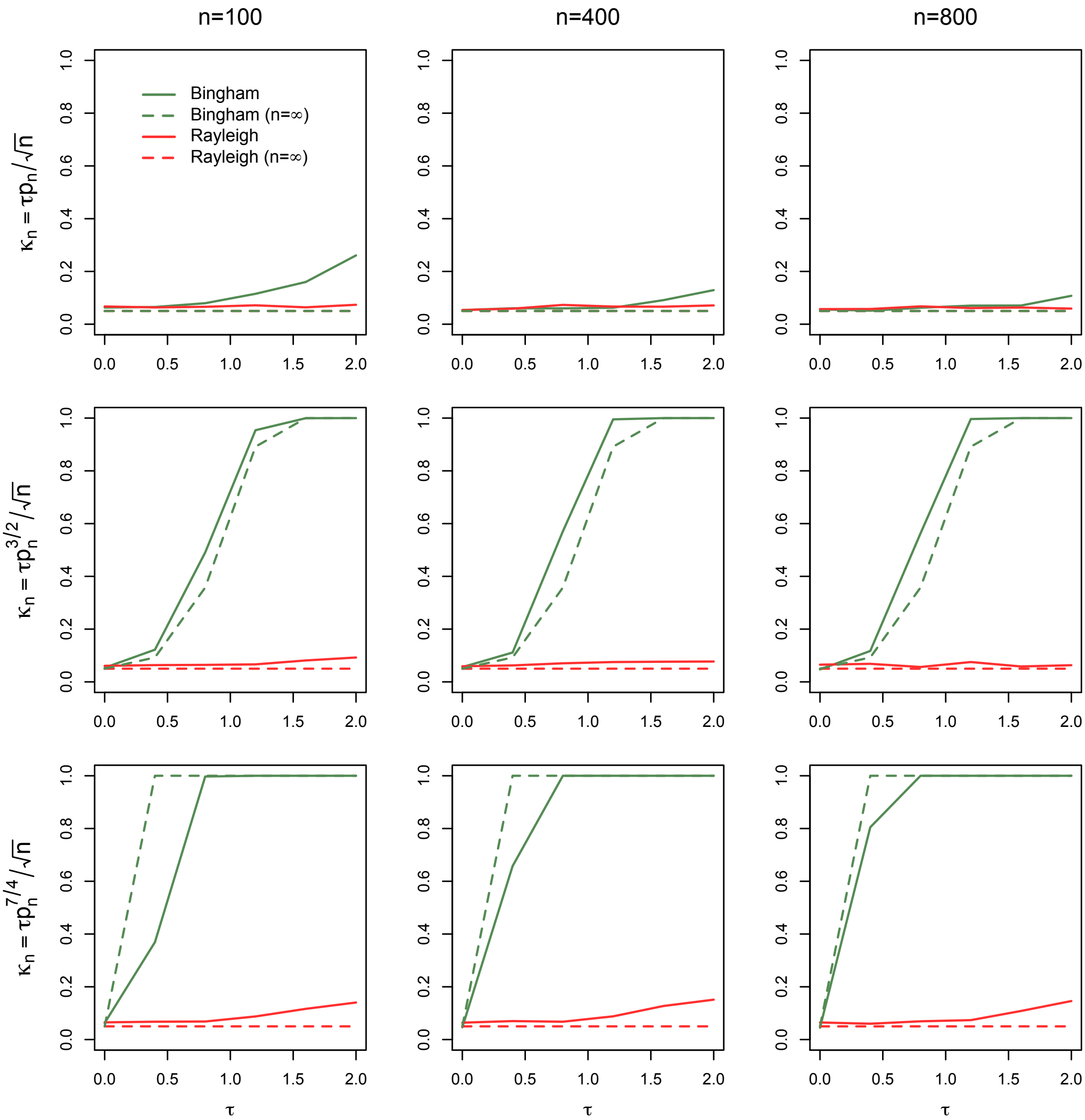


Figure 4.3 – For any $n \in \{100, 400, 800\}$, rejection frequencies of the Bingham test (solid green curve) and of the Rayleigh test (solid red curve), obtained from a collection of $\mathrm{M} = 2500$ independent random samples of size n from the Watson distribution with dimension $p_n = \lfloor\sqrt{n}\rfloor$, location $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^{p_n}$, and concentration $\kappa_n = \tau p_n/\sqrt{n}$ (top), $\kappa_n = \tau p_n^{3/2}/\sqrt{n}$ (middle), or $\kappa_n = \tau p_n^{7/4}/\sqrt{n}$ (bottom). The corresponding asymptotic powers are also plotted (dashed curves).

(ii) if $\kappa_n = \tau_n p_n^{3/4}/n^{1/4}$ *with* $(\tau_n) \to \tau (> 0)$ *and* $p_n = o(n)$*, then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$*, as* $n \to \infty$*,*

$$\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathcal{N}\left(\frac{\tau^4}{2} f''(0)^2, 1\right); \tag{4.13}$$

(iii) if $n^{1/4}\kappa_n/p_n^{3/4} \to \infty$, $\kappa_n = o\left(\sqrt{p_n}\right)$ *and* $f''(0) \neq 0$*, then, for any* $\mathrm{M} \in \mathbb{R}$*, as* $n \to \infty$*,*

$$\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{Q}_n^{\mathrm{St}} > \mathrm{M}\right] \to 1.$$

Proof. See Section 4.3.11. □

Interpreting this result in the same way as Theorem 4.2.5, we learn that the detection threshold of the Bingham test under monotone alternatives is $\kappa_n \sim p_n^{3/4}/n^{1/4}$, with resulting asymptotic powers

$$\lim_{n\to\infty} \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{Q}_n^{\mathrm{St}} > z_\alpha\right] = 1 - \Phi\left(z_\alpha - \frac{\tau^4}{2}(f''(0))^2\right)$$

with $\kappa_n = \tau_n p_n^{3/4}/n^{1/4}$ and $(\tau_n) \to \tau$. This should be compared to the detection threshold of the Rayleigh test, that is $\kappa_n \sim p_n^{3/4}/\sqrt{n}$ (see Section 2.3.2). This is in line with the fact that the Bingham test is primarily designed for axial data whereas the Rayleigh one aims at non-axial data. As mentioned in Section 4.2.2, however, the Rayleigh test will be blind to arbitrarily severe axial alternatives, whereas the Bingham test will show power under both monotone and axial alternatives.

In order to illustrate these results, we performed the following FvML version of the Watson simulation exercise conducted in Section 4.2.2. For any $n \in \{100, 400, 800\}$, we generated $\mathrm{M} = 2500$ independent random samples of size n from the FvML distribution with dimension $p_n = \lfloor\sqrt{n}\rfloor$, location $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^{p_n}$, and concentration

- (i) $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$,
- (ii) $\kappa_n = \tau p_n^{3/4}/n^{1/4}$,
- (iii) $\kappa_n = \tau p_n^{5/4}/n^{1/4}$,

still with $\tau = 0, 0.4, \ldots, 2$ in each case. Figure 4.4 reports the rejection frequencies of the Bingham test and of the Rayleigh test, as well as the corresponding asymptotic powers (obtained from Theorem 4.2.7 for the Bingham test and from Theorem 4.1.2 for the Rayleigh test). The results fully support the comments from the previous paragraph, for both tests. Strictly speaking, the consistency result in Theorem 4.2.7(iii) does not apply in the concentration scheme (iii) above, as the condition $\kappa_n = o\left(\sqrt{p_n}\right)$ is not met there; in Appendix C, however, we show that this condition is superfluous in the FvML case, so that our theoretical results imply consistency in the concentration scheme (iii), too.

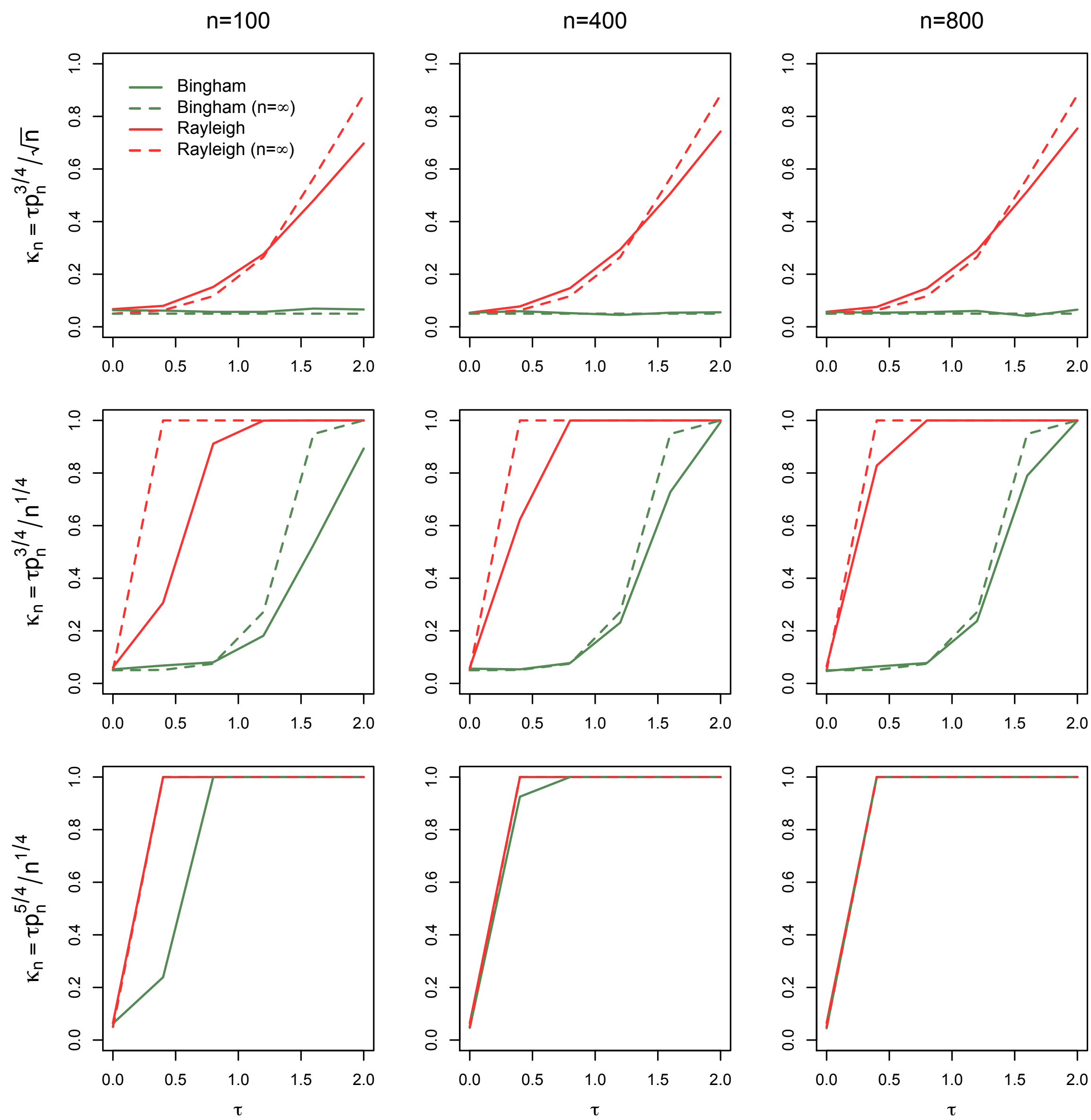


Figure 4.4 – For any $n \in \{100, 400, 800\}$, rejection frequencies of the Bingham test (solid green curve) and of the Rayleigh test (solid red curve), obtained from a collection of $M = 2500$ independent random samples of size n from the FvML distribution with dimension $p_n = \lfloor\sqrt{n}\rfloor$, location $\boldsymbol{\theta}_n = (1, 0, \ldots, 0)' \in \mathbb{R}^{p_n}$, and concentration $\kappa_n = \tau p_n^{3/4}/\sqrt{n}$ (top), $\kappa_n = \tau p_n^{3/4}/n^{1/4}$ (middle), or $\kappa_n = \tau p_n^{5/4}/n^{1/4}$ (bottom). The corresponding asymptotic powers are also plotted (dashed curves).

4.3 Proofs

4.3.1 Preliminary lemmas

Using the tangent-normal decomposition in (1.9), we can write under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

$$\mathbf{X}_{ni} = u_{ni}\boldsymbol{\theta}_n + v_{ni}\mathbf{S}_{ni}$$

where $u_{ni} := \mathbf{X}'_{ni}\boldsymbol{\theta}_n$, $v_{ni} := \sqrt{1-u^2_{ni}}$ and $\mathbf{S}_{ni} := (\mathbf{X}_{ni} - u_{ni}\boldsymbol{\theta}_n)\|\mathbf{X}_{ni} - u_{ni}\boldsymbol{\theta}_n\|^{-1}$ if $\mathbf{X}_{ni} \neq \boldsymbol{\theta}_n$ and $\mathbf{0}$ otherwise. With this notation,

$$e_{n\ell} := \mathrm{E}\left[\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^{\ell}\right] = \mathrm{E}\left[u^{\ell}_{ni}\right],$$
$$f_{n\ell} := \mathrm{E}\left[\left(1-\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n\right)^2\right)^{\ell/2}\right] = \mathrm{E}\left[v^{\ell}_{ni}\right],$$

Lemma 4.3.1. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

(i) $\mathrm{E}\left[\mathbf{S}_{ni}\mathbf{S}'_{ni}\right] = \frac{1}{p_n-1}\left(\mathbf{I}_{p_n} - \boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)$ *for any* i;

(ii) $\mathrm{E}\left[(\mathbf{S}'_{ni}\mathbf{S}_{nj})^2\right] = \frac{1}{p_n-1}$ *for any* $i \neq j$;

(iii) $\mathrm{E}\left[(\mathbf{S}'_{ni}\mathbf{S}_{nj})^4\right] = \frac{3}{p_n^2-1}$ *for any* $i \neq j$;

(iv) $\mathrm{E}\left[(\mathbf{S}'_{ni}\mathbf{S}_{nj})^8\right] = \frac{105}{(p_n^2-1)(p_n+3)(p_n+5)}$ *for any* $i \neq j$.

Proof.

(i) Let $\mathbf{O}$ be a $p_n \times p_n$ orthogonal matrix such that $\mathbf{O}\boldsymbol{\theta}_n = \mathbf{e}_1$, where $\mathbf{e}_1$ denotes the first vector of the canonical basis of $\mathbb{R}^{p_n}$. Then the random vectors $\mathbf{O}\mathbf{S}_{ni}$, $i = 1, \ldots, n$ form a random sample from the uniform distribution over $\left\{\mathbf{x} \in \mathcal{S}^{p_n-1} : \mathbf{e}'_1\mathbf{x} = 0\right\}$. Consequently, $\mathbf{O}\mathrm{E}\left[\mathbf{S}_{ni}\mathbf{S}'_{ni}\right]\mathbf{O}' = \frac{1}{p_n-1}\left(\mathbf{I}_{p_n} - \mathbf{e}_1\mathbf{e}'_1\right)$, which yields the result.

(ii)-(iv) It follows from the joint distribution of the $\mathbf{O}\mathbf{S}_{ni}$'s just derived that, for any $i \neq j$, $\mathbf{S}'_{ni}\mathbf{S}_{nj} = (\mathbf{O}\mathbf{S}_{ni})'(\mathbf{O}\mathbf{S}_{nj})$ is equal in distribution to $\mathbf{U}'\mathbf{V}$, where the independent random (p_n-1)-vectors $\mathbf{U}$, $\mathbf{V}$ are uniformly distributed over $\mathcal{S}^{p_n-2}$. The result then follows from Lemma A.1(iii) in [Paindaveine and Verdebout, 2016].

□

Lemma 4.3.2. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

(i) $\mathrm{E}\left[\mathbf{X}'_{ni}\mathbf{X}_{nj}\right] = e^2_{n1}$ *for any* $i \neq j$;

(ii) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{nj}\right)^2\right] = e^2_{n2} + \frac{f^2_{n2}}{p_n-1} = \frac{p_n}{p_n-1}g^2_{n2} + \frac{1}{p_n}$ *for any* $i \neq j$;

(iii) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{nj}\right)^4\right] = e^2_{n4} + \frac{6(e_{n2}-e_{n4})^2}{p_n-1} + \frac{3f^2_{n4}}{p_n^2-1}$ *for any* $i \neq j$;

(iv) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{nj}\right)\left(\mathbf{X}'_{nk}\mathbf{X}_{n\ell}\right)\right] = e_{n2}e^2_{n1}$ *for any* $i \neq j$ *and* $k \neq \ell$ *such that* $\{i,j,k,\ell\}$ *contains exactly three different indices;*

(v) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{nj}\right)\left(\mathbf{X}'_{nk}\mathbf{X}_{n\ell}\right)\right] = e^4_{n1}$ *if* i,j,k,ℓ *are distinct;*

(vi) $\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)^2\right] = e_{n4}e_{n2}^2 + \frac{2e_{n2}(e_{n2}-e_{n4})f_{n2}}{p_n-1} + \frac{f_{n2}^2 f_{n4}}{(p_n-1)^2}$ *for any* $i \neq j$ *and* $k \neq \ell$ *such that* $\{i,j,k,\ell\}$ *contains exactly three different indices;*

(vii) $\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)^2\right] = \left(\frac{p_n}{p_n-1}g_{n2}^2 + \frac{1}{p_n}\right)^2$ *if* i,j,k,ℓ *are distinct.*

Proof.

(i) The tangent-normal decomposition provides

$$\mathbf{X}_{ni}'\mathbf{X}_{nj} = u_{ni}u_{nj} + v_{ni}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right).$$

As $\mathbf{S}_{ni}$ and $\mathbf{S}_{nj}$ are independent and such that $\mathrm{E}[\mathbf{S}_{ni}] = \mathrm{E}[\mathbf{S}_{nj}] = 0$, the result follows.

(ii) The previous decomposition yields

$$\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2 = u_{ni}^2u_{nj}^2 + v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 + 2u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right). \tag{4.14}$$

Lemma 4.3.1(ii) then shows that

$$\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\right] = \mathrm{E}\left[u_{n1}^2\right]^2 + \frac{\mathrm{E}\left[v_{n1}^2\right]^2}{p_n-1} = e_{n2}^2 + \frac{f_{n2}^2}{p_n-1}$$

The second equality is proven by using successively the identities $f_{n2} = 1 - e_{n2}$ and $e_{n2} = g_{n2} + 1/p_n$.

(iii) Using (4.14) and the fact that $\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^\ell\right] = 0$ for any odd positive integer ℓ, we obtain

$$\begin{aligned}\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^4\right] &= \mathrm{E}\left[u_{ni}^4u_{nj}^4 + 6u_{ni}^2u_{nj}^2v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 + v_{ni}^4v_{nj}^4\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^4\right] \\ &= \mathrm{E}[u_{n1}^4]^2 + \frac{6}{p_n-1}\mathrm{E}\left[u_{n1}^2(1-u_{n1}^2)\right]^2 + \frac{3}{p_n^2-1}\mathrm{E}[v_{n1}^4]^2,\end{aligned}$$

which establishes the result.

(iv)-(v) Note that, for $i<j$ and $k<\ell$,

$$\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)\right] = \mathrm{E}\left[u_{ni}u_{nj}u_{nk}u_{n\ell}\right] + \mathrm{E}\left[v_{ni}v_{nj}v_{nk}v_{n\ell}\right]\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\left(\mathbf{S}_{nk}'\mathbf{S}_{n\ell}\right)\right].$$

There is always one of the indices i,j,k,ℓ that is different from the other three indices, which implies that $\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\left(\mathbf{S}_{nk}'\mathbf{S}_{n\ell}\right)\right] = 0$. The result readily follows.

(vi) Without any loss of generality, assume that, in $\{i,j,k,\ell\}$, only j and k are equal to each other. Proceeding as in (iii), we then have

$$\begin{aligned}&\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)^2\right] \\ &= \mathrm{E}\left[u_{ni}^2u_{nj}^2u_{nk}^2u_{n\ell}^2 + u_{ni}^2u_{nj}^2v_{nk}^2v_{n\ell}^2\left(\mathbf{S}_{nk}'\mathbf{S}_{n\ell}\right)^2\right. \\ &\qquad \left. + u_{nk}^2u_{n\ell}^2v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 + v_{ni}^2v_{nj}^2v_{nk}^2v_{n\ell}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2\left(\mathbf{S}_{nk}'\mathbf{S}_{n\ell}\right)^2\right] \\ &= \mathrm{E}\left[u_{ni}^2\right]^2\mathrm{E}\left[u_{nk}^4\right] + \frac{2}{p_n-1}\mathrm{E}\left[u_{ni}^2\right]\mathrm{E}\left[v_{ni}^2\right]\mathrm{E}\left[u_{ni}^2(1-u_{ni}^2)\right] + \frac{1}{(p_n-1)^2}\mathrm{E}\left[v_{ni}^2\right]^2\mathrm{E}\left[v_{ni}^4\right],\end{aligned}$$

where we used the fact that

$$\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2\left(\mathbf{S}_{nj}'\mathbf{S}_{n\ell}\right)^2\right] = \mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2\left(\mathbf{S}_{nj}'\mathbf{S}_{n\ell}\right)^2 \middle| \mathbf{S}_{nj}\right]\right] = \frac{1}{(p_n-1)^4}$$

since $\mathrm{E}\left[\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 \middle| \mathbf{S}_{nj}\right] = 1/(p_n-1)$.

