Tests for the concentration based on LAN properties

Thomas Verdebout (Université Lille 3)

joint work with Ch. Ley

ADISTA 14

May 20–22 2014

Outline

Outline

Definition

Throughout, the data points $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d with a FvML distribution characterized by a density function (with respect to the usual surface area measure on spheres) of the form

$$f_{\boldsymbol{\theta}}(\mathbf{x}) = c_{k,\kappa} \, \exp(\kappa \mathbf{x}' \boldsymbol{\theta}), \tag{1.1}$$

where $\mathbf{x} \in \mathcal{S}_{k-1}$, $\boldsymbol{\theta} \in \mathcal{S}^{k-1}$ is a location parameter, $\kappa > 0$ is a concentration parameter and $c_{k,\kappa}$ is a normalizing constant.

Definition

If $\mathbf{X}_1, \ldots, \mathbf{X}_n$ are i.i.d. with density (1.1), then the cosines $\mathbf{X}_1'\boldsymbol{\theta}, \ldots, \mathbf{X}_n'\boldsymbol{\theta}$ are i.i.d. with density

$$t \mapsto \tilde{f}_\kappa(t) := C_{k,f_1,\kappa} \exp(\kappa t)(1-t^2)^{(k-3)/2}, \quad -1 \leq t \leq 1.$$

As a direct consequence, the parameter κ is clearly identified using the identity

$$\mathrm{E}[\mathbf{X}] = \mathrm{E}[\mathbf{X}'\boldsymbol{\theta}]\boldsymbol{\theta} =: A_k(\kappa)\boldsymbol{\theta} = \left(\frac{\int_{-1}^{1} t e^{\kappa t}(1-t^2)^{\frac{k-3}{2}}}{\int_{-1}^{1} e^{\kappa t}(1-t^2)^{\frac{k-3}{2}}} \right) \boldsymbol{\theta},$$

where, letting $I_q(v)$ stand for the modified Bessel function of first kind and of order q, A_k is defined by $A_k(.) := I_{k/2}(.)/I_{k/2-1}(.)$; one readily obtains that $\kappa := A_k^{-1}(\mathrm{E}[\mathbf{X}'\boldsymbol{\theta}])$.

ULAN property

In the sequel, we write $\mathrm{P}_{\boldsymbol{\vartheta}}^{(n)}$ or when it is relevant $\mathrm{P}_{(\boldsymbol{\theta},\kappa)}^{(n)}$ for the joint cdf of $\mathbf{X}_1, \ldots, \mathbf{X}_n$ with parameter $\boldsymbol{\vartheta} = (\boldsymbol{\theta}', \kappa)' \in \boldsymbol{\Theta} := \mathcal{S}^{k-1} \times \mathbb{R}^+$. The model $\mathrm{P}_{\boldsymbol{\vartheta}}^{(n)}$ is called ULAN if for any sequence $\boldsymbol{\vartheta}^{(n)} \in \boldsymbol{\Theta}$ such that $\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta} = O(n^{-1/2})$, the likelihood ratio between $\mathrm{P}_{\boldsymbol{\vartheta}^{(n)} + n^{-1/2}\boldsymbol{\tau}^{(n)}}$ and $\mathrm{P}_{\boldsymbol{\vartheta}^{(n)}}$ allows a specific form of (probabilistic) Taylor expansion as a function of the perturbation $\boldsymbol{\tau}^{(n)}$. Therefore, to provide such a property, we have to clearly define the local perturbations $\boldsymbol{\tau}^{(n)}$; $\boldsymbol{\tau}^{(n)} =: ((\mathbf{t}^{(n)})', c^{(n)})'$.

ULAN property

The perturbations $\boldsymbol{\tau}^{(n)} =: ((\mathbf{t}^{(n)})', c^{(n)})'$ must be chosen so that $(\boldsymbol{\theta}', \kappa)' + n^{-1/2}((\mathbf{t}^{(n)})', c^{(n)})'$ remains on $\boldsymbol{\Theta} = \mathcal{S}^{k-1} \times \mathbb{R}^{+}$. Thus, in particular, $\mathbf{t}^{(n)}$ need to satisfy

$$\begin{aligned} 0 &= (\boldsymbol{\theta} + n^{-1/2}\mathbf{t}^{(n)})'(\boldsymbol{\theta} + n^{-1/2}\mathbf{t}^{(n)}) - 1 \\ &= 2n^{-1/2}\boldsymbol{\theta}'\mathbf{t}^{(n)} + n^{-1}(\mathbf{t}^{(n)})'\mathbf{t}^{(n)}. \end{aligned} \tag{1.2}$$

Consequently, $\mathbf{t}^{(n)}$ must be such that $2n^{-1/2}\boldsymbol{\theta}'\mathbf{t}^{(n)} + o(n^{-1/2}) = 0$: for $\boldsymbol{\theta} + n^{-1/2}\mathbf{t}^{(n)}$ to remain in $\mathcal{S}^{k-1}$, the perturbation $\mathbf{t}^{(n)}$ must belong, up to a $o(n^{-1/2})$ quantity, to the tangent space to $\mathcal{S}^{k-1}$ at $\boldsymbol{\theta}$. For the "κ-part" of the perturbation, we simply restrict to sequences $c^{(n)}$ such that $\kappa + n^{-1/2}c^{(n)}$ remains strictly positive. We have the following result.

