Optimal tests for symmetry on the torus

Sophia Loizidou

University of Luxembourg
sophia.loizidou@uni.lu

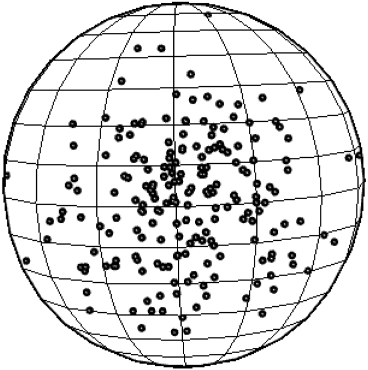
Joint work with A. Anastasiou and C. Ley

September 5, 2025

FNR PRIDE/21/16747448/MATHCODA

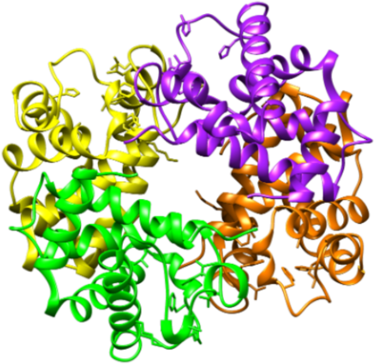
Introduction

Directional statistics deals with data on **non-linear** manifolds.

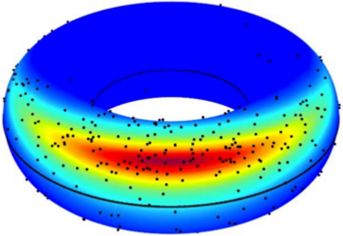


This data arises in earth sciences, meteorology, astronomy, life sciences, etc.

Protein data

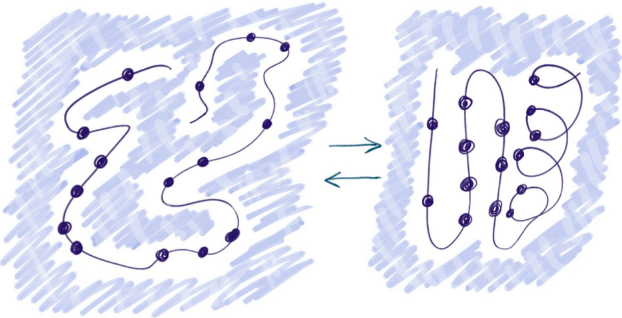


Jacobsen et al. (2023)



García-Portugués et al. (2015)

Protein structure



Jacobsen et al. (2023)

Protein structure

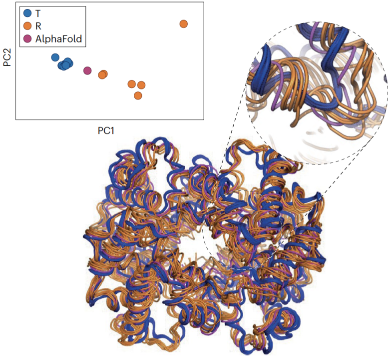


Fig. 2 | AlphaFold2's prediction of hemoglobin lies between the R and T states. Shown are experimental structures of O_2- or CO-bound (R, orange)

Lane (2023)

- Demis Hassabis and John Jumper won the Nobel Prize in Chemistry 2024 for protein structure prediction.

Protein structure

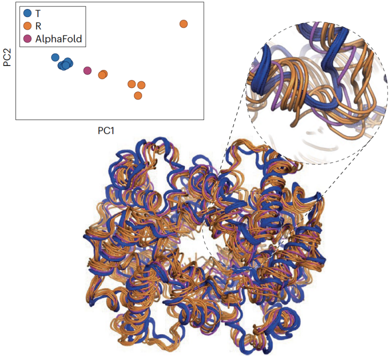


Fig. 2 | AlphaFold2's prediction of hemoglobin lies between the R and T states. Shown are experimental structures of O_2- or CO-bound (R, orange)

Lane (2023)

- Demis Hassabis and John Jumper won the Nobel Prize in Chemistry 2024 for protein structure prediction.
- Open questions in the areas of dynamics, mutants, accuracy and RNA folding.

Protein structure

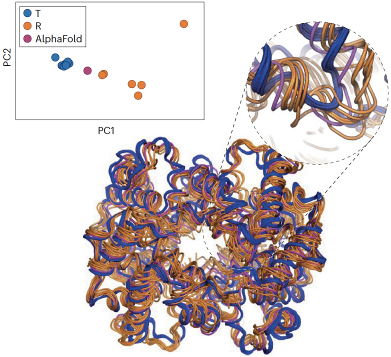


Fig. 2 | AlphaFold2's prediction of hemoglobin lies between the R and T states. Shown are experimental structures of O_2- or CO-bound (R, orange)

Lane (2023)

- Demis Hassabis and John Jumper won the Nobel Prize in Chemistry 2024 for protein structure prediction.