(vii) Since i, j, k, ℓ are pairwise different, $\mathbf{X}'_{ni}\mathbf{X}_{nj}$ and $\mathbf{X}'_{nk}\mathbf{X}_{n\ell}$ are mutually independent, so that the result directly follows from Part (ii) of the lemma.

□

Lemma 4.3.3. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

(i) $\mathrm{E}\left[\left(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n\right)\left(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n\right)'\right]=\tilde{e}_{n2}\boldsymbol{\theta}_n\boldsymbol{\theta}'_n+\frac{f_{n2}}{p_n-1}\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)$;

(ii) $\mathrm{Var}\left[\left(\mathbf{X}'_{ni}\boldsymbol{\theta}_n-e_{n1}\right)\left(\mathbf{X}'_{nj}\boldsymbol{\theta}_n-e_{n1}\right)\right]=\tilde{e}_{n4}-\tilde{e}^2_{n2}$ *for* $i=j$ *and* $\tilde{e}^2_{n2}$ *for* $i\neq j$;

(iii) $\mathrm{E}\left[\mathbf{X}'_{ni}\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)\mathbf{X}_{nj}\right]=f_{n2}$ *for* $i=j$ *and* 0 *for* $i\neq j$;

(iv) $\mathrm{Var}\left[\mathbf{X}'_{ni}\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)\mathbf{X}_{nj}\right]=f_{n4}-f^2_{n2}$ *for* $i=j$ *and* $f^2_{n2}/(p_n-1)$ *for* $i\neq j$.

Proof.

(i) Using the tangent-normal decomposition and Lemma 4.3.1(i), we obtain

$$\begin{aligned}\mathrm{E}\left[(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)'\right]&=\mathrm{E}\left[\left((u_{ni}-e_{n1})\boldsymbol{\theta}_n+v_{ni}\mathbf{S}_{ni}\right)\left((u_{ni}-e_{n1})\boldsymbol{\theta}_n+v_{ni}\mathbf{S}_{ni}\right)'\right]\\&=\mathrm{E}\left[(u_{ni}-e_{n1})^2\right]\boldsymbol{\theta}_n\boldsymbol{\theta}'_n+f_{n2}\,\mathrm{E}\left[\mathbf{S}_{ni}\mathbf{S}'_{ni}\right]\\&=\tilde{e}_{n2}\boldsymbol{\theta}_n\boldsymbol{\theta}'_n+\frac{f_{n2}}{p_n-1}\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right).\end{aligned}$$

(ii)-(iv) The results readily follow from the fact that $\mathbf{X}'_{ni}\boldsymbol{\theta}_n-e_{n1}=u_{ni}-e_{n1}$,

$$\mathbf{X}'_{ni}\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)\mathbf{X}_{nj}=v_{ni}v_{nj}\mathbf{S}'_{ni}\mathbf{S}_{nj},$$

and from Lemma 4.3.1(ii).

□

Lemma 4.3.4. *Consider expectations of the form* $c_{ijrs}=\mathrm{E}\left[\Delta_{i\ell}\Delta_{j\ell}\Delta_{r\ell}\Delta_{s\ell}\right]$ *taken under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$, *with* $\Delta_{i\ell}:=\left(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n\right)'\left(\mathbf{X}_{n\ell}-e_{n1}\boldsymbol{\theta}_n\right)$ *and* $i\leq j\leq r\leq s<\ell$. *Then*

(i) $c_{ijrs}=\tilde{e}^2_{n4}+\frac{6}{p_n-1}\mathrm{E}\left[v^2_{ni}(u_{ni}-e_{n1})^2\right]^2+\frac{3f^2_{n4}}{p^2_n-1}$ *if* $i=j=r=s$;

(ii) $c_{ijrs}=\tilde{e}^2_{n2}\tilde{e}_{n4}+\frac{2\tilde{e}_{n2}f_{n2}}{p_n-1}\mathrm{E}\left[v^2_{ni}(u_{ni}-e_{n1})^2\right]+\frac{f^2_{n2}f_{n4}}{(p_n-1)^2}$ *if* $i=j<r=s$;

(iii) $c_{ijrs}=0$ *otherwise.*

Proof.

(i)-(ii) We use the tangent-normal decomposition again to write

$$\Delta_{j\ell}=\left(u_{nj}-e_{n1}\right)\left(u_{n\ell}-e_{n1}\right)+v_{nj}v_{n\ell}\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right).$$

Since $\mathrm{E}\left[\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^k\right]=0$ for any odd integer k, this leads to decomposing c_{jjrr} into

$$\begin{aligned}c_{jjrr}=\;&\mathrm{E}\left[\left(u_{nj}-e_{n1}\right)^2\left(u_{nr}-e_{n1}\right)^2\left(u_{n\ell}-e_{n1}\right)^4\right]\\&+4\mathrm{E}\left[\left(u_{nj}-e_{n1}\right)\left(u_{nr}-e_{n1}\right)\left(u_{n\ell}-e_{n1}\right)^2v_{nj}v_{nr}v^2_{n\ell}\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)\left(\mathbf{S}'_{nr}\mathbf{S}_{n\ell}\right)\right]\\&+2\mathrm{E}\left[\left(u_{nr}-e_{n1}\right)^2\left(u_{n\ell}-e_{n1}\right)^2v^2_{nj}v^2_{n\ell}\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^2\right]+\mathrm{E}\left[v^2_{nj}v^2_{nr}v^4_{n\ell}\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^2\left(\mathbf{S}'_{nr}\mathbf{S}_{n\ell}\right)^2\right].\end{aligned}$$

The result then follows from Lemma 4.3.1(ii)-(iii).

(iii) Assume that $j = r$, so that we are not in case (ii). Since case (i) is excluded, we have $i < j$ or $r < s$. In both cases, one of the four indices i, j, r, s is different from the other three indices. Since $\mathrm{E}[\Delta_{i\ell}] = 0$, we obtain that $c_{ijrs} = 0$, which establishes (iii).

□

Lemma 4.3.5. *Let* $\mathbf{U}$ *be uniformly distributed on* $\mathcal{S}^{p-1}$*. Then,*

$$\mathrm{E}\left[(\mathbf{v}'\mathbf{U})^2(\mathbf{w}'\mathbf{U})^2\right] = \frac{2(\mathbf{v}'\mathbf{w})^2+1}{p(p+2)}$$

for any $\mathbf{v}, \mathbf{w} \in \mathcal{S}^{p-1}$.

Proof. Let $\mathbf{K}_\ell$ be the $\ell^2 \times \ell^2$ commutation matrix and define $\mathbf{J}_\ell := (\mathrm{vec}\,\mathbf{I}_\ell)(\mathrm{vec}\,\mathbf{I}_\ell)'$, where $\mathrm{vec}\,\mathbf{A}$ is the vector stacking the columns of $\mathbf{A}$ on top of each other. Then,

$$\mathrm{E}\left[(\mathbf{v}'\mathbf{U})^2(\mathbf{w}'\mathbf{U})^2\right] = \mathrm{E}\left[\mathbf{v}'\mathbf{U}\mathbf{U}'\mathbf{v}\mathbf{w}'\mathbf{U}\mathbf{U}'\mathbf{w}\right] = (\mathbf{v}\otimes\mathbf{v})'\mathrm{E}\left[\mathrm{vec}(\mathbf{U}\mathbf{U}')\mathrm{vec}'(\mathbf{U}\mathbf{U}')\right](\mathbf{w}\otimes\mathbf{w})$$

Lemma A.2(iii) from [Paindaveine and Verdebout, 2016] then yields

$$\mathrm{E}\left[(\mathbf{v}'\mathbf{U})^2(\mathbf{w}'\mathbf{U})^2\right] = \frac{1}{p(p+2)}(\mathbf{v}\otimes\mathbf{v})'(\mathbf{I}_{p^2}+\mathbf{J}_p+\mathbf{K}_p)(\mathbf{w}\otimes\mathbf{w}) = \frac{2(\mathbf{v}'\mathbf{w})^2+1}{p(p+2)},$$

where we used $(\mathrm{vec}\,\mathbf{A})'(\mathrm{vec}\,\mathbf{B}) = \mathrm{tr}[\mathbf{A}'\mathbf{B}]$ and $\mathbf{K}_p(\mathbf{w}\otimes\mathbf{w}) = (\mathbf{w}\otimes\mathbf{w})\mathbf{K}_1 = (\mathbf{w}\otimes\mathbf{w})$ (see for example [Magnus and Neudecker, 2007]). □

Lemma 4.3.6. *Under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

(i) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{n\ell}\right)^2|\mathbf{X}_{ni}\right] = \frac{p_n}{p_n-1}g_{n2}u_{ni}^2 + \frac{f_{n2}}{p_n-1}$ *for any* $i \neq \ell$;

(ii) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{n\ell}\right)^2\left(\mathbf{X}'_{nj}\mathbf{X}_{n\ell}\right)^2|\mathbf{X}_{ni},\mathbf{X}_{nj}\right] = \frac{e_{n2}-e_{n4}}{p_n-1} + \frac{p_ne_{n4}-e_{n2}}{p_n-1}u_{ni}^2u_{nj}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{e_{n2}-e_{n4}}{p_n-1}\right)v_{ni}^2v_{nj}^2$ $+ \frac{2f_{n4}}{p_n^2-1}v_{ni}^2v_{nj}^2\left(\mathbf{S}'_{ni}\mathbf{S}_{nj}\right)^2 + \frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}'_{ni}\mathbf{S}_{nj}\right)$ *for any indices* i, j, ℓ *such that* $\ell \notin \{i, j\}$;

(iii) $\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{n\ell}\right)^2u_{n\ell}^2|\mathbf{X}_{ni}\right] = \frac{e_{n2}-e_{n4}}{p_n-1} + \frac{p_ne_{n4}-e_{n2}}{p_n-1}u_{ni}^2$ *for any* $i \neq \ell$.

Proof.

(i) Using (4.14) and the fact that $\mathrm{E}\left[\left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)^2|\mathbf{X}_{ni}\right] = 1/(p_n-1)$ for any $i \neq \ell$,

$$\begin{aligned}\mathrm{E}\left[\left(\mathbf{X}'_{ni}\mathbf{X}_{n\ell}\right)^2|\mathbf{X}_{ni}\right] &= \mathrm{E}\left[u_{ni}^2u_{n\ell}^2 + v_{ni}^2v_{n\ell}^2\left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)^2 + 2u_{ni}v_{ni}u_{n\ell}v_{n\ell}\left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)|\mathbf{X}_{ni}\right]\\ &= e_{n2}u_{ni}^2 + v_{ni}^2\frac{f_{n2}}{p_n-1} = \left(e_{n2} - \frac{f_{n2}}{p_n-1}\right)u_{ni}^2 + \frac{f_{n2}}{p_n-1}\\ &= \frac{p_n}{p_n-1}g_{n2}u_{ni}^2 + \frac{f_{n2}}{p_n-1},\end{aligned}$$

where we used the identities $v_{ni}^2 = 1-u_{ni}^2$, $f_{n2} = 1-e_{n2}$ and $g_2 = e_2 - 1/p_n$.

(ii) Fix i, j, ℓ with $\ell \notin \{i, j\}$. Since $\mathrm{E}\left[\left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^2 |\mathbf{X}_{ni}, \mathbf{X}_{nj}\right] = 0$, we have

$$
\begin{aligned}
&\mathrm{E}[(\mathbf{X}'_{ni}\mathbf{X}_{n\ell})^2 \left(\mathbf{X}'_{nj}\mathbf{X}_{n\ell}\right)^2 |\mathbf{X}_{ni}, \mathbf{X}_{nj}] \\
&= \mathrm{E}\Big[u_{ni}^2 u_{nj}^2 u_{n\ell}^4 + u_{ni}^2 v_{nj}^2 u_{n\ell}^2 v_{n\ell}^2 \left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^2 + v_{ni}^2 u_{nj}^2 u_{n\ell}^2 v_{n\ell}^2 \left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)^2 \\
&\quad + v_{ni}^2 v_{nj}^2 v_{n\ell}^4 \left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)^2 \left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right)^2 + 4 u_{ni} v_{ni} u_{nj} v_{nj} u_{n\ell}^2 v_{n\ell}^2 \left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)\left(\mathbf{S}'_{nj}\mathbf{S}_{n\ell}\right) |\mathbf{X}_{ni}, \mathbf{X}_{nj}\Big].
\end{aligned}
$$

Therefore, applying Lemma 4.3.5 (in the fourth term of the righthand side) and Lemma 4.3.1(i) (in the fifth one) provides

$$
\begin{aligned}
\mathrm{E}\left[(\mathbf{X}'_{ni}\mathbf{X}_{n\ell})^2 \left(\mathbf{X}'_{nj}\mathbf{X}_{n\ell}\right)^2 |\mathbf{X}_{ni}, \mathbf{X}_{nj}\right] &= e_{n4} u_{ni}^2 u_{nj}^2 + \frac{e_{n2} - e_{n4}}{p_n - 1}\left(u_{ni}^2 v_{nj}^2 + v_{ni}^2 u_{nj}^2\right) \\
&+ \frac{f_{n4}}{p_n^2 - 1} v_{ni}^2 v_{nj}^2 \left(1 + 2\left(\mathbf{S}'_{ni}\mathbf{S}_{nj}\right)^2\right) + \frac{4(e_{n2} - e_{n4})}{p_n - 1} u_{ni} v_{ni} u_{nj} v_{nj} \left(\mathbf{S}'_{ni}\mathbf{S}_{nj}\right).
\end{aligned}
$$

The result then follows by using the identity $u_{ni}^2 v_{nj}^2 + v_{ni}^2 u_{nj}^2 = 1 - u_{ni}^2 u_{nj}^2 - v_{ni}^2 v_{nj}^2$ (which result from the fact that $u_{ni}^2 + v_{ni}^2 = 1$ for any i).

(iii) Finally,

$$
\begin{aligned}
\mathrm{E}\left[(\mathbf{X}'_{ni}\mathbf{X}_{n\ell})^2 u_{n\ell}^2 |\mathbf{X}_{ni}\right] &= \mathrm{E}\left[\left(u_{ni}^2 u_{n\ell}^2 + v_{ni}^2 v_{n\ell}^2 \left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)^2 + 2 u_{ni} v_{ni} u_{n\ell} v_{n\ell} \left(\mathbf{S}'_{ni}\mathbf{S}_{n\ell}\right)\right) u_{n\ell}^2 |\mathbf{X}_{ni}\right] \\
&= e_{n4} u_{ni}^2 + \frac{e_{n2} - e_{n4}}{p_n - 1} v_{ni}^2 = \frac{e_{n2} - e_{n4}}{p_n - 1} + \frac{p_n e_{n4} - e_{n2}}{p_n - 1} u_{ni}^2,
\end{aligned}
$$

where we used again the identity $v_{ni}^2 = 1 - u_{ni}^2$.

□

4.3.2 Proof of Proposition 4.1.1

Since the expectation readily follows from Lemma 4.3.2(i), we can focus on the variance. Using Lemma 4.3.2(i) again, we obtain

$$
\mathrm{Var}_{\mathrm{F}_n}[\mathrm{R}_n^{\mathrm{St}}] = \frac{2p_n}{n^2} \sum_{1 \le i < j \le n} \sum_{1 \le k < \ell \le n} \left(\mathrm{E}\left[(\mathbf{X}'_{ni}\mathbf{X}_{nj})(\mathbf{X}'_{nk}\mathbf{X}_{n\ell})\right] - e_{n1}^4\right).
$$

In this sum, there are $\binom{n}{2}$ terms corresponding to Lemma 4.3.2(ii) and $6\binom{n}{4}$ terms (not contributing to the sum) corresponding to Lemma 4.3.2(v). Therefore, there are $\binom{n}{2}^2 - \binom{n}{2} - 6\binom{n}{4} = n(n-1)(n-2)$ terms corresponding to Lemma 4.3.2(iv). Consequently,

$$
\begin{aligned}
\mathrm{Var}_{\mathrm{F}_n}[\mathrm{R}_n^{\mathrm{St}}] &= \frac{2p_n}{n^2}\left\{\binom{n}{2}\left(e_{n2}^2 + \frac{f_{n2}^2}{p_n - 1} - e_{n1}^4\right) + n(n-1)(n-2)\left(e_{n2}e_{n1}^2 - e_{n1}^4\right)\right\} \\
&= \frac{p_n(n-1)}{n}\left\{\left(e_{n2}^2 - e_{n1}^4\right) + 2(n-2)e_{n1}^2\left(e_{n2} - e_{n1}^2\right) + \frac{f_{n2}^2}{p_n - 1}\right\}
\end{aligned}
$$

which, since $\tilde{e}_{n2} = e_{n2} - e_{n1}^2$, establishes the result. □

4.3.3 Proof of Theorem 4.1.2

This proof is based on Theorem 1.4.1. Let $\mathscr{F}_{n\ell}$ the σ-algebra generated by $\mathbf{X}_{n1},\ldots,\mathbf{X}_{n\ell}$ and $\mathrm{E}_{n\ell}$ the conditional expectation with respect to $\mathscr{F}_{n\ell}$. Define

$$\mathrm{Z}_{n\ell} := \mathrm{E}_{n\ell}\left[\frac{\mathrm{R}_n^{\mathrm{St}} - \mathrm{E}\left[\mathrm{R}_n^{\mathrm{St}}\right]}{\sigma_n}\right] = \frac{\sqrt{2p_n}}{n\sigma_n}\left\{\sum_{1\leq i<j\leq \ell}(\mathbf{X}'_{ni}\mathbf{X}_{nj} - e_{n1}^2) + (n-\ell)e_{n1}\sum_{i=1}^{\ell}(\mathbf{X}'_{ni}\boldsymbol{\theta}_n - e_{n1})\right\}.$$

Note that

$$\frac{\mathrm{R}_n^{\mathrm{St}} - \mathrm{E}\left[\mathrm{R}_n^{\mathrm{St}}\right]}{\sigma_n} = \sum_{\ell=1}^{n}\mathrm{D}_{n\ell},$$

where $\mathrm{D}_{n\ell} := \mathrm{Z}_{n\ell} - \mathrm{Z}_{n,\ell-1}$ rewrites

$$\mathrm{D}_{n\ell} = \frac{\sqrt{2p_n}}{n\sigma_n}\left\{\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni} - e_{n1}\boldsymbol{\theta}_n) + (n-1)e_{n1}\boldsymbol{\theta}_n\right\}'(\mathbf{X}_{n\ell} - e_{n1}\boldsymbol{\theta}_n),\ \ \ell = 1,2,\ldots \tag{4.15}$$

Throughout, sums over empty set of indices are defined as being equal to zero. Conditions (1.20) and (1.21) will be established thanks to Lemmas 4.3.7 and 4.3.8.
In the rest of the proof, C is a positive constant that may change from line to line.

Lemma 4.3.7. *Let the assumptions of Theorem 4.1.2 hold. Recall that* $\sigma_{n\ell}^2 := \mathrm{E}_{n,\ell-1}\left[\mathrm{D}_{n\ell}^2\right]$. *Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$,

(i) $\sum_{\ell=1}^{n}\mathrm{E}\left[\sigma_{n\ell}^2\right]$ *converges to one as* $n\to\infty$;

(ii) $\mathrm{Var}\left[\sum_{\ell=1}^{n}\sigma_{n\ell}^2\right]$ *converges to zero as* $n\to\infty$.