ULAN property

Proposition

The family $\left\{\mathrm{P}_{\boldsymbol{\vartheta}}^{(n)} \mid \boldsymbol{\vartheta} \in \boldsymbol{\vartheta}\right\}$ *is ULAN; that is for any sequence* $\boldsymbol{\vartheta}^{(n)} \in \boldsymbol{\Theta}$ *such that* $\boldsymbol{\vartheta}^{(n)} - \boldsymbol{\vartheta} = O(n^{-1/2})$ *and any bounded sequence* $\boldsymbol{\tau}^{(n)}$ *as described in (1.2), under* $\mathrm{P}_{\boldsymbol{\vartheta}^{(n)}}$ *as* $n \to \infty$*. The central sequence* $\boldsymbol{\Delta}_{\boldsymbol{\vartheta}}^{(n)} := \left(\boldsymbol{\Delta}_{\boldsymbol{\vartheta}}^{(\mathrm{I})\prime}, \Delta_{\boldsymbol{\vartheta}}^{(\mathrm{II})}\right)'$ *is defined by*

$$\boldsymbol{\Delta}_{\boldsymbol{\vartheta}}^{(\mathrm{I})} := \kappa n^{-1/2} \sum_{i=1}^{n} (1 - (\mathbf{X}_i'\boldsymbol{\theta})^2)^{1/2} \mathbf{S}_{\boldsymbol{\theta}}(\mathbf{X}_i),$$

and

$$\Delta_{\boldsymbol{\vartheta}}^{(\mathrm{II})} := n^{-1/2} \sum_{i=1}^{n} \mathbf{X}_i'\boldsymbol{\theta} - \mathrm{E}\left[\mathbf{X}_i'\boldsymbol{\theta}\right].$$

The associated Fisher information is given by $\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}} := \mathrm{diag}\left(\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}}^{(\mathrm{I})}, \Gamma_{\boldsymbol{\vartheta}}^{(\mathrm{II})}\right)$*, where, putting* $\mathcal{J}_k(\kappa) := \int_{-1}^{1} (1 - u^2) \tilde{f}_\kappa(u) du$*,*

$$\boldsymbol{\Gamma}_{\boldsymbol{\vartheta}}^{(\mathrm{I})} := \frac{\kappa^2 \mathcal{J}_k(\kappa)}{k-1} (\mathbf{I}_k - \boldsymbol{\theta}\boldsymbol{\theta}') \quad \text{and} \quad \Gamma_{\boldsymbol{\vartheta}}^{(\mathrm{II})} := \mathrm{E}[(\mathbf{X}_i'\boldsymbol{\theta})^2] - \mathrm{E}^2[(\mathbf{X}_i'\boldsymbol{\theta})].$$

Outline

One-sample case

The score test of Watamori and Jupp (2005) for the null hypothesis $\mathcal{H}_0 : \kappa = \kappa_0$ rejects the null at asymptotic nominal level α when $(\bar{\mathbf{X}} := n^{-1/2} \sum_{i=1}^{n} \mathbf{X}_i)$

$$Q_{\kappa_0}^{(n)} := \frac{\left(n\|\bar{\mathbf{X}}\| - A_k^{-1}(\kappa_0)\right)^2}{n(1 - \frac{k-1}{\kappa_0} A_k(\kappa_0) - (A_k(\kappa_0))^2)}$$

exceeds the α-upper quantile of the chi-square distribution with 1 degree of freedom.