- Open questions in the areas of dynamics, mutants, accuracy and RNA folding.

$\Rightarrow$ Probability distributions are required.

Skewed distributions

Some of the most famous toroidal distributions in the literature share a common feature: symmetry around the location on the torus. However, several datasets are not symmetric!

Skewed distributions

Some of the most famous toroidal distributions in the literature share a common feature: symmetry around the location on the torus. However, several datasets are not symmetric!

In Ameijeiras-Alonso and Ley (2022), sine-skewed distributions are proposed. For $\boldsymbol{\theta} \in \mathbb{R}^d$ the base density $f(\boldsymbol{\theta} - \boldsymbol{\mu}; \boldsymbol{\vartheta})$ is transformed into

$$\boldsymbol{\theta} \mapsto f(\boldsymbol{\theta} - \boldsymbol{\mu}; \boldsymbol{\vartheta}) \left(1 + \sum_{j=1}^{d} \lambda_j \sin(\theta_j - \mu_j) \right),$$

where $\boldsymbol{\lambda} \in [-1, 1]^d$ plays the role of skewness parameter and satisfies $\sum_{j=1}^{d} |\lambda_j| \leq 1$.

Skewed distributions

Some of the most famous toroidal distributions in the literature share a common feature: symmetry around the location on the torus. However, several datasets are not symmetric!

In Ameijeiras-Alonso and Ley (2022), sine-skewed distributions are proposed. For $\boldsymbol{\theta} \in \mathbb{R}^d$ the base density $f(\boldsymbol{\theta} - \boldsymbol{\mu}; \boldsymbol{\vartheta})$ is transformed into

$$\boldsymbol{\theta} \mapsto f(\boldsymbol{\theta} - \boldsymbol{\mu}; \boldsymbol{\vartheta}) \left(1 + \sum_{j=1}^{d} \lambda_j \sin(\theta_j - \mu_j)\right),$$

where $\boldsymbol{\lambda} \in [-1, 1]^d$ plays the role of skewness parameter and satisfies $\sum_{j=1}^{d} |\lambda_j| \leq 1$.

Likelihood ratio tests were given in Ameijeiras-Alonso and Ley (2022), but they are parametric
$\Rightarrow$ efficient tests which are generally valid are needed.

Optimal tests for symmetry

We wish to test

$$H_0 : \boldsymbol{\lambda} = 0 \quad \text{vs} \quad H_1 : \boldsymbol{\lambda} \neq 0$$

over all possible symmetric f_0 in

$$f_{\boldsymbol{\mu},\boldsymbol{\lambda}}(\boldsymbol{\theta}) = f_0(\boldsymbol{\theta} - \boldsymbol{\mu}) \left(1 + \sum_{j=1}^{d} \lambda_j \sin(\theta_j - \mu_j)\right), \quad \boldsymbol{\lambda} \in [-1, 1]^d, \sum_{j=1}^{d} |\lambda_j| \leq 1.$$

We shall proceed in two steps:

1. build an efficient parametric test under specified f
2. render it semi-parametric

To achieve these goals, we use the Le Cam theory of asymptotic experiments.

The LAN property and limit experiments

We want to construct inferential procedures for a certain parameter $\boldsymbol{\eta}$ in a parametric family $f_{\boldsymbol{\eta}}$.

LAN property of this family in a neighborhood $\boldsymbol{\eta}_0^{(n)}$ of a fixed value $\boldsymbol{\eta}_0$:

$$L^{(X_1,\ldots,X_n)}_{\boldsymbol{\eta}_0^{(n)}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\eta}_0^{(n)};f} = \boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)}) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau} + o_P(1)$$

for $n \to \infty$ under $f_{\boldsymbol{\eta}_0}$, and $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)})$ is asymptotically normal.

The LAN property and limit experiments

We want to construct inferential procedures for a certain parameter $\boldsymbol{\eta}$ in a parametric family $f_{\boldsymbol{\eta}}$.

LAN property of this family in a neighborhood $\boldsymbol{\eta}_0^{(n)}$ of a fixed value $\boldsymbol{\eta}_0$:

$$L^{(X_1,\ldots,X_n)}_{\boldsymbol{\eta}_0^{(n)}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\eta}_0^{(n)};f} = \boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)}) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau} + o_P(1)$$

for $n \to \infty$ under $f_{\boldsymbol{\eta}_0}$, and $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)})$ is asymptotically normal.

