Online Change Detection in Exponential Families with Unknown Parameters

Arnaud Dessein and Arshia Cont

August 28th 2013
Introduction

- From information geometry theory:
  - Study of statistics with concepts from differential geometry and information theory.
  - Parametric statistical models possess an intrinsic geometrical structure.
Introduction

From information geometry theory:
- Study of statistics with concepts from differential geometry and information theory.
- Parametric statistical models possess an intrinsic geometrical structure.

To computational information geometry:
- Broad community around the development and application of computational methods based on information geometry theory.
- Many techniques in machine learning and signal processing rely on statistical models or distance functions: principal component analysis, independent component analysis, centroid computation, $k$-means, expectation-maximization, nearest neighbor search, range search, smallest enclosing balls, Voronoi diagrams.
Introduction

- From information geometry theory:
  - Study of statistics with concepts from differential geometry and information theory.
  - Parametric statistical models possess an intrinsic geometrical structure.

- To computational information geometry:
  - Broad community around the development and application of computational methods based on information geometry theory.
  - Many techniques in machine learning and signal processing rely on statistical models or distance functions: principal component analysis, independent component analysis, centroid computation, $k$-means, expectation-maximization, nearest neighbor search, range search, smallest enclosing balls, Voronoi diagrams.

- Objectives of this work:
  - Employ this framework for audio signal processing.
  - Primary motivations from real-time machine listening.
  - Focus on the fundamental task of audio segmentation.
Outline

1. Preliminaries on Information Geometry
   - Exponential families of probability distributions
   - Dually flat geometry

2. Sequential Change Detection with Exponential Families

3. Real-Time Audio Segmentation
Basic notions and properties

- **Exponential family**: \( p_\theta(x) = \exp(\theta^T x - \psi(\theta)) \).
  - The sufficient observations \( x \) belong to \( \mathbb{R}^m \).
  - The natural parameters \( \theta \) belong to a convex set \( \mathcal{N} \subseteq \mathbb{R}^m \).
  - The log-normalizer \( \psi \) is convex on \( \mathcal{N} \) and smooth on \( \text{int} \mathcal{N} \).
Basic notions and properties

- **Exponential family**: \( p_{\theta}(x) = \exp(\theta^T x - \psi(\theta)) \).
  - The sufficient observations \( x \) belong to \( \mathbb{R}^m \).
  - The natural parameters \( \theta \) belong to a convex set \( \mathcal{N} \subseteq \mathbb{R}^m \).
  - The log-normalizer \( \psi \) is convex on \( \mathcal{N} \) and smooth on \( \text{int} \mathcal{N} \).

- **Many common models**: Bernoulli, Dirichlet, Gaussian, Laplace, Pareto, Poisson, Rayleigh, von Mises-Fisher, Weibull, Wishart, log-normal, exponential, beta, gamma, geometric, binomial, negative binomial, categorical, multinomial.
Basic notions and properties

- **Exponential family**: $p_\theta(x) = \exp(\theta^T x - \psi(\theta))$.
  - The sufficient observations $x$ belong to $\mathbb{R}^m$.
  - The natural parameters $\theta$ belong to a convex set $\mathcal{N} \subseteq \mathbb{R}^m$.
  - The log-normalizer $\psi$ is convex on $\mathcal{N}$ and smooth on $\text{int} \mathcal{N}$.

- **Many common models**: Bernoulli, Dirichlet, Gaussian, Laplace, Pareto, Poisson, Rayleigh, von Mises-Fisher, Weibull, Wishart, log-normal, exponential, beta, gamma, geometric, binomial, negative binomial, categorical, multinomial.

