Robust Likelihood Ratio Test using $\alpha-$ Divergence

ICASSP 2020

Aref Miri Rekavandi, Abd-Krim Seghouane, Robin J. Evans

The university of Melbourne

May, 2020

Outline

1. Problem Definition
2. Matched Subspace Detector
3. Proposed Detector
4. Results

Problem Definition

Observations follow a general linear model (GLM)as

$$\mathbf{y} = \mathbf{H}\theta + \mathbf{B}\phi + \boldsymbol{\xi}$$

where

- $\mathbf{y} \in \mathbb{R}^N$
- $\mathbf{H} \in \mathbb{R}^{N \times p}$ and $\mathbf{B} \in \mathbb{R}^{N \times t}$ [known]
- $\theta \in \mathbb{R}^p$ and $\phi \in \mathbb{R}^t$ [unknown]
- $\boldsymbol{\xi} \in \mathbb{R}^N$: Nominally Gaussian with unknown variance

Let's define $\boldsymbol{\omega} = [\boldsymbol{\beta}^\top, \sigma^2]^\top = [\boldsymbol{\theta}^\top, \boldsymbol{\phi}^\top, \sigma^2]^\top = [\boldsymbol{\theta}^\top, \boldsymbol{\lambda}^\top]^\top$ and $\mathbf{C} = [\mathbf{H}, \mathbf{B}]$.
The goal is to decide between:

$$\mathcal{H}_0 : \mathbf{y} : \mathbf{B}\phi + \boldsymbol{\xi} \quad \text{vs} \quad \mathcal{H}_1 : \mathbf{y} : \mathbf{H}\theta + \mathbf{B}\phi + \boldsymbol{\xi}$$

Matched Subspace Detector(MSD)

Using the GLRT framework

$$T_G(\mathbf{y}) = \frac{2}{N}[\sup_{\omega\in\boldsymbol{\Omega}_1} \sum_{i=1}^{N} \log\left(f(y_i, \boldsymbol{\omega})\right) - \sup_{\omega\in\boldsymbol{\Omega}_0} \sum_{i=1}^{N} \log\left(f(y_i, \boldsymbol{\omega})\right)] \geq \gamma,$$

MSD[1] is given by

$$L_{MSD}(\mathbf{y}) = \frac{\hat{\sigma}_0^2 - \hat{\sigma}_1^2}{\hat{\sigma}_1^2} = \frac{\mathbf{y}^\top \mathbf{P}_{\mathbf{B}^\perp} \mathbf{P}_{\mathbf{G}} \mathbf{P}_{\mathbf{B}^\perp} \mathbf{y}}{\mathbf{y}^\top \mathbf{P}_{\mathbf{B}^\perp} \mathbf{P}_{\mathbf{G}^\perp} \mathbf{P}_{\mathbf{B}^\perp} \mathbf{y}},$$

where $\mathbf{G} = \mathbf{P}_{\mathbf{B}^\perp}\mathbf{H}$ and $\mathbf{P}_{\mathbf{D}} = \mathbf{D}(\mathbf{D}^\top\mathbf{D})^{-1}\mathbf{D}^\top$ is the projection matrix.

Disadvantageous

The MSD is not robust against deviations in hypotheses.

1. Scharf-1994: Matched Subspace Detectors

From KL divergece to $\alpha-$divergence

MSD is based on the maximum likelihood solution which in large samples regime ($N \to \infty$) is equivalent to

$$\hat{\boldsymbol{\omega}}_{ML} = \arg\max_{\boldsymbol{\omega}} \frac{1}{N} \sum_{i=1}^{N} \log\left(f(y_i, \omega)\right) = \arg\min_{\boldsymbol{\omega}} KL\left(f(\mathbf{y}, \boldsymbol{\omega}^*), f(\mathbf{y}, \boldsymbol{\omega})\right).$$

As an alternative, we propose to use the $\alpha-$divergence defined as

$$D_\alpha\left(g(\mathbf{y}, \omega^*) \parallel f(\mathbf{y}, \omega)\right) = \frac{1}{\alpha(\alpha-1)}\left[\int g(\mathbf{y}, \omega^*)^\alpha f(\mathbf{y}, \omega)^{1-\alpha} d\mathbf{y} - 1\right],$$

to develop the test. In above, $g(\mathbf{y}, \omega^*) = (1-\epsilon) f(\mathbf{y}, \omega^*) + \epsilon h$. When $\alpha \to 1$, $D_\alpha(.) \to KL(.)$.

