Non-standard Hypotheses in Directional Statistics

SPP Seminar

V. Meurice

Université libre de Bruxelles

2023

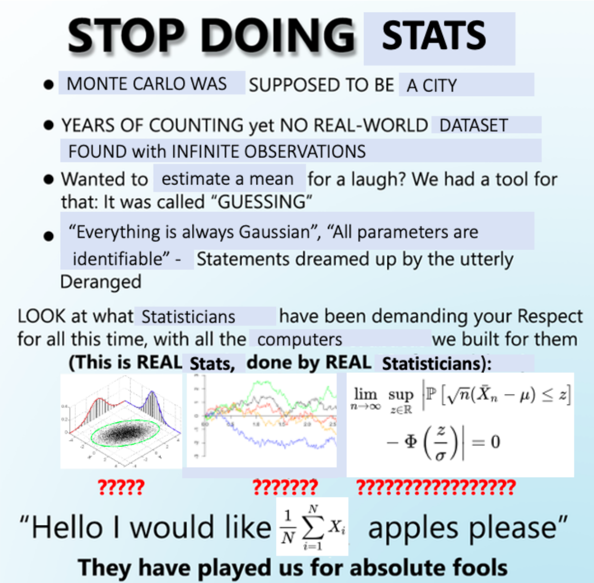
Table of Contents

1. (Non-)standard hypotheses in classical statistics
2. About directional statistics
3. Standard hypotheses in directional statistics
4. $\mathcal{E}$-symmetry
5. Weak identifiability

Table of Contents

1. (Non-)standard hypotheses in classical statistics
2. About directional statistics
3. Standard hypotheses in directional statistics
4. $\mathcal{E}$-symmetry
5. Weak identifiability

STOP DOING STATISTICS



Standard hypotheses in classical statistics

Many classical methods are built upon *standard* hypotheses, such as:

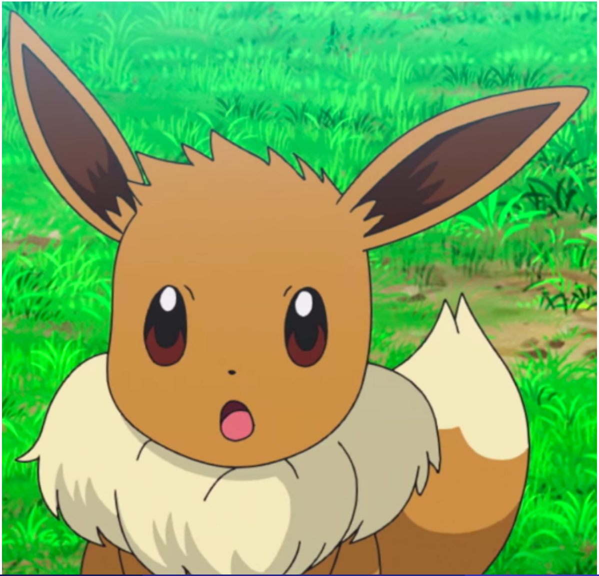
- The data is normally distributed
- All observations are independent and identically distributed (iid)
- All parameters of interest can be properly identified

Violating one or multiple of those hypotheses would then often invalidate the results of the statistical analysis at hand.

In real life however, things might not appear so well behaved...

Real life be like

Statistician finding out his data is not normally distributed be like:



Relaxed hypotheses in classical statistics

A lot of theory has now been developed based on weaker hypotheses, such as:

- The data is distributed (a)symmetrically around the mean
- Observations are weakly correlated
- Individuals have the same mean but possibly different variances (heteroskedasticity)

Practitioners are (finally!) beginning to recognise those imperfect situations, and to use more specific methods accordingly.

The field of directional statistics being newer and a lot more niche, standard assumptions still prevail considerably.