Likelihood ratio (for a single observation $\boldsymbol{\Delta}$) of the Gaussian shift model with shift $\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau}$: $\boldsymbol{\tau}'\boldsymbol{\Delta} - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau}$.

The LAN property and limit experiments

We want to construct inferential procedures for a certain parameter $\boldsymbol{\eta}$ in a parametric family $f_{\boldsymbol{\eta}}$.

LAN property of this family in a neighborhood $\boldsymbol{\eta}_0^{(n)}$ of a fixed value $\boldsymbol{\eta}_0$:

$$L^{(X_1,\dots,X_n)}_{\boldsymbol{\eta}_0^{(n)}+n^{-1/2}\boldsymbol{\tau}/\boldsymbol{\eta}_0^{(n)};f} = \boldsymbol{\tau}'\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)}) - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau} + o_P(1)$$

for $n \to \infty$ under $f_{\boldsymbol{\eta}_0}$, and $\boldsymbol{\Delta}_f^{(n)}(\boldsymbol{\eta}_0^{(n)})$ is asymptotically normal.

Likelihood ratio (for a single observation $\boldsymbol{\Delta}$) of the Gaussian shift model with shift $\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau}$: $\boldsymbol{\tau}'\boldsymbol{\Delta} - \frac{1}{2}\boldsymbol{\tau}'\boldsymbol{\Gamma}_f(\boldsymbol{\eta}_0)\boldsymbol{\tau}$.

Similarities, locally (in $\boldsymbol{\eta}$) and asymptotically (in n), with the Gaussian shift model!

$\Rightarrow$ import optimal procedures from the (well-known) Gaussian world to the initial setting in order to obtain (asymptotically) optimal parametric procedures!

The (Uniformly)LAN property

Under very mild regularity assumptions on f_0 we get

Property

The sine-skewed family is ULAN at $(\boldsymbol{\mu}^{(n)}, \mathbf{0}) = (\mu_1^{(n)}, \cdots, \mu_d^{(n)}, 0, \cdots, 0)$. *More precisely, for any* $\mu_i^{(n)} = \mu_i + O(n^{-1/2}), i = 1, \ldots, d$, *and for any bounded sequence* $\boldsymbol{\tau}^{(n)} := (\boldsymbol{\tau}_1^{(n)}, \cdots, \boldsymbol{\tau}_{2d}^{(n)})' \in \mathbb{R}^{2d}$, *we have*

$$\Lambda^{(n)} := \log\left(\frac{P^{(n)}_{(\boldsymbol{\mu}^{(n)},\mathbf{0})+n^{-1/2}\boldsymbol{\tau}^{(n)};f_0}}{P^{(n)}_{(\boldsymbol{\mu}^{(n)},\mathbf{0});f_0}}\right) = \boldsymbol{\tau}^{(n)\prime}\Delta^{(n)}_{f_0}(\boldsymbol{\mu}^{(n)}) - \frac{1}{2}\boldsymbol{\tau}^{(n)\prime}\Gamma_{f_0}\boldsymbol{\tau}^{(n)} + o_P(1)$$

and the central sequence

$$\Delta^{(n)}_{f_0}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\phi_1^{f_0}(\boldsymbol{\theta}-\boldsymbol{\mu})\\ \vdots \\ \phi_d^{f_0}(\boldsymbol{\theta}-\boldsymbol{\mu})\\ \sin(\theta_{1i}-\mu_1)\\ \vdots \\ \sin(\theta_{di}-\mu_d)\end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\mathbf{0}, \Gamma_{f_0}\right),$$

all under H_0 *when* $n \to \infty$, *where* $\phi_j^{f_0}(\boldsymbol{\theta}-\boldsymbol{\mu}) = -\frac{\frac{\partial}{\partial\theta_j}f_0(\boldsymbol{\theta}-\boldsymbol{\mu})}{f_0(\boldsymbol{\theta}-\boldsymbol{\mu})}$, *for* $j \in \{1, \ldots, d\}$, *and* Γ_{f_0} *is the Fisher Information matrix.*

Efficient tests around a specified symmetry center

For $I^{f_0}_{\lambda_j\lambda_k} = \int_{[-\pi,\pi)^d} \sin(\theta_j - \mu_j)\sin(\theta_k - \mu_k) f_0(\boldsymbol{\theta} - \boldsymbol{\mu}) d\boldsymbol{\theta}$ define

$$\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \sin(\theta_{1i} - \mu_1) \\ \vdots \\ \sin(\theta_{di} - \mu_d) \end{pmatrix}, \quad \Gamma_{f_0;\lambda} = \begin{pmatrix} I^{f_0}_{\lambda_1\lambda_1} & \dots & I^{f_0}_{\lambda_1\lambda_d} \\ \vdots & \ddots & \vdots \\ I^{f_0}_{\lambda_1\lambda_d} & \dots & I^{f_0}_{\lambda_d\lambda_d} \end{pmatrix}.$$

