SAND2019-8217PE

TENSOR-BASED INFERENTIAL PROCEDURES FOR VIDEO DATA

PRESENTED BY

Maximillian Chen (mgchen@sandia.gov)

Data Science & Applications Department

Systems Engineering Group, Inc.

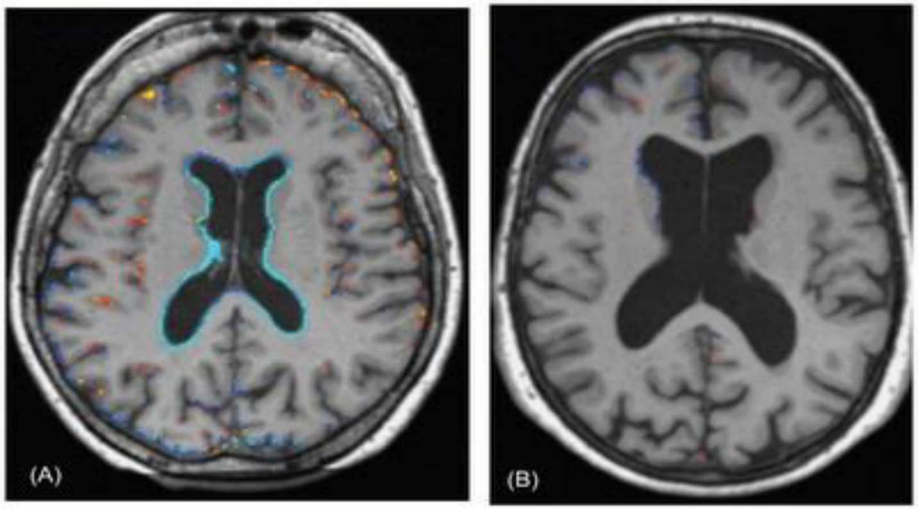
July 22, 2019

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

SAND2019

APPLICATION: ANALYSIS OF POPULATION OF IMAGES

- Suppose we want to test a treatment for Alzheimer's disease.
- Data: Brain images of two groups of Alzheimer's patients with multiple images per patient taken over time
- Two groups: treatment and control



Two questions:

- How do we analyze these high-dimensional images (120 x ~277,000) collectively?
- How do we determine if the treatment is effective?

APPLICATION TO TRAIN VIDEO DATA

BNSF train car

Single-stack train car

Double-stack train car

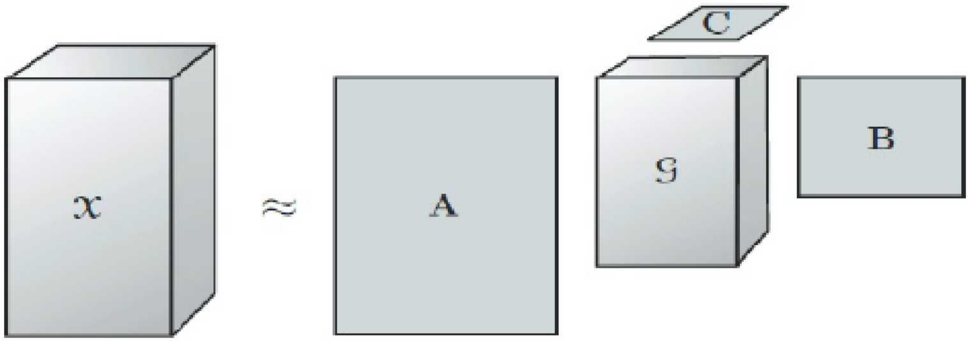
Two questions:

- How do we reduce the dimensions of these video frames?
- How do we detect significant differences between and within video segments?

OUTLINE

1. Three-way Tucker Decomposition and Tensor Preliminaries
2. Inferential Procedures
   - General Framework
   - Likelihood-Ratio Test
   - Score Test
   - Regression Based Inference
   - Simulations
   - Application to Train Video Data
3. Future Work

Tucker decomposition of a three-way array:



$$\mathfrak{X} \approx \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} \, \mathbf{a}_p \circ \mathbf{b}_q \circ \mathbf{c}_r = [[\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]]$$

$$x_{ijk} \approx \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr} a_{ip} b_{jq} c_{kr} \quad \text{for} \quad i = 1, \ldots, I, j = 1, \ldots, J, k = 1, \ldots, K$$

Question: Can we develop hypothesis testing procedures for this framework, where $\mathbf{C}$ measures the temporal correlation between frames?

MODE-n UNFOLDINGS

Tensors can be unfolded into various matrices that contain all of the elements of the tensor.

The best way to understand how this works is by considering an example. Let the frontal slides of a tensor $\mathcal{X} \in \mathbb{R}^{3\times 4\times 2}$ be

$$\mathbf{X}_1 = \begin{bmatrix} 1 & 4 & 7 & 10 \\ 2 & 5 & 8 & 11 \\ 3 & 6 & 9 & 12 \end{bmatrix}, \mathbf{X}_2 = \begin{bmatrix} 13 & 16 & 19 & 22 \\ 14 & 17 & 20 & 23 \\ 15 & 18 & 21 & 24 \end{bmatrix}.$$

Then the three mode-n unfoldings are

$$\mathbf{X}_{(1)} = \begin{bmatrix} 1 & 4 & 7 & 10 & 13 & 16 & 19 & 22 \\ 2 & 5 & 8 & 11 & 14 & 17 & 20 & 23 \\ 3 & 6 & 9 & 12 & 15 & 18 & 21 & 24 \end{bmatrix},$$

$$\mathbf{X}_{(2)} = \begin{bmatrix} 1 & 2 & 3 & 13 & 14 & 15 \\ 4 & 5 & 6 & 16 & 17 & 18 \\ 7 & 8 & 9 & 19 & 20 & 21 \\ 10 & 11 & 12 & 22 & 23 & 24 \end{bmatrix},$$

$$\mathbf{X}_{(3)} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & \dots & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & \dots & 21 & 22 & 23 & 24 \end{bmatrix}.$$

GAUSSIAN DISTRIBUTIONS

- Scalar: x
  - Distribution Name: Univariate Normal
  - Parameters: $\mathbf{X} \sim N(\mu, \Sigma)$, $\mu \in \mathbb{R}^k$, $\Sigma \in \mathbb{R}^{k \times k}$
  - Probability Distribution Function (PDF):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Vector: $\mathbf{x} = (x_1, ..., x_k)$
  - Distribution Name: Multivariate Normal
  - Parameters: $X \sim N(\mu, \sigma^2)$, $\mu \in (-\infty, \infty)$, $\Sigma \in \mathbb{R}^{k \times k}$, positive semi-definite
  - Probability Distribution Function (PDF):

$$f(\mathbf{x}) = (2\pi)^{\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)}$$

MATRIX NORMAL DISTRIBUTION

A random matrix X of dimensions $T \times F$ that follows the matrix normal distribution has the pdf

$$p(X|M, \Sigma, \Omega) = \frac{\exp(-\frac{1}{2}\text{tr}[\Sigma^{-1}(X - M)\Omega^{-1}(X - M)'])}{(2\pi)^{TF/2}|\Omega|^{T/2}|\Sigma|^{F/2}},$$

where M is the $T \times F$ mean matrix, Σ is the $T \times T$ row covariance matrix, and Ω is the $F \times F$ column covariance matrix.