Table of Contents

1 (Non-)standard hypotheses in classical statistics

2 About directional statistics

3 Standard hypotheses in directional statistics

4 $\mathcal{E}$-symmetry

5 Weak identifiability

Directional statistics

In directional statistics, we consider realisations of p-dimensional random variables whose values lie on $p-1$-(hyper)spheres; i.e., any variable $\boldsymbol{X}$ takes values in

$$\mathbb{S}^{p-1} := \{\boldsymbol{x} \in \mathbb{R}^p : ||\boldsymbol{x}|| = 1\}.$$

The case $p = 2$ (circular data) is usually treated separately.

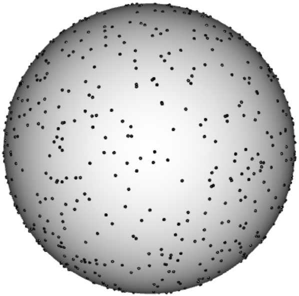


Figure: Uniformly distributed data on a 2-sphere.

Mean direction

The expected value of a spherical random variable usually does not lie on the sphere, but inside of it.

One is therefore often more interested in doing inference on the *mean direction*, that is, the normed expectation $\frac{\mathbb{E}[\boldsymbol{X}]}{||\mathbb{E}[\boldsymbol{X}]||}$, often called $\boldsymbol{\theta}$.

We then focus on how the data is distributed around that pole, with questions such as:

- How concentrated is it around the mean?
- Is the spread symmetrical in some way?

The Fisher-von Mises-Langevin (FvML) distribution)

The spherical equivalent of the Gaussian distribution is the Fisher-von Mises-Langevin distribution.

A variable $\boldsymbol{X} \sim \text{FvML}(\boldsymbol{\theta}, \kappa, p)$ has density

$$f(\boldsymbol{x}) = C_{p,\kappa} \exp(\kappa \boldsymbol{x}' \boldsymbol{\theta})$$

where $C_{p,\kappa}$ is a normalisation constant (this depends on the measure used, Lebesgue or sphere surface area).

$\boldsymbol{\theta}$ is a location parameter (the mean direction), and κ is a concentration parameter around $\boldsymbol{\theta}$. Note that $\kappa = 0$ induces a uniform distribution on the sphere.

Rotational symmetry of the FvML distribution

The FvML density only depends on $\boldsymbol{x}'\boldsymbol{\theta}$, that is, on the *angle* formed with the mean direction $\boldsymbol{\theta}$. It is therefore rotationally symmetric around $\boldsymbol{\theta}$.

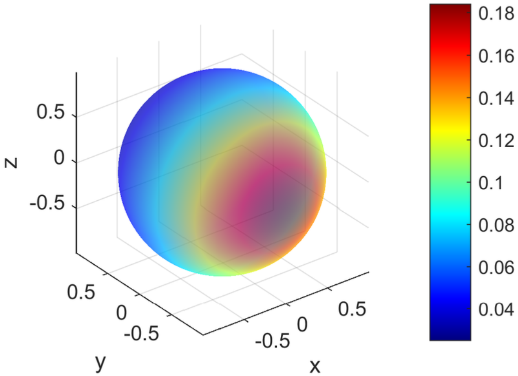


Figure: FvML density with $\boldsymbol{\theta} = (1, 0, 0)$.

Table of Contents

1. (Non-)standard hypotheses in classical statistics
2. About directional statistics
3. Standard hypotheses in directional statistics
4. $\mathcal{E}$-symmetry
5. Weak identifiability

Standard hypotheses

Similarly as for general multivariate data, one often imposes quite strict assumptions when dealing with directional data. Some of them are the following:

- Everything is i.i.d FvML.
- At the very least, the data follows a rotationally symmetric distribution around its mean.
- The concentration around $\boldsymbol{\theta}$ is strong.
- All parameters can be properly identified, and consistently estimated asymptotically.

Real life be like

Statistician finding out his data is not FvML-distributed be like:

Examples given

In the sequel, we will focus on two hypotheses that we would want to relax:

- Rotational symmetry
- Easy identification of $\boldsymbol{\theta}$

Table of Contents

1. (Non-)standard hypotheses in classical statistics
2. About directional statistics
3. Standard hypotheses in directional statistics
4. $\mathcal{E}$-symmetry
5. Weak identifiability

Rotationally symmetric distributions I

While ideally, one would always prefer to deal with FvML data, this may be too strict of an assumption to make.