The parametric test statistic is given by

$$Q^{(n);\boldsymbol{\mu}}_{f_0} = \left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu})\right)^{\mathsf{T}} \left(\Gamma_{f_0;\lambda}\right)^{-1} \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}).$$

Efficient tests around a specified symmetry center

For $I^{f_0}_{\lambda_j\lambda_k} = \int_{[-\pi,\pi)^d} \sin(\theta_j - \mu_j)\sin(\theta_k - \mu_k) f_0(\boldsymbol{\theta} - \boldsymbol{\mu}) d\boldsymbol{\theta}$ define

$$\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \sin(\theta_{1i} - \mu_1) \\ \vdots \\ \sin(\theta_{di} - \mu_d) \end{pmatrix}, \quad \Gamma_{f_0;\lambda} = \begin{pmatrix} I^{f_0}_{\lambda_1\lambda_1} & \dots & I^{f_0}_{\lambda_1\lambda_d} \\ \vdots & \ddots & \vdots \\ I^{f_0}_{\lambda_1\lambda_d} & \dots & I^{f_0}_{\lambda_d\lambda_d} \end{pmatrix}.$$

The parametric test statistic is given by

$$Q^{(n);\boldsymbol{\mu}}_{f_0} = \left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu})\right)^{\mathsf{T}} \left(\Gamma_{f_0;\lambda}\right)^{-1} \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}).$$

The semi-parametric test statistic is

$$Q^{*(n);\boldsymbol{\mu}} = \left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu})\right)^{\mathsf{T}} \left(\tilde{\Gamma}_{\lambda}\right)^{-1} \Delta^{(n)}_{\lambda}(\boldsymbol{\mu})$$

for $\tilde{\Gamma}_{\lambda} = \begin{pmatrix} \tilde{I}_{\lambda_1\lambda_1} & \dots & \tilde{I}_{\lambda_1\lambda_d} \\ \vdots & \ddots & \vdots \\ \tilde{I}_{\lambda_1\lambda_d} & \dots & \tilde{I}_{\lambda_d\lambda_d}, \end{pmatrix}$ where $\tilde{I}_{\lambda_j\lambda_k} = \frac{1}{n}\sum_{i=1}^{n} \sin(\theta_{ji} - \mu_j)\sin(\theta_{ki} - \mu_k)$.

Asymptotic results

Denoting by $P^{(n)}_{(\boldsymbol{\mu},\boldsymbol{\lambda})',f_0}$ the joint distribution of $\{\boldsymbol{\theta}_i\}_{i=1}^n$, we show that:

- It holds that under $\cup_{f_0\in\mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},\mathbf{0})',f_0}$,

$$Q^{*(n);\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} \chi^2_d$$

  as $n \to \infty$.

Asymptotic results

Denoting by $P^{(n)}_{(\boldsymbol{\mu},\boldsymbol{\lambda})',f_0}$ the joint distribution of $\{\boldsymbol{\theta}_i\}_{i=1}^n$, we show that:

- It holds that under $\cup_{f_0 \in \mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},\mathbf{0})',f_0}$,

$$Q^{*(n);\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} \chi^2_d$$

  as $n \to \infty$.

- It holds that under $P^{(n)}_{(\boldsymbol{\mu},n^{-1/2}\boldsymbol{\tau}^{(n)})',f_0}$,

$$Q^{*(n);\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} \chi^2_d(\boldsymbol{\tau}^\intercal \Gamma_{f_0;\lambda} \boldsymbol{\tau})$$

  as $n \to \infty$ with $\boldsymbol{\tau} = \lim_{n\to\infty}(\tau^{(n)}_{d+1}, \ldots, \tau^{(n)}_{2d})$.

Asymptotic results

Denoting by $P^{(n)}_{(\boldsymbol{\mu},\boldsymbol{\lambda})',f_0}$ the joint distribution of $\{\boldsymbol{\theta}_i\}_{i=1}^n$, we show that:

- It holds that under $\cup_{f_0\in\mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},\mathbf{0})',f_0}$,

$$Q^{*(n);\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} \chi^2_d$$

  as $n \to \infty$.