The matrix normal distribution is related to the multivariate normal distribution.

$$X \sim MN_{T\times F}(M, \Sigma, \Omega)$$

if and only if

$$vec(X) \sim N_{TF}(vec(M), \Omega \otimes \Sigma).$$

TENSOR NORMAL DISTRIBUTION

Suppose $Y_1, ..., Y_r$ are dependent images that make up the *slices* of a tensor of order 3 that follows a tensor normal distribution with the following parameters:

$$\mathfrak{X} = Y_1, ..., Y_r, \mathcal{G} = V_1, ..., V_r, \mathbf{A}, \mathbf{B}, \mathbf{C} = \Omega,$$

where $\mathfrak{X} = Y_1, ..., Y_r$ is of size $T \times F \times r$, $\mathcal{G} = V_1, ..., V_r$ is of size $t \times f \times r$, $\mathbf{A} = P$ is of size $T \times t$, $\mathbf{B} = D$ is of size $f \times F$, and $\mathbf{C} = \Omega$ is of size $r \times r$.
Additionally, there will be the row and covariance matrices of Y_i, Σ and Ψ, respectively.

We assume that $\mathfrak{X}$ follows a *tensor normal distribution*, written as
$\mathfrak{X} \sim N_{T,F,r}(\mathfrak{B} \times \mathbf{A}, \mathbf{C}, \mathbf{D}, \Sigma, \Psi, \Omega)$.
The probability density function is

$$f_{\mathfrak{X}}(\mathfrak{X}) = (2\pi)^{-\frac{TFn}{2}} |\Sigma|^{-\frac{-Fn}{2}} |\Psi|^{-\frac{-Tn}{2}} |\Omega|^{-\frac{-TF}{2}}$$
$$\exp\{-\frac{1}{2}(\mathfrak{X} - \mathbf{A}\mathcal{B}\mathbf{C}) \times_{1\ldots3} (\circ_{j=1}^{3} U_j^{-1}) \times_{1\ldots3} (\mathfrak{X} - \mathbf{A}\mathcal{B}\mathbf{C})\},$$

where $\circ$ denotes the outer product and $\times$ denotes the tensor product.
A useful property of the tensor normal distribution:

$$\text{vec}(\mathfrak{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}) = (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathfrak{B}) \quad (1)$$
$$\sim N_{TFr}((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})vec(\mathfrak{B}), \Omega \otimes \Psi \otimes \Sigma) \quad (2)$$

ONE-SAMPLE PROBLEM

Suppose we have n i.i.d. tensors, denoted as $\mathfrak{X}_i$, $i = 1, \ldots, n$, that follow a tensor normal distribution. For a one-sample hypothesis testing problem, we assume that $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$ are fixed and computed.

Assumptions:

- One population:

$$\begin{aligned} \mathfrak{X}_i &= \mathfrak{B} \times \mathbf{A}, \mathbf{C}, \mathbf{D} + \mathfrak{E}_i, i = 1, \ldots, n \\ \mathfrak{X} &\sim N_{T,F,r}(\mathfrak{B} \times \mathbf{A}, \mathbf{C}, \mathbf{D}, \Sigma, \Psi, \Omega) \\ \Rightarrow \text{vec}(\mathfrak{B} \times \mathbf{A}, \mathbf{C}, \mathbf{D}) &\sim N_{TFr}((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A}) vec(\mathfrak{B}), \Omega \otimes \Psi \otimes \Sigma) \\ &= (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A}) \text{vec}(\mathfrak{B}) \end{aligned}$$

- Row covariance matrix Σ is $T \times T$, fixed, p.d., but unknown
- Column covariance matrix Ψ is $F \times F$, fixed, p.d., but unknown
- Slice covariance matrix Ω is $r \times r$, fixed, p.d., but unknown

We are testing

$$\begin{aligned} H_0 &: \mathcal{B} = \mathcal{B}_0 \\ H_a &: \mathcal{B} \neq \mathcal{B}_0 \end{aligned}$$

MAXIMUM-LIKELIHOOD ESTIMATION

We assume that $\mathbf{A}$ and $\mathbf{B}$ are calculated and fixed. We assume the covariances matrices Σ, Ψ, and Ω are unknown.

Nzabanita et al (2015) derive the maximum likelihood estimates for all of the parameters of the tensor normal distribution when the mean has the structure $\mathcal{M} = \mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$.

These methods *unfold* the tensor into matrices. The model

$$\mathcal{X} \sim N_{T,F,r}(\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}, \Sigma, \Psi, \Omega) \tag{3}$$

in matrix form using three different modes as

$$\mathbf{X}_{(1)} \sim N_{T,Fr}(\mathbf{A}\mathbf{B}_{(1)}(\mathbf{D} \otimes \mathbf{C})', \Sigma, \Omega \otimes \Psi) \tag{4}$$

$$\mathbf{X}_{(2)} \sim N_{F,Tr}(\mathbf{C}\mathbf{B}_{(2)}(\mathbf{D} \otimes \mathbf{A})', \Psi, \Omega \otimes \Sigma) \tag{5}$$

$$\mathbf{X}_{(3)} \sim N_{r,TF}(\mathbf{D}\mathbf{B}_{(3)}(\mathbf{C} \otimes \mathbf{A})', \Omega, \Psi \otimes \Sigma). \tag{6}$$

The likelihoods for each of the three modes are equivalent. MLEs for all of these parameters ($\mathbf{B}_{(1)}$, $\mathbf{B}_{(2)}$, $\mathbf{B}_{(3)}$, Σ, Ψ, Ω) are derived.

ONE-SAMPLE PROBLEM: LIKELIHOOD-RATIO TEST

The likelihood using the mode-1 model with

$$\mathbf{X}_{(1)} \sim N_{T,Fr}(\mathbf{A}\mathbf{B}_{(1)}(\mathbf{D} \otimes \mathbf{C})', \Sigma, \Omega \otimes \Psi)$$

is

$$L(B_{(1)}|\mathbf{A}, \mathbf{C}, \mathbf{D}, x_{(1),1}, ..., x_{(1),n})$$
$$= \frac{\exp(-\frac{1}{2}\sum_{i=1}^{n} \operatorname{tr}\{(\Omega \otimes \Psi)^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1)}(\mathbf{D} \otimes \mathbf{C})']'\Sigma^{-1}[(\mathbf{D} \otimes \mathbf{C})(\mathbf{D} \otimes \mathbf{C})']\})}{(2\pi)^{TFrn/2}|\Omega \otimes \Psi|^{nT/2}|\Sigma|^{TFr/2}}.$$

We want to compute the likelihood-ratio test statistic

$$\Lambda = \frac{L_{H_0}}{L_{H_a}}$$
$$= (\frac{\hat{\Omega}_A \otimes \hat{\Psi}_A}{\hat{\Omega}_0 \otimes \hat{\Psi}_0})^{\frac{nT}{2}} (\frac{\hat{\Sigma}_A}{\hat{\Sigma}_0})^{\frac{nFr}{2}}.$$