It is useful to define the set of rotationally symmetric distributions, as the one with densities of the form

$$f(\boldsymbol{x}) = c_{p,\kappa,f}\, f(\kappa \boldsymbol{x}'\boldsymbol{\theta}),$$

with location parameter $\boldsymbol{\theta} \in \mathbb{S}^{p-1}$ and concentration parameter $\kappa \geq 0$, where $c_{p,\kappa,f}$ is a normalisation constant and where $f : \mathbb{R} \to \mathbb{R}_0^+$ is (strictly) monotone increasing.

The distribution is then invariant to all orthogonal transformations around $\boldsymbol{\theta}$.

Rotationally symmetric distributions II

Considering rotationally symmetric densities is a nice way to avoid focusing on FvML, and generally works well, especially in an asymptotic setting.

That being said, real-life phenomena have no reason to be symmetric around their mean...

There are a lot of ways to relax such an hypothesis, and of them is to define so-called $\mathcal{E}$-symmetric distributions.

The idea is to transform random vectors generated by a rotationally-symmetric distribution, by stretching the ambient space along the axes of a chosen basis, and then norming the resulting vector.

Definition: $\mathcal{E}$-symmetry

Suppose the random vector $\boldsymbol{y} \in \mathbb{S}^{p-1}$ follows a rotationally-symmetric distribution $G_{\boldsymbol{\theta}}$ centered around $\boldsymbol{\theta} \in \mathbb{S}^{p-1}$, with density $g(\boldsymbol{y})$, where $\boldsymbol{\theta}$ is the first vector of an orthonormal basis $\{\boldsymbol{\theta}, \gamma_2, \ldots, \gamma_p\}$.
Let $\boldsymbol{Q} = (\boldsymbol{\theta}, \gamma_2, \ldots, \gamma_p)$ and $\boldsymbol{\Lambda} = \text{diag}(a_1^2, \ldots, a_p^2)$, with $a_1 > 0$, $a_2 \geq a_3 \geq \ldots \geq a_p > 0$ and $\prod_{j=2}^p a_j = 1$.
Then the random vector $\boldsymbol{x}$ given by

$$\boldsymbol{x} := \frac{\boldsymbol{T}\boldsymbol{y}}{||\boldsymbol{T}\boldsymbol{y}||}$$

where $\boldsymbol{T} = \boldsymbol{Q}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{Q}'$, follows an $\mathcal{E}$-symmetric distribution.

Note that in the case of the canonical basis, we would get

$$\boldsymbol{x} := \frac{(a_1 y_1, a_2 y_2, \ldots, a_p y_p)}{\sqrt{\sum_{j=1}^p (a_j y_j)^2}}.$$

Theorem: $\mathcal{E}$-symmetric density

The $\mathcal{E}$-symmetric distribution defined in the previous slide possesses the following probability density:

$$
\begin{aligned}
h(\boldsymbol{x}) &= a_1^{-1}\left[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right]^{-\frac{p-1}{2}} \times g\left(\boldsymbol{x}'\boldsymbol{\theta}(\det\boldsymbol{\Sigma})^{-\frac{1}{2}}\left[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right]^{-\frac{1}{2}}\right) \\
&= a_1^{-1}\left[\left(\frac{\boldsymbol{x}'\boldsymbol{\theta}}{a_1}\right)^2 + \sum_{j=2}^{p}\left(\frac{\boldsymbol{x}'\gamma_j}{a_j}\right)^2\right]^{-\frac{p-1}{2}} \\
&\quad \times g\left(\frac{\boldsymbol{x}'\boldsymbol{\theta}}{a_1}\left[\left(\frac{\boldsymbol{x}'\boldsymbol{\theta}}{a_1}\right)^2 + \sum_{j=2}^{p}\left(\frac{\boldsymbol{x}'\gamma_j}{a_j}\right)^2\right]^{-\frac{1}{2}}\right)
\end{aligned}
$$

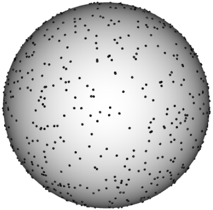
where $\boldsymbol{\Sigma} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}'$.