Proof. (i) Note that

$$\sigma_{n\ell}^2 = \frac{2p_n}{n^2\sigma_n^2}\left\{\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni} - e_{n1}\boldsymbol{\theta}_n) + (n-1)e_{n1}\boldsymbol{\theta}_n\right\}'\mathrm{E}\left[(\mathbf{X}_{n\ell} - e_{n1}\boldsymbol{\theta}_n)(\mathbf{X}_{n\ell} - e_{n1}\boldsymbol{\theta}_n)'\right]$$
$$\times\left\{\sum_{j=1}^{\ell-1}(\mathbf{X}_{nj} - e_{n1}\boldsymbol{\theta}_n) + (n-1)e_{n1}\boldsymbol{\theta}_n\right\}.$$

By using Lemma 4.3.3(i), we obtain

$$\sigma_{n\ell}^2 = \frac{2p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\left\{\sum_{i,j=1}^{\ell-1}(\mathbf{X}'_{ni}\boldsymbol{\theta}_n - e_{n1})\left(\mathbf{X}'_{nj}\boldsymbol{\theta}_n - e_{n1}\right) + 2(n-1)e_{n1}\sum_{i=1}^{\ell-1}(\mathbf{X}'_{ni}\boldsymbol{\theta}_n - e_{n1})\right.$$
$$\left. + (n-1)^2e_{n1}^2\right\} + \frac{2p_nf_{n2}}{(p_n-1)n^2\sigma_n^2}\sum_{i,j=1}^{\ell-1}\mathbf{X}'_{ni}\left(\mathbf{I}_{p_n} - \boldsymbol{\theta}_n\boldsymbol{\theta}'_n\right)\mathbf{X}_{nj}. \tag{4.16}$$

Therefore

$$\mathrm{E}\left[\sigma_{n\ell}^2\right] = \frac{2p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\left\{(\ell-1)\tilde{e}_{n2} + 0 + (n-1)^2e_{n1}^2\right\} + \frac{2p_n(\ell-1)f_{n2}^2}{(p_n-1)n^2\sigma_n^2}, \tag{4.17}$$

where we have used Lemma 4.3.3(iii). This yields

$$s_n^2 := \sum_{\ell=1}^{n}\mathrm{E}\left[\sigma_{n\ell}^2\right] = \frac{(n-1)p_n\tilde{e}_{n2}^2}{n\sigma_n^2} + \frac{2p_n\tilde{e}_{n2}}{n\sigma_n^2}(n-1)^2e_{n1}^2 + \frac{(n-1)p_nf_{n2}^2}{(p_n-1)n\sigma_n^2} \to 1$$

as $n\to\infty$, as was to be shown.

(ii) From (4.16), we obtain

$$\mathrm{Var}\left[\sum_{\ell=1}^{n}\sigma_{n\ell}^2\right] \le \mathrm{C}\left(\mathrm{Var}\left[\mathrm{A}_n\right]+\mathrm{Var}\left[\mathrm{B}_n\right]+\mathrm{Var}\left[\mathrm{C}_n\right]\right),$$

where

$$\mathrm{A}_n := \frac{p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{i,j=1}^{\ell-1}\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{nj}'\boldsymbol{\theta}_n - e_{n1}\right),$$

$$\mathrm{B}_n := \frac{p_n e_{n1}\tilde{e}_{n2}}{n\sigma_n^2}\sum_{\ell=1}^{n}\sum_{i=1}^{\ell-1}\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right),$$

$$\mathrm{C}_n := \frac{f_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{i,j=1}^{\ell-1}\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n} - \boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{nj}.$$

We establish the result by showing that, under the assumptions considered, $\mathrm{Var}[\mathrm{A}_n]$, $\mathrm{Var}[\mathrm{B}_n]$ and $\mathrm{Var}[\mathrm{C}_n]$ all are $o(1)$ as $n\to\infty$. We start with A_n, which we split into

$$\begin{aligned}\mathrm{A}_n &= \frac{p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{i=1}^{\ell-1}\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)^2 + \frac{2p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{1\le i<j\le \ell-1}\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{nj}'\boldsymbol{\theta}_n - e_{n1}\right)\\ &= \frac{p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\sum_{i=1}^{n-1}(n-i)\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)^2 + \frac{2p_n\tilde{e}_{n2}}{n^2\sigma_n^2}\sum_{1\le i<j\le n-1}(n-j)\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{nj}'\boldsymbol{\theta}_n - e_{n1}\right),\end{aligned}$$

that is, into $\mathrm{A}_n^{(1)}+\mathrm{A}_n^{(2)}$, say. Clearly,

$$\begin{aligned}\mathrm{Var}\left[\mathrm{A}_n^{(1)}\right] &= \frac{p_n^2\tilde{e}_{n2}^2}{n^4\sigma_n^4}\sum_{i=1}^{n-1}(n-i)^2\,\mathrm{Var}\left[\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)^2\right] \le \mathrm{C}\,\frac{p_n^2\tilde{e}_{n2}^2\left(\tilde{e}_{n4}-\tilde{e}_{n2}^2\right)}{n\sigma_n^4}\\ &\le \mathrm{C}\,\frac{p_n^2\tilde{e}_{n2}^2\left(\tilde{e}_{n4}-\tilde{e}_{n2}^2\right)}{n(p_n\tilde{e}_{n2}^2)^2} = \mathrm{C}\left(\frac{\tilde{e}_{n4}}{n\tilde{e}_{n2}^2}-\frac{1}{n}\right),\end{aligned}$$

which, by assumption, is $o(1)$ as $n\to\infty$. Since $\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{nj}'\boldsymbol{\theta}_n - e_{n1}\right)$, $i<j$, and $\left(\mathbf{X}_{nk}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{n\ell}'\boldsymbol{\theta}_n - e_{n1}\right)$, $k<\ell$, are uncorrelated as soon as $(i,j)\neq(k,\ell)$, we obtain

$$\mathrm{Var}\left[\mathrm{A}_n^{(2)}\right] = \frac{4p_n^2\tilde{e}_{n2}^2}{n^4\sigma_n^4}\sum_{1\le i<j\le n-1}(n-j)^2\,\mathrm{Var}\left[\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\left(\mathbf{X}_{nj}'\boldsymbol{\theta}_n - e_{n1}\right)\right] \le \mathrm{C}\,\frac{p_n^2\tilde{e}_{n2}^4}{\sigma_n^4}.$$

In view of the upper bounds

$$\frac{p_n^2\tilde{e}_{n2}^4}{\sigma_n^4} \le \mathrm{C}\frac{p_n^2\tilde{e}_{n2}^4}{(2np_ne_{n1}^2\tilde{e}_{n2})^2} = \mathrm{C}\left(\frac{\tilde{e}_{n2}}{ne_{n1}^2}\right)^2 \quad\text{and}\quad \frac{p_n^2\tilde{e}_{n2}^4}{\sigma_n^4} \le \mathrm{C}\left(\frac{p_n\tilde{e}_{n2}^2}{f_{n2}^2}\right)^2,$$

$\mathrm{Var}\left[\mathrm{A}_n^{(2)}\right]$, by assumption, is $o(1)$ as $n\to\infty$. So $\mathrm{Var}[\mathrm{A}_n]$ is indeed $o(1)$ as $n\to\infty$.

Turning to B_n,

$$\begin{aligned}\mathrm{Var}[\mathrm{B}_n] &= \frac{p_n^2e_{n1}^2\tilde{e}_{n2}^2}{n^2\sigma_n^4}\mathrm{Var}\left[\sum_{i=1}^{n-1}(n-i)\left(\mathbf{X}_{ni}'\boldsymbol{\theta}_n - e_{n1}\right)\right] = \frac{p_n^2e_{n1}^2\tilde{e}_{n2}^2}{n^2\sigma_n^4}\sum_{i=1}^{n-1}(n-i)^2\tilde{e}_{n2}\\ &\le \mathrm{C}\frac{np_n^2e_{n1}^2\tilde{e}_{n2}^3}{\sigma_n^4},\end{aligned}$$

which is $o(1)$ as $n\to\infty$ since it can be bounded above by

$$\mathrm{C}\frac{np_n^2e_{n1}^2\tilde{e}_{n2}^3}{(2np_ne_{n1}^2\tilde{e}_{n2})^2}=\mathrm{C}\frac{\tilde{e}_{n2}}{ne_{n1}^2}\quad\text{and by}\quad\mathrm{C}\frac{np_n^2e_{n1}^2\tilde{e}_{n2}^3}{np_ne_{n1}^2\tilde{e}_{n2}f_{n2}^2}=\mathrm{C}\frac{p_n\tilde{e}_{n2}^2}{f_{n2}^2}.$$

Finally, we consider C_n. Proceeding as for A_n, we split C_n into

$$\begin{aligned}\mathrm{C}_n &= \frac{f_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{i=1}^{\ell-1}\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{ni}+\frac{2f_{n2}}{n^2\sigma_n^2}\sum_{\ell=1}^{n}\sum_{1\le i<j\le\ell-1}\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{nj}\\ &= \frac{f_{n2}}{n^2\sigma_n^2}\sum_{i=1}^{n-1}(n-i)\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{ni}+\frac{2f_{n2}}{n^2\sigma_n^2}\sum_{1\le i<j\le n-1}(n-j)\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{nj},\end{aligned}$$

that is, into $\mathrm{C}_n^{(1)}+\mathrm{C}_n^{(2)}$, say. Clearly,

$$\mathrm{Var}\left[\mathrm{C}_n^{(1)}\right]=\frac{f_{n2}^2}{n^4\sigma_n^4}\sum_{i=1}^{n-1}(n-i)^2\,\mathrm{Var}\left[\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{ni}\right]\le\mathrm{C}\frac{f_{n2}^2(f_{n4}-f_{n2}^2)}{n\sigma_n^4}\le\mathrm{C}\frac{f_{n4}-f_{n2}^2}{nf_{n2}^2},$$

so that $\mathrm{Var}\left[\mathrm{C}_n^{(1)}\right]$ is $o(1)$ as $n\to\infty$. Since $\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{nj}$, $i<j$, and $\mathbf{X}_{nk}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{n\ell}$, $k<\ell$, are uncorrelated as soon as $(i,j)\neq(k,\ell)$, we obtain

$$\mathrm{Var}\left[\mathrm{C}_n^{(2)}\right]=\frac{4f_{n2}^2}{n^4\sigma_n^4}\sum_{1\le i<j\le n-1}(n-j)^2\,\mathrm{Var}\left[\mathbf{X}_{ni}'\left(\mathbf{I}_{p_n}-\boldsymbol{\theta}_n\boldsymbol{\theta}_n'\right)\mathbf{X}_{nj}\right]\le\mathrm{C}\frac{f_{n2}^4}{\sigma_n^4(p_n-1)}\le\frac{\mathrm{C}}{p_n}.$$

Therefore, $\mathrm{Var}[\mathrm{C}_n]$ is also $o(1)$ as $n\to\infty$, which establishes the result.

□

Lemma 4.3.8. *Let the assumptions of Theorem 4.1.2 hold and fix* $\varepsilon>0$. *Then, under* $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}$, $\sum_{\ell=1}^n\mathrm{E}\left[\mathrm{D}_{n\ell}^2\mathbb{I}_{\{|\mathrm{D}_{n\ell}|>\varepsilon\}}\right]\to0$ *as* $n\to\infty$.

Proof. The Cauchy–Schwarz and Chebychev inequalities yield

$$\sum_{\ell=1}^{n}\mathrm{E}\left[\mathrm{D}_{n\ell}^2\mathbb{I}_{\{|\mathrm{D}_{n\ell}|>\varepsilon\}}\right]\le\sum_{\ell=1}^{n}\sqrt{\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]\mathrm{P}\left[|\mathrm{D}_{n\ell}|>\varepsilon\right]}\le\frac{1}{\varepsilon}\sum_{\ell=1}^{n}\sqrt{\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]\mathrm{Var}\left[\mathrm{D}_{n\ell}\right]}.\tag{4.18}$$

Recalling that $\sigma_{n\ell}^2=\mathrm{E}\left[\mathrm{D}_{n\ell}^2\,\middle|\,\mathscr{F}_{n,\ell-1}\right]$, (4.17) provides

$$\mathrm{Var}\left[\mathrm{D}_{n\ell}\right]\le\mathrm{E}\left[\mathrm{D}_{n\ell}^2\right]=\mathrm{E}\left[\sigma_{n\ell}^2\right]\le\frac{2p_n}{n\sigma_n^2}\left(\tilde{e}_{n2}^2+ne_{n1}^2\tilde{e}_{n2}+\frac{f_{n2}^2}{p_n-1}\right)\le\frac{\mathrm{C}}{n}.$$

Using (4.15) and the inequalities $(a+b)^4\le8(a^4+b^4)$ and $\sigma_n^2\ge2np_ne_{n1}^2\tilde{e}_{n2}$ then yields

$$\begin{aligned}\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right] &\le \frac{\mathrm{C}p_n^2}{n^4\sigma_n^4}\left(\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)'(\mathbf{X}_{n\ell}-e_{n1}\boldsymbol{\theta}_n)\right)^4\right]+n^4e_{n1}^4\,\mathrm{E}\left[\left(\mathbf{X}_{n\ell}'\boldsymbol{\theta}_n-e_{n1}\right)^4\right]\right)\\ &\le \frac{\mathrm{C}p_n^2}{n^4\sigma_n^4}\,\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)'(\mathbf{X}_{n\ell}-e_{n1}\boldsymbol{\theta}_n)\right)^4\right]+\frac{\mathrm{C}\tilde{e}_{n4}}{n^2\tilde{e}_{n2}^2}.\end{aligned}\tag{4.19}$$

Applying Lemma 4.3.4, we have

$$\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)'(\mathbf{X}_{n\ell}-e_{n1}\boldsymbol{\theta}_n)\right)^4\right]=(\ell-1)\left(\tilde{e}_{n4}^2+\frac{6}{p_n-1}\mathrm{E}\left[v_{ni}^2(u_{ni}-e_{n1})^2\right]^2+\frac{3f_{n4}^2}{p_n^2-1}\right)$$
$$+3(\ell-1)(\ell-2)\left(\tilde{e}_{n2}^2\tilde{e}_{n4}+\frac{2\tilde{e}_{n2}f_{n2}}{p_n-1}\mathrm{E}\left[v_{ni}^2(u_{ni}-e_{n1})^2\right]+\frac{f_{n2}^2f_{n4}}{(p_n-1)^2}\right),$$

By Cauchy–Schwarz, this yields

$$\frac{p_n^2}{n^4\sigma_n^4}\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}(\mathbf{X}_{ni}-e_{n1}\boldsymbol{\theta}_n)'(\mathbf{X}_{n\ell}-e_{n1}\boldsymbol{\theta}_n)\right)^4\right]$$
$$\leq\frac{1}{n^3\sigma_n^4}\left(p_n^2\tilde{e}_{n4}^2+6p_nf_{n4}\tilde{e}_{n4}+3f_{n4}^2\right)+\frac{3}{n^2\sigma_n^4}\left(p_n^2\tilde{e}_{n2}^2\tilde{e}_{n4}+2p_n\tilde{e}_{n2}f_{n2}f_{n4}^{1/2}\tilde{e}_{n4}^{1/2}+f_{n2}^2f_{n4}\right)$$
$$\leq\frac{\mathrm{C}}{n^3}\left(\frac{\tilde{e}_{n4}^2}{\tilde{e}_{n2}^4}+\frac{f_{n4}\tilde{e}_{n4}}{f_{n2}^2\tilde{e}_{n2}^2}+\frac{f_{n4}^2}{f_{n2}^4}\right)+\frac{\mathrm{C}}{n^2}\left(\frac{\tilde{e}_{n4}}{\tilde{e}_{n2}^2}+\left(\frac{f_{n4}\tilde{e}_{n4}}{f_{n2}^2\tilde{e}_{n2}^2}\right)^{1/2}+\frac{f_{n4}}{f_{n2}^2}\right).$$

Plugging into (4.19), we conclude that

$$\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]\leq\frac{\mathrm{C}}{n^3}\left(\frac{\tilde{e}_{n4}^2}{\tilde{e}_{n2}^4}+\frac{f_{n4}\tilde{e}_{n4}}{f_{n2}^2\tilde{e}_{n2}^2}+\frac{f_{n4}^2}{f_{n2}^4}\right)+\frac{\mathrm{C}}{n^2}\left(\frac{\tilde{e}_{n4}}{\tilde{e}_{n2}^2}+\left(\frac{f_{n4}\tilde{e}_{n4}}{f_{n2}^2\tilde{e}_{n2}^2}\right)^{1/2}+\frac{f_{n4}}{f_{n2}^2}\right)\leq\frac{\mathrm{C}}{n}\left(\frac{\tilde{e}_{n4}}{n\tilde{e}_{n2}^2}+\frac{f_{n4}}{nf_{n2}^2}\right),$$

which, by assumption, is $o(1/n)$ as $n\to\infty$.
All majorations and o's above being uniform in ℓ, we finally obtain that

$$\sum_{\ell=1}^{n}\sqrt{\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]\mathrm{Var}\left[\mathrm{D}_{n\ell}\right]}\leq\mathrm{C}\left(n\max_{\ell=1,\ldots,n}\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]\right)^{1/2}\to0$$

as $n\to\infty$, which, in view of (4.18), establishes the result. □

4.3.4 Proof of Theorem 4.1.3

From Theorem 4.1.2, we have that, as $n\to\infty$,

$$\left|\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}\left[\mathrm{R}_n^{\mathrm{St}}>z_\alpha\right]-\left(1-\Phi\left(z_\alpha-\frac{\tau^2}{\sqrt{2}}\right)\right)\right|=\left|\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}\left[\mathrm{R}_n^{\mathrm{St}}\leq z_\alpha\right]-\Phi\left(z_\alpha-\frac{\tau^2}{\sqrt{2}}\right)\right|$$
$$\leq\sup_{z\in\mathbb{R}}\left|\mathrm{P}_{\boldsymbol{\theta}_n,\mathrm{F}_n}^{(n)}\left[\frac{\mathrm{R}_n^{\mathrm{St}}-\mathrm{E}\left[\mathrm{R}_n^{\mathrm{St}}\right]}{\sigma_n}\leq z\right]-\Phi(z)\right|+\left|\Phi\left(\frac{z_\alpha-\mathrm{E}\left[\mathrm{R}_n^{\mathrm{St}}\right]}{\sigma_n}\right)-\Phi\left(z_\alpha-\frac{\tau^2}{\sqrt{2}}\right)\right|\to0,$$

where we used Lemma 2.11 from [van der Vaart, 1998]. □

4.3.5 Proof of Proposition 4.2.1

Using Lemma 4.3.2(ii), we readily obtain

$$\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right]=\frac{p_n}{n}\sum_{\substack{i<j\\i,j=1}}^{n}\left(\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\right]-\frac{1}{p_n}\right)=\frac{p_n}{n}\times\frac{n(n-1)}{2}\times\frac{p_n}{p_n-1}g_{n2}^2=\frac{p_n^2(n-1)}{2(p_n-1)}g_{n2}^2.$$

Turning to the variance, we have

$$\begin{aligned}\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] &= \frac{p_n^2}{n^2}\sum_{1\le i<j\le n}\sum_{1\le k<\ell\le n}\left\{\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)^2\right]-\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\right]^2\right\}\\ &= \frac{p_n^2}{n^2}\sum_{1\le i<j\le n}\sum_{1\le k<\ell\le n}\left\{\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\left(\mathbf{X}_{nk}'\mathbf{X}_{n\ell}\right)^2\right]-\left(\frac{p_n}{p_n-1}g_{n2}^2+\frac{1}{p_n}\right)^2\right\}.\end{aligned}$$

In this last sum, $\binom{n}{2}$ terms correspond to Lemma 4.3.2(iii) and $6\binom{n}{4}$terms (all equal to zero) correspond to Lemma 4.3.2(vii). Thus, $\binom{n}{2}^2-\binom{n}{2}-6\binom{n}{4}=n(n-1)(n-2)$ terms correspond to Lemma 4.3.2(vi), which leads to

$$\begin{aligned}\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] &= \frac{p_n^2}{n^2}\left[\frac{n(n-1)}{2}\left\{e_{n4}^2+\frac{6(e_{n2}-e_{n4})^2}{p_n-1}+\frac{3f_{n4}^2}{p_n^2-1}-\left(\frac{p_n}{p_n-1}g_{n2}^2+\frac{1}{p_n}\right)^2\right\}\right.\\ &\qquad\left.+n(n-1)(n-2)\left\{e_{n4}e_{n2}^2+\frac{2e_{n2}(e_{n2}-e_{n4})f_{n2}}{p_n-1}+\frac{f_{n2}^2f_{n4}}{(p_n-1)^2}-\left(\frac{p_n}{p_n-1}g_{n2}^2+\frac{1}{p_n}\right)^2\right\}\right]\\ &= \frac{(n-1)p_n^2}{2n}\left\{e_{n4}^2+\frac{6}{p_n-1}(e_{n2}-e_{n4})^2+\frac{3f_{n4}^2}{p_n^2-1}-\left(\frac{p_n}{p_n-1}g_{n2}^2+\frac{1}{p_n}\right)^2\right\}\\ &\qquad+\frac{(n-1)(n-2)p_n^4}{n(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2,\end{aligned}$$

which establishes the result. □

4.3.6 Proof of Theorem 4.2.2

First note that under the assumptions of this theorem, we have $e_{n2}=o(1/\sqrt{p_n})$, which, jointly with Assumptions (a) and (c), entails that $\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right]=1+o(1)$. Therefore, it is sufficient to prove that

$$\mathrm{Q}_n^{\mathrm{St}}-\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right]\xrightarrow{\mathcal{D}}\mathcal{N}(0,1).$$

To do so, let $\mathscr{F}_{n\ell}$, for $\ell=1,\ldots,n$, be the σ-algebra generated by $\mathbf{X}_{n1},\ldots,\mathbf{X}_{n\ell}$. Denoting as $\mathrm{E}_{n\ell}$ the conditional expectation with respect to $\mathscr{F}_{n\ell}$, define

$$\mathrm{Z}_{n\ell}:=\mathrm{E}_{n\ell}\left[\mathrm{Q}_n^{\mathrm{St}}\right]=\frac{p_n}{n}\sum_{\substack{i,j=1\\ i<j}}^{n}\mathrm{E}_{n\ell}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2-\frac{1}{p_n}\right].$$

Note that $\mathrm{Q}_n^{\mathrm{St}}-\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right]=\sum_{\ell=1}^n\mathrm{D}_{n\ell}$, where $\mathrm{D}_{n\ell}:=\mathrm{Z}_{n\ell}-\mathrm{Z}_{n,\ell-1}$ rewrites

$$\begin{aligned}\mathrm{D}_{n\ell} &= \frac{p_n}{n}\sum_{\substack{i,j=1\\ i<j}}^{n}\left\{\mathrm{E}_{n\ell}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\right]-\mathrm{E}_{n,\ell-1}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{nj}\right)^2\right]\right\}\\ &= \frac{p_n}{n}\sum_{i=1}^{\ell-1}\left\{\left(\mathbf{X}_{ni}'\mathbf{X}_{n\ell}\right)^2-\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{n\ell}\right)^2\,|\mathbf{X}_i\right]\right\}\\ &\quad+\frac{p_n}{n}\sum_{j=\ell+1}^{n}\sum_{\ell+1\le j\le n}\left\{\mathrm{E}\left[\left(\mathbf{X}_{n\ell}'\mathbf{X}_{nj}\right)^2\,|\mathbf{X}_\ell\right]-\mathrm{E}\left[\left(\mathbf{X}_{n\ell}'\mathbf{X}_{nj}\right)^2\right]\right\}\\ &= \frac{p_n}{n}\sum_{i=1}^{\ell-1}\left\{\left(\mathbf{X}_{ni}'\mathbf{X}_{n\ell}\right)^2-\left(\frac{p_n}{p_n-1}g_{n2}u_{ni}^2+\frac{f_{n2}}{p_n-1}\right)\right\}+\frac{(n-\ell)p_n^2}{n(p_n-1)}g_{n2}\left(u_{n\ell}^2-e_{n2}\right),\end{aligned}\qquad(4.20)$$

where the last equality follows from Lemma 4.3.2(ii) and Lemma 4.3.6(i).