One-sample case

Proposition

We have that

(i) $Q_{\kappa_0}^{(n)}$ *is asymptotically chi-square with* 1 *degree of freedom under* $\cup_{\boldsymbol{\theta} \in \mathcal{S}^{k-1}} \mathrm{P}_{(\kappa_0, \boldsymbol{\theta})}^{(n)}$ *;*

(ii) $Q_{\kappa_0}^{(n)}$ *is asymptotically non-central chi-square with* 1 *degree of freedom and non-centrality parameter* $(1 - \frac{k-1}{\kappa_0} A_k(\kappa_0) - (A_k(\kappa_0))^2) c^2$ *under* $\cup_{\boldsymbol{\theta} \in \mathcal{S}^{k-1}} \mathrm{P}_{(\kappa_0 + n^{-1/2} c^{(n)}, \boldsymbol{\theta})}^{(n)}$*, where* $c := \lim_{n \to \infty} c^{(n)}$*.*

Multi-sample case

Let us assume that the samples $(\mathbf{X}_{i1}, \ldots, \mathbf{X}_{in_i})$, $i = 1, \ldots, m$, are independent samples of i.i.d. random vectors such that the n_i observations $\mathbf{X}_{ij}$, $j = 1, \ldots, n_i$, in sample i have a FvML density with concentration κ_i and location $\boldsymbol{\theta}_i$. We denote this time by $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^{(m)}}$ the joint distribution of $(\mathbf{X}_{11}, \ldots, \mathbf{X}_{mn_m})$, with $\boldsymbol{\vartheta}^{(m)} := (\kappa_1, \ldots, \kappa_m, \boldsymbol{\theta}'_1, \ldots, \boldsymbol{\theta}'_m)' \in (\mathbb{R}_0^+)^m \times (\mathcal{S}^{k-1})^m$.

It is easy to show that under some mild assumptions, this model is also ULAN.

Multi-sample case

The score test of Watamori and Jupp (2005) for the null hypothesis $\mathcal{H}_0 : \kappa_1 = \ldots = \kappa_m$ rejects the null at asymptotic nominal level α when ($\hat{D}_k := 1 - \frac{k-1}{\hat{\kappa}} A_k(\hat{\kappa}) - (A_k(\hat{\kappa}))^2$)

$$Q_{\mathrm{Hom}}^{(n)} \quad := \quad \hat{D}_k^{-1} \left(\sum_{i=1}^{m} n_i \|\bar{\mathbf{X}}_i\|^2 - \frac{1}{n} \left(\sum_{i=1}^{m} n_i \|\bar{\mathbf{X}}_i\| \right)^2 \right)$$

exceeds the α-upper quantile of the chi-square distribution with $m-1$ degrees of freedom.

Multi-sample case

Proposition

Let Assumption A. Then

(i) $Q^{(n)}_{\mathrm{Hom}}$ *is asymptotically chi-square with* $m-1$ *degrees of freedom under* $\bigcup_{\boldsymbol{\vartheta}^{(m)} \in \mathcal{H}_0^{\mathrm{Hom}}} \mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^m}$ *;*

(ii) *Letting* $\mathbf{c} := \lim_{n\to\infty}(c_1^{(n)}, \ldots, c_m^{(n)})'$, $Q^{(n)}_{\mathrm{Hom}}$ *is asymptotically non-central chi-square with* $m-1$ *degrees of freedom and non-centrality parameter*

$$l_{\mathrm{Hom}} = \mathbf{c}'\boldsymbol{\Gamma}^{(\mathrm{II})}_{\boldsymbol{\vartheta}^{(m)}}(\boldsymbol{\Gamma}^{(\mathrm{II})}_{\boldsymbol{\vartheta}^{(m)}})^{\perp}\boldsymbol{\Gamma}^{(\mathrm{II})}_{\boldsymbol{\vartheta}^{(m)}}\mathbf{c}$$

under $\mathrm{P}^{(n)}_{\boldsymbol{\vartheta}^m + n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)}}$ *with* $\boldsymbol{\vartheta}^m \in \mathcal{H}_0^{\mathrm{Hom}}$ *and* $\boldsymbol{\vartheta}^m + n^{-1/2}\boldsymbol{\nu}^{(n)}\boldsymbol{\tau}^{(n)} \notin \mathcal{H}_0^{\mathrm{Hom}}$

Outline

Using ranks

The Watamori and Jupp test for the homogeneity of concentrations $\mathcal{H}_0 : \kappa_1 = \ldots = \kappa_m$ is based on the FvML assumption. Clearly, testing $\mathcal{H}_0 : \kappa_1 = \ldots = \kappa_m$ is completely equivalent to test $\mathcal{H}_0 : \mathrm{E}[\mathbf{X}_1'\boldsymbol{\theta}_1] = \ldots = \mathrm{E}[\mathbf{X}_m'\boldsymbol{\theta}_m]$.

Now, simple computations entail that

$$Q_{\mathrm{Hom}}^{(n)} = D_k^{-1} \left(\sum_{i=1}^{m} n_i (\boldsymbol{\theta}_i' \bar{\mathbf{X}}_i)^2 - \frac{1}{n} \left(\sum_{i=1}^{m} n_i \boldsymbol{\theta}_i' \bar{\mathbf{X}}_i \right)^2 \right) + o_{\mathrm{P}}(1).$$

If you replace the "$\mathbf{X}_{ij}'\boldsymbol{\theta}_i$" in the formula above by their ranks, you obtain a Kruskal-Wallis type test for the homogeneity of concentrations.