$\mathcal{E}$-symmetry

Scealy & Wood (2019, 2020)

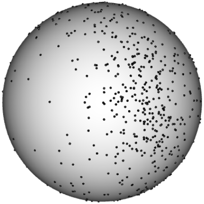
Example # 1: When $g(\boldsymbol{y})$ is the uniform density on the sphere, $h(\boldsymbol{x})$ becomes the angular central gaussian density (see Tyler (1987)).

Below, 1000 realisations from the uniform distribution (left), then transformed (right) with $T = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$.



$$\boldsymbol{y} \mapsto \frac{\boldsymbol{Ty}}{||\boldsymbol{Ty}||}$$

with $\boldsymbol{y} \sim$ Unif



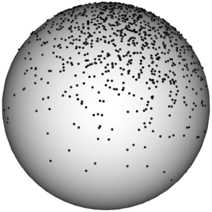
Scaled von-Mises Fisher distribution

Scealy & Wood (2019, 2020)

Example # 2: When going from the famous Fisher von-Mises Langevin distribution and stretching it, we end up with what Scealy and Woods call the *Scaled von-Mises Fisher* distribution, with density

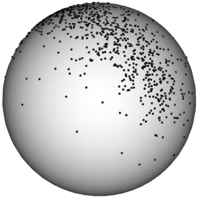
$$c_p(\kappa) \times (\det \boldsymbol{\Sigma})^{-\frac{1}{2}} \left[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right]^{-\frac{p-1}{2}} \times \exp\left(\kappa \boldsymbol{x}'\boldsymbol{\theta}(\det \boldsymbol{\Sigma})^{-\frac{1}{2}} \left[\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}\right]^{-\frac{1}{2}}\right),$$

where $c_p(\kappa)$ is a normalising constant. The example below uses the same $\boldsymbol{T}$ matrix as the previous one.



$$\boldsymbol{y} \mapsto \frac{\boldsymbol{T}\boldsymbol{y}}{||\boldsymbol{T}\boldsymbol{y}||}$$

with $\boldsymbol{y} \sim$ FvML



Future developments

The point of researching this topic is to design statistical procedures that remain valid under $\mathcal{E}$-symmetry.

Estimating $\boldsymbol{\theta}, \gamma_2, \ldots, \gamma_p$ does not seem to be a problem. However,

- Maximum likelihood estimation of both κ and a_1 at the same time is impossible. It is also still unclear as to what their relationship is.
- Estimating $a_2, \ldots, a_p$ seems very complicated.
- Testing procedures made for rotationally symmetric distributions appear to suffer from elliptical distortions of this kind.

Table of Contents

1. (Non-)standard hypotheses in classical statistics
2. About directional statistics
3. Standard hypotheses in directional statistics
4. $\mathcal{E}$-symmetry
5. Weak identifiability

Intuition

Finding $\boldsymbol{\theta}$ is easiest when the concentration around it is strong.

Additionally, estimating it should become easier when sample size n grows.

But what if κ is very small, even relatively to the sample size? Can we still estimate $\boldsymbol{\theta}$ consistently? How small is too small for κ?

Setting

Paindaveine & Verdebout (2017)

First, we assume the data is rotationally symmetric (sadly).

We consider the asymptotic setting where the sample size $n \to \infty$ grows to infinity, and make sure κ_n depends on n.

Specifically, we let $\kappa_n = \sqrt{p}\eta_n\xi + o(\eta_n)$ as $n \to \infty$.

We put aside the presence of p and ξ, and focus on η_n.

η_n is the rate at which κ_n can converge to 0, meaning the distribution will possibly be close to uniformity for large n. Do things break for some specific value of η_n?

Spoiler: Central Limit Theorem

Is there any convergence rate that plays a major role in asymptotic statistics? I wonder...