We establish Theorem 4.2.2 by proving Lemmas 4.3.9 and 4.3.10 (so that Theorem 1.4.1 applies).

Lemma 4.3.9. *Under the assumptions of Theorem 4.2.2, as $n \to \infty$,*

(i) $\sum_{\ell=1}^{n} \mathrm{E}\left[\sigma_{n\ell}^2\right] = 1$,

(ii) $\mathrm{Var}\left[\sum_{\ell=1}^{n} \sigma_{n\ell}^2\right] = o(1)$,

so that $\sum_{\ell=1}^{n} \sigma_{n\ell}^2 = 1 + o_{\mathrm{P}}(1)$.

Proof.

(i) Using (4.20), $\sigma_{n\ell}^2 = \mathrm{E}_{n,\ell-1}\left[\mathrm{D}_{n\ell}^2\right]$ takes the form

$$\begin{aligned}\sigma_{n\ell}^2 = \frac{p_n^2}{n^2} \sum_{i,j=1}^{\ell-1} \Bigg\{ & \mathrm{E}_{n,\ell-1}\left[(\mathbf{X}_{ni}'\mathbf{X}_{n\ell})^2 (\mathbf{X}_{nj}'\mathbf{X}_{n\ell})^2 \right] \\ & - \left(\frac{p_n}{p_n-1} g_{n2} u_{ni}^2 + \frac{f_{n2}}{p_n-1}\right)\left(\frac{p_n}{p_n-1} g_{n2} u_{nj}^2 + \frac{f_{n2}}{p_n-1}\right)\Bigg\} \\ & + \frac{(n-\ell)^2 p_n^4}{n^2(p_n-1)^2} g_{n2}^2 \mathrm{Var}\left[u_{n\ell}^2\right] + \frac{2(n-\ell)p_n^3}{n^2(p_n-1)} g_{n2} \sum_{i=1}^{\ell-1} \mathrm{E}_{n,\ell-1}\left[(\mathbf{X}_{ni}'\mathbf{X}_{n\ell})^2 (u_{n\ell}^2 - e_{n2}) \right] \\ =: & \frac{p_n^2}{n^2} \sum_{i,j=1}^{\ell-1} \mathrm{A}_{n,ij} + \frac{(n-\ell)^2 p_n^4}{n^2(p_n-1)^2} g_{n2}^2 \mathrm{Var}\left[u_{n\ell}^2\right] + \frac{2(n-\ell)p_n^3}{n^2(p_n-1)} g_{n2} \sum_{i=1}^{\ell-1} \mathrm{B}_{n,i}.\end{aligned}$$

Using the identity $u_{ni}^2 + u_{nj}^2 = 1 + u_{ni}^2 u_{nj}^2 - v_{ni}^2 v_{nj}^2$, Lemma 4.3.6(ii) implies that

$$\begin{aligned}\mathrm{A}_{n,ij} = & \frac{e_{n2}-e_{n4}}{p_n-1} + \frac{p_n e_{n4} - e_{n2}}{p_n-1} u_{ni}^2 u_{nj}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{e_{n2}-e_{n4}}{p_n-1}\right) v_{ni}^2 v_{nj}^2 \\ & + \frac{2 f_{n4}}{p_n^2-1} v_{ni}^2 v_{nj}^2 (\mathbf{S}_{ni}'\mathbf{S}_{nj})^2 + \frac{4(e_{n2}-e_{n4})}{p_n-1} u_{ni} v_{ni} u_{nj} v_{nj} (\mathbf{S}_{ni}'\mathbf{S}_{nj}) \\ & - \frac{p_n^2}{(p_n-1)^2} g_{n2}^2 u_{ni}^2 u_{nj}^2 - \left(\frac{f_{n2}}{p_n-1}\right)^2 - \frac{p_n f_{n2} g_{n2}}{(p_n-1)^2}\left(1 + u_{ni}^2 u_{nj}^2 - v_{ni}^2 v_{nj}^2\right),\end{aligned}$$

which, after some algebra, rewrites

$$\begin{aligned}\mathrm{A}_{n,ij} = & \frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1} u_{ni}^2 u_{nj}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right) v_{ni}^2 v_{nj}^2 \\ & + \frac{2 f_{n4}}{p_n^2-1} v_{ni}^2 v_{nj}^2 (\mathbf{S}_{ni}'\mathbf{S}_{nj})^2 + \frac{4(e_{n2}-e_{n4})}{p_n-1} u_{ni} v_{ni} u_{nj} v_{nj} (\mathbf{S}_{ni}'\mathbf{S}_{nj}) - \frac{e_{n4}-e_{n2}^2}{p_n-1}.\end{aligned}$$

Similarly, after some algebra, Lemma 4.3.6(i) and (iii) yield

$$\begin{aligned}\mathrm{B}_{n,i} & = \frac{e_{n2}-e_{n4}}{p_n-1} + \frac{p_n e_{n4} - e_{n2}}{p_n-1} u_{ni}^2 - e_{n2}\left(\frac{p_n}{p_n-1} g_{n2} u_{ni}^2 + \frac{f_{n2}}{p_n-1}\right) \\ & = \frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}\left(u_{ni}^2 - \frac{1}{p_n}\right).\end{aligned}$$

Therefore, we conclude that

$$\sigma^2_{n\ell} = \frac{p_n^2}{n^2}\sum_{i,j=1}^{\ell-1}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{ni}^2u_{nj}^2+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{ni}^2v_{nj}^2$$
$$+\frac{2f_{n4}}{p_n^2-1}v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2+\frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\Bigg\}$$
$$-\frac{(\ell-1)^2p_n^2}{n^2(p_n-1)}(e_{n4}-e_{n2}^2)+\frac{(n-\ell)^2p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2$$
$$+\frac{2(n-\ell)p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}\sum_{i=1}^{\ell-1}\left(u_{ni}^2-\frac{1}{p_n}\right). \tag{4.21}$$

Splitting the double sum over i, j according to whether $i = j$ or $i \neq j$, taking expectation yields

$$\mathrm{E}\left[\sigma^2_{n\ell}\right] = (\ell-1)(\ell-2)\frac{p_n^2}{n^2}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}e_{n2}^2+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)f_{n2}^2$$
$$+\frac{2f_{n4}f_{n2}^2}{(p_n-1)^2(p_n+1)}\Bigg\}+(\ell-1)\frac{p_n^2}{n^2}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}e_{n4}+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)f_{n4}$$
$$+\frac{2f_{n4}^2}{p_n^2-1}+\frac{4(e_{n2}-e_{n4})^2}{p_n-1}\Bigg\}-\frac{(\ell-1)^2p_n^2}{n^2(p_n-1)}(e_{n4}-e_{n2}^2)+\frac{(n-\ell)^2p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2$$
$$+\frac{2(\ell-1)(n-\ell)p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2,$$

which eventually provides

$$\sum_{\ell=1}^{n}\mathrm{E}\left[\sigma^2_{n\ell}\right] = \frac{(n-1)(n-2)p_n^2}{3n}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}e_{n2}^2+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)f_{n2}^2$$
$$+\frac{2f_{n4}f_{n2}^2}{(p_n-1)^2(p_n+1)}\Bigg\}+\frac{(n-1)p_n^2}{2n}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}e_{n4}+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)f_{n4}$$
$$+\frac{2f_{n4}^2}{p_n^2-1}+\frac{4(e_{n2}-e_{n4})^2}{p_n-1}\Bigg\}-\frac{(n-1)(2n-1)p_n^2}{6n(p_n-1)}(e_{n4}-e_{n2}^2)+\frac{(n-1)(2n-1)p_n^4}{6n(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2$$
$$+\frac{(n-1)(n-2)p_n^4}{3n(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}^2=1,$$

where the last equality is obtained after painful, yet straightforward, algebra.

(ii) Note that (4.21) implies that

$$\sigma^2_{n\ell} = \frac{p_n^2}{n^2}\sum_{i,j=1}^{\ell-1}\Bigg\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{ni}^2u_{nj}^2+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{ni}^2v_{nj}^2$$
$$+\frac{2f_{n4}}{p_n^2-1}v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2+\frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\Bigg\}$$
$$+\frac{2(n-\ell)p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}\sum_{i=1}^{\ell-1}u_{ni}^2+\mathrm{C}_{n\ell}$$

for some real constant $C_{n\ell}$. Therefore,

$$\mathrm{Var}\left[\sum_{\ell=1}^{n} \sigma_{n\ell}^2\right] \le 2\left(\mathrm{Var}[T_{1n}] + \mathrm{Var}[T_{2n}]\right),$$

where we let

$$T_{1n} := \frac{p_n^2}{n^2}\sum_{\ell=1}^{n}\sum_{i,j=1}^{\ell-1}\left\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{ni}^2u_{nj}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{ni}^2v_{nj}^2\right.$$
$$\left. + \frac{2f_{n4}}{p_n^2-1}v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 + \frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\right\}$$

and

$$T_{2n} := \frac{2p_n^4}{n^2(p_n-1)^2}(e_{n4}-e_{n2}^2)g_{n2}\sum_{\ell=1}^{n}(n-\ell)\sum_{i=1}^{\ell-1}u_{ni}^2.$$

In the rest of the proof, C is a positive constant that may change from line to line. Let us start with the variance of T_{2n}. Since

$$\sum_{\ell=1}^{n}(n-\ell)\sum_{i=1}^{\ell-1}u_{ni}^2 = \sum_{i=1}^{n-1}c_{ni}u_{ni}^2,$$

for positive constants c_{ni} that are upper-bounded by n^2, we have

$$\mathrm{Var}\,[T_{2n}] \le C\frac{p_n^8}{n^4(p_n-1)^4}(e_{n4}-e_{n2}^2)^2g_{n2}^2(n-1)n^4(e_{n4}-e_{n2}^2) \le Cnp_n^4(e_{n4}-e_{n2}^2)^3g_{n2}^2 = o(1)$$

by Assumptions (a) and (c). We turn to T_{1n}, that can be split into $T_{1n}^{(a)} + T_{1n}^{(b)}$, where

$$T_{1n}^{(a)} = \frac{p_n^2}{n^2}\sum_{i=1}^{n-1}(n-i)\left\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{ni}^4\right.$$
$$\left. + \left(\frac{3f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{ni}^4 + \frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}^2v_{ni}^2\right\}$$

and

$$T_{1n}^{(b)} = \frac{2p_n^2}{n^2}\sum_{\substack{i<j\\ i,j=1}}^{n-1}(n-j)\left\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{ni}^2u_{nj}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{ni}^2v_{nj}^2\right.$$
$$\left. + \frac{2f_{n4}}{p_n^2-1}v_{ni}^2v_{nj}^2\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)^2 + \frac{4(e_{n2}-e_{n4})}{p_n-1}u_{ni}v_{ni}u_{nj}v_{nj}\left(\mathbf{S}_{ni}'\mathbf{S}_{nj}\right)\right\}$$
$$=: \frac{p_n^2}{n^2}\sum_{\substack{i<j\\ i,j=1}}^{n-1}(n-j)\,w_{n,ij}.$$

Firstly, since $\mathrm{Var}[u_{n1}^r v_{n1}^s] \leq \mathrm{E}[u_{n1}^{2r} v_{n1}^{2s}] \leq e_{n,2r}$ for any $r, s \geq 0$ (here, we let $e_{n0} = 1$),

$$\begin{aligned}
\mathrm{Var}\left[\mathrm{T}_{1n}^{(a)}\right] &\leq \frac{p_n^4}{n}\mathrm{Var}\Bigg[\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}u_{n1}^4+\left(\frac{3f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)v_{n1}^4 \\
&\quad +\frac{4(e_{n2}-e_{n4})}{p_n-1}u_{n1}^2v_{n1}^2\Bigg] \\
&\leq \frac{\mathrm{C}p_n^4}{n}\left\{\frac{p_n^2(e_{n4}-e_{n2}^2)^2}{(p_n-1)^2}e_{n8}+\left(\frac{3f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)^2+\frac{16(e_{n2}-e_{n4})^2}{(p_n-1)^2}e_{n4}\right\} \\
&\leq \frac{\mathrm{C}}{n}\left(p_n^4(e_{n4}-e_{n2}^2)^2e_{n8}+f_{n4}^2+f_{n2}^4+p_n^2(e_{n4}-e_{n2}^2)^2+p_n^2(e_{n2}-e_{n4})^2e_{n4}\right)=o\left(\frac{1}{n^{1/3}}\right)
\end{aligned}$$

under Assumptions (a)–(b). Secondly,

$$\mathrm{Var}\left[\mathrm{T}_{1n}^{(b)}\right] = \frac{\mathrm{C}p_n^4}{n^4}\sum_{\substack{i<j\\ i,j=1}}^{n}\sum_{\substack{k<\ell\\ k,\ell=1}}^{n}(n-j)(n-\ell)\mathrm{Cov}\left[w_{n,ij},w_{n,k\ell}\right].$$

Using the same argument as in the proof of Proposition 4.2.1, this sum over i,j,k,ℓ contains $\binom{n}{2}$ terms corresponding to $\mathrm{Var}\left[w_{n,ij}\right]$ and $n(n-1)(n-2)$ terms corresponding to $\mathrm{Cov}\left[w_{n,ij},w_{n,k\ell}\right]$ with $i\neq j$, $k\neq\ell$ and $\#\{i,j,k,\ell\}=3$ (terms for which the four indices are pairwise different are equal to zero since $w_{n,ij}$ and $w_{n,k\ell}$ are then mutually independent). Therefore,

$$\mathrm{Var}\left[\mathrm{T}_{1n}^{(b)}\right] \leq \mathrm{C}p_n^4\left(\mathrm{Var}\left[w_{n,12}\right]+n\mathrm{Cov}\left[w_{n,12},w_{n,13}\right]\right). \tag{4.22}$$

Since $\mathrm{Var}[\mathrm{Z}] \leq \mathrm{E}\left[\mathrm{Z}^2\right]$ for any random variable Z,

$$\begin{aligned}
&\mathrm{Var}\left[w_{n,12}\right] \\
&\leq \mathrm{C}\Bigg\{\frac{p_n^2(e_{n4}-e_{n2}^2)^2}{(p_n-1)^2}\mathrm{Var}\left[u_{n1}^2u_{n2}^2\right]+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)^2\mathrm{Var}\left[v_{n1}^2v_{n2}^2\right] \\
&\qquad +\frac{4f_{n4}^2}{(p_n^2-1)^2}\mathrm{Var}\left[v_{n1}^2v_{n2}^2(\mathbf{S}_{n1}'\mathbf{S}_{n2})^2\right]+\frac{16(e_{n2}-e_{n4})^2}{(p_n-1)^2}\mathrm{Var}\left[u_{n1}v_{n1}u_{n2}v_{n2}(\mathbf{S}_{n1}'\mathbf{S}_{n2})\right]\Bigg\} \\
&\leq \mathrm{C}\Bigg\{\frac{p_n^2(e_{n4}-e_{n2}^2)^2}{(p_n-1)^2}e_{n4}^2+\left(\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}+\frac{e_{n4}-e_{n2}^2}{p_n-1}\right)^2f_{n4}^2+\frac{12f_{n4}^4}{(p_n^2-1)^3} \\
&\qquad +\frac{16(e_{n2}-e_{n4})^4}{(p_n-1)^3}\Bigg\},
\end{aligned}$$

which is $o(1/p_n^4)$, since

$$\frac{f_{n4}}{p_n^2-1}-\frac{f_{n2}^2}{(p_n-1)^2}=\frac{4e_{n2}-2+(p-1)e_{n4}-(p+1)e_{n2}^2}{(p^2-1)(p-1)}=o\left(\frac{1}{p_n^2}\right)$$

under Assumption (a). Now,

$$\mathrm{Cov}\left[w_{n,12}, w_{n,13}\right] = \frac{p_n^2(e_{n4}-e_{n2}^2)^2}{(p_n-1)^2} e_{n4}e_{n2}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)^2 f_{n4}f_{n2}^2$$

$$+\frac{4f_{n4}^3f_{n2}^2}{(p_n^2-1)^2(p_n-1)^2} + \frac{2p_n(e_{n4}-e_{n2}^2)}{p_n-1}\left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)(e_{n2}-e_{n4})e_{n2}f_{n2}$$

$$+\frac{4p_n(e_{n4}-e_{n2}^2)f_{n4}}{(p_n-1)^2(p_n^2-1)}(e_{n2}-e_{n4})e_{n2}f_{n2} + \frac{4f_{n4}^2f_{n2}^2}{(p_n^2-1)(p_n-1)}\left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)$$

$$-\left\{\frac{p_n(e_{n4}-e_{n2}^2)}{p_n-1}e_{n2}^2 + \left(\frac{f_{n4}}{p_n^2-1} - \frac{f_{n2}^2}{(p_n-1)^2} + \frac{e_{n4}-e_{n2}^2}{p_n-1}\right)f_{n2}^2 + \frac{2f_{n4}f_{n2}^2}{(p_n^2-1)(p_n-1)}\right\}^2.$$

Tedious computations provide

$$\mathrm{Cov}\left[w_{n,12}, w_{n,13}\right] = \frac{p_n^4}{(p_n-1)^4}(e_{n4}-e_{n2}^2)^3 g_{n2}^2,$$

which is $o(n^{-1}p_n^{-4})$ under Assumptions (a) and (c). From (4.22), we conclude that $\mathrm{Var}\left[\mathrm{T}_{1n}^{(b)}\right]$ is $o(1)$, which establishes the result.

□

Lemma 4.3.10. *Under the assumptions of Theorem 4.2.2,* $\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^2 \mathbb{I}_{\{|\mathrm{D}_{n\ell}|>\varepsilon\}}\right] \to 0$ *for any* $\varepsilon > 0$.