- It holds that under $P^{(n)}_{(\boldsymbol{\mu},n^{-1/2}\boldsymbol{\tau}^{(n)})',f_0}$,

$$Q^{*(n);\boldsymbol{\mu}} \xrightarrow{\mathcal{D}} \chi^2_d(\boldsymbol{\tau}^\intercal \Gamma_{f_0;\lambda}\boldsymbol{\tau})$$

  as $n \to \infty$ with $\boldsymbol{\tau} = \lim_{n\to\infty}(\tau^{(n)}_{d+1}, \ldots, \tau^{(n)}_{2d})$.

- The test is uniformly (in f_0) locally and asymptotically maximin optimal against sine-skewed alternatives.

Error bounds using Stein's method

What about a limited sample size?

Error bounds using Stein's method

What about a limited sample size? For $h \in C_b^6(\mathbb{R})$,

$$\left| \mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right) \right|$$

Error bounds using Stein's method

What about a limited sample size? For $h \in C_b^6(\mathbb{R})$,

$$\left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| \leq \left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}})\right| \tag{1}$$

$$+ \left|\mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| \tag{2}$$

Error bounds using Stein's method

What about a limited sample size? For $h \in C_b^6(\mathbb{R})$,

$$\left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| \leq \left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}})\right| \tag{1}$$

$$+ \left|\mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| \tag{2}$$

- (2) can be bounded using Theorem 2.4 of Gaunt and Reinert (2023)

$$(2) \leq C_1/n$$

for a constant C_1.

Error bounds using Stein's method

What about a limited sample size? For $h \in C_b^6(\mathbb{R})$,

$$\begin{aligned}\left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q^{*(n);\boldsymbol{\mu}}) - \mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}})\right| && (1)\\ &\quad + \left|\mathbb{E}h(Q_{f_0}^{(n);\boldsymbol{\mu}}) - \mathbb{E}h\left(\chi_d^2\right)\right| && (2)\\ &\leq C/\sqrt{n}\end{aligned}$$

- (2) can be bounded using Theorem 2.4 of Gaunt and Reinert (2023)

$$(2) \leq C_1/n$$

  for a constant C_1.

- (1) requires much more work

$$(1) \leq C_2/\sqrt{n}$$

  for a constant C_2.

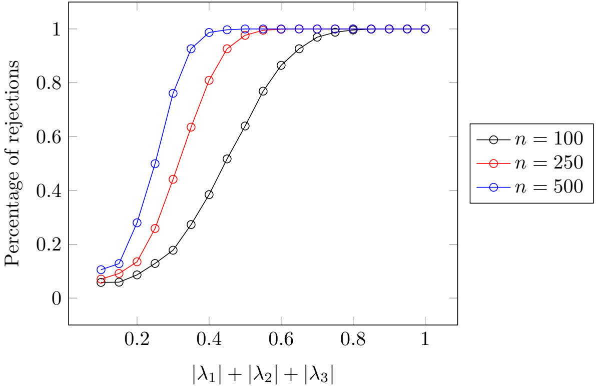
Simulations

Simulations framework: increasing $\|\boldsymbol{\lambda}\|_1$

- The percentage of rejections of the test for increasing $\|\boldsymbol{\lambda}\|_1$ at the 5% level of significance is plotted.
- 5000 samples were generated for each value of $\|\boldsymbol{\lambda}\|_1$.
- $\lambda_1 = \lambda_2 = 0.05$ and λ_3 is increasing.

Simulations framework: increasing $\|\boldsymbol{\lambda}\|_1$

- The percentage of rejections of the test for increasing $\|\boldsymbol{\lambda}\|_1$ at the 5% level of significance is plotted.
- 5000 samples were generated for each value of $\|\boldsymbol{\lambda}\|_1$.
- $\lambda_1 = \lambda_2 = 0.05$ and λ_3 is increasing.



Symmetry tests for unspecified location

Symmetry tests for unspecified location - parametric test

We need a new central sequence that is orthogonal to

$$\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \phi_1^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \\ \vdots \\ \phi_d^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \end{pmatrix}.$$

Symmetry tests for unspecified location - parametric test

We need a new central sequence that is orthogonal to

$$\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \phi_1^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \\ \vdots \\ \phi_d^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \end{pmatrix}.$$

This can be calculated by

$$\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right) \mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1} \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}).$$

Symmetry tests for unspecified location - parametric test

We need a new central sequence that is orthogonal to

$$\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \begin{pmatrix} \phi_1^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \\ \vdots \\ \phi_d^{f_0}(\boldsymbol{\theta} - \boldsymbol{\mu}) \end{pmatrix}.$$

This can be calculated by

$$\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right) \mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1} \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}).$$

For estimators $\hat{\boldsymbol{\mu}}^{(n)}$ of $\boldsymbol{\mu}$ that satisfy mild assumptions, the parametric test is given by:

$$Q^{(n)}_{f_0} = \left(\Delta^{(n)*}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{\mathsf{T}} \mathrm{Var}\left(\Delta^{(n)*}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{-1} \left(\Delta^{(n)*}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)})\right).$$

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)?