Proposed Detector

Minimum of $\alpha-$divergence is

$$\sup_{\omega} \frac{1}{N} \sum_{i=1}^{N} \log_{\alpha}\left(f(y_i, \omega)\right),$$

where $\log_{\alpha}(x) = \frac{x^{1-\alpha}-1}{1-\alpha}$. Using the new $\alpha-$logarithm function in GLRT framework gives

$$T_{G_r}(\mathbf{y}) = \frac{2}{N}\Big[\sup_{\boldsymbol{\omega}\in\boldsymbol{\Omega}_1} \sum_{i=1}^{N} \log_{\alpha}\left(f(y_i, \boldsymbol{\omega})\right) - \sup_{\boldsymbol{\omega}\in\boldsymbol{\Omega}_0} \sum_{i=1}^{N} \log_{\alpha}\left(f(y_i, \omega)\right)\Big] \geq \eta,$$

$$L_{\alpha}(\mathbf{y}) = \frac{1}{N}\Big[\sum_{i=1}^{N} \frac{w_{i,1}}{\hat{\sigma}^{\tau}_{\alpha,1}} - \sum_{i=1}^{N} \frac{w_{i,0}}{\hat{\sigma}^{\tau}_{\alpha,0}}\Big] \geq \eta'$$

where $\tau = 1 - \alpha$.

Parameters of Proposed Detector

For example, in the case of $\mathcal{H}_1$

$$w_{i,1} = exp\{-\frac{\tau}{2\sigma^2}(y_i - \mathbf{c}_i\beta)^2\},$$

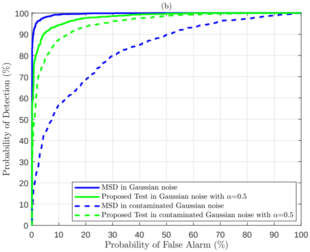
$$\hat{\sigma}^2_{\alpha,1} = \frac{\sum_{i=1}^N w_{i,1}(y_i - \mathbf{c}_i\beta)^2}{\sum_{i=1}^N w_{i,1}}.$$

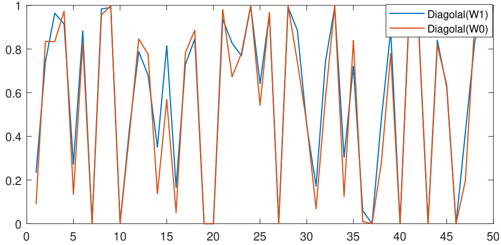
$$\hat{\beta}_{\alpha,1} = \left(\mathbf{C}^\top \mathbf{W}_1 \mathbf{C}\right)^{-1} \mathbf{C}^\top \mathbf{W}_1 \mathbf{y}.$$

Similar expressions can be used in the case of $\mathcal{H}_0$.

Results(Simulated data)

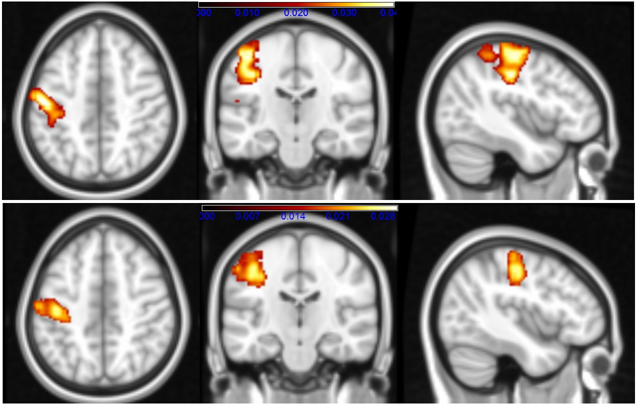
The performance is compared to the MSD in point mass noise, $p = t = 2, N = 200, \epsilon = 20\%, \alpha = 0.5$ and SNR $= -10dB$. Receiver operating characteristic (left) and wieghts of first 50 observations (right) are presented.





Results(Real fMRI data)

Active area of brain using Proposed detector on BPRFT (top) and ERRFT (bottom) data[1]. $\alpha = 0.9$, **H** is rank-1 signal (conv(HRF,stimulus)).



1. Seghouane et al.-2017: Sequential dictionary learning from correlated data: application to fMRI data analysis.

Questions?

Thanks for your attention!