Proof. Since $\mathrm{Var}\left[\mathrm{D}_{n\ell}\right] \leq \mathrm{E}\left[\mathrm{D}_{n\ell}^2\right] = \mathrm{E}\left[\sigma_{n\ell}^2\right]$, Lemma 4.3.9(i) shows that

$$\sum_{\ell=1}^n \mathrm{Var}\left[\mathrm{D}_{n\ell}\right] \leq \sum_{\ell=1}^n \mathrm{E}\left[\sigma_{n\ell}^2\right] = 1.$$

Therefore, applying Cauchy–Schwarz inequality, Chebychev inequality, then Cauchy–Schwarz inequality again, yields

$$\begin{aligned}\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^2 \mathbb{I}_{\{|\mathrm{D}_{n\ell}|>\varepsilon\}}\right] &\leq \sum_{\ell=1}^n \sqrt{\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]}\sqrt{\mathrm{P}\left[|\mathrm{D}_{n\ell}|>\varepsilon\right]} \leq \frac{1}{\varepsilon}\sum_{\ell=1}^n \sqrt{\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]}\sqrt{\mathrm{Var}\left[\mathrm{D}_{n\ell}\right]} \\ &\leq \frac{1}{\varepsilon}\sqrt{\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]}\sqrt{\sum_{\ell=1}^n \mathrm{Var}\left[\mathrm{D}_{n\ell}\right]} \leq \frac{1}{\varepsilon}\sqrt{\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]}.\end{aligned} \tag{4.23}$$

Letting

$$\mathrm{Y}_{ni\ell} := (\mathbf{X}_{ni}'\mathbf{X}_{n\ell})^2 - \frac{p_n}{p_n-1}g_{n2}u_{ni}^2 - \frac{f_{n2}}{p_n-1},$$

(4.20) provides

$$\mathrm{E}\left[\mathrm{D}_{n\ell}^4\right] \leq \frac{\mathrm{C}p_n^4}{n^4}\mathrm{E}\left[\left(\sum_{i=1}^{\ell-1}\mathrm{Y}_{ni\ell}\right)^4\right] + \frac{\mathrm{C}(n-\ell)^4p_n^8}{n^4(p_n-1)^4}g_{n2}^4\mathrm{E}\left[\left(u_{n\ell}^2-e_{n2}\right)^4\right],$$

hence

$$\begin{aligned}\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^4\right] &\leq \frac{\mathrm{C}p_n^4}{n^4}\sum_{\ell=1}^n\sum_{i,j,r,s=1}^{\ell-1}\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right] + \mathrm{C}np_n^4g_{n2}^4e_{n8} \\ &= \frac{\mathrm{C}p_n^4}{n^4}\sum_{\ell=1}^n\sum_{i,j,r,s=1}^{\ell-1}\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right] + o\left(\frac{1}{n^{1/3}}\right).\end{aligned} \tag{4.24}$$

Now, in the sum over i, j, r, s, there are $\ell - 1 \le n$ terms for which $\{i, j, r, s\}$ has cardinality one; for these terms, (4.14) yields

$$\begin{aligned}
\mathrm{E}\left[\left|\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right|\right] &= \mathrm{E}\left[\mathrm{Y}_{n1\ell}^4\right] \le \mathrm{CE}\left[(\mathbf{X}_{n1}'\mathbf{X}_{n\ell})^8 + \frac{p_n^4}{(p_n-1)^4} g_{n2}^4 u_{n1}^8 + \frac{f_{n2}^4}{(p_n-1)^4}\right] \\
&\le \mathrm{C}\left\{\mathrm{E}\left[u_{n1}^8 u_{n\ell}^8 + v_{n1}^8 v_{n\ell}^8 (\mathbf{S}_{n1}'\mathbf{S}_{n\ell})^8 + u_{n1}^4 v_{n1}^4 u_{n\ell}^4 v_{n\ell}^4 (\mathbf{S}_{n1}'\mathbf{S}_{n\ell})^4\right] + g_{n2}^4 e_{n8} + \frac{f_{n2}^4}{p_n^4}\right\} \\
&\le \mathrm{C}\left\{e_{n8}^2 + \frac{105}{(p_n^2-1)(p_n+3)(p_n+5)} + \frac{3e_{n4}^2}{p_n^2-1} + g_{n2}^4 e_{n8} + \frac{f_{n2}^4}{p_n^4}\right\} = o\left(\frac{n^{4/3}}{p_n^4}\right), \qquad (4.25)
\end{aligned}$$

where we used Lemma 4.3.1(iv), the identities $v_{n1}^4, v_{n\ell}^4 \le 1$, and Assumptions (a)–(c). In the sum over i, j, r, s, there are $3(\ell-1)(\ell-2) \le 3n^2$ for which $\{i, j, r, s\}$ has cardinality two and contains two pairs of equal indices. For such terms, Lemma 4.3.2(iii) yields

$$\begin{aligned}
\mathrm{E}\left[|\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}|\right] = \mathrm{E}\left[\mathrm{Y}_{n1\ell}^2\right]^2 &\le \mathrm{C}\left\{\mathrm{E}\left[\left(\mathbf{X}_{ni}'\mathbf{X}_{n\ell}\right)^4\right] + \frac{p_n^2}{(p_n-1)^2} g_{n2}^2 e_{n4} + \frac{f_{n2}^2}{(p_n-1)^2}\right\}^2 \\
&\le \mathrm{C}\left\{e_{n4}^2 + \frac{2(e_{n2}-e_{n4})^2}{p_n-1} + \frac{3f_{n4}^2}{p_n^2-1} + g_{n2}^2 e_{n4} + \frac{f_{n2}^2}{p_n^2}\right\}^2 = \mathrm{O}\left(\frac{1}{p_n^4}\right).
\end{aligned}$$

Similarly, the sum over i, j, r, s in (4.24) contains no more than $\mathrm{C}_1 n^2$ terms (where C_1 does not depend on ℓ) such that $\{i, j, r, s\}$ has cardinality two and contains a triple of equal indices. For such terms, Lemma 4.3.6(i) yields

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right] &= \mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}^3\mathrm{Y}_{n2\ell}\middle|\mathbf{X}_{n\ell}\right]\right] = \mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}^3\middle|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n2\ell}\middle|\mathbf{X}_{n\ell}\right]\right] \\
&= \frac{p_n}{p_n-1} g_{n2}\mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}^3\middle|\mathbf{X}_{n\ell}\right]\left(u_{n\ell}^2 - e_{n2}\right)\right] \\
&= \frac{p_n}{p_n-1} g_{n2}\mathrm{E}\left[\mathrm{Y}_{n1\ell}^3\left(u_{n\ell}^2 - e_{n2}\right)\right],
\end{aligned}$$

so that the Hölder inequality provides

$$\begin{aligned}
\left|\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right]\right| &\le \mathrm{C}g_{n2}\mathrm{E}\left[\mathrm{Y}_{n1\ell}^3\left(u_{n\ell}^2 - e_{n2}\right)\right] \le \mathrm{C}g_{n2}\mathrm{E}\left[\mathrm{Y}_{nr\ell}^4\right]^{3/4}\mathrm{E}\left[\left(u_{n\ell}^2 - e_{n2}\right)^4\right]^{1/4} \\
&\le \mathrm{C}g_{n2}\mathrm{E}\left[\mathrm{Y}_{nr\ell}^4\right]^{3/4} e_{n8}^{1/4} = \mathrm{O}\left(\frac{1}{\sqrt{np_n}}\right) o\left(\frac{n}{p_n^3}\right) o\left(\frac{n^{1/6}}{\sqrt{p_n}}\right) \\
&= o\left(\frac{n^{2/3}}{p_n^4}\right),
\end{aligned}$$

where we used (4.25). The sum over i, j, r, s in (4.24) contains no more than $\mathrm{C}_2 n^3$ (where C_2 does not depend on ℓ) terms such that $\{i, j, r, s\}$ has cardinality three. Proceeding as above, the corresponding terms are seen to satisfy

$$\begin{aligned}
\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right] &= \mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}^2\middle|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n2\ell}\middle|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n3\ell}\middle|\mathbf{X}_{n\ell}\right]\right] \\
&= \frac{p_n^2}{(p_n-1)^2} g_{n2}^2 \mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}^2\middle|\mathbf{X}_{n\ell}\right]\left(u_{n\ell}^2 - e_{n2}\right)^2\right] \\
&= \frac{p_n}{p_n-1} g_{n2}^2 \mathrm{E}\left[\mathrm{Y}_{n1\ell}^2\left(u_{n\ell}^2 - e_{n2}\right)^2\right],
\end{aligned}$$

which, by using the Cauchy–Schwarz inequality, yields

$$\left|\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right]\right| \leq \mathrm{C}g_{n2}^2\mathrm{E}\left[\mathrm{Y}_{n1\ell}^2\left(u_{n\ell}^2-e_{n2}\right)^2\right] \leq \mathrm{C}g_{n2}^2\sqrt{\mathrm{E}\left[\mathrm{Y}_{n1\ell}^4\right]}\sqrt{\mathrm{E}\left[\left(u_{n\ell}^2-e_{n2}\right)^4\right]}$$
$$\leq \mathrm{C}g_{n2}^2\sqrt{\mathrm{E}\left[\mathrm{Y}_{nr\ell}^4\right]}\sqrt{e_{n8}} = \mathrm{O}\left(\frac{1}{np_n}\right)o\left(\frac{n^{2/3}}{p_n^2}\right)o\left(\frac{n^{1/3}}{p_n}\right) = o\left(\frac{1}{p_n^4}\right).$$

Finally, there obviously are less than $(\ell-1)^4 \leq n^4$ terms such that $\{i,j,r,s\}$ has cardinality four, and these terms are such that

$$\left|\mathrm{E}\left[\mathrm{Y}_{ni\ell}\mathrm{Y}_{nj\ell}\mathrm{Y}_{nr\ell}\mathrm{Y}_{ns\ell}\right]\right| = \left|\mathrm{E}\left[\mathrm{E}\left[\mathrm{Y}_{n1\ell}|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n2\ell}|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n3\ell}|\mathbf{X}_{n\ell}\right]\mathrm{E}\left[\mathrm{Y}_{n4\ell}|\mathbf{X}_{n\ell}\right]\right]\right|$$
$$= \frac{p_n^4}{(p_n-1)^4}g_{n2}^4\mathrm{E}\left[\left(u_{n\ell}^2-e_{n2}\right)^4\right] \leq \mathrm{C}g_{n2}^4 e_{n8} = \mathrm{O}\left(\frac{1}{n^2p_n^2}\right)o\left(\frac{n^{2/3}}{p_n^2}\right)$$
$$= o\left(\frac{1}{n^{4/3}p_n^4}\right).$$

Altogether, (4.24) thus yields

$$\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^4\right] = \frac{\mathrm{C}np_n^4}{n^4}\left\{no\left(\frac{n^{4/3}}{p_n^4}\right)+3n^2\mathrm{O}\left(\frac{1}{p_n^4}\right)+\mathrm{C}_1n^2o\left(\frac{n^{2/3}}{p_n^4}\right)\right.$$
$$\left.+\mathrm{C}_2n^3o\left(\frac{1}{p_n^4}\right)+n^4o\left(\frac{1}{n^{4/3}p_n^4}\right)\right\}+o\left(\frac{1}{n^{1/3}}\right)=o(1).$$

From (4.23), we thus conclude that

$$\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^2\mathbb{I}_{\{|\mathrm{D}_{n\ell}|>\varepsilon\}}\right] \leq \frac{1}{\varepsilon}\sqrt{\sum_{\ell=1}^n \mathrm{E}\left[\mathrm{D}_{n\ell}^4\right]} = o(1),$$

which establishes the result. □

4.3.7 Proof of Theorem 4.2.3

(i)-(ii) In these cases, we have $\sqrt{np_n}g_{n2} \to \xi$, with $\xi = 0$ in case (i) and $\xi \neq 0$ in case (ii). Under Assumption (a) and $g_{n2} = \mathrm{O}\left(1/\sqrt{np_n}\right)$, Proposition 4.2.1 yields

$$\mu_n := \mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = \frac{p_n^2(n-1)}{2(p_n-1)}g_{n2}^2 = \frac{\xi^2}{2}+o(1)$$

and

$$\sigma_n^2 := \mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right] = \frac{(n-1)p_n^2}{2n}\left\{\frac{3f_{n4}^2}{p_n^2-1}-\left(\mathrm{O}\left(\frac{1}{np_n}\right)+\frac{1}{p_n}\right)^2\right\}+o(1) = 1+o(1),$$

so that Theorem 4.2.2 and Slutzky's lemma provide

$$\mathrm{Q}_n^{\mathrm{St}} = \sigma_n\left(\frac{\mathrm{Q}_n^{\mathrm{St}}-\mathrm{E}\left[\mathrm{Q}_n^{\mathrm{St}}\right]}{\sqrt{\mathrm{Var}\left[\mathrm{Q}_n^{\mathrm{St}}\right]}}\right)+\mu_n \xrightarrow{\mathcal{D}} \mathcal{N}\left(\frac{\xi^2}{2},1\right),$$

as was to be shown.

(iii) Assume that $\sqrt{np_n}|g_{n2}| \to \infty$ and fix $\mathrm{M} > 0$ (clearly, it is enough to prove the result for $\mathrm{M} > 0$). Proposition 4.2.1 then ensures that μ_n diverges to infinity, so that there exists n_0 such that $\mu_n > \mathrm{M}$ for any $n \geq n_0$. For any such n, the Chebychev inequality yields

$$\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\mathrm{Q}^{\mathrm{St}}_n \leq \mathrm{M}\right] = \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\mathrm{Q}^{\mathrm{St}}_n - \mu_n \leq \mathrm{M} - \mu_n\right] \leq \mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\left|\mathrm{Q}^{\mathrm{St}}_n - \mu_n\right| \geq \mu_n - \mathrm{M}\right]$$
$$\leq \frac{\sigma_n^2}{(\mu_n - \mathrm{M})^2} = \frac{\sigma_n^2}{\mu_n^2}(1 + o(1)).$$

Now, if $e_{n4} = o\left(ng_{n2}^2\right)$, then we also have $p_n e_{n2}^2 = o\left(np_n g_{n2}^2\right)$ and $p_n e_{n2}^2 = o\left(\left(np_n g_{n2}^2\right)^2\right)$. Therefore,

$$\frac{\sigma_n^2}{\mu_n^2} \leq \frac{\mathrm{C}p_n^2}{(np_n g_{n2}^2)^2}\left\{e_{n4}^2 + \frac{6}{p_n - 1}(e_{n2} - e_{n4})^2 + \frac{3f_{n4}^2}{p_n^2 - 1} - \left(\frac{p_n}{p_n - 1}g_{n2}^2 + \frac{1}{p_n}\right)^2\right\}$$
$$+ \frac{\mathrm{C}np_n^2}{(np_n g_{n2}^2)^2}(e_{n4} - e_{n2}^2)g_{n2}^2$$
$$\leq \frac{\mathrm{C}}{(np_n g_{n2}^2)^2}\left\{p_n^2 e_{n4}^2 + 12p_n(e_{n2} - e_{n4})^2\right\} + \frac{\mathrm{C}p_n}{np_n g_{n2}^2}(e_{n4} - e_{n2}^2) + o(1) = o(1),$$

which implies that $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\mathrm{F}_n}\left[\mathrm{Q}^{\mathrm{St}}_n \leq \mathrm{M}\right] \to 0$, hence establishes the result.

□

4.3.8 Proof of Proposition 4.2.4

Since κ_n is assumed to be $o(p_n)$ as $n \to \infty$, Lemma 3.7.1 provides

$$e_{n2} = \frac{1}{\kappa_n}\left(\frac{c_{p_n}}{\check{c}_{p_n,\kappa_n,f}}\right)^{-1} c_{p_n}\int_{-1}^{1}\left(1 - s^2\right)^{(p_n-3)/2}\left(\kappa_n s^2\right) f\left(\kappa_n s^2\right) ds$$
$$= \frac{\frac{1}{\kappa_n}\left(\frac{\kappa_n}{p_n} + \frac{3\kappa_n^2}{p_n^2} + o\left(\frac{\kappa_n^2}{p_n^2}\right)\right)}{1 + \frac{\kappa_n}{p_n} + o\left(\frac{\kappa_n}{p_n}\right)} = \frac{\frac{1}{p_n} + \frac{3\kappa_n}{p_n^2} + o\left(\frac{\kappa_n}{p_n^2}\right)}{1 + \frac{\kappa_n}{p_n} + o\left(\frac{\kappa_n}{p_n}\right)} = \frac{1}{p_n} + \frac{2\kappa_n}{p_n^2} + o\left(\frac{\kappa_n}{p_n^2}\right),$$

which proves the result for e_{n2}. The same lemma also yields

$$e_{n4} = \frac{1}{\kappa_n^2}\left(\frac{c_{p_n}}{\check{c}_{p_n,\kappa_n,f}}\right)^{-1} c_{p_n}\int_{-1}^{1}\left(1 - s^2\right)^{(p_n-3)/2}\left(\kappa_n s^2\right)^2 f\left(\kappa_n s^2\right) ds$$
$$= \frac{\frac{1}{\kappa_n^2}\left(\frac{3\kappa_n^2}{p_n^2} + o\left(\frac{\kappa_n^2}{p_n^2}\right)\right)}{1 + \frac{\kappa_n}{p_n} + o\left(\frac{\kappa_n}{p_n}\right)} = \frac{\frac{3}{p_n^2} + o\left(\frac{1}{p_n^2}\right)}{1 + \frac{\kappa_n}{p_n} + o\left(\frac{\kappa_n}{p_n}\right)} = \frac{3}{p_n^2} + o\left(\frac{1}{p_n^2}\right).$$

The claim for e_{n8} directly follows from the identity $e_{n8} \leq e_{n4}$. □

4.3.9 Proof of Theorem 4.2.5

First note that, in all cases, we have $\kappa_n = o(p_n)$, so that Proposition 4.2.4 applies and ensures that conditions (a)–(b) in Theorem 4.2.3 are fulfilled. Let us then treat cases (i)–(iii) separately.

(i) Since $\kappa_n = o\big(p_n^{3/2}/\sqrt{n}\big)$, Proposition 4.2.4 implies that

$$g_{n2} = e_{n2} - \frac{1}{p_n} = \mathrm{O}\left(\frac{\kappa_n}{p_n^2}\right) = o\left(\frac{1}{\sqrt{np_n}}\right),$$

so that Theorem 4.2.3(i) shows that $\mathrm{Q}_n^{\mathrm{St}}$ is asymptotically standard normal.

(ii) Since $\sqrt{n}\kappa_n/p_n^{3/2} \to \tau(\neq 0)$ and $p_n = o(n)$, Proposition 4.2.4 provides

$$\sqrt{np_n}g_{n2} = \frac{2\sqrt{n}\kappa_n}{p_n^{3/2}} + o\left(\frac{\sqrt{n}\kappa_n}{p_n^{3/2}}\right) = 2\tau + o(1),$$

Theorem 4.2.3(ii) shows that $\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathscr{D}} \mathscr{N}(\xi^2/2, 1)$, with $\xi = 2\tau$, which establishes the result.

(iii) The claim follows from Theorem 4.2.3(iii) since

$$\sqrt{np_n}|g_{n2}| = \frac{2\sqrt{n}\kappa_n}{p_n^{3/2}} + o\left(\frac{\sqrt{n}\kappa_n}{p_n^{3/2}}\right)$$

diverges to infinity.

□

4.3.10 Proof of Proposition 4.2.6

The following result needed to prove Proposition 4.2.6 is a higher-order extension of Lemma 2.5.1 and can be shown in a similar fashion.

Lemma 4.3.11. *Let $g : \mathbb{R} \to \mathbb{R}$ be four times differentiable at 0. Let (p_n) be a sequence of positive integers diverging to ∞ and (κ_n) be a real sequence that is $o\big(\sqrt{p_n}\big)$. Then,*

$$\mathrm{R}_n(g) := c_{p_n}\int_{-1}^{1}\big(1-s^2\big)^{(p_n-3)/2} g(\kappa_n s)\, ds = g(0) + \frac{\kappa_n^2}{2p_n}g''(0) + \frac{\kappa_n^4}{8p_n^2}g''''(0) + o\left(\frac{\kappa_n^4}{p_n^2}\right)$$

as $n \to \infty$, still with $c_p := \left(\int_{-1}^{1}\big(1-s^2\big)^{(p-3)/2}\, ds\right)^{-1}$.

