Testing uniformity on the sphere: from pairs to m-tuples

Alberto Fernández-de-Marcos[1], **Eduardo García-Portugués**[1], **and Thomas Verdebout**[2]

[1]Department of Statistics, Universidad Carlos III de Madrid (Spain)

[2]Department of Mathematics and ECARES, Université libre de Bruxelles (Belgium)

↑ Paper ↑

ADISTA LUXEMBOURG 10-12 SEPTEMBER

Abstract [2]

- We define a class of U and V-tests of uniformity on $\mathbb{S}^q$ with square-integrable kernels of arbitrary degree m.
- The class of m-points tests **generalizes the Sobolev class** of tests, and **outperforms** it in terms of **power** under several scenarios.
- **Asymptotic** distributions involve **random Hermite polynomials**.
- We establish **consistency against fixed alternatives** and asymptotic distributions under **local alternatives**.
- Our V-statistics can be **computed in** $O(n)$ time, regardless of m.

1 Uniformity testing on the sphere: Sobolev tests

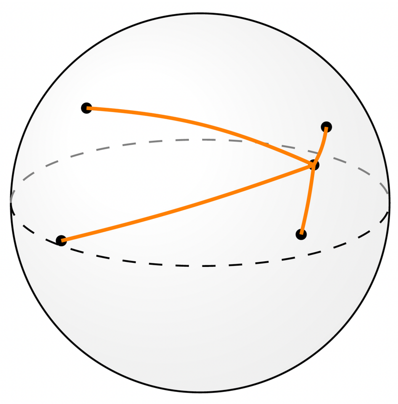
- Given a sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ on $\mathbb{S}^q$, $q \geq 1$, with $\mathbf{X} \sim \mathrm{P}$, we test:

$$\mathcal{H}_0 : \mathrm{P} = \nu_q \quad \text{vs.} \quad \mathcal{H}_1 : \mathrm{P} \neq \nu_q,$$

with ν_q being the uniform distribution on $\mathbb{S}^q$.

- The central class of **Sobolev tests** [1, 3] has been particularly influential:
- Based on **V-statistics** with a kernel $\phi \in L^2_q[-1, 1]$,

$$S_\phi^{(n)} := \frac{1}{n} \sum_{i,j=1}^{n} \phi(\mathbf{X}_i{}'\mathbf{X}_j) = \frac{1}{n} \sum_{i,j=1}^{n} \sum_{k=1}^{\infty} b_{q,k}(\phi)\, h_{q,k}(\mathbf{X}_i{}'\mathbf{X}_j),$$



representing ϕ as a Fourier series, with

$$h_{q,k}(x) := \begin{cases} \cos\left(k \cos^{-1}(x)\right), & q = 1, \\ C_k^{(q-1)/2}(x), & q > 1. \end{cases}$$

- "Average" of **distances** $\mathbf{X}_i{}'\mathbf{X}_j$.
- Depends on kernel ϕ.

2 From pairs...

- Sobolev kernels ϕ can be seen as a function Φ defined on $L^2(\mathbb{S}^q \times \mathbb{S}^q, \nu_q)$.
- **Addition formula** yields **spherical harmonics**:

$$\Phi(\mathbf{X}_1, \mathbf{X}_2) = \begin{cases} \sum_{k=1}^{\infty} b_{q,k}(\phi) \left(\cos k\theta_1 \cos k\theta_2 + \sin k\theta_1 \sin k\theta_2\right), & q = 1, \\ \sum_{k=1}^{\infty} w_k \left(\sum_{r=1}^{d_{p,k}} g_{k,r}(\mathbf{X}_1) g_{k,r}(\mathbf{X}_2)\right), & q > 1. \end{cases} \quad \mathbf{X} = (\cos\theta, \sin\theta)$$

- $w_k := b_{q,k}(\phi)(1 + 2k/(q-1))^{-1}$.
- Spherical harmonics $\{g_{k,r}\}$ are an orthonormal basis of $L^2(\mathbb{S}^q, \nu_q)$.
- For $q = 1$, $g_{k,1}(\mathbf{x}) = \sqrt{2}\cos\theta$ and $g_{k,2}(\mathbf{x}) = \sqrt{2}\sin\theta$.