- **Legendre-Fenchel conjugate**: $\phi(\eta) = \sup_{\theta \in \mathbb{R}^m} \theta^T \eta - \psi(\theta)$.
  - The expectation parameters $\eta$ belong to the convex set $\text{int} \mathcal{K}$.
  - We have duality between natural and expectation parameters through $\nabla \psi$ and $\nabla \phi$.
Basic notions and properties

- **Exponential family**: \( p_\theta(x) = \exp(\theta^T x - \psi(\theta)) \).
  - The sufficient observations \( x \) belong to \( \mathbb{R}^m \).
  - The natural parameters \( \theta \) belong to a convex set \( \mathcal{N} \subseteq \mathbb{R}^m \).
  - The log-normalizer \( \psi \) is convex on \( \mathcal{N} \) and smooth on \( \text{int}\mathcal{N} \).

- **Many common models**: Bernoulli, Dirichlet, Gaussian, Laplace, Pareto, Poisson, Rayleigh, von Mises-Fisher, Weibull, Wishart, log-normal, exponential, beta, gamma, geometric, binomial, negative binomial, categorical, multinomial.

- **Legendre-Fenchel conjugate**: \( \phi(\eta) = \sup_{\theta \in \mathbb{R}^m} \theta^T \eta - \psi(\theta) \).
  - The expectation parameters \( \eta \) belong to the convex set \( \text{int}\mathcal{K} \).
  - We have duality between natural and expectation parameters through \( \nabla \psi \) and \( \nabla \phi \).

- **Maximum likelihood**: \( \hat{\eta}_{ml}(x_1, \ldots, x_n) = \frac{1}{n} \sum_{j=1}^{n} x_j \).
  - Simple arithmetic mean in expectation parameters.
  - Natural parameters obtained by convex duality.
Elements of Bregman geometry

- **Canonical divergences:**
  - Kullback-Leibler divergence: \( D_{KL}(P_\theta \| P_{\theta'}) = \int p_\theta \log(p_\theta/p_{\theta'}) \, d\nu. \)
  - Bregman divergences: \( B_\phi(\xi \| \xi') = \phi(\xi) - \phi(\xi') - (\xi - \xi')^T \nabla \phi(\xi'). \)
  - Relation: \( D_{KL}(P_\theta \| P_{\theta'}) = B_\psi(\theta' \| \theta) = B_\phi(\eta(\theta) \| \eta(\theta')). \)
Outline

1 Preliminaries on Information Geometry

2 Sequential Change Detection with Exponential Families
   - Context
   - Statistical framework
   - Methods for exponential families
   - Sample examples
   - Discussion

3 Real-Time Audio Segmentation
Background

- **Principle:**
  - Decide whether the process presents some structural modifications along time.
  - Find the time instants corresponding to the different change points.
  - Characterize the properties within the respective segments.
Background

- **Principle:**
  - Decide whether the process presents some structural modifications along time.
  - Find the time instants corresponding to the different change points.
  - Characterize the properties within the respective segments.

- **Applications:**
  - Quality control in industrial production.
  - Fault detection in technological processes.
  - Automatic surveillance for intrusion and abnormal behavior in security monitoring.
  - Signal processing in geophysics, econometrics, audio, medicine, image.
Background

- **Principle:**
  - Decide whether the process presents some structural modifications along time.
  - Find the time instants corresponding to the different change points.
  - Characterize the properties within the respective segments.

- **Applications:**
  - Quality control in industrial production.
  - Fault detection in technological processes.
  - Automatic surveillance for intrusion and abnormal behavior in security monitoring.
  - Signal processing in geophysics, econometrics, audio, medicine, image.

- **Approaches:**
Motivations and contributions

- Issues of statistical online approaches:
  - Either approximations of the exact statistics with unknown parameters for tractability.
  - Or restrictions on the data and scenarios [Siegmund & Venkatraman, 1995, Mei, 2006].
  - With the exception of a full Bayesian framework for exponential families [Lai & Xing, 2010].
Motivations and contributions

- **Issues of statistical online approaches:**
  - Either approximations of the exact statistics with unknown parameters for tractability.
  - Or restrictions on the data and scenarios [Siegmund & Venkatraman, 1995, Mei, 2006].
  - With the exception of a full Bayesian framework for exponential families [Lai & Xing, 2010].