Proof of Proposition 4.2.6. Since κ_n is assumed to be $o\big(\sqrt{p_n}\big)$ as $n \to \infty$, Lemma 4.3.11 provides

$$\begin{aligned} e_{n2} &= \frac{1}{\kappa_n^2}\left(\frac{c_{p_n}}{c_{p_n,\kappa_n,f}}\right)^{-1} c_{p_n}\int_{-1}^{1}\big(1-s^2\big)^{(p_n-3)/2}(\kappa_n s)^2 f(\kappa_n s)\, ds \\ &= \frac{\frac{1}{\kappa_n^2}\left(\frac{\kappa_n^2}{p_n} + \frac{12\kappa_n^4}{8p_n^2}f''(0) + o\left(\frac{\kappa_n^4}{p_n^2}\right)\right)}{1 + \frac{\kappa_n^2}{2p_n}f''(0) + \frac{\kappa_n^4}{8p_n^2}f''''(0) + o\left(\frac{\kappa_n^4}{p_n^2}\right)} = \frac{\frac{1}{p_n} + \frac{3\kappa_n^2}{2p_n^2}f''(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right)}{1 + \frac{\kappa_n^2}{2p_n}f''(0) + \frac{\kappa_n^4}{8p_n^2}f''''(0) + o\left(\frac{\kappa_n^4}{p_n^2}\right)} \\ &= \frac{1}{p_n} + \frac{\kappa_n^2}{p_n^2}f''(0) + o\left(\frac{\kappa_n^2}{p_n^2}\right), \end{aligned}$$

which proves the result for e_{n2}. The same lemma also yields

$$
\begin{aligned}
e_{n4} &= \frac{1}{\kappa_n^4}\left(\frac{c_{p_n}}{c_{p_n,\kappa_n,f}}\right)^{-1} c_{p_n}\int_{-1}^{1}\left(1-s^2\right)^{(p_n-3)/2}(\kappa_n s)^4 f(\kappa_n s)\,ds \\
&= \frac{\frac{1}{\kappa_n^4}\left(\frac{24\kappa_n^4}{8p_n^2}+o\left(\frac{\kappa_n^4}{p_n^2}\right)\right)}{1+\frac{\kappa_n^2}{2p_n}f''(0)+\frac{\kappa_n^4}{8p_n^2}f''''(0)+o\left(\frac{\kappa_n^4}{p_n^2}\right)} = \frac{\frac{3}{p_n^2}+o\left(\frac{1}{p_n^2}\right)}{1+\frac{\kappa_n^2}{2p_n}f''(0)+\frac{\kappa_n^4}{8p_n^2}f''''(0)+o\left(\frac{\kappa_n^4}{p_n^2}\right)} \\
&= \frac{3}{p_n^2}+o\left(\frac{1}{p_n^2}\right).
\end{aligned}
$$

The claim for e_{n8} directly follows from the identity $e_{n8} \le e_{n4}$. □

4.3.11 Proof of Theorem 4.2.7

First note that, in all cases, we have $\kappa_n = o\left(\sqrt{p_n}\right)$, so that Proposition 4.2.6 applies and ensures that conditions (a)–(b) in Theorem 4.2.3 are fulfilled. Let us then treat cases (i)–(iii) separately.

(i) Since $\kappa_n = o\left(p_n^{3/4}/n^{1/4}\right)$, Proposition 4.2.6 implies that

$$
g_{n2} = e_{n2} - \frac{1}{p_n} = \mathrm{O}\left(\frac{\kappa_n^2}{p_n^2}\right) = o\left(\frac{1}{\sqrt{np_n}}\right),
$$

so that Theorem 4.2.3(i) shows that $\mathrm{Q}_n^{\mathrm{St}}$ is asymptotically standard normal.

(ii) Since $n^{1/4}\kappa_n/p_n^{3/4} \to \tau(\neq 0)$ and $p_n = o(n)$, Proposition 4.2.6 provides

$$
\sqrt{np_n}\,g_{n2} = \frac{\sqrt{n}\kappa_n^2}{p_n^{3/2}}f''(0) + o\left(\frac{\sqrt{n}\kappa_n^2}{p_n^{3/2}}\right) = \tau^2 f''(0) + o(1),
$$

Theorem 4.2.3(ii) shows that $\mathrm{Q}_n^{\mathrm{St}} \xrightarrow{\mathcal{D}} \mathcal{N}(\xi^2/2, 1)$, with $\xi = \tau^2 f''(0)$, which establishes the result.

(iii) The claim follows from Theorem 4.2.3(iii) since

$$
\sqrt{np_n}|g_{n2}| = \frac{2\sqrt{n}\kappa_n}{p_n^{3/2}} + o\left(\frac{\sqrt{n}\kappa_n}{p_n^{3/2}}\right)
$$

diverges to infinity.

□

Conclusion

In Chapter 2 monotone rotationally symmetric alternatives with modal location $\boldsymbol{\theta}_n$, concentration κ_n and functional parameter f were considered. We showed that if the concentration parameter is such that $\kappa_n \sim \sqrt{p_n/n}$, the null of uniformity and the sequence of alternatives are too close to be told apart, they are contiguous. For specified $\boldsymbol{\theta}_n$ and at this rate, a Local Asymptotic Normality result was established, which allowed us, both in low and in high dimensions, to define locally asymptotically most powerful tests for the specified-$\boldsymbol{\theta}_n$ problem, and implied that the Rayleigh test shows power only in the low-dimensional case.

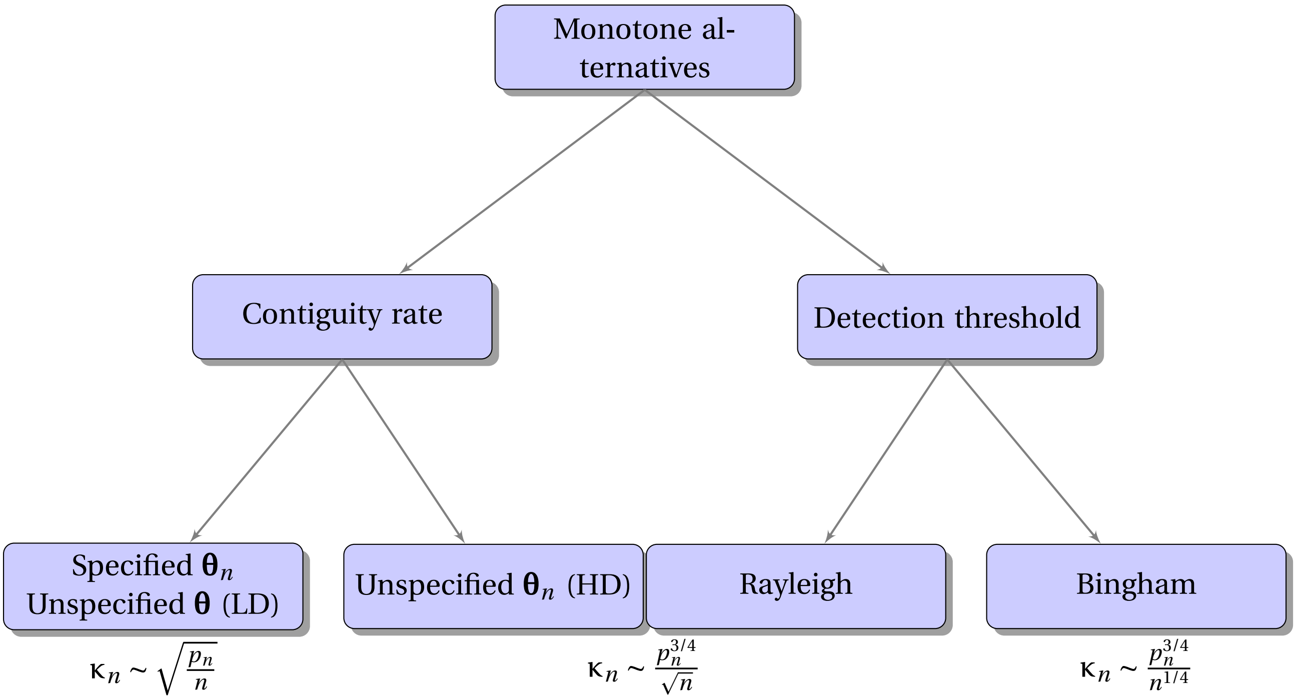
In practice, however, $\boldsymbol{\theta}_n$ may rarely be assumed to be known. In the corresponding unspecified-$\boldsymbol{\theta}_n$ problem, we showed that the Rayleigh test enjoys nice asymptotic optimality properties, both in the low- and high-dimensional cases. In low dimensions, it is locally asymptotically maximin and locally most powerful invariant, irrespective of f. In high dimensions, it is locally asymptotically most powerful invariant in the FvML case, and a conjecture — that is strongly supported by a fourth-order expansion of the relevant f-based local log-likelihood ratio and by the computation of asymptotic powers in Section 4.1 — states that, provided that $p_n = o(n^2)$, this optimality holds for any f that is five times differentiable at 0.

Our results fully characterize the cost of $\boldsymbol{\theta}_n$ not being specified. In low dimensions, this cost is in terms of asymptotic powers but not in terms of rate. In high-dimensions, however, there is a cost in terms of rate, as optimal tests cannot detect the contiguous alternatives in $\kappa_n \sim \sqrt{p_n/n}$, but only the more severe alternatives in $\kappa_n \sim p_n^{3/4}/\sqrt{n}$. Simulation results are in remarkable agreement with our asymptotic results, irrespective of the relative magnitude of n and p — which illustrates the robustness of most of our results in the rate at which p_n goes to infinity with n. A real data example illustrated the usefulness of the high-dimensional Rayleigh test in the framework of testing for sphericity.

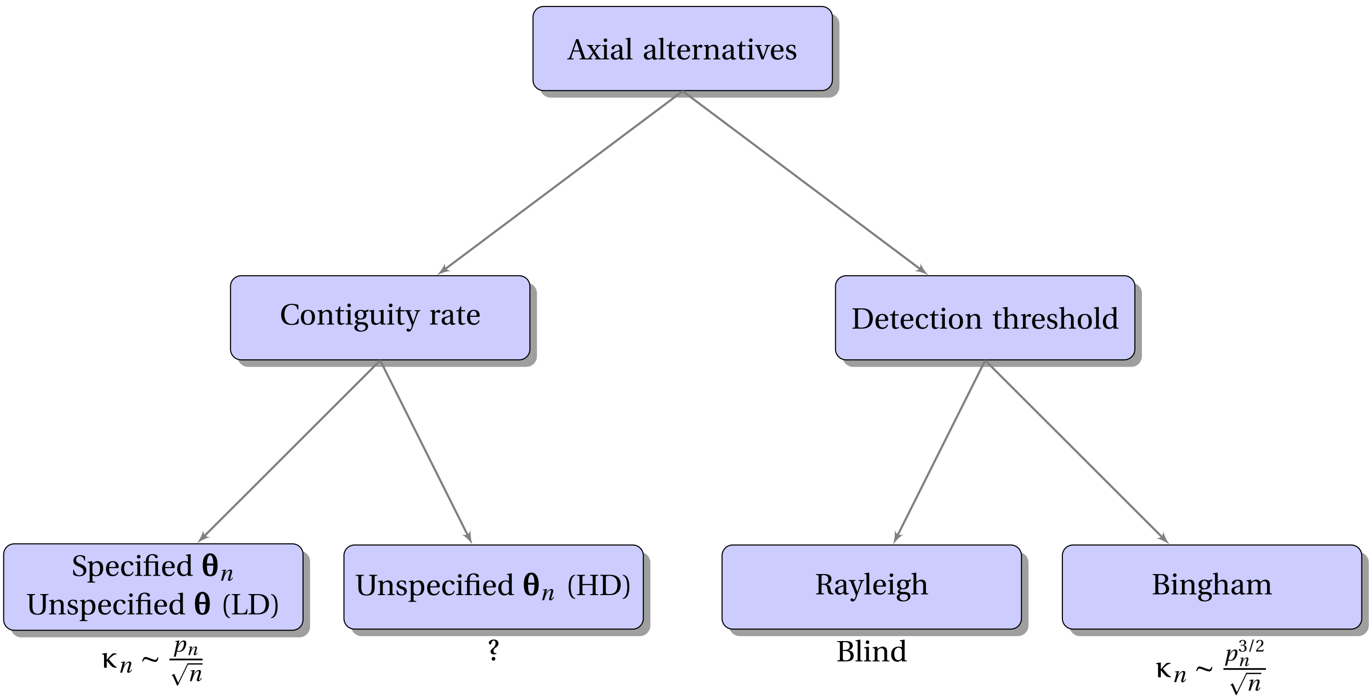
In Chapter 3 we undertook the same work for axial rotationally symmetric alternatives. The problem turned out to be more complex than the monotone one as the higher contiguity rate, $\kappa_n \sim p_n/\sqrt{n}$, suggested. A LAN result at this rate made it again possible to define locally asymptotically most powerful tests for the specified-$\boldsymbol{\theta}_n$ problem; in low dimensions and when $\boldsymbol{\theta}_n$ is not specified, this result can be rewritten thanks to a new parametrisation and the Bingham test emerges as a natural test for uniformity. It is nonetheless not optimal and would be more suitable for multi-spiked alternatives. This is why we considered single-spiked tests based on the extreme eigenvalues of the sign covariance matrix and characterised their asymptotic behaviour under contiguous axial rotationally symmetric alternatives.

In high dimensions, it is extremely challenging to derive the non-null asymptotic powers of these tests and the Bingham test under suitable local alternatives. For the single-spiked tests, for instance, this is due to the fact that eigenvalues of sample covariance matrices suffer complicated phase transition phenomena which, close to uniformity, results in a lack of consistency.

In Chapter 4, by relying on martingale central limit theorems, we showed that after appropriate standardisation, the asymptotic distributions of the Rayleigh and Bingham tests are normal under broad classes of rotationally symmetric distributions. We identified the rotationally symmetric alternatives under which the Bingham test will show non-trivial asymptotic powers. We proved that this test will be blind to less severe alternatives and consistent under more severe ones. Our results impose only very mild assumptions on the considered rotationally symmetric alternatives. In particular, they apply to both the axial and non-axial cases, which allowed us to determine the detection threshold of the Bingham test in each case. Our results reveal that although it exhibits slower consistency rates than the Rayleigh test in the non-axial case, the Bingham test can detect both types of alternatives, whereas the Rayleigh test will be blind to arbitrarily severe axial alternatives. These results are summarized in Diagram 4.5.



(a) Contiguity rate associated with the semiparametric class of distributions considered in Chapter 2 and Section 4.2.3, as well as the corresponding detection thresholds of the Rayleigh and Bingham tests.



(b) Contiguity rate associated with the semiparametric class of distributions considered in Chapter 3 and Section 4.2.2, as well as the corresponding detection thresholds of the Rayleigh and Bingham tests.

Figure 4.5

Appendix A

Asymptotic distribution under $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}$, with $\kappa_n = \tau\sqrt{p/n}$, of the Rayleigh test statistic in the fixed-p case

We focus on the fixed-p case ($p_n = p$ for any n) and derive the asymptotic distribution of the Rayleigh test statistic R_n under sequences of alternatives of the form $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$, with $\kappa_n = \tau_n/\sqrt{n}$, where the sequence (τ_n) converges to some $\tau \in (0,\infty)$.

Thanks to decomposition (1.9) we can rewrite the Rayleigh test statistic R_n as

$$\begin{aligned}\mathrm{R}_n &= \frac{p}{n}\sum_{i,j=1}^{n}\mathbf{X}'_{ni}\mathbf{X}_{nj} = \frac{p}{n}\sum_{i,j=1}^{n}(u_{ni}\boldsymbol{\theta}+v_{ni}\mathbf{S}_{ni})'(u_{nj}\boldsymbol{\theta}+v_{nj}\mathbf{S}_{nj})\\ &= \frac{p}{n}\sum_{i,j=1}^{n}(u_{ni}u_{nj}+v_{ni}v_{nj}\mathbf{S}'_{ni}\mathbf{S}_{nj}) = \mathrm{Y}_n^2 + \frac{1}{p}\mathbf{Z}'_n\mathbf{A}\mathbf{Z}_n,\end{aligned}$$

where we let

$$\begin{aligned}\mathbf{A} &:= p\left(\mathbf{I}_p - \boldsymbol{\theta}\boldsymbol{\theta}'\right),\\ \mathrm{Y}_n &:= \frac{\sqrt{p}}{\sqrt{n}}\sum_{i=1}^{n}u_{ni},\\ \mathbf{Z}_n &:= \frac{\sqrt{p}}{\sqrt{n}}\sum_{i=1}^{n}v_{ni}\mathbf{S}_{ni}.\end{aligned}$$

Under $\mathrm{P}^{(n)}_0$, the multivariate CLT, along with Lemma 4.3.1(i) and identity (1.5) provide

$$\begin{pmatrix}\mathrm{Y}_n\\ \mathbf{Z}_n\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Sigma}\end{pmatrix}\right), \quad \text{with } \boldsymbol{\Sigma} := \frac{1}{p}(\mathbf{I}_p - \boldsymbol{\theta}\boldsymbol{\theta}'). \tag{A.1}$$

Clearly, Y_n^2 is asymptotically χ^2_1 under $\mathrm{P}^{(n)}_0$. By using the identities $\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}$ and $\mathrm{tr}[\boldsymbol{\Sigma}\mathbf{A}] = p-1$, Theorem 9.2.1 in [Rao and Mitra, 1971] shows that $(1/p)\mathbf{Z}'_n\mathbf{A}\mathbf{Z}_n$ is asymptotically χ^2_{p-1} under $\mathrm{P}^{(n)}_0$. Since the joint asymptotic normality result in (A.1) ensures asymptotic independence of Y_n^2 and $(1/p)\mathbf{Z}'_n\mathbf{A}\mathbf{Z}_n$, this confirms that R_n is asymptotically χ^2_p under $\mathrm{P}^{(n)}_0$.

Let us now turn to the sequence of alternatives $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$ considered above. In view of Theorem 2.2.1, Le Cam's Third Lemma directly shows that under $\mathrm{P}^{(n)}_{\boldsymbol{\theta},\kappa_n,f}$,

$$\begin{pmatrix}\mathrm{Y}_n\\ \mathbf{Z}_n\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(\begin{pmatrix}\tau\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}1 & \mathbf{0}'\\ \mathbf{0} & \boldsymbol{\Sigma}\end{pmatrix}\right)$$

as $n \to \infty$. Under the same sequence of alternatives, Y_n^2 is therefore asymptotically $\chi_1^2(\tau^2)$ (that is, non-central χ_1^2 with non-centrality parameter τ^2) and $\frac{1}{p}\mathbf{Z}_n'\mathbf{A}\mathbf{Z}_n$ is still asymptotically χ_{p-1}^2. Asymptotic independence between Y_n and $\mathbf{Z}_n$ still holds under contiguous alternatives, which shows that under $\mathrm{P}_{\boldsymbol{\theta},\kappa_n,f}^{(n)}$

$$\mathrm{R}_n \xrightarrow{\mathscr{D}} \chi_1^2(\tau^2) + \chi_{p-1}^2,$$

where the χ^2 terms are independent. This establishes the asymptotic result in (2.12). □

Appendix B

Universality of the asymptotic high-dimensional distribution of the Rayleigh test in the FvML case

We prove in this section that conditions (i)-(iii) of Theorem 4.1.2 universally hold under FvML distributions, in the sense that, if F_n is FvML for any n, then the theorem does not require any condition on the dependence of p_n and κ_n on n (but for the fact that $p_n \to \infty$ as $n \to \infty$). The proof requires several preliminary results.