ONE-SAMPLE PROBLEM: MAXIMUM-LIKELIHOOD ESTIMATION

$$\hat{\mathbf{B}}_{(1)} = \frac{1}{N}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\sum_{i=1}^{n}\mathbf{A}'\Sigma^{-1}\mathbf{X}_{(1),i}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1}$$

$$= (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1}\bar{\mathbf{X}}_{(1)}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1}$$

$$\hat{\Sigma} = \frac{\sum_{i=1}^{n}(\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1)}(\mathbf{D}\otimes\mathbf{C})')'(\hat{\Omega}\otimes\hat{\Psi})^{-1}(\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1)}(\mathbf{D}\otimes\mathbf{C})')}{nFr}$$

$$\hat{\Psi} = \frac{\sum_{i=1}^{n}(\mathbf{X}_{(2),i} - \mathbf{C}\mathbf{B}_{(2)}(\mathbf{D}\otimes\mathbf{A})')'(\hat{\Omega}\otimes\hat{\Sigma})^{-1}(\mathbf{X}_{(2),i} - \mathbf{C}\mathbf{B}_{(2)}(\mathbf{D}\otimes\mathbf{A})')}{nTr}$$

$$\hat{\Omega} = \frac{\sum_{i=1}^{n}(\mathbf{X}_{(3),i} - \mathbf{D}\mathbf{B}_{(3)}(\mathbf{C}\otimes\mathbf{A})')'(\hat{\Psi}\otimes\hat{\Sigma})^{-1}(\mathbf{X}_{(3),i} - \mathbf{D}\mathbf{B}_{(3)}(\mathbf{C}\otimes\mathbf{A})')}{nTF}$$

An iterative algorithm is used to estimate $\hat{\mathbf{B}}_{(1)}$, $\hat{\Sigma}$, $\hat{\Psi}$, and $\hat{\Omega}$.

ASYMPTOTIC DISTRIBUTION OF -2 LOG Λ

Because we are testing

$$H_0 : \mathcal{B} = \mathcal{B}_0$$
$$H_a : \mathcal{B} \neq \mathcal{B}_0$$

we have a simple null hypothesis. Therefore, by Wilks' Theorem, as $n \to \infty$,

$$-2 \log \Lambda \sim \chi^2_{tfr}.$$

ONE-SAMPLE PROBLEM: FORMULATION AS REGRESSION PROBLEM

Recall

$$\hat{\mathbf{B}}_{(1)} = (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\sum_{i=1}^{n}\mathbf{A}'\sigma^{-1}\bar{\mathbf{X}}_{(1)}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1}$$

Then we can formulate the following regression problem:

$$\underbrace{\text{vec}(\bar{\mathbf{X}}_{(1)})}_{Y} = \underbrace{(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{A})}_{X}\underbrace{\text{vec}(\mathbf{B}_{(1)})}_{\beta} + \underbrace{\text{vec}(\mathbf{E})}_{\epsilon}$$

If the errors are not homoscedastic ($\text{vec}(\mathbf{E}) \sim N(0, \frac{1}{n}\Omega\otimes\Psi\otimes\Sigma)$, where none of Σ, Ψ, and Ω are equal to $\sigma^2 I$), then we take the Cholesky decomposition to get a matrix C such that

$$C'C = (\frac{1}{n}\Omega\otimes\Psi\otimes\Sigma)^{-1} = n\Omega^{-1}\otimes\Psi^{-1}\otimes\Sigma^{-1}$$

and process like we would for generalized least squares.

$$C\text{vec}(\bar{\mathbf{X}}_{(1)}) = C(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{A})\text{vec}(\mathbf{B}_{(1)}) + C\text{vec}(\mathbf{E})$$

Test statistic: $F \sim F_{tfr,TFr-tfr}$

ONE-SAMPLE PROBLEM: SCORE TEST

Under H_0, $\mathcal{X}_i = \mathcal{B}_0 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\} + \mathcal{E}$

$$L(\mathbf{B}_{(1)}|\mathbf{A}, \mathbf{C}, \mathbf{D}, x_{(1),1}, ..., x_{(1),n})$$

$$= \frac{\exp(-\frac{1}{2}\sum_{i=1}^{n} \mathrm{tr}\{(\Omega \otimes \Psi)^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1)}(\mathbf{D}\otimes\mathbf{C})']'\Sigma^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1)}(\mathbf{D}\otimes\mathbf{C})']\})}{(2\pi)^{TFrn/2}|\Omega\otimes\Psi|^{nT/2}|\Sigma|^{TFr/2}}$$

$$l(\mathbf{B}_{(1)}|\mathbf{A}, \mathbf{C}, \mathbf{D}, x_{(1),1}, ..., x_{(1),n}) = -\frac{1}{2}\sum_{i=1}^{n}[\{\mathrm{vec}(\mathbf{X}_{(1),i})' - [(\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{A})\mathrm{vec}(\mathbf{B}_{(1)})]'\}\times$$

$$(\Omega^{-1}\otimes\Psi^{-1}\otimes\Sigma^{-1})\times\{\mathrm{vec}(\mathbf{X}'_{(1),i}) - (\mathbf{A}\otimes\mathbf{D}\otimes\mathbf{C})\mathrm{vec}(\mathbf{B}'_{(1)})\}]$$

$$-\frac{nTFr}{2}\log(2\pi) - \frac{nT}{2}\log|\Omega\otimes\Psi| - \frac{nFr}{2}\log|\Sigma|$$

$$U(\mathbf{B}_{(1)}) = \frac{\partial l}{\partial \mathbf{B}_{(1)}} = (\mathbf{D}'\Omega^{-1}\otimes\mathbf{C}'\Psi^{-1}\otimes\mathbf{A}'\Sigma^{-1})\sum_{i=1}^{n}[\mathrm{vec}(\mathbf{X}_{(1),i}) - (\mathbf{D}\otimes\mathbf{C}\otimes\mathbf{A})\mathrm{vec}(\mathbf{B}_{(1)})]$$

$$\frac{\partial^2 l}{\partial \mathbf{B}_{(1)}^2} = -n(\mathbf{D}'\Omega^{-1}\mathbf{D}\otimes\mathbf{C}'\Psi^{-1}\mathbf{C}\otimes\mathbf{A}'\Sigma^{-1}\mathbf{A})$$

$$I(\mathbf{B}_{(1)}) = -E[-n(\mathbf{D}'\Omega^{-1}\mathbf{D}\otimes\mathbf{C}'\Psi^{-1}\mathbf{C}\otimes\mathbf{A}'\Sigma^{-1}\mathbf{A})] = n(\mathbf{D}'\Omega^{-1}\mathbf{D}\otimes\mathbf{C}'\Psi^{-1}\mathbf{C}\otimes\mathbf{A}'\Sigma^{-1}\mathbf{A}$$

ONE-SAMPLE PROBLEM: SCORE TEST STATISTIC

$$U(\mathbf{B}_{(1),0})' I(\mathbf{B}_{(1),0})^{-1} U(\mathbf{B}_{(1),0})$$

$$= \{\sum_{i=1}^{n}[\text{vec}(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}_{(1),0})]\}' \times$$

$$\frac{1}{n}(\mathbf{D}'\Omega^{-1}(\mathbf{D}'^{\Omega^{-1}\mathbf{D}})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'\Sigma^{-1}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1}$$

$$\{\sum_{i=1}^{n}[\text{vec}(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}_{(1),0})]\}$$

$$\sim \chi^2_{tfr}$$

ONE-SAMPLE TESTS: SIMULATIONS

Model:

$$\mathcal{X} = \mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\} + \mathcal{E}$$