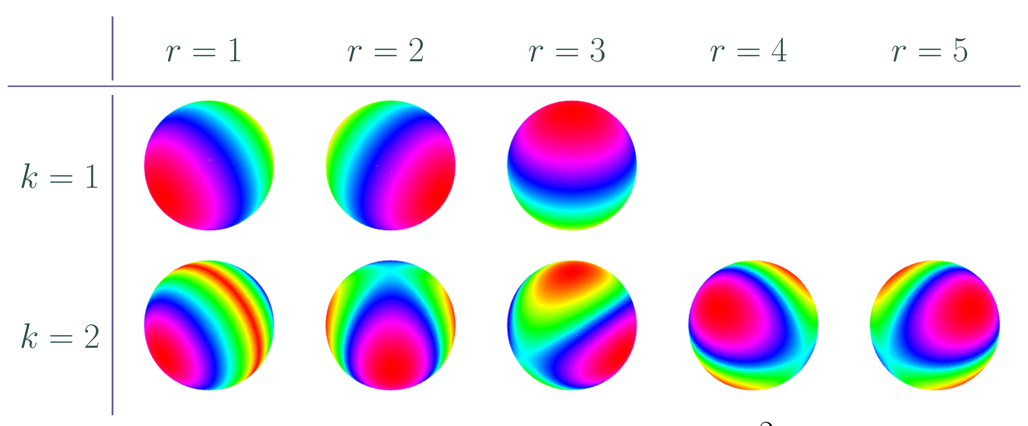


Table 1: $g_{k,r}$ for $k \in \{1, 2\}$ on $\mathbb{S}^2$.

Core idea: Build statistics with a more **flexible** kernel $\Phi \in L^2((\mathbb{S}^q)^m, \nu_q)$ to capture interactions among m-tuples of observations.

- Rely on the Fourier expansion of Φ

$$\Phi(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \sum_{k_1, \ldots, k_m = 1}^{\infty} \sum_{r_1 = 1}^{d_{q,k_1}} \cdots \sum_{r_m = 1}^{d_{q,k_m}} \overbrace{\langle \Phi, g_{k_1,r_1} \cdots g_{k_m,r_m} \rangle_m}^{w_{\mathbf{k},\mathbf{r}}} g_{k_1,r_1}(\mathbf{X}_1) \cdots g_{k_m,r_m}(\mathbf{X}_m),$$

with inner product $\langle \cdot, \cdot \rangle_m$ on $L^2((\mathbb{S}^q)^m, \nu_q)$.

- A given sequence of coefficients $w_{\mathbf{k},\mathbf{r}}$ characterizes the kernel Φ_w.
- We can **truncate** Φ up to K.
- We impose simplifications on $w_{\mathbf{k},\mathbf{r}}$: **diagonal and homogeneous** weights $(w_k)_{k=1}^{\infty}$.

3 ...to m-tuples

Definition (m-points test statistics)

Let $m \geq 2$, and $w := (w_k)_{k=1}^{\infty}$ be a real sequence such that $\sum_{k=1}^{\infty} w_k^2 d_{q,k} < \infty$. The *$m$-points test statistic* is given by the V- and U-statistics based on the kernel

$$\Phi_{m,w,K}(\mathbf{X}_1, \ldots, \mathbf{X}_m) := \sum_{k=1}^{K} w_k \sum_{r=1}^{d_{q,k}} \prod_{j=1}^{m} g_{k,r}(\mathbf{X}_j).$$

That is,

$$V_{m,w,K}^{(n)} := n^{-m/2} \sum_{i_1, \ldots, i_m = 1}^{n} \Phi_w(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_m}) \text{ and } U_{m,w,K}^{(n)} := n^{m/2} \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \Phi_w(\mathbf{X}_{i_1}, \ldots, \mathbf{X}_{i_m}).$$

- **Closed-form** expressions for classical kernels in $\mathbb{S}^1$.
- Can be **extended** to $K = \infty$ (technical).

Computational **cost independent of** m and $\mathcal{O}(n)$ in practice:

$$V_{m,w}^{(n)} = n^{-m/2} \sum_{k=1}^{K_{\max}} w_k \sum_{r=1}^{d_{q,k}} \sum_{i_1, \ldots, i_m = 1}^{n} \prod_{j=1}^{m} g_{k,r}(\mathbf{X}_{i_j}) = \sum_{k=1}^{K_{\max}} w_k \sum_{r=1}^{d_{q,k}} \left(n^{-1/2} \sum_{i=1}^{n} g_{k,r}(\mathbf{X}_i)\right)^m.$$

4 Null asymptotics

Theorem (Asymptotic distribution under $\mathcal{H}_0$)