Lemma B.0.1. *Write* $e_\ell := e_{\ell;p,\kappa} := \mathrm{E}\left[(\mathbf{X}'\boldsymbol{\theta})^\ell\right]$, *where* $\mathbf{X}$ *follows a* p*-dimensional FvML distribution with concentration* $\kappa(>0)$ *and location* $\boldsymbol{\theta}(\in \mathcal{S}^{p-1})$. *Then*

$$e_1 = r, \qquad e_2 = -\frac{p-1}{\kappa} r + 1, \qquad e_3 = \frac{p(p-1)+\kappa^2}{\kappa^2} r - \frac{p-1}{\kappa}$$

and

$$e_4 = -\frac{(p-1)((p+1)p+2\kappa^2)}{\kappa^3} r + \frac{(p-1)(p+1)+\kappa^2}{\kappa^2},$$

where we let

$$r := r_{p,\kappa} := \frac{\mathcal{I}_{\frac{p}{2}}(\kappa)}{\mathcal{I}_{p/2-1}(\kappa)},$$

where $\mathcal{I}_\nu(\cdot)$ *stands here for the order-*ν *modified Bessel function of the first kind.*

Proof of Lemma B.0.1. Using integration by parts in the representation result

$$\mathcal{I}_\nu(z) = \frac{(z/2)^\nu}{\sqrt{\pi}\,\Gamma\left(\nu+\frac{1}{2}\right)} \int_{-1}^{1} (1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\, ds \tag{B.1}$$

(see, e.g., (10.32.2) in [Olver et al., 2010]) provides

$$\begin{aligned}\int_{-1}^{1} s(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\, ds &= \frac{z}{2\nu+1} \int_{-1}^{1} (1-s^2)^{(\nu+1)-\frac{1}{2}} \exp(zs)\, ds \\ &= \frac{z\sqrt{\pi}\,\Gamma(\nu+\frac{3}{2})\mathcal{I}_{\nu+1}(z)}{(2\nu+1)(z/2)^{\nu+1}},\end{aligned} \tag{B.2}$$

which readily leads to

$$e_1 = \frac{(\kappa/2)^{(p-2)/2}}{\sqrt{\pi}\Gamma\left(\frac{p-1}{2}\right)\mathcal{I}_{p/2-1}(\kappa)} \int_{-1}^{1} s(1-s^2)^{\frac{p-3}{2}} \exp(\kappa s)\, ds = \frac{\Gamma\left(\frac{p+1}{2}\right)\mathcal{I}_{\frac{p}{2}}(\kappa)}{\frac{p-1}{2}\Gamma\left(\frac{p-1}{2}\right)\mathcal{I}_{p/2-1}(\kappa)} = \frac{\mathcal{I}_{\frac{p}{2}}(\kappa)}{\mathcal{I}_{p/2-1}(\kappa)}.$$

Turning to e_2, (B.1) above yields

$$\int_{-1}^{1} s^2(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds \;=\; \int_{-1}^{1} (1-(1-s^2))(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds$$
$$= \frac{\sqrt{\pi}\,\Gamma\left(\nu+\frac{1}{2}\right)\mathscr{I}_\nu(z)}{(z/2)^\nu} - \frac{\sqrt{\pi}\,\Gamma(\nu+\frac{3}{2})\mathscr{I}_{\nu+1}(z)}{(z/2)^{\nu+1}}.$$

Hence,

$$e_2 \;=\; \frac{(\kappa/2)^{(p-2)/2}}{\sqrt{\pi}\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)} \int_{-1}^{1} s^2(1-s^2)^{\frac{p-3}{2}} \exp(\kappa s)\,ds$$
$$= \frac{(\kappa/2)^{(p-2)/2}}{\sqrt{\pi}\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)} \left[\frac{\sqrt{\pi}\,\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)}{(\kappa/2)^{\frac{p}{2}-1}} - \frac{\sqrt{\pi}\,\Gamma\left(\frac{p+1}{2}\right)\mathscr{I}_{\frac{p}{2}}(\kappa)}{(\kappa/2)^{p/2}} \right]$$
$$= 1 - \frac{\Gamma\left(\frac{p+1}{2}\right)\mathscr{I}_{\frac{p}{2}}(\kappa)}{(\kappa/2)\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)} = 1 - \frac{(p-1)\mathscr{I}_{\frac{p}{2}}(\kappa)}{\kappa\mathscr{I}_{p/2-1}(\kappa)},$$

as was to be shown. The results for

$$e_\ell = \frac{(\kappa/2)^{(p-2)/2}}{\sqrt{\pi}\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)} \int_{-1}^{1} s^2(1-s^2)^{\frac{p-3}{2}} \exp(\kappa s)\,ds, \qquad \ell = 3,4,$$

follow similarly by using the expressions obtained by plugging (B.1)-(B.2) into

$$\int_{-1}^{1} s^3(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds = \int_{-1}^{1} s(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds - \int_{-1}^{1} s(1-s^2)^{\nu+1-\frac{1}{2}} \exp(zs)\,ds$$

and

$$\int_{-1}^{1} s^4(1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds = \int_{-1}^{1} (1-s^2)^{\nu+2-\frac{1}{2}} \exp(zs)\,ds$$
$$-2\int_{-1}^{1} (1-s^2)^{\nu+1-\frac{1}{2}} \exp(zs)\,ds + \int_{-1}^{1} (1-s^2)^{\nu-\frac{1}{2}} \exp(zs)\,ds,$$

along with the well-known recurrence relation

$$\mathscr{I}_{\nu+1}(z) = \mathscr{I}_{\nu-1}(z) - \frac{2\nu}{z}\mathscr{I}_\nu(z); \tag{B.3}$$

see (10.29.1) in [Olver et al., 2010]. □

Note that closed form expressions for $f_\ell := f_{\ell;p,\kappa} = \mathrm{E}\left[\left(1-(\mathbf{X}'\boldsymbol{\theta})^2\right)^{\ell/2}\right]$, where $\mathbf{X}$ still follows a p-dimensional FvML distribution with concentration $\kappa(>0)$ and location $\boldsymbol{\theta}$, can be obtained much more directly than in Lemma B.0.1, as (B.1) readily yields

$$f_\ell = \frac{(\kappa/2)^{(p-2)/2}}{\sqrt{\pi}\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)} \int_{-1}^{1} (1-s^2)^{\frac{p+\ell-3}{2}} \exp(\kappa s)\,ds = \frac{\Gamma\left(\frac{p+\ell-1}{2}\right)\mathscr{I}_{\frac{p+\ell}{2}-1}(\kappa)}{(\kappa/2)^{\frac{\ell}{2}}\Gamma\left(\frac{p-1}{2}\right)\mathscr{I}_{p/2-1}(\kappa)}. \tag{B.4}$$

The following theorem implies that, under FvML distributions, Theorem 4.1.2 does not impose any condition on the way p_n should go to infinity as a function of n.

Theorem B.0.2. *Let us still write* $e_\ell := e_{\ell;p,\kappa} := \mathrm{E}\left[(\mathbf{X}'\boldsymbol{\theta})^\ell\right]$ *and* $f_\ell := f_{\ell;p,\kappa} = \mathrm{E}\left[\left(1-(\mathbf{X}'\boldsymbol{\theta})^2\right)^{\ell/2}\right]$, *where* $\mathbf{X}$ *follows a* p*-dimensional FvML distribution with concentration* $\kappa(>0)$ *and location* $\boldsymbol{\theta}$. *Further let* $\tilde{e}_\ell := \tilde{e}_{\ell;p,\kappa} := \mathrm{E}\left[\left(\mathbf{X}'\boldsymbol{\theta}-e_1\right)^\ell\right]$. *Then there exist a positive integer* p_0 *and a real constant* C *such that*

$$\text{(i) } \frac{p\tilde{e}_2^2}{f_2^2} \leq \mathrm{C}, \qquad \text{(ii) } \frac{\tilde{e}_4}{\tilde{e}_2^2} \leq \mathrm{C}, \quad \text{and} \quad \text{(iii) } \frac{f_4}{f_2^2} \leq \mathrm{C},$$

for any $p \geq p_0$ *and any* $\kappa > 0$.

The proof requires both following lemmas on the modified Bessel functions ratio

$$\mathscr{R}_\nu(z) := \frac{\mathscr{I}_{\nu+1}(z)}{\mathscr{I}_\nu(z)},$$

where we adopt the same notation as in [Hornik and Grün, 2013].

Lemma B.0.3. *Fix* $\nu > 0$ *and* $z > 0$, *and let* $\mathrm{G}_{\alpha,\beta}(t) = t/(\alpha+\sqrt{t^2+\beta^2})$. *Then*

(i) $\mathscr{R}_\nu(z) \geq \mathrm{G}_{\nu+1,\nu+1}(z)$,

(ii) $\mathscr{R}_\nu(z) \geq \mathrm{G}_{\nu+1/2,\nu+3/2}(z)$,

(iii) $\mathscr{R}_\nu(z) \leq \mathrm{G}_{\nu+1/2,\nu+1/2}(z)$,

(iv) $\mathscr{R}_\nu(z) \leq \mathrm{G}_{\nu,\nu+2}(z)$,

(v) $\mathscr{R}_\nu(z) \leq \mathrm{G}_{\nu,\nu}(z)$,

(vi) $\mathscr{R}_\nu(z) \leq \mathrm{G}_{\nu+1/2,\sqrt{(\nu+1/2)(\nu+3/2)}}(z)$.

These bounds, that have been obtained in [Amos, 1974] ((i)-(v)) and [Simpson and Spector, 1984] ((vi)), are actually sufficient to establish Theorem B.0.2(i) and (iii). To prove Theorem B.0.2(ii), however, we will need the following reinforcement of the bounds in Lemma B.0.3(ii)-(iii) and an appropriate control of the resulting approximation error; see [Paindaveine, 2016] for a proof.

Lemma B.0.4. *Fix* $\nu > 0$ *and* $z \geq 0$, *and let*

$$a_\nu(z) := \frac{\left(\nu+\frac{3}{2}\right)(\nu+4)+z^2}{\left(\nu+\frac{3}{2}\right)\left(\nu+\frac{7}{2}\right)+z^2}, \quad b_\nu(z) := \frac{\left(\nu+\frac{5}{2}\right)(\nu+2)+z^2}{\left(\nu+\frac{3}{2}\right)\left(\nu+\frac{7}{2}\right)+z^2},$$

$$c_\nu(z) := \frac{\left(\nu+\frac{3}{2}\right)\left(\nu+\frac{5}{2}\right)+z^2}{\left(\nu+\frac{3}{2}\right)\left(\nu+\frac{7}{2}\right)+z^2}, \quad d_\nu(z) := \frac{(\nu+1)\left(\nu+\frac{3}{2}\right)+z^2}{\left(\nu+\frac{1}{2}\right)\left(\nu+\frac{3}{2}\right)+z^2}$$

and

$$e_\nu(z) := \frac{\left(\nu+\frac{1}{2}\right)\left(\nu+\frac{5}{2}\right)+z^2}{\left(\nu+\frac{1}{2}\right)\left(\nu+\frac{3}{2}\right)+z^2}.$$

Then

(i) $\mathrm{L}_\nu(z) \leq \mathrm{R}_\nu(z) \leq \mathrm{U}_\nu(z)$, *with*

$$\mathrm{L}_\nu(z) := \frac{z}{a_\nu(z)\left(\nu+\frac{1}{2}\right)+\sqrt{\left(b_\nu(z)\left(\nu+\frac{3}{2}\right)\right)^2+c_\nu(z)z^2}}$$

and

$$\mathrm{U}_\nu(z) := \frac{z}{d_\nu(z)\left(\nu+\frac{1}{2}\right)+\sqrt{\left(d_\nu(z)\left(\nu+\frac{1}{2}\right)\right)^2+e_\nu(z)z^2}};$$

(ii) there exists $\nu_0 > 0$ *such that*

$$\frac{\nu^7 + z^7}{z^3}\left(\mathrm{U}_\nu(z) - \mathrm{L}_\nu(z)\right) \le \frac{3(2\nu+3)^2}{8}$$

for any $\nu \ge \nu_0$ *and any* $z > 0$.

Proof of Theorem B.0.2 (i). From Lemma B.0.1, we obtain

$$\frac{p\tilde{e}_2^2}{f_2^2} = \frac{p\tilde{e}_2^2}{(1-e_2)^2} = \frac{\kappa^2 p\left(r^2 + \frac{p-1}{\kappa}r - 1\right)^2}{(p-1)^2 r^2} = \frac{\kappa^2 p\left(g_a(r)g_b(r)\right)^2}{(p-1)^2 r^2}, \tag{B.5}$$

where we let

$$g_a(x) := \mathrm{G}_{-\frac{p-1}{2},\frac{p-1}{2}}(\kappa) + x,$$
$$g_b(x) := \mathrm{G}_{\frac{p-1}{2},\frac{p-1}{2}}(\kappa) - x.$$

We need to control both $g_a(r)$ and $g_b(r)$, which can be achieved by using Lemma B.0.3. Starting with $g_a(r)$, Lemma B.0.3(iii) readily yields that

$$g_a(r) \le \mathrm{G}_{-\frac{p-1}{2},\frac{p-1}{2}}(\kappa) + \mathrm{G}_{\frac{p-1}{2},\frac{p-1}{2}}(\kappa) = \frac{2}{\kappa}\sqrt{\left(\tfrac{p-1}{2}\right)^2 + \kappa^2}.$$

As for $g_b(r)$, Lemma B.0.3(ii) entails

$$g_b(r) \le \mathrm{G}_{\frac{p-1}{2},\frac{p-1}{2}}(\kappa) - \mathrm{G}_{\frac{p-1}{2},\frac{p+1}{2}}(\kappa) = \frac{\kappa\left(\sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2} - \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)}{\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)\left(\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2}\right)}$$

$$= \frac{\kappa p}{\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)\left(\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2}\right)\left(\sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)}$$

$$\le \frac{\kappa p}{2\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)^2\sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}}.$$

Plugging into (B.5) provides

$$\frac{p\tilde{e}_2^2}{f_2^2} \le \frac{\kappa^2 p^3}{(p-1)^2\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)^4 r^2}.$$

Using Lemma B.0.3(ii) again then yields

$$\frac{p\tilde{e}_2^2}{f_2^2} \le \frac{\kappa^2 p^3}{(p-1)^2\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)^4\left(\mathrm{G}_{\frac{p-1}{2},\frac{p+1}{2}}(\kappa)\right)^2} = \frac{p^3\left(\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2}\right)^2}{(p-1)^2\left(\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}\right)^4}$$

$$\le \frac{p^3}{(p-1)^4}\left(\frac{\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2+\kappa^2}}{\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}}\right)^2 \le \frac{8}{p-1}\left(\frac{\frac{p-1}{2} + \sqrt{9\left(\frac{p-1}{2}\right)^2+\kappa^2}}{\frac{p-1}{2} + \sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2}}\right)^2 \le 24,$$

for any $p \ge 2$ and any $\kappa > 0$, which establishes the result. □

Proof of Theorem B.0.2 (iii). From (B.4), we obtain

$$\frac{f_4}{f_2^2} = \frac{(p+1)\mathscr{I}_{p/2-1}(\kappa)\mathscr{I}_{p/2+1}(\kappa)}{(p-1)\mathscr{I}_{p/2}^2(\kappa)} \le \frac{3\mathscr{I}_{p/2-1}(\kappa)\mathscr{I}_{p/2+1}(\kappa)}{\mathscr{I}_{p/2}^2(\kappa)},$$

for any $p \ge 2$ and any $\kappa > 0$. By using (B.3), this provides

$$\frac{f_4}{3f_2^2} \le \frac{\mathscr{I}_{p/2-1}(\kappa)(\mathscr{I}_{p/2-1}(\kappa) - (p/\kappa)\mathscr{I}_{\frac{p}{2}}(\kappa))}{\mathscr{I}_{p/2}^2(\kappa)} = \frac{1}{r^2} - \frac{p}{\kappa r} = \frac{\kappa - pr}{\kappa r^2}$$

(note that Lemma B.0.3(iv) yields $r \le \mathrm{G}_{p/2-1,p/2+1}(\kappa) \le \kappa/p$). Lemma B.0.3(i) then entails

$$\begin{aligned}
\frac{f_4}{3f_2^2} &\le \frac{\kappa - p\mathrm{G}_{\frac{p-1}{2},\frac{p+1}{2}}(\kappa)}{\kappa\left(\mathrm{G}_{\frac{p-1}{2},\frac{p+1}{2}}(\kappa)\right)^2} = \frac{1}{\kappa^2}\left(\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}\right)^2\left(1 - \frac{p}{\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}}\right) \\
&\le \frac{1}{\kappa^2}\left(\frac{p-1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}\right)\left(-\frac{p+1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}\right) \\
&= 1 + \frac{1}{\kappa^2}\left(\frac{p+1}{2} - \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}\right) = 1 - \frac{1}{\frac{p+1}{2} + \sqrt{\left(\frac{p+1}{2}\right)^2 + \kappa^2}} \le 1,
\end{aligned}$$

which proves the result. □

Proof of Theorem B.0.2 (ii). Plugging the expressions of e_ℓ, $\ell = 1,2,3,4$, from Lemma B.0.1 into

$$\frac{\tilde{e}_{n4}}{\tilde{e}_{n2}^2} = \frac{e_4 - 4e_3e_1 + 6e_2e_1^2 - 4e_1^4 + e_1^4}{(e_2 - e_1^2)^2} = \frac{e_4 - 4e_3e_1 + 6e_2e_1^2 - 3e_1^4}{e_2^2 - 2e_1^2e_2 + e_1^4} \tag{B.6}$$

yields (after tedious computations)

$$\frac{\tilde{e}_{n4}}{\tilde{e}_{n2}^2} = \frac{-(p^2 + 2p + 4\kappa^2 - 3)x^2 - \frac{p-1}{\kappa}(p^2 + p + 4\kappa^2)x + (p^2 + 4\kappa^2 - 1)}{\kappa^2\left(x^2 + \frac{p-1}{\kappa}x - 1\right)^2} - 3 =: h(r) - 3,$$

We need to show that $h(r)$ is bounded in (p,κ) for p large enough, which will be done on the basis of the factorization

$$h(r) = \frac{(p^2 + 2p + 4\kappa^2 - 3)f_a(r)f_b(r)}{\kappa^2\left(g_a(r)g_b(r)\right)^2}, \tag{B.7}$$

where we let

$$f_a(x) := \left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2} + \sqrt{\left(\left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2}\right)^2 + \frac{\left(\frac{p-1}{2}\right)\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2}} + x,$$

$$f_b(x) := -\left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2} + \sqrt{\left(\left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2}\right)^2 + \frac{\left(\frac{p-1}{2}\right)\left(\frac{p+1}{2}\right) + \kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right) + \kappa^2}} - x,$$

and where $g_a(x)$ and $g_b(x)$ are the functions already considered in the proof of Theorem B.0.2(i).

We start with $g_a(r)$, which, in view of Lemma B.0.3(i), satisfies

$$g_a(r) \geq \mathrm{G}_{-\frac{p-1}{2},\frac{p-1}{2}}(\kappa)+\mathrm{G}_{\frac{p}{2},\frac{p}{2}}(\kappa) \geq \sqrt{1+\left(\frac{p-1}{2\kappa}\right)^2}+\tfrac{p-1}{2\kappa}+\sqrt{1+\left(\tfrac{p}{2\kappa}\right)^2}-\tfrac{p}{2\kappa}$$
$$\geq 2\sqrt{1+\left(\frac{p-1}{2\kappa}\right)^2}-\tfrac{1}{2\kappa} \geq \sqrt{2}\left(1+\left(\tfrac{p-1}{2\kappa}\right)\right)-\tfrac{1}{2\kappa} \geq \sqrt{2}\left(1+\left(\tfrac{p-2}{2\kappa}\right)\right) \geq \frac{\mathrm{C}(p+\kappa)}{\kappa},$$

where C stands for a positive real constant (that may change from line to line in the rest of the proof). Turning to $g_b(r)$, Lemma B.0.3(vi) yields

$$g_b(r) = \mathrm{G}_{\frac{p-1}{2},\frac{p-1}{2}}(\kappa)-\mathrm{G}_{\frac{p-1}{2},\sqrt{(p^2-1)/4}}(\kappa) = \frac{\kappa(\sqrt{\frac{p^2-1}{4}+\kappa^2}-\sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2})}{(\frac{p-1}{2}+\sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2})(\frac{p-1}{2}+\sqrt{\frac{p^2-1}{4}+\kappa^2})}$$
$$= \frac{\kappa(p-1)}{2(\frac{p-1}{2}+\sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2})(\frac{p-1}{2}+\sqrt{\frac{p^2-1}{4}+\kappa^2})(\sqrt{\frac{p^2-1}{4}+\kappa^2}+\sqrt{\left(\frac{p-1}{2}\right)^2+\kappa^2})}$$
$$\geq \frac{\kappa(p-1)}{4(p+\sqrt{p^2+\kappa^2})^3} \geq \frac{\mathrm{C}\kappa p}{p^3+\kappa^3}.$$

Now, by applying Lemma B.0.3(v), we obtain

$$f_a(r) \leq \tfrac{p-1}{2\kappa}+\sqrt{\left(\tfrac{p-1}{2\kappa}\right)^2+1}+r \leq \tfrac{p-1}{2\kappa}+\sqrt{\left(\tfrac{p-1}{2\kappa}\right)^2+1}+\mathrm{G}_{\frac{p}{2}-1,\frac{p}{2}-1}(\kappa)$$
$$= \frac{\frac{p-1}{2}+\sqrt{(\frac{p-1}{2})^2+\kappa^2}}{\kappa}+\frac{\sqrt{(\frac{p}{2}-1)^2+\kappa^2}-(\frac{p}{2}-1)}{\kappa}$$
$$\leq \frac{\frac{1}{2}+2\sqrt{(\frac{p-1}{2})^2+\kappa^2}}{\kappa} \leq \frac{\mathrm{C}(p+\kappa)}{\kappa}.$$

Finally, using the notation and results from Lemma B.0.4, we obtain

$$\begin{aligned} f_b(r) &= -\left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right)+\kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right)+\kappa^2}+\sqrt{\left(\left(\frac{p-1}{2\kappa}\right)\frac{\frac{p}{2}\left(\frac{p+1}{2}\right)+\kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right)+\kappa^2}\right)^2+\frac{\left(\frac{p-1}{2}\right)\left(\frac{p+1}{2}\right)+\kappa^2}{\left(\frac{p-1}{2}\right)\left(\frac{p+3}{2}\right)+\kappa^2}}-r \\ &= \mathrm{U}_{\frac{p}{2}-1}-r \leq \mathrm{U}_{\frac{p}{2}-1}-\mathrm{L}_{\frac{p}{2}-1} \leq \frac{\mathrm{C}\kappa^3(\frac{p+1}{2})^2}{(\frac{p}{2}-1)^7+\kappa^7} \leq \frac{\mathrm{C}\kappa^3 p^2}{p^7+\kappa^7}, \end{aligned}$$

for p large enough and any $\kappa > 0$.