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)? Then,

$$\mathrm{Cov}_{g_0}\left(\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right) \neq 0.$$

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)? Then,

$$\mathrm{Cov}_{g_0}\left(\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right) \neq 0.$$

Solution: re-define the central sequence

$$\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)\mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})$$

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)? Then,

$$\mathrm{Cov}_{g_0}\left(\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right) \neq 0.$$

Solution: re-define the central sequence

$$\begin{array}{c}
\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \color{red}{\mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)}\color{blue}{\mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1}}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) \\
\color{red}{\downarrow} \qquad\qquad \color{blue}{\downarrow} \\
\tilde{\Delta}^{(n)}_{\lambda;f_0;g_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \color{red}{\mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)}\color{blue}{\mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)^{-1}}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})
\end{array}$$

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)? Then,

$$\mathrm{Cov}_{g_0}\left(\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right) \neq 0.$$

Solution: re-define the central sequence

$$\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)\mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})$$

$$\downarrow \qquad \downarrow$$

$$\tilde{\Delta}^{(n)}_{\lambda;f_0;g_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)\mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)^{-1}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})$$

$$\downarrow \qquad \downarrow$$

$$\tilde{\Delta}^{*(n)}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)}) = \Delta^{(n)}_{\lambda}(\hat{\boldsymbol{\mu}}^{(n)}) - \tilde{I}^{f_0}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}^{(n)})\left(\tilde{I}^{f_0}_{\boldsymbol{\mu}\boldsymbol{\lambda}}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{-1}\Delta^{(n)}_{\mu;f_0}(\hat{\boldsymbol{\mu}}^{(n)})$$

Semi-parametric test

The expression was calculated assuming f_0. What if the true underlying density is g_0 (that satisfies the same assumptions)? Then,

$$\mathrm{Cov}_{g_0}\left(\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right) \neq 0.$$

Solution: re-define the central sequence

$$\begin{array}{c}
\Delta^{(n)*}_{\lambda;f_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \color{red}{\mathrm{Cov}_{f_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)}\color{blue}{\mathrm{Var}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu})\right)^{-1}}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) \\
\color{red}{\downarrow} \qquad\qquad \color{blue}{\downarrow} \\
\tilde{\Delta}^{(n)}_{\lambda;f_0;g_0}(\boldsymbol{\mu}) = \Delta^{(n)}_{\lambda}(\boldsymbol{\mu}) - \color{red}{\mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\lambda}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)}\color{blue}{\mathrm{Cov}_{g_0}\left(\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}), \Delta^{(n)}_{\mu;g_0}(\boldsymbol{\mu})\right)^{-1}}\Delta^{(n)}_{\mu;f_0}(\boldsymbol{\mu}) \\
\color{red}{\downarrow} \qquad\qquad \color{blue}{\downarrow} \\
\tilde{\Delta}^{*(n)}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)}) = \Delta^{(n)}_{\lambda}(\hat{\boldsymbol{\mu}}^{(n)}) - \color{red}{\tilde{I}^{f_0}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\hat{\boldsymbol{\mu}}^{(n)})}\color{blue}{\left(\tilde{I}^{f_0}_{\boldsymbol{\mu}\boldsymbol{\lambda}}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{-1}}\Delta^{(n)}_{\mu;f_0}(\hat{\boldsymbol{\mu}}^{(n)})
\end{array}$$

The test statistic is given by

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) = \left(\tilde{\Delta}^{*(n)}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{\intercal}\left(\hat{C}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)})\right)^{-1}\tilde{\Delta}^{*(n)}_{\lambda;f_0}(\hat{\boldsymbol{\mu}}^{(n)}).$$

Semi-parametric test

Asymptotic results

- Under $\cup_{\boldsymbol{\mu}\in[-\pi,\pi)^d} \cup_{g_0\in\mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},0);g_0}$ as $n \to \infty$

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) \xrightarrow{\mathcal{D}} \chi^2_d.$$

Semi-parametric test

Asymptotic results

- Under $\cup_{\boldsymbol{\mu}\in[-\pi,\pi)^d} \cup_{g_0\in\mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},0);g_0}$ as $n \to \infty$