Simulated Data (under $H_0 : \mathcal{B} = V_0, \ldots V_0$):

$$\mathcal{X} \sim N_{T,F,r}(\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}, \Sigma, \Psi, \Omega)$$

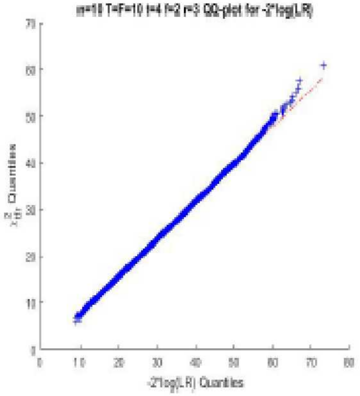
where

- $\mathcal{X}$ is a $10 \times 10 \times 3$ tensor ($T, F = 10$, $r = 3$),
- $\mathbf{A}$ is a 10×4 arbitrary, orthogonal matrix ($t = 4$),
- $\mathcal{B}_0$ is a $4 \times 2 \times 3$ tensor, with each slice consisting of the 4×2 matrix B_0
- B_0 is a 4×2 matrix consisting of independent $N(0, 10^2)$ observations,
- $\mathbf{C}$ is a 2×10 arbitrary, orthogonal matrix ($f = 2$),
- $\mathbf{D}$ is a 3×3 arbitrary, orthogonal matrix,
- $\mathcal{E}$ is a $10 \times 10 \times 3$ tensor with $N_{T,F,r}(0, \Sigma, \Psi, \Omega)$ distribution,
- Σ is a 10×10 arbitrary symmetric, positive-definite covariance matrix,
- Ψ is a 10×10 arbitrary symmetric, positive-definite covariance matrix,
- Ω is a 3×3 arbitrary symmetric, positive-definite covariance matrix.

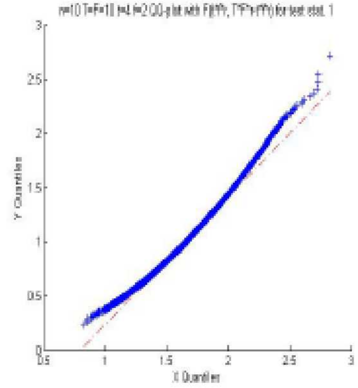
We run 10,000 simulations under $H_0 : \mathcal{B} = B_0, \ldots B_0$ using MATLAB.

ONE-SAMPLE TESTS: SIMULATION RESULTS

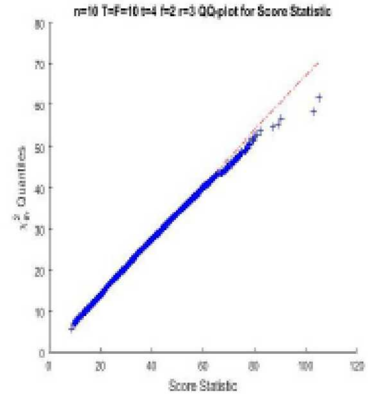
LRT $(-2\log\Lambda \sim \chi^2_{tfr})$:



Regression Test $(F \sim F_{tfr,TFr-tfr})$:



Score Test $(U(B_0)'I(B_0)^{-1}U(B_0) \sim \chi^2_{tfr})$:



k-SAMPLE PROBLEM

Assumptions:

- Population g: $g = 1, ..., k$ (We have k independent populations)

$$\mathfrak{X}_i^g = \mathfrak{B}_g \times \mathbf{A}, \mathbf{C}, \mathbf{D} + \mathfrak{E}_i, i = 1, ..., N_g(\sum_{i=1}^{g} N_g = N)$$

$$\mathfrak{X}_i^g \sim N_{p,q,r}(\mathfrak{B}_g \times \mathbf{A}, \mathbf{C}, \mathbf{D}, \Sigma, \Psi, \Omega)$$

$$\Rightarrow \text{vec}(\mathfrak{B}_i^g \times \mathbf{A}, \mathbf{C}, \mathbf{D}) \sim N_{prq}(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})vec(\mathfrak{B}_g), \Omega \otimes \Psi \otimes \Sigma)$$

$$= (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathfrak{B}_g)$$

- Common $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$ for all populations
- Row covariance matrix Σ is $T \times T$, fixed, p.d., but unknown
- Column covariance matrix Ψ is $F \times F$, fixed, p.d., but unknown
- Slice covariance matrix Ω is $r \times r$, fixed, p.d., but unknown

We are testing

$$H_0 : \mathcal{B}_1 = \mathcal{B}_2 = ... = \mathcal{B}_k$$

$$H_a : \text{At least one of } \mathcal{B}_1, ..., \mathcal{B}_k \text{ is different.}$$

k-SAMPLE PROBLEM: MAXIMUM-LIKELIHOOD ESTIMATION

Let $n_1, ..., n_k$ denote the total *cumulative* sample size up to and including sample k. $n = \sum_{i=1}^{k} n_i$, and $\mathbf{B}_{(1),g}$, $g = 1, ..., k$ denote the $\mathbf{B}_{(1)}$ value corresponding to group g.

$$\hat{\mathbf{B}}_{(1),1} = \frac{1}{n_1}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\sum_{i=1}^{n_1}\mathbf{A}'\Sigma^{-1}\mathbf{X}_{(1),i}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-}$$

$$= (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1}\bar{\mathbf{X}}_{(1),1}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1}$$

$$\hat{\mathbf{B}}_{(1),g} = \frac{1}{n_g - n_{g-1}}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\sum_{i=n_{g-1}+1}^{n_g}\mathbf{A}'\Sigma^{-1}\mathbf{X}_{(1),i}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})\times$$

$$[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1}$$

$$= (\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1}\bar{\mathbf{X}}_{(1),g}(\Omega\otimes\Psi)^{-1}(\mathbf{D}\otimes\mathbf{C})[(\mathbf{D}\otimes\mathbf{C})'(\Omega\otimes\Psi)(\mathbf{D}\otimes\mathbf{C})]^{-1},$$

$g = 2, ..., k$

k-SAMPLE PROBLEM: DISTRIBUTION OF TEST STATISTICS

- Likelihood-ratio test (asymptotic distribution):

$$-2\log\Lambda \sim \chi^2_{(k-1)tfr}.$$

- Regression problem framework: Test statistic $F \sim F_{ktfr,kTFr-ktfr}$
- Score test: Cannot conclude theoretically that $U(\hat{\mathbf{B}_{(1)}})' I^{-1}_{\hat{\mathbf{B}_{(1)}}} U(\hat{\mathbf{B}_{(1)}}) \sim \chi^2_{tfr}$ exactly or $U(\hat{\mathbf{B}_{(1)}})' I^{-1}_{\hat{\mathbf{B}_{(1)}}} U(\hat{\mathbf{B}_{(1)}}) \sim \chi^2_{(k-1)tfr}$ asymptotically as $n_k \to \infty$

TRAIN VIDEO DATASET

- Publicly available YouTube video at the following link: https://www.youtube.com/watch?v=tNT2iQZ1WiI.
- 32-minute video consists of Amtrak, BNSF, and Metrolink trains in Santa Fe Springs, CA taken on 12/13/14.
- We take three segments of the video consisting of a BNSF train, single-stack train, and double-stack train.
- Each segment is 10 frames, and each image, X_i, $i = 1, ..., n$, is 71×101 in size.
- Following the work of Lock et al (2011), we scale our data so that all 30 observations have the same total variability. Letting $\bar{x}_i$ be the mean and s_i be the standard deviation of the entries of X_i, define

$$X_i^{\text{scaled}} = \frac{X_i - \bar{x}_i}{s_i}.$$

We scale all of our 30 images based on the above definition.