Central Limit Theorem

For $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ i.i.d. (any distribution) with mean $\mathbb{E}[\boldsymbol{X}_1] =: \boldsymbol{\mu}$, there exists $\boldsymbol{\Sigma}$ such that

$$\sqrt{n}\left(\bar{\boldsymbol{X}} - \boldsymbol{\mu}\right) \xrightarrow{d} \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$$

as $n \to \infty$.

Clearly, the rate $\sqrt{n}$ (or in the case of κ_n, $\frac{1}{\sqrt{n}}$) is very important.

Regimes

Paindaveine & Verdebout (2017)

It is now useful to consider different regimes for η_n, for which we can expect different results:

(i) $\eta_n = 1$

(ii) $\eta_n = o(1)$ with $\sqrt{n}\eta_n \to \infty$

(iii) $\eta_n = \frac{1}{\sqrt{n}}$

(iv) $\eta_n = o(1)$ with $\sqrt{n}\eta_n \to 0$

What does inference on $\boldsymbol{\theta}$ look like in each of these regimes?

Estimating $\boldsymbol{\theta}$

Paindaveine & Verdebout (2017)

How does the usual estimator of $\boldsymbol{\theta}$, that is,

$$\hat{\boldsymbol{\theta}}_n := \frac{\bar{\boldsymbol{X}}}{||\bar{\boldsymbol{X}}||},$$

react when in a *neighbourhood of uniformity*?

Theorem: convergence of $\hat{\boldsymbol{\theta}}_n$

In regimes (i) and (ii), $\hat{\boldsymbol{\theta}}_n$ converges in probability to $\boldsymbol{\theta}$.

In regime (iii), $\hat{\boldsymbol{\theta}}_n \xrightarrow{d} \frac{\boldsymbol{Z}}{||\boldsymbol{Z}||}$, with $\boldsymbol{Z} \sim \mathcal{N}_p(\boldsymbol{\theta}, \boldsymbol{I}_p)$.

In regime (iv), $\hat{\boldsymbol{\theta}}_n \xrightarrow{d} \text{Unif}(\mathbb{S}^{p-1})$, the uniform distribution on the sphere.

Consequences I

On top of struggling for point estimation, $\hat{\boldsymbol{\theta}}_n$ is obviously used in many other statistical procedures, such as hypotheses tests.

For example, consider the problem of testing

$$H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0 \quad \text{against} \quad H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$$

for some $\boldsymbol{\theta}_0 \in \mathbb{S}^{p-1}$.
Paindaveine and Verdebout (2017) have shown that the usual Watson score test remained valid under the null, but severely underperformed power-wise (i.e. the test does not reject as often as expected when it needs to) in regimes (iii) and (iv).

Consequences II

This kind of issue is bound to arise in all sorts of problems involving the estimation of $\boldsymbol{\theta}$.

We are currently researching how robust current methods are to this issue, and how we can improve them if needed.

Conclusions

Many statistical tools have been built for somewhat strict standard hypotheses.

Real-life random processes are often more complex than those hypotheses, and practitioners need be aware of it.

Similarily as in other sub-fields, a lot of research is still needed for directional statistics methods that work well in less-than-ideal situations.

References

- Scealy, J. L., & Wood, A. T. A. (2019), Scaled von Mises-Fisher Distributions and Regression Models for Paleomagnetic Directional Data, *Journal of the American Statistical Association*, **114**, 1547–1560.
- Scealy, J. L., & Wood, A. T. A. (2020). Analogues on the Sphere of the Affine-Equivariant Spatial Median. *Journal of the American Statistical Association*, **116**(535), 1457–1471
- Paindaveine, D., & Verdebout, T. (2017). Inference on the mode of weak directional signals: A Le Cam perspective on hypothesis testing near singularities. *The Annals of Statistics*, **45**(2).
- Watson, G. S. (1983). *Statistics on Spheres*. Wiley.
- Tyler, D. E. (1987). Statistical analysis for the angular central Gaussian distribution on the sphere. *Biometrika*, **74**(3), 579–589.