Let $q \geq 1$, $\mathbf{X}_1, \ldots, \mathbf{X}_n$ be iid, $m \geq 2$, and w a real sequence. Under $\mathcal{H}_0$, as $n \to \infty$,

$$U_{m,w,K}^{(n)} \overset{d}{\rightsquigarrow} \sum_{k=1}^{K} \sum_{r=1}^{d_{q,k}} w_k H_m(Z_{k,r}), \quad \text{and} \quad V_{m,w,K}^{(n)} \overset{d}{\rightsquigarrow} \sum_{k=1}^{K} \sum_{r=1}^{d_{q,k}} w_k Z_{k,r}^m.$$

with $\{Z_{k,r}\}$ being independent $\mathcal{N}(0, 1)$, and H_m being Hermite polynomials.

- Extended results for $K = \infty$ involve additional conditions on the summability of w.

5 Non-null asymptotics

5.1 Fixed alternatives

$\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathrm{P}$ is an iid sample under a fixed alternative that has the density $h \in L^2(\mathbb{S}^q, \nu_q)$

$$h(\mathbf{x}) := 1 + \sum_{k=1}^{\infty} \mathbf{h}_k{}'\mathbf{g}_k(\mathbf{x}), \quad \mathbf{x} \in \mathbb{S}^q,$$

where $\mathbf{h}_k := (h_{k,1}, \ldots, h_{k,d_{q,k}})' \in \mathbb{R}^{d_{q,k}}$ and $\mathbf{g}_k := (g_{k,1}, \ldots, g_{k,d_{q,k}})'$.

Proposition (Consistency under fixed alternatives)

Let $m \geq 2$ be an even integer, and w a real sequence such that $w_k > 0$ for all $k \geq 1$. Let $\mathcal{S}_{\neq} = \{(k, r) : h_{k,r} \neq 0\}$. Assume $\mathcal{S}_{\neq}$ is non-empty. Then, under P and as $n \to \infty$:

(*i*) $U_{m,w,K}^{(n)} \overset{p}{\to} +\infty$ and $V_{m,w,K}^{(n)} \overset{p}{\to} +\infty$, for $K > \min\{k : (k, r) \in \mathcal{S}_{\neq}\}$;

(*ii*) $V_{m,w,\infty}^{(n)} \overset{p}{\to} +\infty$, provided w fulfills a certain summability condition.

5.2 Local alternatives

$\mathbf{X}_1, \ldots, \mathbf{X}_n \sim \mathrm{P}^{(n)}$ is an iid sample under a local alternative that admits the density

$$h_n(\mathbf{x}) = \frac{1}{\omega_q}\left(1 - n^{-1/2}\right) + \frac{n^{-1/2}}{\omega_q}\left\{1 + \sum_{k=1}^{\infty} \mathbf{h}_k{}'\mathbf{g}_k(\mathbf{x})\right\}, \quad \mathbf{x} \in \mathbb{S}^q.$$

Proposition (Asymptotic distribution under local alternatives)

Let $m \geq 2$, and w be a real sequence. Under $\mathrm{P}^{(n)}$, as $n \to \infty$,

$$U_{m,w,K}^{(n)} \overset{d}{\rightsquigarrow} \sum_{k=1}^{K} \sum_{r=1}^{d_{q,k}} w_k H_m\left(Z_{k,r} + h_{k,r}\right), \quad \text{and} \quad V_{m,w,K}^{(n)} \overset{d}{\rightsquigarrow} \sum_{k=1}^{K} \sum_{r=1}^{d_{q,k}} w_k \left(Z_{k,r} + h_{k,r}\right)^m.$$

6 m-points tests

How to **reject**? Depends on:

		U	V
$m = 2$		one-sided	
$m > 2$	m odd	two-sided	
	m even	two-sided	one-sided

- Sobolev tests are one-sided.
- Numerical evidence of two-tailed behavior.
- V-statistics with even m:
  - Nonnegative.
  - Alternatives shift them to the right: $\mathrm{E}_{H_n}[V_{m,w,K}^{(\infty)}] > \mathrm{E}_{\mathcal{H}_0}[V_{m,w,K}^{(\infty)}]$.

7 Numerical experiments: Power under fixed alternatives

In general, any $m > 2$ achieves **higher power** than $m = 2$.