APPLICATION TO TRAIN VIDEO DATA: DIMENSION REDUCTION AND INFERENTIAL PROCEDURES

- For each video segment, we have the tensor $\mathcal{X}_i, i = 1, ..., 40$ of size $71 \times 101 \times 10$ which incorporates all 10 frames from each video. Thus, we have three tensors.
- Using obtained values of $t = 25$ and $f = 30$ using previously established methods, we compute the Tucker decomposition on the first tensor (segment 1's images) to calculate $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$.
- With these fixed and estimated, apply inferential procedures on dataset of 3 tensors.
  - Three types of problems: one-, two-, and three-sample problems

APPLICATION TO TRAIN VIDEO DATA: ONE-SAMPLE TESTS

We wish to determine if all three video segments have the same mean, i.e. have the same mean of $\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$. With $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$ being estimated and fixed, we want to see if they all have the same value of $\mathcal{B}$. To make this determination, we test the hypotheses

$$H_0 : \mathcal{B} = \mathcal{B}_0$$
$$H_0 : \mathcal{B} \neq \mathcal{B}_0,$$

where we will set $\mathcal{B}_0$ to be the set of images for video segment 1 (BNSF train).

APPLICATION TO TRAIN VIDEO DATA: ONE-SAMPLE TESTS

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (asympt. dist.)	$-2 \log \Lambda \sim \chi^2_{tfr}$	7.7026×10^3	5.7513×10^4	Reject H_0
Score	$U(\mathbf{B}_0)' I(\mathbf{B}_0)^{-1} U(\mathbf{B}_0) \sim \chi^2_{tfr}$	7.7026×10^3	1.4488×10^5	Reject H_0
Regression	$F \sim F_{tfr, TFr-tfr}$	1.0286	7.4041	Reject H_0

- $\mathbf{B}_0 = \mathbf{B}_{0,(1)}$
- $T = 71,\ F = 101,\ r = 10$
- $t = 25,\ f = 30,\ r = 10$
- $n = 3$

APPLICATION TO TRAIN VIDEO DATA: TWO-SAMPLE TESTS

We seek to determine if there is a significant difference in the means of the images for the video segment of the BNSF train, and the video segment of the single-stack train.

- Population 1: images of BNSF train video segment ($n_1 = 1$)
- Population 2: images of single-stack video segment ($n_2 = 2$)

With $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$ being estimated and fixed, if the mean for population 1 is $\mathcal{B}_1 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$ and the mean for population 2 is $\mathcal{B}_2 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$, then we want to see if $\mathcal{B}_1 = \mathcal{B}_2$. Therefore, we test the hypotheses

$$H_0 : \mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}$$
$$H_0 : \mathcal{B}_1 \neq \mathcal{B}_2.$$

APPLICATION TO TRAIN VIDEO DATA: TWO-SAMPLE TESTS

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (asympt. dist.)	$-2 \log \Lambda \sim \chi^2_{tfr}$	7.7026×10^3	1.1928×10^5	Reject H_0
Regression	$F \sim F_{2tf, 2TF-2tf}$	1.0202	8.4154	Reject H_0

- $\mathbf{B}_0 = \mathbf{B}_{0,(1)}$
- $T = 71, \ F = 101, \ r = 10$
- $t = 25, \ f = 30, \ r = 10$
- $n_2 = 2, \ k = 2$

APPLICATION TO TRAIN VIDEO DATA: THREE-SAMPLE TESTS

We seek to determine if there is a significant difference in the means of the images for:

- Population 1: BNSF train video segment ($n_1 = 1$)
- Population 2: single-stack video segment ($n_2 = 2$)
- Population 3: double-stack video segment ($n_3 = 3$)

With $\mathbf{A}$, $\mathbf{C}$, and $\mathbf{D}$ being estimated and fixed, if the mean for population 1 is $\mathcal{B}_1 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$, the mean for population 2 is $\mathcal{B}_2 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$, and the mean for population 3 is $\mathcal{B}_3 \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}$, then we want to see if $\mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}_3$. Therefore, we test the hypotheses

$$H_0 : \mathcal{B}_1 = \mathcal{B}_2 = \mathcal{B}_3 = \mathcal{B}$$

$$H_0 : \text{At least one of } \mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3 \text{ is not equal.}$$

APPLICATION TO TRAIN VIDEO DATA: THREE-SAMPLE TESTS

Test	Dist. of Test Statistic	Critical Value ($\alpha = 0.05$)	Test Statistic	Decision
LRT (asympt. dist.)	$-2 \log \Lambda \sim \chi^2_{(k-1)tfr}$	1.5286×10^4	6.5350×10^4	Reject H_0
Regression	$F \sim F_{3tfr, 3TFr-3tfr}$	1.0165	6.5293	Reject H_0

- $\mathbf{B}_0 = \mathbf{B}_{0,(1)}$
- $T = 71,\ F = 101,\ r = 10$
- $t = 25,\ f = 30,\ r = 10$
- $n_3 = 3,\ k = 3$