Plugging in (B.7) the bounds just obtained on $g_a(r)$, $g_b(r)$, $f_a(r)$ and $f_b(r)$ entails

$$\begin{aligned} \frac{\tilde{e}_{n4}}{\tilde{e}_{n2}^2}+3 = h(r) &\leq \mathrm{C}\frac{p^2+2p+4\kappa^2-3}{\kappa^2}\times\frac{\kappa^2}{(p+\kappa)^2}\times\frac{(p^3+\kappa^3)^2}{\kappa^2p^2}\times\frac{p+\kappa}{\kappa}\times\frac{\kappa^3p^2}{p^7+\kappa^7} \\ &\leq \mathrm{C}\frac{(p^2+\kappa^2)(p^3+\kappa^3)^2}{(p+\kappa)(p^7+\kappa^7)} \leq \mathrm{C}, \end{aligned}$$

for p large enough and any $\kappa > 0$, as was to be proved. □

Appendix C

Consistency of the Bingham test when $n^{1/4}\kappa_n/p_n^{3/4} \to \infty$ in the high-dimensional FvML case

In this appendix, we show that the constraint $\kappa_n = o\left(\sqrt{p_n}\right)$ in Theorem 4.2.7(iii) is superfluous in the FvML case, which validates the concentration scheme (iii) in the simulation exercise we conducted in Section 4.2.3 (recall that the condition $\kappa_n = o\left(\sqrt{p_n}\right)$ is not met in this concentration scheme).

Proposition C.0.1. *Let (p_n) be a sequence of positive integers diverging to ∞ and $(\boldsymbol{\theta}_n)$ be a sequence such that $\boldsymbol{\theta}_n \in \mathcal{S}^{p_n-1}$ for any n. Fix $f(z) = \exp(z)$. Assume that the real non-negative sequence (κ_n) satisfies $n^{1/4}\kappa_n/p_n^{3/4} \to \infty$. Then,*

(i) $\sqrt{np_n}\, g_{n2} \to \infty$ *as* $n \to \infty$,

(ii) $e_{n4} = o(ng_{n2}^2)$ *as* $n \to \infty$,

so that

(iii) for any real number M, $\mathrm{P}^{(n)}_{\boldsymbol{\theta}_n,\kappa_n,f}\left[\mathrm{Q}_n^{\mathrm{St}} > \mathrm{M}\right] \to 1$ *as* $n \to \infty$.

Proof.

(i) As seen from (2)–(3) in [Schou, 1978] or from Lemma B.0.1,

$$e_{n1} = \frac{\mathscr{I}_{p_n/2}(\kappa_n)}{\mathscr{I}_{p_n/2-1}(\kappa_n)}, \qquad e_{n2} = -\frac{p_n-1}{\kappa_n}e_{n1} + 1 \tag{C.1}$$

and

$$e_{n4} = -\frac{(p_n-1)\left(p_n(p_n+1)+2\kappa_n^2\right)}{\kappa_n^3}e_{n1} + \frac{(p_n-1)(p_n+1)+\kappa_n^2}{\kappa_n^2}, \tag{C.2}$$

where $\mathscr{I}_\nu(\cdot)$ stands for the order-ν modified Bessel function of the first kind. It follows from (11) in [Amos, 1974] that

$$\frac{\kappa_n}{\frac{p_n}{2} + \sqrt{\frac{p_n^2}{4} + \kappa_n^2}} \le e_{n1} \le \frac{\kappa_n}{\frac{p_n}{2} - 1 + \sqrt{\left(\frac{p_n}{2}+1\right)^2 + \kappa_n^2}}. \tag{C.3}$$

Now, note that (C.1) provides (for $p_n \ge 2$)

$$g_{n2} = e_{n2} - \frac{1}{p_n} = 1 - \frac{1}{p_n} - \frac{p_n-1}{\kappa_n}e_{n1} = \frac{p_n-1}{p_n}\left(1 - \frac{e_{n1}}{\kappa_n/p_n}\right) \ge \frac{1}{2}\left(1 - \frac{e_{n1}}{\kappa_n/p_n}\right).$$

Since (C.3) implies in particular that

$$1-\frac{e_{n1}}{\kappa_n/p_n}\geq 1-\frac{1}{\frac{1}{2}-\frac{1}{p_n}+\sqrt{\left(\frac{1}{2}+\frac{1}{p_n}\right)^2+\frac{\kappa_n^2}{p_n^2}}}=\frac{\kappa_n^2/p_n^2}{\frac{1}{2}+\frac{1}{p_n}+\frac{\kappa_n^2}{p_n^2}+\sqrt{\left(\frac{1}{2}+\frac{1}{p_n}\right)^2+\frac{\kappa_n^2}{p_n^2}}},$$

we thus have

$$2\sqrt{np_n}g_{n2}\geq\frac{\sqrt{n}\kappa_n^2/p_n^{3/2}}{\frac{1}{2}+\frac{1}{p_n}+\frac{\kappa_n^2}{p_n^2}+\sqrt{\left(\frac{1}{2}+\frac{1}{p_n}\right)^2+\frac{\kappa_n^2}{p_n^2}}}.$$

Therefore, by using the identity $\sqrt{a^2+b^2}\leq|a|+|b|$, we obtain that (still for $p_n\geq 2$)

$$2\sqrt{np_n}g_{n2}\geq\frac{\sqrt{n}\kappa_n^2/p_n^{3/2}}{1+\frac{2}{p_n}+\frac{\kappa_n^2}{p_n^2}+\frac{\kappa_n}{p_n}}\geq\frac{\sqrt{n}\kappa_n^2/p_n^{3/2}}{2\left(1+\frac{\kappa_n}{p_n}\right)^2}\geq\frac{\sqrt{n}\kappa_n^2/p_n^{3/2}}{2\left(2\max\left(1,\frac{\kappa_n}{p_n}\right)\right)^2}$$

$$=\frac{\sqrt{n}\kappa_n^2}{8p_n^{3/2}}\min\left(1,\frac{p_n^2}{\kappa_n^2}\right)=\frac{1}{8}\min\left(\left(\frac{n^{1/4}\kappa_n}{p_n^{3/4}}\right)^2,\sqrt{np_n}\right),\qquad\text{(C.4)}$$

which diverges to infinity by assumption. Part (i) of the result follows.

(ii) From (C.4), we have

$$\frac{e_{n4}}{ng_{n2}^2}=\frac{4p_ne_{n4}}{\left(2\sqrt{np_n}g_{n2}\right)^2}\leq 256p_ne_{n4}\max\left(\frac{p_n^3}{n\kappa_n^4},\frac{1}{np_n}\right)\leq 256\max\left(\frac{p_n^4e_{n4}}{n\kappa_n^4},\frac{1}{n}\right).\qquad\text{(C.5)}$$

Now, by using (C.2) and (C.3), we obtain

$$e_{n4}\leq-\frac{(p_n-1)\left(p_n(p_n+1)+2\kappa_n^2\right)}{\kappa_n^3}\times\frac{\kappa_n/p_n}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}+\frac{(p_n-1)(p_n+1)+\kappa_n^2}{\kappa_n^2}$$

$$=1+\frac{p_n^2-1}{\kappa_n^2}-\frac{p_n^2-1}{\kappa_n^2\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)}-\frac{2(p_n-1)}{p_n\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)}$$

$$=1+\frac{p_n^2-1}{\kappa_n^2}\left(1-\frac{1}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}\right)-\frac{2}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}+\frac{2}{p_n\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)}$$

$$=1+\frac{p_n^2-1}{p_n^2\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)^2}-\frac{2}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}+\frac{2}{p_n\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)},$$

which entails

$$e_{n4}\leq 1+\frac{1}{\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)^2}-\frac{2}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}+\frac{2}{p_n}=\left(1-\frac{1}{\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}}\right)^2+\frac{2}{p_n}$$

$$=\frac{\kappa_n^4/p_n^4}{\left(\frac{1}{2}+\sqrt{\frac{1}{4}+\frac{\kappa_n^2}{p_n^2}}\right)^4}+\frac{2}{p_n}\leq\frac{\kappa_n^4}{p_n^4}+\frac{2}{p_n}.$$

Therefore,

$$\frac{e_{n4}p_n^4}{n\kappa_n^4} \leq \frac{1}{n} + \frac{2p_n^3}{n\kappa_n^4} = o(1)$$

by assumption, so that Part (ii) of the result follows from (C.5).

(iii) In view of Parts (i)–(ii) of the result, the claim directly follows from Theorem 4.2.3(iii).

□

Bibliography

[Amos, 1974] Amos, D. E. (1974). Computation of modified bessel functions and their ratios. *Mathematics of Computation*, 28(125):239–251. 121, 125

[Anderson, 2003] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York, 3rd edition. 72, 73

[Anderson and Stephens, 1972] Anderson, T. W. and Stephens, M. A. (1972). Tests for randomness of directions against equatorial and bimodal alternatives. *Biometrika*, 43:613–621. 12, 44, 53, 55

[Arfken et al., 2013] Arfken, G. B., Weber, H. J., and Harris, F. E. (2013). Chapter 1 - mathematical preliminaries. In Arfken, G. B., Weber, H. J., and Harris, F. E., editors, *Mathematical Methods for Physicists (Seventh Edition)*, pages 1 – 82. Academic Press, Boston, seventh edition edition. 39, 64

[Azzalini and Capitanio, 1999] Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *J. R. Stat. Soc. Ser. B*, 61:579–602. 60

[Banerjee et al., 2003] Banerjee, A., Dhillon, I., Ghosh, J., and Sra, S. (2003). Generative model-based clustering of directional data. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 19–28. 9

[Banerjee and Ghosh, 2004] Banerjee, A. and Ghosh, J. (2004). Frequency sensitive competitive learning for scalable balanced clustering on high-dimensional hyperspheres. *IEEE T. Neural Networ.*, 15:702–719. 9

[Bernoulli, 1735] Bernoulli, D. (1735). Quelle est la cause physique de l'inclinaison des plans des Orbites des Planètes par rapport au plan de l'Equateur de la révolution du Soleil autour de son axe ; et d'où vient que les inclinaisons de ces Orbites sont différentes entre elles. In *Recueil des pièces qui ont remporté le prix de l'Académie Royale des Sciences de Paris 1734*, pages 93–122. Académie Royale des Sciences de Paris, Paris. Reprinted in *Daniel Bernoulli, Werke*, Vol. 3, 226–303, Birkhäuser, Basel (1982). 9

[Bhattacharya, 2019] Bhattacharya, B. B. (2019). Asymptotic distribution and detection thresholds for two-sample tests based on geometric graphs. *Ann. Statist., to appear*. 84

[Bijral et al., 2007] Bijral, A. S., Breitenbach, M., and Grudic, G. (2007). Mixture of watson distributions: a generative model for hyperspherical embeddings. In *Artificial Intelligence and Statistics*, pages 35–42. 11

[Billingsley, 1995] Billingsley, P. (1995). *Probability and Measure*. Wiley, New York, Chichester, 3rd edition edition. 24

[Bingham, 1974] Bingham, C. (1974). An antipodally symmetric distribution on the sphere. *Ann. Statist.*, 2:1201–1225. 12, 17, 44

[Cai et al., 2013] Cai, T., Fan, J., and Jiang, T. (2013). Distributions of angles in random packing on spheres. *J. Mach. Learn. Res.*, 14:1837–1864. 10, 60

[Chaudhuri, 1992] Chaudhuri, P. (1992). Multivariate location estimation using extension of r-estimates through u-statistics type approach. *Ann. Statist.*, 20:897–916. 9

[Chikuse, 1991] Chikuse, Y. (1991). High dimensional limit theorems and matrix decompositions on the stiefel manifold. *J. Multivariate anal.*, 36:145–162. 10

[Chikuse, 1993] Chikuse, Y. (1993). High dimensional asymptotic expansions for the matrix langevin distributions on the stiefel manifold. *J. Multivariate Anal.*, 44:82–101. 10

[Chikuse, 2003] Chikuse, Y. (2003). Concentrated matrix langevin distributions. *J. Multivariate Anal.*, 85:375–394. 34, 36, 46

[Cuesta-Albertos et al., 2009] Cuesta-Albertos, J. A., Cuevas, A., and Fraiman, R. (2009). On projection-based tests for directional and compositional data. *Stat. Comput.*, 19(4):367–380. 9, 60

[Davies, 1977] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64:247–254. 32, 52

[Davies, 1987] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74:33–43. 32, 52

[Davies, 2002] Davies, R. B. (2002). Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika*, 89:484–489. 32, 52

[Dortet-Bernadet and Wicker, 2007] Dortet-Bernadet, J.-L. and Wicker, N. (2007). Model-based clustering on the unit sphere with an illustration using gene expression profiles. *Biostatistics*, 9(1):66–80. 62

[Dryden, 2005] Dryden, I. L. (2005). Statistical analysis on high-dimensional spheres and shape spaces. *Ann. Statist.*, 33:1643–1665. 9, 11

[Dürre et al., 2016] Dürre, A., Tyler, D. E., and Vogel, D. (2016). On the eigenvalues of the spatial sign covariance matrix in more than two dimensions. *Statist. Probab. Lett.*, 111:80–85. 44

[Eisen et al., 1998] Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA*, 95:14863–14868. 62

[Fisher et al., 1987] Fisher, N. I., Lewis, T., and Embleton, B. J. (1987). *Statistical analysis of spherical data*. Cambridge Univ. Press press, Cambridge. 9

[Fuller, 1995] Fuller, W. (1995). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. Wiley. 42

[García-Portugués et al., 2019] García-Portugués, E., Paindaveine, D., and Verdebout, T. (2019). On optimal tests for rotational symmetry against new classes of hyperspherical distributions. *J. Amer. Statist. Assoc. To appear.* 9

[García-Portugués and Verdebout, 2018] García-Portugués, E. and Verdebout, T. (2018). An overview of uniformity tests on the hypersphere. *arXiv preprint arXiv:1804.00286.* 9

[Giri, 1996] Giri, N. C. (1996). *Group Invariance in Statistical Inference.* World Scientific Publishing Company, Singapore. 24

[Hallin and Paindaveine, 2006] Hallin, M. and Paindaveine, D. (2006). Semiparametrically efficient rank-based inference for shape. i. optimal rank-based tests for sphericity. *Ann. Statist.*, 34:2707–2756. 9

[Hornik and Grün, 2013] Hornik, K. and Grün, B. (2013). Amos-type bounds for modified bessel function ratios. *J. Math. Anal. Appl.*, 408:91–101. 121

[Hornik and Grün, 2014] Hornik, K. and Grün, B. (2014). movmf: An r package for fitting mixtures of von mises-fisher distributions. *J. Statist. Softw.*, 58. 42

[John, 1972] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika*, 59:169–173. 10, 60

[Juan and Prieto, 2001] Juan, J. and Prieto, F. J. (2001). Using angles to identify concentrated multivariate outliers. *Technometrics*, 43:311–322. 9, 60

[Jupp, 2001] Jupp, P. (2001). Modifications of the rayleigh and bingham tests for uniformity of directions. *J. Multivariate Anal.*, 77(1):1–20. 12

[Jupp, 2008] Jupp, P. E. (2008). Data-driven Sobolev tests of uniformity on compact Riemannian manifolds. *Ann. Statist.*, 36(3):1246–1260. 9

[Le Cam, 1960] Le Cam, L. (1960). *Locally Asymptotically Normal Families of Distributions. Certain Approximations to Families of Distributions and Their Use in the Theory of Estimation and Testing Hypotheses.* University of California Publications in Statistics. vol. 3. no. 2. Berkeley & Los Angeles. 20

[Le Cam and Yang, 2000] Le Cam, L. and Yang, G. (2000). *Asymptotics in Statistics.* Springer New York. 17, 19

[Ledoit and Wolf, 2002] Ledoit, O. and Wolf, M. (2002). Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size. *Ann. Statist.*, 30:1081–1102. 10, 11, 60

[Lehmann and Romano, 2005] Lehmann, E. and Romano, J. (2005). *Testing Statistical Hypotheses.* Springer, New York. 20, 23, 31

[Ley and Verdebout, 2017] Ley, C. and Verdebout, T. (2017). *Modern Directional Statistics.* CRC Press, Boca Raton. 9

[Liese and Miescke, 2008] Liese, F. and Miescke, K.-J. (2008). *Statistical Decision Theory: Estimation, Testing, and Selection.* Springer, New York. 17, 19, 23

[Magnus and Neudecker, 2007] Magnus, J. R. and Neudecker, H. (2007). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, Chichester, 3rd edition edition. 14, 95

[Mardia and Jupp, 2000] Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. John Wiley & Sons, Chichester. 9, 11, 15, 16, 17

[Moreira, 2009] Moreira, M. (2009). A maximum likelihood method for the incidental parameter problem. *Ann. Statist.*, 37:3660–3696. 34

[Möttönen and Oja, 1995] Möttönen, J. and Oja, H. (1995). Multivariate spatial sign and rank methods. *J. Nonparametric Stat.*, 5(2):201–213. 9

[Muirhead, 1982] Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. John Wiley & Sons. 14, 69

[Oja, 2010] Oja, H. (2010). *Multivariate Nonparametric Methods with R. An Approach Based on Spatial Signs and Ranks*. Springer-Verlag. 9

[Olver et al., 2010] Olver, F. W. J., Lozier, D. W., Boisvert, R. F., and Clark, C. W. (2010). *NIST Handbook of Mathematical Functions*. Cambridge Univ. Press. 119, 120

[Paindaveine, 2016] Paindaveine, D. (2016). Generalized amos-type bounds for ratios of modified bessel functions. *Manuscript in preparation*. 121

[Paindaveine et al., 2018] Paindaveine, D., Remy, J., and Verdebout, T. (2018). Sign tests for weak principal directions. *arXiv preprint arXiv:1812.09367*. 11

[Paindaveine and Verdebout, 2016] Paindaveine, D. and Verdebout, T. (2016). On high-dimensional sign tests. *Bernoulli*, 22:1745–1769. 10, 11, 12, 28, 44, 60, 92, 95

[Rao and Mitra, 1971] Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverses of Matrices and its Applications*. J. Wiley, New York. 117

[Rayleigh, 1919] Rayleigh, Lord. (1919). On the problem of random vibrations and random flights in one, two and three dimensions. *Phil. Mag.*, 37:321–346. 11, 28

[Saw, 1978] Saw, J. G. (1978). A family of distributions on the m-sphere and some hypothesis tests. *Biometrika*, 65:69–73. 15

[Schou, 1978] Schou, G. (1978). Estimation of the concentration parameter in von mises-fisher distributions. *Biometrika*, 65(2):369–377. 125

[Shao, 2003] Shao, J. (2003). *Mathematical Statistics*. Springer, New York. 23

[Simpson and Spector, 1984] Simpson, H. C. and Spector, S. J. (1984). Some monotonicity results for ratios of modified bessel functions. *Quart. Appl. Math.*, 42(1):95–98. 121

[Sra and Karp, 2013] Sra, S. and Karp, D. (2013). The multivariate watson distribution: Maximum-likelihood estimation and other aspects. *J. Multivariate Anal.*, 114:256–269. 11

[Sun and Lockhart, 2019] Sun, S. Z. and Lockhart, R. A. (2019). Bayesian optimality for Beran's class of tests of uniformity around the circle. *J. Statist. Plann. Inference*, 198:79–90. 9

[Tyler, 1987a] Tyler, D. (1987a). A distribution-free m-estimator of multivariate scatter. *Ann. Statist.*, 15:234–251. 14

[Tyler, 1987b] Tyler, D. E. (1987b). Statistical analysis for the angular central gaussian distribution on the sphere. *Biometrika*, 74(3):579–589. 11

[van der Vaart, 1998] van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge Univ. Press, Cambridge. 19, 20, 21, 100

[Wang et al., 2015] Wang, L., Peng, B., and Li, R. (2015). A high-dimensional nonparametric multivariate test for mean vector. *J. Amer. Statist. Assoc.*,, page to appear. 9

[Watson, 1965] Watson, G. (1965). Equatorial distributions on a sphere. *Biometrika*, 52(1/2):193–201. 11

[Watson, 1988] Watson, G. S. (1988). The langevin distribution on high dimensional spheres. *J. Appl. Statist.*, 15:123–130. 30

[Zou et al., 2014] Zou, C., Peng, L., Feng, L., and Wang, Z. (2014). Multivariate-sign-based high-dimensional tests for sphericity. *Biometrika*, 101:229–236. 9