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) \xrightarrow{\mathcal{D}} \chi^2_d.$$

- Under $\cup_{\mu\ \in[-\pi,\pi)^d} P^{(n)}_{(\boldsymbol{\mu},n^{-1/2}\boldsymbol{\tau}^{(n)})',g_0}$

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) \xrightarrow{\mathcal{D}} \chi^2_d\left(\boldsymbol{\tau}^{\intercal} C^{f_0}_{g_0}(\boldsymbol{\mu}) V^{f_0}_{g_0}(\boldsymbol{\mu})^{-1} C^{f_0}_{g_0}(\boldsymbol{\mu})\boldsymbol{\tau}\right)$$

as $n \to \infty$ with

$$V^{f_0}_{g_0}(\boldsymbol{\mu}) = \mathrm{Var}_{g_0}\left(\tilde{\Delta}_{\lambda;f_0;g_0}(\boldsymbol{\mu})\right)$$

and

$$C^{f_0}_{g_0}(\boldsymbol{\mu}) = \mathrm{Cov}_{g_0}\left(\tilde{\Delta}_{\lambda;f_0;g_0}(\boldsymbol{\mu}), \Delta_{\lambda}(\boldsymbol{\mu})\right).$$

Semi-parametric test

Asymptotic results

- Under $\cup_{\boldsymbol{\mu}\in[-\pi,\pi)^d} \cup_{g_0\in\mathcal{F}} P^{(n)}_{(\boldsymbol{\mu},0);g_0}$ as $n \to \infty$

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) \xrightarrow{\mathcal{D}} \chi^2_d.$$

- Under $\cup_{\mu\ \in[-\pi,\pi)^d} P^{(n)}_{(\boldsymbol{\mu},n^{-1/2}\boldsymbol{\tau}^{(n)})',g_0}$

$$Q^{*(n)}_{f_0}(\hat{\boldsymbol{\mu}}^{(n)}) \xrightarrow{\mathcal{D}} \chi^2_d\left(\boldsymbol{\tau}^{\intercal} C^{f_0}_{g_0}(\boldsymbol{\mu}) V^{f_0}_{g_0}(\boldsymbol{\mu})^{-1} C^{f_0}_{g_0}(\boldsymbol{\mu})\boldsymbol{\tau}\right)$$

  as $n \to \infty$ with

$$V^{f_0}_{g_0}(\boldsymbol{\mu}) = \mathrm{Var}_{g_0}\left(\tilde{\Delta}_{\lambda;f_0;g_0}(\boldsymbol{\mu})\right)$$

  and

$$C^{f_0}_{g_0}(\boldsymbol{\mu}) = \mathrm{Cov}_{g_0}\left(\tilde{\Delta}_{\lambda;f_0;g_0}(\boldsymbol{\mu}), \Delta_{\lambda}(\boldsymbol{\mu})\right).$$

- The test is locally and asymptotically maximin when testing $\mathcal{H}_0$ against $\cup_{\mu\ \in[-\pi,\pi)^d} P^{(n)}_{(\boldsymbol{\mu},n^{-1/2}\boldsymbol{\tau}^{(n)})',f_0}$.

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right|$$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\begin{aligned}\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| \leq &\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\ &+ \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\ &+ \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right|\end{aligned}$$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\begin{aligned}
\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right| \\
&\leq \qquad\qquad M_3/n
\end{aligned}$$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$
\begin{aligned}
\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right| \\
&\leq \qquad M_2/\sqrt{n} \ + \ M_3/n
\end{aligned}
$$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\begin{aligned}\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\ &\quad + \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\ &\quad + \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right| \\ &\leq M_2/\sqrt{n} + M_3/n\end{aligned}$$

Further assumption:

Assumption

$$\mathbb{E}_{g_0}\left[(\hat{\mu}_j^{(n)} - \mu_j)^2\right] = O\left(\tfrac{1}{n}\right)$$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\begin{aligned}\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\ &\quad + \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\ &\quad + \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right| \\ &\leq M_1/\sqrt{n} \;+\; M_2/\sqrt{n} \;+\; M_3/n\end{aligned}$$

Further assumption:

Assumption

$\mathbb{E}_{g_0}\left[(\hat{\mu}_j^{(n)} - \mu_j)^2\right] = O\left(\frac{1}{n}\right)$