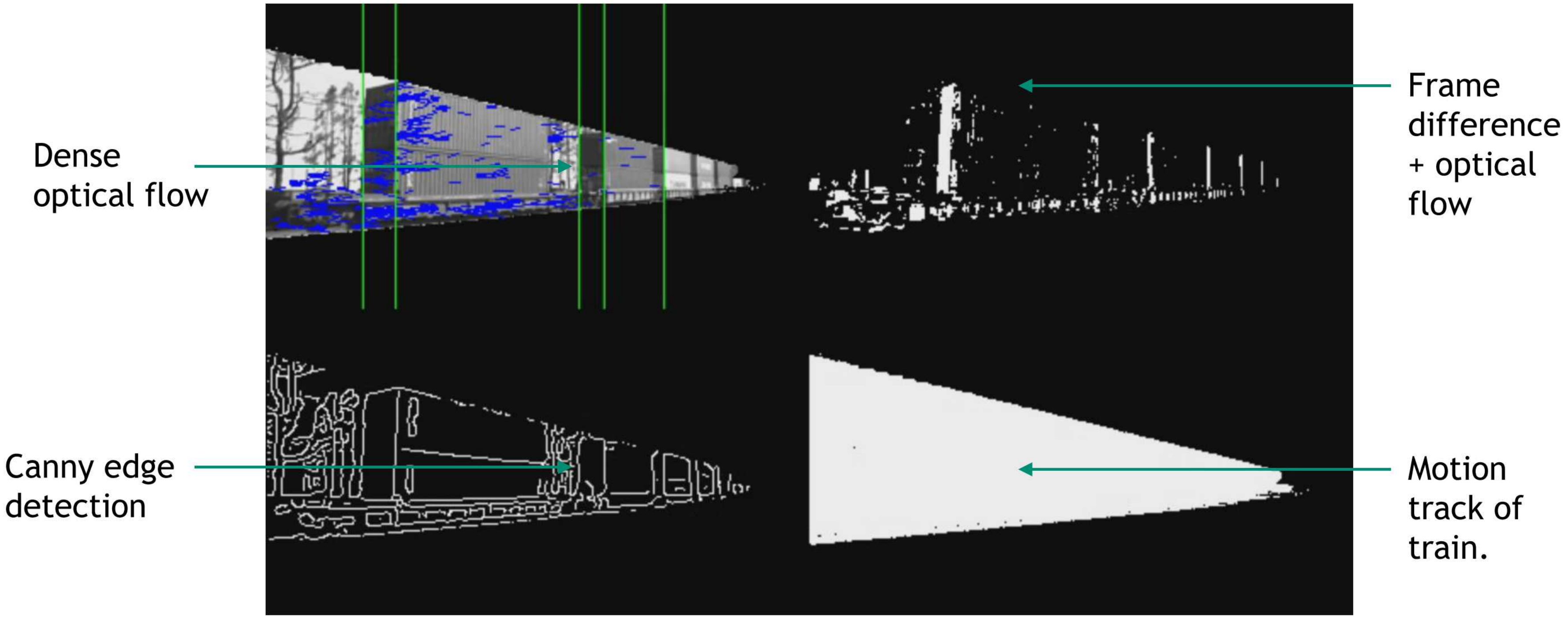
DISCUSSION OF RESULTS

- Regression-based inference test, which has the most solid mathematical support, serves as the reference test. This test rejects H_0 for one-, two-, and three-sample problems, which is the expected result.
- Asymptotic distribution for LRT and score test (one-sample) reject H_0.

FUTURE WORK

Integration of methods with feature detection methods in computer vision

- Deep learning
- Unsupervised machine learning



FUTURE WORK

- Develop hypothesis testing procedures without unfolding tensors
- Computational issues, especially with Kronecker products of covariance matrices
- Inferential procedures for other tensor distributions, including nonparametric methods
- Goodness-of-fit tests for tensor distributions
- Using other tensor decomposition methods, such as CANDLECOMP and PARAFRAC
- Hierarchical hypothesis testing

ACKNOWLEDGEMENTS

- This work is funded by the Sandia National Laboratories Laboratory Directed Research and Development (LDRD) program as part of the Inferential and Feature Selection Methods for Video Imaging project.
- Thank you to Aleksander Bapst, Daniel Dunlavy, and Tamara Kolda for helpful discussions.

APPENDIX SLIDES

ONE-SAMPLE LIKELIHOOD-RATIO TEST: DERIVING DISTRIBUTION OF TEST STATISTIC

Test statistic:

$$\Lambda = \frac{\sup_{\mathbf{B}_{(1),0}} L(\theta|\mathcal{X}_{(1),i})}{\sup_{\mathbf{B}_{(1)}} L(\theta|\mathcal{X}_{(1),i})} = (\frac{\hat{\Omega}_A \otimes \hat{\Psi}_A}{\hat{\Omega}_0 \otimes \hat{\Psi}_0})^{\frac{nT}{2}} (\frac{\hat{\Sigma}_A}{\hat{\Sigma}_0})^{\frac{nFr}{2}}$$

where

$$\hat{\Sigma}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i} - \mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D} \otimes \mathbf{C})')](\hat{\Omega}_A \otimes \hat{\Psi}_A)^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D} \otimes \mathbf{C})')]'}{nFr}$$

$$\hat{\Psi}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(2),i} - \mathbf{C}\hat{\mathbf{B}}_{(2)}(\mathbf{D} \otimes \mathbf{A})')](\hat{\Omega}_A \otimes \hat{\Sigma}_A)^{-1}[\mathbf{X}_{(2),i} - \mathbf{C}\hat{\mathbf{B}}_{(2)}(\mathbf{D} \otimes \mathbf{A})')]'}{nTr}$$

$$\hat{\Omega}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(3),i} - \mathbf{D}\hat{\mathbf{B}}_{(3)}(\mathbf{C} \otimes \mathbf{A})')](\hat{\Psi}_A \otimes \hat{\Sigma}_A)^{-1}[\mathbf{X}_{(3),i} - \mathbf{D}\hat{\mathbf{B}}_{(3)}(\mathbf{C} \otimes \mathbf{A})')]'}{nTF}$$

$$\hat{\Sigma}_0 = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D} \otimes \mathbf{C})')](\hat{\Omega}_0 \otimes \hat{\Psi}_0)^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D} \otimes \mathbf{C})')]'}{nFr}$$

$$\hat{\Psi}_0 = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(2),i} - \mathbf{C}\mathbf{B}_{(2),0}(\mathbf{D} \otimes \mathbf{A})')](\hat{\Omega}_0 \otimes \hat{\Sigma}_0)^{-1}[\mathbf{X}_{(2),i} - \mathbf{C}\mathbf{B}_{(2),0}(\mathbf{D} \otimes \mathbf{A})')]'}{nTr}$$

$$\hat{\Omega}_0 = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(3),i} - \mathbf{D}\mathbf{B}_{(3),0}(\mathbf{C} \otimes \mathbf{A})')](\hat{\Psi}_0 \otimes \hat{\Sigma}_0)^{-1}[\mathbf{X}_{(3),i} - \mathbf{D}\mathbf{B}_{(3),0}(\mathbf{C} \otimes \mathbf{A})')]'}{nTF}.$$

ONE-SAMPLE LIKELIHOOD-RATIO TEST: DERIVING DISTRIBUTION OF TEST STATISTIC

Dependency issues in numerator and denominator:

$$\left(\frac{\hat{\Sigma}_A}{\hat{\Sigma}_0}\right)^{\frac{nFr}{2}} = \left(\frac{\frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i}-\mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D}\otimes\mathbf{C})')](\hat{\Omega}_A\otimes\hat{\Psi}_A)^{-1}[\mathbf{X}_{(1),i}-\mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D}\otimes\mathbf{C})')]'}{nFr}}{\frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i}-\mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D}\otimes\mathbf{C})')](\hat{\Omega}_0\otimes\hat{\Psi}_0)^{-1}[\mathbf{X}_{(1),i}-\mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D}\otimes\mathbf{C})')]'}{nFr}}\right)^{\frac{nFr}{2}}$$

$$= \left(\frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i}-\mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D}\otimes\mathbf{C})')](\hat{\Omega}_A\otimes\hat{\Psi}_A)^{-1}[\mathbf{X}_{(1),i}-\mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D}\otimes\mathbf{C})')]'}{\sum_{i=1}^{n}[\mathbf{X}_{(1),i}-\mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D}\otimes\mathbf{C})')](\hat{\Omega}_0\otimes\hat{\Psi}_0)^{-1}[\mathbf{X}_{(1),i}-\mathbf{A}\mathbf{B}_{(1),0}(\mathbf{D}\otimes\mathbf{C})')]'}\right)^{\frac{nFr}{2}}$$

RELATIONSHIP TO WISHART DISTRIBUTION

Theorem 7.8.4 of Gupta and Nagar (2000): Let $S = XAX'$, where $X \sim N_{p,n}(M, \Sigma, \Omega)$. The necessary and sufficient condition for S to be distributed as $W_p(t, \Sigma, \Sigma^{-1}MAM')$ is that $A\Omega A = A$ and rank(A)=$t \geq p$.