- **Goals in the non-Bayesian framework:**
  - Known or unknown parameters.
  - Additive or non-additive changes.
  - Topology of the parameters and data.
  - Exact inference for online schemes.
Motivations and contributions

- **Issues of statistical online approaches:**
  - Either approximations of the exact statistics with unknown parameters for tractability.
  - Or restrictions on the data and scenarios [Siegmund & Venkatraman, 1995, Mei, 2006].
  - With the exception of a full Bayesian framework for exponential families [Lai & Xing, 2010].

- **Goals in the non-Bayesian framework:**
  - Known or unknown parameters.
  - Additive or non-additive changes.
  - Topology of the parameters and data.
  - Exact inference for online schemes.

- **Contributions in this context:**
  - Study of the generalized likelihood ratios within the dually flat information geometry.
  - Estimation with arbitrary estimators compared to maximum likelihood.
  - Alternative expression of the statistics through convex duality.
  - Attractive simplification for exact inference with maximum likelihood.
Multiple hypothesis

- Problem formulation:
  - $X_1, \ldots, X_n$ are mutually independent from $\mathcal{P} = \{P_\xi\}_{\xi \in \Xi}$.
  - Observe $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$.
  - Decide whether $X_1, \ldots, X_n$ are i.i.d. or not.
Multiple hypothesis

- Problem formulation:
  - $X_1, \ldots, X_n$ are mutually independent from $\mathcal{P} = \{P_{\xi}\}_{\xi \in \Xi}$.
  - Observe $\mathbf{x} = (x_1, \ldots, x_n) \in \mathcal{X}^n$.
  - Decide whether $X_1, \ldots, X_n$ are i.i.d. or not.

- Multiple hypotheses:
  - $H_0 : X_1, \ldots, X_n \sim P_{\xi_0}, \; \xi_0 \in \Xi_0$.
  - $H_1 : X_1, \ldots, X_i \sim P_{\xi_i}, \; \xi_i \in \Xi_i, \; X_{i+1}, \ldots, X_n \sim P_{\xi_1}, \; \xi_1 \in \Xi_1, \; i \in [1, n-1]$.
  - $H^i_1 : X_1, \ldots, X_i \sim P_{\xi_i}, \; \xi_i \in \Xi_i, \; X_{i+1}, \ldots, X_n \sim P_{\xi_1}, \; \xi_1 \in \Xi_1$.
Test statistics and decision rules

- Likelihood ratio for known parameters: \( \Lambda^i(\bar{x}) = -2 \log \frac{\prod_{j=1}^{i} p_{\xi_{bef}}(x_j)}{\prod_{j=1}^{i} p_{\xi_{bef}}(x_j) \prod_{j=i+1}^{n} p_{\xi_{aft}}(x_j)} \).

- Simplification as cumulative sum statistics: \( \frac{1}{2} \Lambda^i(\bar{x}) = \sum_{j=i+1}^{n} \log \frac{p_{\xi_{aft}}(x_j)}{p_{\xi_{bef}}(x_j)} \).

- Efficient recursive implementation for online procedures.
Test statistics and decision rules

- Likelihood ratio for known parameters: \( \Lambda^i(\bar{x}) = -2 \log \frac{\prod_{j=1}^i p_{\xi_{bef}}(x_j)}{\prod_{j=1}^i p_{\xi_{bef}}(x_j) \prod_{j=i+1}^n p_{\xi_{aft}}(x_j)} \).
  - Simplification as cumulative sum statistics: \( \frac{1}{2} \Lambda^i(\bar{x}) = \sum_{j=i+1}^n \log \frac{p_{\xi_{aft}}(x_j)}{p_{\xi_{bef}}(x_j)} \).
  - Efficient recursive implementation for online procedures.
- Generalized likelihood ratio: \( \hat{\Lambda}^i(\bar{x}) = -2 \log \frac{\prod_{j=i+1}^n p_{\hat{\xi}_0(\bar{x})}(x_j)}{\prod_{j=1}^i p_{\hat{\xi}_0}(x_j) \prod_{j=i+1}^n p_{\hat{\xi}_1(\bar{x})}(x_j)} \).
  - Two cumulative sums: \( \frac{1}{2} \hat{\Lambda}^i(\bar{x}) = \sum_{j=1}^i \log \frac{p_{\hat{\xi}_0}(x_j)}{p_{\hat{\xi}_0(\bar{x})}(x_j)} + \sum_{j=i+1}^n \log \frac{p_{\hat{\xi}_1}(x_j)}{p_{\hat{\xi}_0}(x_j)} \).
  - Computationally more demanding so usually approximated for online procedures.
Test statistics and decision rules