Error bounds using Stein's method

For $h \in C_b^6(\mathbb{R})$, and constants M_1, M_2, M_3 and M,

$$\begin{aligned}
\left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h\left(\chi_d^2\right)\right| &\leq \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\hat{\boldsymbol{\mu}}^{(n)})) - \mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{*(n)}(\boldsymbol{\mu})) - \mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))\right| \\
&\quad + \left|\mathbb{E}h(Q_{f_0}^{(n)}(\boldsymbol{\mu}))) - \mathbb{E}h\left(\chi_d^2\right)\right| \\
&\leq M_1/\sqrt{n} \; + \; M_2/\sqrt{n} \; + \; M_3/n \\
&\leq M/\sqrt{n}
\end{aligned}$$

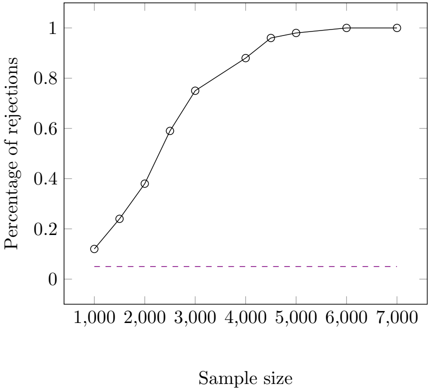
Further assumption:

Assumption

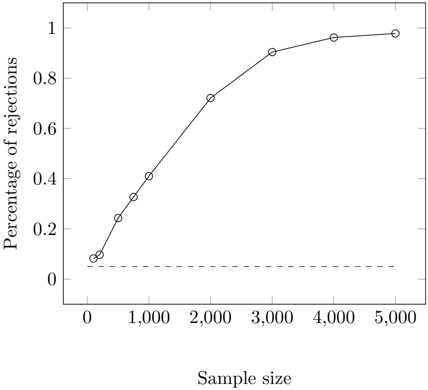
$\mathbb{E}_{g_0}\left[(\hat{\mu}_j^{(n)} - \mu_j)^2\right] = O\left(\frac{1}{n}\right)$

Simulations: increasing n

$$f_0 = f_{bwc}(\cdot; \kappa_1 = 0.5, \kappa_2 = 0.5, \rho = 0.7)$$
$$g_0 = f_{bwc}(\cdot; \kappa_1 = 0.5, \kappa_2 = 0.5, \rho = 0.7)$$
$$\lambda_1 = 0.15, \lambda_2 = 0.15$$



$$f_0 = f_{wc}(\cdot; \rho_1 = 0.1) f_{wc}(\cdot; \rho_2 = 0.1)$$
$$g_0 = f_{bwc}(\cdot; \kappa_1 = 0.5, \kappa_2 = 0.5, \rho = 0.3)$$
$$\lambda_1 = 0.15, \lambda_2 = 0.15$$



Real data

- We consider the molecular dynamic trajectory of the SARS-CoV-2 spike domain, in a position where an α-helix occurs $\Rightarrow$ unimodal data.

- It is a known fact in biology that the α helix occurs when the consecutive $(\phi,\ \psi)$ angle pairs are around (-60,-50) (Tooze (1998)).

- In this case $\omega = 0$. So we check for symmetry around $(-60, -50, 0)$ and for an unspecified symmetry center.

Real data

- We consider the molecular dynamic trajectory of the SARS-CoV-2 spike domain, in a position where an α-helix occurs $\Rightarrow$ unimodal data.

- It is a known fact in biology that the α helix occurs when the consecutive (ϕ, ψ) angle pairs are around (-60,-50) (Tooze (1998)).

- In this case $\omega = 0$. So we check for symmetry around $(-60, -50, 0)$ and for an unspecified symmetry center.

- Both tests reject the null hypothesis of symmetry at all commonly used levels of significance.

Thank you!

sophia.loizidou@uni.lu

References

References I

Ameijeiras-Alonso, J. and Ley, C. (2022). Sine-skewed toroidal distributions and their application in protein bioinformatics. *Biostatistics*, 23(3):685–704.

García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2015). Central Limit Theorems for Directional and Linear Random Variables with Applications. *Statistica Sinica*, 25(3):1207–1229. Publisher: Institute of Statistical Science, Academia Sinica.

Gaunt, R. E. and Reinert, G. (2023). The rate of convergence of some asymptotically chi-square distributed statistics by Stein's method. arXiv:1603.01889 [math].

Jacobsen, A., van Dijk, E., Mouhib, H., Stringer, B., Ivanova, O., Gavaldá-Garciá, J., Hoekstra, L., Feenstra, K. A., and Abeln, S. (2023). Introduction to Protein Structure. arXiv:2307.02169 [q-bio].

Lane, T. J. (2023). Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173. Number: 2 Publisher: Nature Publishing Group.

Tooze, John, C. I. B. (1998). *Introduction to Protein Structure*. Garland Science, New York, 2 edition.