Corollary 7.8.4.1 of Gupta and Nagar (2000): The necessary and sufficient condition for $S = XAX'$, where $X \sim N_{p,n}(0, \Sigma, \Omega)$ to be distributed as $W_p(t, \Sigma)$ is that $A\Omega A = A$ and rank(A)=$t \geq p$.

ONE-SAMPLE SCORE TEST: DISTRIBUTION OF TEST STATISTIC

Theorem

The score statistic

$$U(\mathbf{B}_{(1),0})' I(\mathbf{B}_{(1),0})^{-1} U(\mathbf{B}_{(1),0})$$

$$= \{\sum_{i=1}^{n}[vec(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})vec(\mathbf{B}_{(1)})]\}' \times$$

$$\frac{1}{n}(\Omega^{-1}\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \Psi^{-1}\mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \Sigma^{-1}\mathbf{A}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1})$$

$$\sum_{i=1}^{n}[vec(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})vec(\mathbf{B}_{(1)})]$$

follows a χ^2_{tfr} *distribution.*

ONE-SAMPLE SCORE TEST: DISTRIBUTION OF TEST STATISTIC

Proof.
Because $\mathbf{X}_{(1),i} \sim N_{T,F,r}(\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}, \Sigma, \Psi, \Omega)$ under H_0, we know that

$$\text{vec}(\mathbf{X}_{(1),i}) \sim N_{TFr}((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}_{(1),0}), \Omega \otimes \Psi \otimes \Sigma)$$

$$\text{vec}(\mathbf{X}_{(1),i}) - ((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}_{(1),0}) \sim N_{TF}(0, \Omega \otimes \Psi \otimes \Sigma)$$

$$\sum_{i=1}^{n}[\text{vec}(\mathbf{X}_{(1),i}) - ((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathbf{B}_{(1),0})] \sim N_{TF}(0, \Omega \otimes \Psi \otimes n\Sigma).$$

Let A denote the constant term in the middle of the score statistic, and Ψ denote the column covariance matrix of $\sum_{i=1}^{n}[\text{vec}(\mathbf{X}_{(1),i}) - ((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathfrak{B}_{(1),0})]$. By Theorem 7.8.4 in Gupta and Nagar (2000), because we set

$$A = \frac{1}{n}(\Omega^{-1}\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \Psi^{-1}\mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \Sigma^{-1}\mathbf{A}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma$$

$$\Psi = \Omega \otimes \Psi \otimes n\Sigma,$$

then

$$A\Psi A = A,$$

and we can conclude that $U(\mathbf{B}_{(1),0})'I(\mathbf{B}_{(1),0})^{-1}U(\mathbf{B}_{(1),0}) \sim \chi^2_{tfr}$. □

k-SAMPLE SCORE TEST: DISTRIBUTION OF TEST STATISTIC

The score statistic $U(\hat{\mathbf{B}}_{(1),0})' I(\hat{\mathbf{B}}_{(1),0})^{-1} U(\hat{\mathbf{B}}_{(1),0})$ is

$$U(\hat{\mathbf{B}}_{(1),0})' I(\hat{\mathbf{B}}_{(1),0})^{-1} U(\hat{\mathbf{B}}_{(1),0})$$
$$= \{\sum_{i=1}^{n_k} [\text{vec}(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\hat{\mathbf{B}}_{(1),0})]\}' \times$$
$$\frac{1}{n_k}(\Omega^{-1}\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \Psi^{-1}\mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \Sigma^{-1}\mathbf{A}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'\Sigma^{-1}$$
$$\sum_{i=1}^{n_k} [\text{vec}(\mathbf{X}_{(1),i}) - (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\hat{\mathbf{B}}_{(1),0})]$$

follows a χ^2_{tfr} distribution exactly.

k-SAMPLE SCORE TEST: DISTRIBUTION OF TEST STATISTIC

Attempting to use Theorem 7.8.4 of Gupta and Nagar (2000), setting

$$A = \frac{1}{n_k}(\Omega^{-1}\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \Psi^{-1}\mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \Sigma^{-1}\mathbf{A}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'$$

$$\Psi = n_k^2(\Omega \otimes \Psi \otimes \Sigma) - n_k[\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'\Sigma^{-1}\mathbf{A})^{-}$$

then $A\Psi A \neq A$, we cannot conclude that the score statistic follows a follows a χ^2_{tfr} distribution.

LIKELIHOOD-RATIO TEST

Likelihood function:

- How likely particular values of statistical parameters are for a given set of observations
- Equal to joint probability distribution function (pdf)

$$L(\theta|x) = P(x|\theta)$$

- Let $x_1, ..., x_n$ be independent and identically distributed (i.i.d.) data with pdf $p(x_i|\theta), i = 1, ..., n$. Then

$$L(\theta|x) = P(x|\theta) = \prod_{i=1}^{n} p(x_i|\theta).$$

LIKELIHOOD-RATIO TEST

Likelihood-Ratio Test:

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)} = \frac{L(\theta_0|x)}{L(\theta|x)}$$

Reject H_0 if $\lambda(x) < c$.

Wilks' Theorem:
Under H_0, as $n \to \infty$,

$$-2 \log \lambda \sim \chi^2_{df},$$

where df = dim(Θ) - dim(Θ_0).

REGRESSION-BASED TEST

Model:

$$\underbrace{\text{vec}(\bar{\mathbf{X}}_{(1)})}_{Y} = \underbrace{(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})}_{X} \underbrace{\text{vec}(\mathbf{B}_{(1)})}_{\beta} + \underbrace{\text{vec}(\mathbf{E})}_{\epsilon}.$$

We can rewrite $H_0 : \mathcal{B} = \mathcal{B}_0$ as

$$CB = 0$$

$$C = \begin{bmatrix} I_{tfr} & -I_{tfr} \end{bmatrix}$$

$$B = \begin{bmatrix} \text{vec}(\mathcal{B}) \\ \text{vec}(\mathcal{B}_0) \end{bmatrix}$$

$$CB = \text{vec}(\mathcal{B}) - \text{vec}(\mathcal{B}_0) = \text{vec}(\mathcal{B} - \mathcal{B}_0) = 0.$$

REGRESSION-BASED TEST

Suppose, under H_0,

$$\mathcal{X}_i \sim N_{T,F,r}(\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}, \Sigma, \Psi, \Omega),$$

where Σ, Ψ, and Ω are positive-definite, Then

$$\bar{\mathcal{X}} \sim N_{T,F,r}(\mathcal{B} \times \{\mathbf{A}, \mathbf{C}, \mathbf{D}\}, \frac{1}{n}\Sigma, \Psi, \Omega)$$
$$\Rightarrow \text{vec}(\bar{\mathcal{X}}) \sim N_{TFr}((\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathcal{B}), \Omega \otimes \Psi \otimes \frac{1}{n}\Sigma)$$
$$\Rightarrow \text{vec}(E) \sim N_{TFr}(0, \Omega \otimes \Psi \otimes \frac{1}{n}\Sigma).$$

