On Distribution Free Inference for Tensors

S.N. Lahiri

Department of Statistics & Data Science
Washington University in St. Louis

s.lahiri@wustl.edu

Outline

- Introduction/Problem Formulation /motivation
- The Empirical Likelihood (EL) method
- The EL in high dimensions
- Main results
- Calibration based on Subsampling/...
- ...

Introduction

- Tensors are multi-dimensional arrays, generalizing the concepts of vectors/matrices.
- Common in many modern applications, including
  - Neuroscience : MRI/ fMRI data
  - Recommender systems
  - Computer vision
  - ...

  as well as in classical Statistics, Psychometrics, Chemometrics, etc.
- Our goal is to develop distribution free inference tools for analyzing such multi-array data.

Introduction

- Let

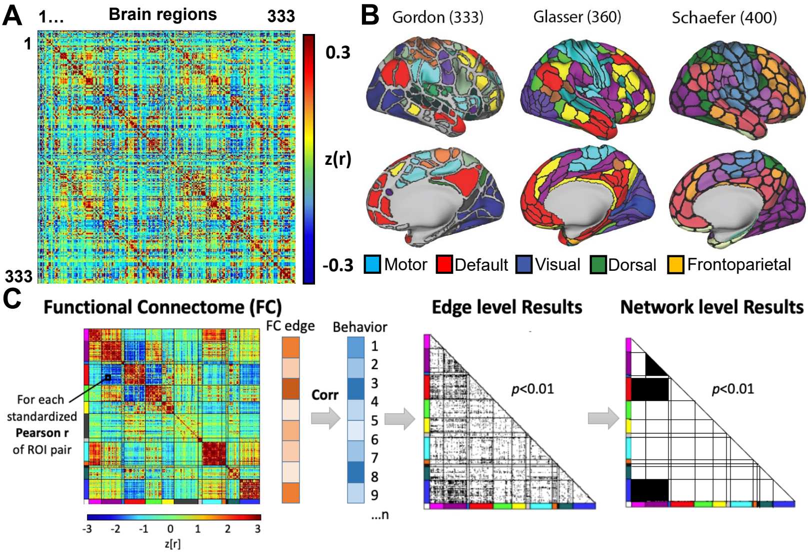
$$\mathbf{X} = \boldsymbol{\Theta} + \boldsymbol{\epsilon}$$

  where $\mathbf{X}, \boldsymbol{\Theta}, \boldsymbol{\epsilon}$ are $p_1 \times p_2 \times p_3$ order-3 tensors, $\boldsymbol{\Theta} \in \mathbb{R}^{p_1 p_2 p_3}$ are unknown parameters and $\boldsymbol{\epsilon}$ is a tensor of zero mean, finite variance random variables $\epsilon_{jk\ell}$, $j = 1, \ldots, p_1$, $k = 1, \ldots, p_2$ and $\ell = 1, \ldots, p_3$.
- Typically, the random variables $\epsilon_{jk\ell}$ are correlated!
- Specification of the covariance structure is important as it must capture the interactions among the components of $\boldsymbol{\epsilon}$.
- In turn, it also determines the distributional properties of estimators and tests.

Covariance structure

- Here we will use the (resting state) fMRI example to motivate the covariance structure.
- This imaging modality detects small changes in the magnetic resonance (MR) of blood vessels near firing neurons as oxygen levels drop.
- The resulting shifts are voxelized into three-dimensional (3D) image volumes and then used to identify neural regions of activity.

Resting state fMRI



Covariance structure

- Let $\mathcal{I} = \{(j_1, j_2, j_3) : 1 \leq j_r \leq p_r, r = 1, 2, 3\}$.
- We shall suppose that there exist a correlation function ρ_0 of a stationary process on $\mathbb{Z}^3$ and a permutation $\pi^* : \mathcal{I} \to \mathcal{I}$ such that

$$\mathrm{Cov}(\epsilon_{j,k,\ell}, \epsilon_{j',k',\ell'}) = \sigma^2 \cdot \rho_0\Big(\pi^*(j,k,\ell) - \pi^*(j',k',\ell')\Big)$$

  for $(j,k,\ell), (j',k',\ell') \in \mathcal{I}$.
- We do not impose any distributional assumptions on $\boldsymbol{\epsilon}$ (otherwise).

The Testing Problem

- Next, for $r = 1, 2, 3$, let A_r be a (known) matrix of order $m_r \times p_r$.