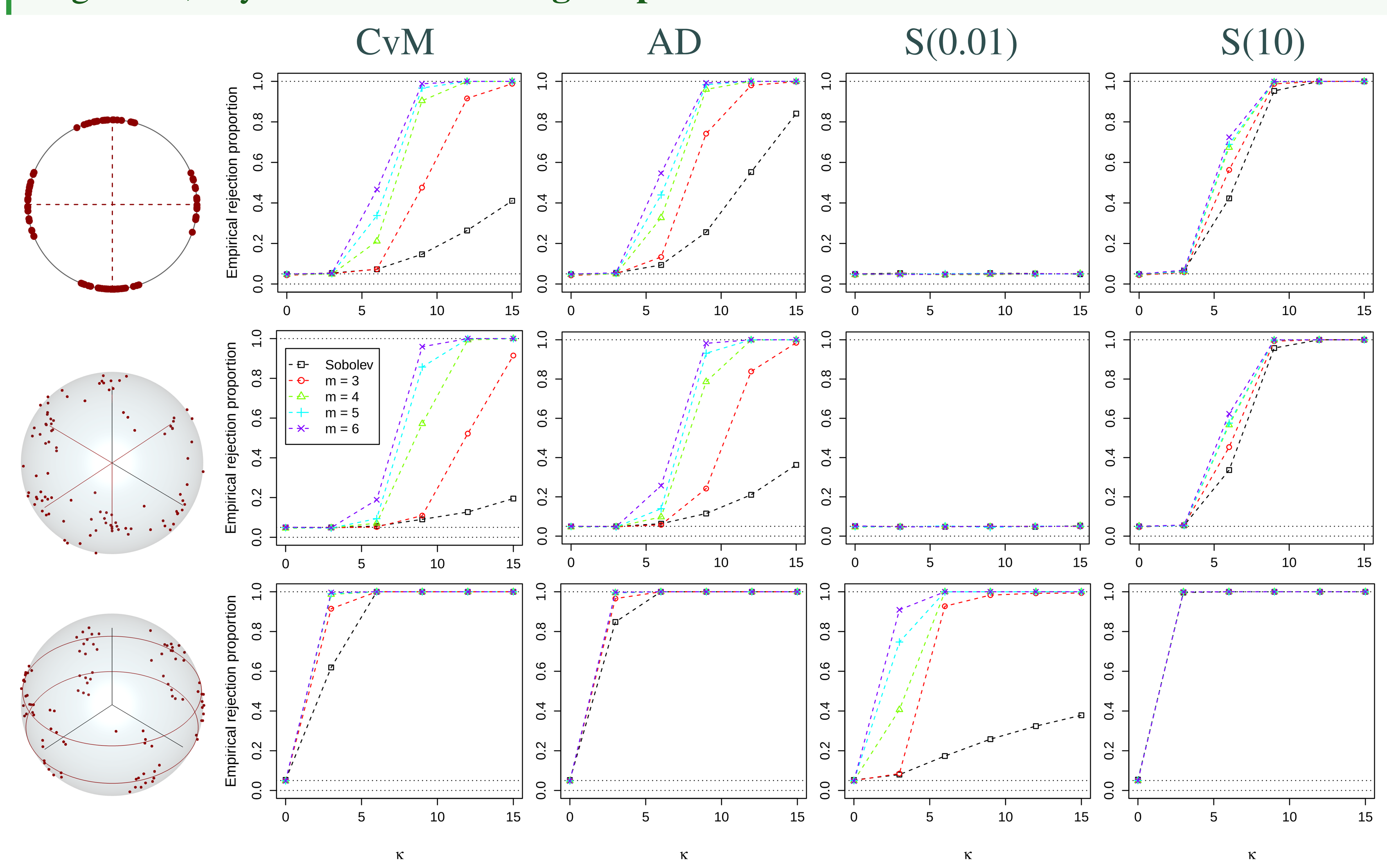


Figure 1: Power of $V_{m,w,10}$ with $m \in \{2, 3, 4, 5, 6\}$.

References

[1] Beran, R. J. (1968). Testing for uniformity on a compact homogeneous space. *J. Appl. Probab.*, 5(1):177–195.

[2] Fernández-de-Marcos, A., García-Portugués, E., and Verdebout, T. (2025). On m-points uniformity tests on hyperspheres. *Submitted*.

[3] Giné, E. (1975). Invariant tests for uniformity on compact Riemannian manifolds based on Sobolev norms. *Ann. Stat.*, 3(6):1243–1266.