Using ranks

More precisely, letting $R_{ij}(\boldsymbol{\theta})$ be the (univariate rank) of $\mathbf{X}'_{ij}\boldsymbol{\theta}_i$ among the cosines

$$\mathbf{X}'_{11}\boldsymbol{\theta}_1, \ldots, \mathbf{X}'_{1n_1}\boldsymbol{\theta}_1, \mathbf{X}'_{21}\boldsymbol{\theta}_2, \ldots, \mathbf{X}'_{mn_m}\boldsymbol{\theta}_m$$

and $\bar{R}_i := n_i^{-1} \sum_{j=1}^{n_i} R_{ij}(\boldsymbol{\theta})$,we can consider the statistic

$$Q^{(n)}_{\mathrm{Hom}}(K_{\mathrm{KW}}) := \frac{12}{(n+1)^2} \sum_{i=1}^{m} n_i \left(\bar{R}_i - \frac{n+1}{2} \right)^2$$

which is nothing but (up to an irrelevant $(n+1)/n$ factor) the traditional rank-based Kruskal-Wallis test statistic (see Kruskal (1952) and Kruskal and Wallis (1952)).

Using ranks

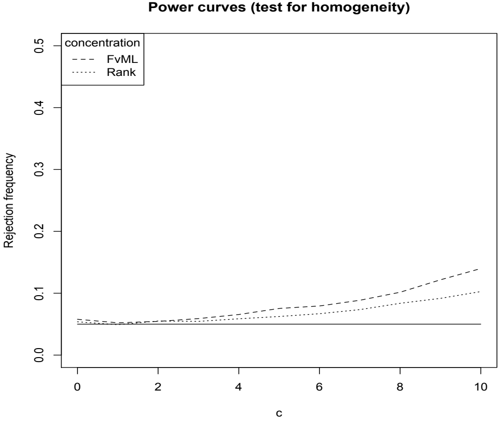
When the location parameters are known, we have the finite-sample distribution of this Kruskal-Wallis type test.

The substitution of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ by root-n consistent estimators do not have any asymptotic cost.

The corresponding test which rejects the null when $Q^{(n)}_{\mathrm{Hom}}(K_{\mathrm{KW}})$ exceeds the α-upper quantile of the chi-square distribution with $m-1$ degrees of freedom is *validity-robust*.

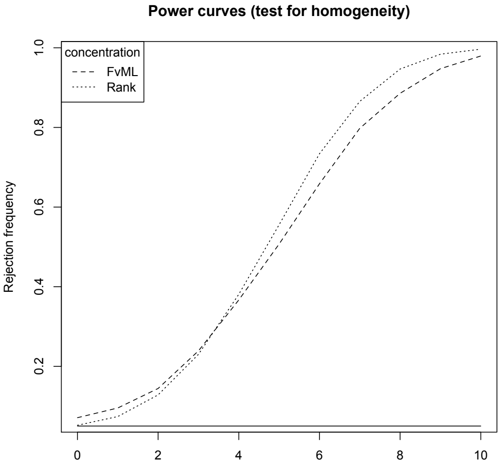
Using ranks

Power curve of the Watamori and Jupp (2005) test and the Kruskal-Wallis type test under Fisher-von Mises-Langevin distributions with $\boldsymbol{\theta}_1 = (1, 0)'$ and $\boldsymbol{\theta}_2 = (-1, 0)'$. The concentration under the null is $\kappa = 2$. Sample sizes are $n_1 = 200$ and $n_2 = 250$.The number of replications is 10000.



Using ranks

Power curve of the Watamori and Jupp (2005) test under wrapped-Cauchy distributions with $\boldsymbol{\theta}_1 = (1, 0)'$ and $\boldsymbol{\theta}_2 = (-1, 0)'$. The concentration under the null is $\kappa = .5$. Sample sizes are $n_1 = 200$ and $n_2 = 250$.The number of replications is 10000.



Outline

- Watamori, Y. and Jupp, P. E. (2005). Improved likelihood ratio and score tests on concentration parameters of von Mises-Fisher distributions. *Statistics and Probability Letters* **72**, 93–102.
- Ch.Ley, Y.Swan, B. Thiam and T.Verdebout (2013). Optimal R-estimation of a spherical location. *Statistica Sinica*, **23**(1), 305-333 (2013).
- Ch. Ley and T. Verdebout (2014). Local powers of optimal one- and multi-sample tests for the concentration of Fisher-von Mises-Langevin distributions. *International Statistical Review*, to appear.
- T.Verdebout (2014). On a Kruskal-Wallis type test for the equality of concentrations. *Work in progress*

Thank you !