- We want to test the hypotheses:

$$H_0 : \mathbf{\Theta} \times_1 A_1 \times_2 A_2 \times_3 A_3 = \mathbf{0} \quad \text{vs.} \quad H_1 : \mathbf{\Theta} \times_1 A_1 \times_2 A_2 \times_3 A_3 \neq \mathbf{0},$$

  based on n iid copies of $\mathbf{X}$.

- In the following, **we allow the dimensions** p_1, p_2, p_3 **and** m_1, m_2, m_3 **to depend on/diverge with** n**.**

The Testing Problem

- How do we test H_0 vs H_1 ?
- First suppose that $m_1, \dots, m_3$ are finite.
  - If we impose a parametric model on ϵ, we can use the LRT !
  - Alternatively, we can use self-normalized estimators of

$$\boldsymbol{\Theta} \times_1 A_1 \times_2 A_2 \times_3 A_3$$

    to test H_0 vs H_1.
  - For this, we need the asymptotic distribution/variance & its estimator.
- Both approaches become more challenging if $m_1, \dots, m_3$ are unbounded !!
- Here we will use the Empirical Likelihood approach of Owen (1988) that bypasses both challenges!!

The Testing Problem

- How do we test H_0 vs H_1 ?
- First suppose that $m_1, \ldots, m_3$ are finite.
  - If we impose a parametric model on $\boldsymbol{\epsilon}$, we can use the LRT !
  - Alternatively, we can use self-normalized estimators of

$$\boldsymbol{\Theta} \times_1 A_1 \times_2 A_2 \times_3 A_3$$

    to test H_0 vs H_1.
  - For this, we need the asymptotic distribution/variance & its estimator.
- Both approaches become more challenging if $m_1, \ldots, m_3$ are unbounded !!
- Here we will use the Empirical Likelihood approach of Owen (1988) that bypasses both challenges!!

Empirical Likelihood

- What is Empirical Likelihood?
- Consider a parametric model $\{f(\cdot;\theta) : \theta \in \Theta\}$ and let $X_1, \ldots, X_n$ be iid, $X_1 \sim f(\cdot;\theta_0)$. Then, the *likelihood function* for θ is

$$L_n(\theta) = \prod_{i=1}^{n} f(X_i;\theta).$$

- Under some regularity conditions, Wilk (1938)'s theorem asserts that

$$-2\log R_n(\theta_0) \to^d \chi_p$$

where $R_n(\theta_0)$ is the *likelihood ratio statistic* (LRT) for testing $H_0 : \theta = \theta_0$.

Empirical Likelihood

- Empirical Likelihood(EL) of Owen (1988) is a method that defines a likelihood for certain population parameters *without* requiring a parametric model.
- Let $X_1, \ldots, X_n$ be iid with mean $\mu \in \mathbb{R}$. The EL for μ is

$$L_n(\mu) = \sup\left\{ \prod_{i=1}^{n} \pi_i : \pi_i \geq 0, \sum \pi_i = 1, \sum \pi_i X_i = \mu \right\}$$

- The unconstrained maximum is at $\pi_i = n^{-1}$ for all i. Thus, the EL ratio statistic for testing $H_0 : \mu = \mu_0$ is $R_n(\mu_0) = \frac{L_n(\theta_0)}{n^{-n}}$.
- Owen (1988) proved a version of Wilk's Theorem:

$$-2 \log R_n(\mu_0) \to^d \chi^2_1.$$

Literature Review: The EL

- Owen (1988, 1990) introduced the Empirical Likelihood (EL) method for independent random variables.
- Extensions and refinements of the EL method to different problems are given by
  - Chen and Hall (1993) : Quantiles
  - Qin and Lawless (1994) : Estimating Equations
  - DiCiccio, Hall and Romano (1996) : Bartlett Corrections
  - Einmahl and McKeague (2003) : Functional Hypothesis Testing
  - Bertail (2006) : Semiparametric models

EL in High Dimensions

- Hjort, McKeague, and Van Keilegom (2009) : Functional nuisance parameters & Increasing dimensions
- Chen, Peng and Qin (2008): Increasing dimensions $p = o(n^{1/2})$
- A remarkable result of Tsao (2004) showed that **for $p > n/2$, there is a nontrivial *positive probability* that true mean will lie outside the convex hull of $\{X_1, \ldots, X_n\}$.**
- Thus, the EL is **not usable** for $p > n/2$.