We assume that Σ, Ψ, and Ω are all unknown. We estimate these covariance matrices using their MLEs.

$$\hat{\Sigma}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(1),i} - \mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D} \otimes \mathbf{C})')](\hat{\Omega}_A \otimes \hat{\Psi}_A)^{-1}[\mathbf{X}_{(1),i} - \mathbf{A}\hat{\mathbf{B}}_{(1)}(\mathbf{D} \otimes \mathbf{C})')]'}{nFr}$$
$$\hat{\Psi}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(2),i} - \mathbf{C}\hat{\mathbf{B}}_{(2)}(\mathbf{D} \otimes \mathbf{A})')](\hat{\Omega}_A \otimes \hat{\Sigma}_A)^{-1}[\mathbf{X}_{(2),i} - \mathbf{C}\hat{\mathbf{B}}_{(2)}(\mathbf{D} \otimes \mathbf{A})')]'}{nTr}$$
$$\hat{\Omega}_A = \frac{\sum_{i=1}^{n}[\mathbf{X}_{(3),i} - \mathbf{D}\hat{\mathbf{B}}_{(3)}(\mathbf{C} \otimes \mathbf{A})')](\hat{\Psi}_A \otimes \hat{\Sigma}_A)^{-1}[\mathbf{X}_{(3),i} - \mathbf{D}\hat{\mathbf{B}}_{(3)}(\mathbf{C} \otimes \mathbf{A})')]'}{nTF}$$

REGRESSION-BASED TEST

Because Σ, Ψ, and Ω are all positive-definite, we can take the Cholesky decomposition of the inverse of the covariance matrix of vec($\bar{Y}$), $\Omega \otimes \Psi \otimes \Sigma$, and get a matrix C such that

$$C'C = (\Omega \otimes \Psi \otimes \frac{1}{n}\Sigma)^{-1} = \Omega^{-1} \otimes \Psi^{-1} \otimes n\Sigma^{-1}.$$

Then, we have

$$\begin{aligned} Y^* &= X^*\beta + u^* \\ C\text{vec}(\bar{\mathcal{X}}) &= C(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathcal{B}) + C\text{vec}(\mathcal{E}). \end{aligned}$$

The generalized least-squares solution of β is

$$\beta^* = (X'V^{-1}X)^{-1}X'V^{-1}Y.$$

Setting $X = (\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})$, $V = \Omega \otimes \Psi \otimes \frac{1}{n}\Sigma$, and $Y = \text{vec}(\bar{\mathcal{X}})$,

$$\begin{aligned} \beta^* &= \text{vec}(\hat{\mathcal{B}}) \\ &= [(\mathbf{D}' \otimes \mathbf{C}' \otimes \mathbf{A}')(\Omega^{-1} \otimes \Psi^{-1} \otimes n\Sigma^{-1})(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})]^{-1}[(\mathbf{D}' \otimes \mathbf{C}' \otimes \mathbf{A}')]\times \\ &\quad [\Omega^{-1} \otimes \Psi^{-1} \otimes n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}})] \\ &= [(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1}[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}}). \end{aligned}$$

REGRESSION-BASED TEST

From (3.37) of Seber and Lee (2003),

$$
\begin{aligned}
\hat{\beta}_H &= \hat{\beta} + (X^{*\prime}X^*)^{-1}A'[A(X^{*\prime}X^*)^{-1}A']^{-1}(c - A\hat{\beta}) \\
&= [(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1}[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}}) \\
&+ [(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1} I_{tfr}[I_{tfr}[(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1} I_{tfr} \\
&[\text{vec}(\mathcal{B}_0) - I_{tfr}[(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1}[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}})] \\
&= [(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1}[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathfrak{X}}) \\
&+ [\text{vec}(\mathcal{B}_0) - [(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1}[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}})] \\
&= \text{vec}(\mathcal{B}_0).
\end{aligned}
$$

Following from Section 4.3 of Seber and Lee (2003), we want to test

$$
H_0 : \underbrace{I_{tfr}}_{A} \underbrace{\text{vec}(\mathcal{B})}_{\beta} = \underbrace{\text{vec}(\mathcal{B}_0)}_{c}.
$$

REGRESSION-BASED TEST

Under H_0,

$$RSS_H = ||Y - X^* \hat{\beta}_H||^2$$
$$= ||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathcal{B}_0)||^2.$$

Under H_a,

$$RSS = ||Y - X^* \hat{\beta}||^2 = (n-p)S^2$$
$$= ||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})[(\mathbf{D}'\Omega^{-1}\mathbf{D} \otimes \mathbf{C}'\Psi^{-1}\mathbf{C} \otimes \mathbf{A}'n\Sigma^{-1}\mathbf{A})]^{-1} \times$$
$$[\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}'\Psi^{-1} \otimes \mathbf{A}'n\Sigma^{-1}]\text{vec}(\bar{\mathcal{X}})||^2$$
$$= ||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'n\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'n\Sigma^{-}$$

Note that in our problem, $X^* = C(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})$ is a $TFr \times tfr$ matrix, so $n = TFr$ and $p = tfr$.
Also, $A = I_{tfr}$, so $q = p = tfr$. We have

$$S^2 = \frac{RSS}{n-p}$$
$$= \frac{||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'n\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'n}{TFr - tfr}$$

REGRESSION-BASED TEST

Therefore, the F-statistic is

$$F = \frac{(RSS_H - RSS)/q}{RSS/(n-q)} = \frac{(A\hat{\beta} - c)'[A(X^{*\prime}X^*)^{-1}A']^{-1}(A\hat{\beta} - c)}{qS^2} \sim F_{q,n-p}$$

$$= \frac{(||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D} \otimes \mathbf{C} \otimes \mathbf{A})\text{vec}(\mathcal{B}_0)||^2/tfr}{||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'n\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'n\Sigma^{-1})\text{vec}(}$$

$$- \frac{||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'n\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'n\Sigma^{-1})}{||\text{vec}(\bar{\mathcal{X}}) - C(\mathbf{D}(\mathbf{D}'\Omega^{-1}\mathbf{D})^{-1}\mathbf{D}'\Omega^{-1} \otimes \mathbf{C}(\mathbf{C}'\Psi^{-1}\mathbf{C})^{-1}\mathbf{C}'\Psi^{-1} \otimes \mathbf{A}(\mathbf{A}'n\Sigma^{-1}\mathbf{A})^{-1}\mathbf{A}'n\Sigma^{-1})\text{vec}(}$$

$$\sim F_{tfr,TFr-tfr}.$$

SCORE TEST

- Assesses constraints on statistical parameters based on the gradient of the likelihood function, known as the *score*, evaluated at the hypothesized parameter value under the null hypothesis.
- **Only the distribution under the null hypothesis is required**

Test Statistic:

$$U'(\hat{\theta}_0)I^{-1}(\hat{\theta}_0)U(\hat{\theta}_0) \sim \chi_k^2,$$

where k is the number of constraints imposed by H_0,

$$\text{Score} : U(\hat{\theta}_0) = \frac{\partial \log L(\hat{\theta}_0|x)}{\partial \theta}$$

$$\text{Fisher Information} : I(\hat{\theta}_0) = -E[\frac{\partial^2 \log L(\hat{\theta}_0|x)}{\partial \theta \partial \theta'}]$$