EL in High Dimensions & alternative approaches

- Chen, Variyath and Abraham (2006) : Adjusted EL (AEL) in the $p > n/2$ case
- Emerson and Owen (2009) further refined the AEL.
- Bartolucci (2007) proposed a **Penalized EL (PEL)**, for $p \leq n$.
- Lahiri & Mukherjee (2012) defined a different version of the PEL that works in $p > n$ case.
- Here we will extend the PEL for the tensor case, allowing the tensor dimensions to diverge, beyond n.

NOTATION & FRAMEWORK:

- Let $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots,$ be iid copies of $\mathbf{X}$.
- Let $\boldsymbol{\eta} = \boldsymbol{\Theta} \times_1 A_1 \times_2 A_2 \times_3 A_3$ be the parameter of interest! Thus, we can restate the testing pronblem as

$$H_0 : \boldsymbol{\eta} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\eta} \neq \mathbf{0}.$$

- Write $Y_{jk\ell}^{(i)}$ for the (j, k, ℓ)th element of $\mathbf{Y}^{(i)} = \mathbf{X}^{(i)} \times_1 A_1 \times_2 A_2 \times_3 A_3$, $1 \leq j \leq m_1$, $1 \leq k \leq m_2$, $1 \leq \ell \leq m_3$, $i \geq 1$.
- Let $\boldsymbol{\Delta}$ be a tensor with (j, k, ℓ)th element

$$\delta_{j,k,\ell,n} = s_{j,k,\ell,n}^{-1} \mathbb{1}(s_{j,k,\ell,n} \neq 0)$$

where $s_{j,k,\ell,n}^2$ is the sample variance of $\{Y_{jk\ell}^{(i)} : i = 1, \ldots, n\}$.

The proposed PEL

- We propose the following version of PEL for testing $H_0 : \boldsymbol{\eta} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\eta} \neq \mathbf{0}$:

$$\begin{aligned} & L_n(\boldsymbol{\eta}) \\ = \; & \sup \Big\{ \Big(\prod_{i=1}^{n} \pi_i \Big) \exp \Big(-\lambda \Big\| \Big[\sum_{i=1}^{n} \pi_i \mathbf{Y}^{(i)} - \boldsymbol{\eta} \Big] * \boldsymbol{\Delta}_n \Big\|^2 \Big) \\ & \qquad : (\pi_1, \ldots, \pi_n) \in \Pi_n \Big\} \end{aligned}$$

where $\|\mathbf{A}\|^2 = \sum_j \sum_k \sum_\ell a_{jkl}^2$ and $*$ denotes the Hadamard product.

The proposed PEL

- Here, $\boldsymbol{\Delta} = (((\delta_{j,k,\ell,n})))$ gives the component specific weights (which are random), and $\lambda \in (0, \infty)$ is an overall penalty factor.

- Note that the unconstrained maximum of $\prod_{i=1}^{n} \pi_i$ is n^{-n}.

- Hence, the PEL ratio statistic for testing $H_0 : \boldsymbol{\eta} = \mathbf{0} \quad \text{vs.} \quad H_1 : \boldsymbol{\eta} \neq \mathbf{0}$ is given by

$$R_n = \frac{L_n(\mathbf{0})}{n^{-n}}.$$

Limit Distribution

Theorem

Let $\lambda = n/[m_1 m_2 m_3]$ *and* $\Big|\rho_0\big((i_1, i_2, i_3)\big)\Big| \leq C\|(i_1, i_2, i_3)\|^{-(1.5+a)}$ *for all* $i_1, i_2, i_3 \in \mathbb{Z}$ *and for some* $a > 0$. *Then, under some moment and additional weak dependence conditions, for* $1 \ll M \equiv m_1 m_2 m_3 \ll n^2$,

$$M^{1/2}\Big[-\log R_n - 1 \Big]/\kappa_n \rightarrow^d N(0,1)$$

under H_0, *where* R_n *is the PEL ratio statistic defined above and where* $\kappa_n = O(1)$ *is a population parameter depending on the ACF* $\rho_0(\cdot)$.

Limit Distribution

- **Remark 1:** A consistent estimator of κ_n^2 can be constructed:
  - using the form of κ_n, if the permutation π^* is known!
  - using the Subsampling method, if the permutation π^* is unknown!
- **Remark 2:** Thus, the PELRT test can be calibrated using the N(0,1) critical values!
- **Remark 3:** In the case, all m_1, m_2, m_3 are fixed and finite, penalization is not needed. The standard EL works.
- **Remark 4:** The PEL method can be used for validating joint significance of a region of interest that may have been identified using marginal analysis.

Simulation

To be done!!

The END!!!

Thank you !!