- Likelihood ratio for known parameters: \( \Lambda^{i}(\bar{x}) = -2 \log \frac{\prod_{j=1}^{i} p_{\xi_{bef}}(x_j)}{\prod_{j=1}^{i-1} p_{\xi_{bef}}(x_j) \prod_{j=i+1}^{n} p_{\xi_{aft}}(x_j)} \).

  - Simplification as cumulative sum statistics: \( \frac{1}{2} \Lambda^{i}(\bar{x}) = \sum_{j=i+1}^{n} \log \frac{p_{\xi_{aft}}(x_j)}{p_{\xi_{bef}}(x_j)} \).

  - Efficient recursive implementation for online procedures.

- Generalized likelihood ratio: \( \hat{\Lambda}^{i}(\bar{x}) = -2 \log \frac{\prod_{j=i+1}^{n} p_{\hat{\xi}_0}(x_j)(x_j)}{\prod_{j=1}^{i} p_{\hat{\xi}_0}(x_j)(x_j) \prod_{j=i+1}^{n} p_{\hat{\xi}_1}(x_j)(x_j)} \).

  - Two cumulative sums: \( \frac{1}{2} \hat{\Lambda}^{i}(\bar{x}) = \sum_{j=1}^{i} \log \frac{p_{\hat{\xi}_0}(x_j)(x_j)}{p_{\hat{\xi}_0}(x_j)(x_j)} + \sum_{j=i+1}^{n} \log \frac{p_{\hat{\xi}_1}(x_j)(x_j)}{p_{\hat{\xi}_0}(x_j)(x_j)} \).

  - Computationally more demanding so usually approximated for online procedures.

- Non-Bayesian decision rule for a change: \( \max_{1 \leq i \leq n-1} \hat{\Lambda}^{i}(\bar{x}) \gtrless \lambda \).

  - Comparison of the maximum statistics to a threshold.
  - Change point estimated as the first time point where the maximum is reached.
For an exponential family, the generalized likelihood ratio satisfies:

\[
\frac{1}{2} \hat{\Lambda}^i(\bar{x}) = i \left\{ D_{KL} \left( P_{\theta_{0m1}}^i(\bar{x}) \middle\| P_{\theta_0}(\bar{x}) \right) - D_{KL} \left( P_{\theta_{0m1}}^i(\bar{x}) \middle\| P_{\theta_0}^i(\bar{x}) \right) \right\} \\
+ (n - i) \left\{ D_{KL} \left( P_{\theta_{1m1}}^i(\bar{x}) \middle\| P_{\theta_0}(\bar{x}) \right) - D_{KL} \left( P_{\theta_{1m1}}^i(\bar{x}) \middle\| P_{\theta_1}(\bar{x}) \right) \right\} .
\]
Specific cases

- Various scenarios:
  - Known parameters before and after change.
  - Known parameter before change, unknown parameter after change.
  - Unknown parameters before and after change.

Example

Exact statistics and maximum likelihood:
\[
\frac{1}{2} \hat{\Lambda}^i(x) = i D_{KL} \left( P_{\hat{\theta}_i \text{ml}(x)} \bigg\| P_{\hat{\theta}_0 \text{ml}(x)} \right) + (n - i) D_{KL} \left( P_{\hat{\theta}_1 \text{ml}(x)} \bigg\| P_{\hat{\theta}_0 \text{ml}(x)} \right).
\]
Specific cases

- Various scenarios:
  - Known parameters before and after change.
  - Known parameter before change, unknown parameter after change.
  - Unknown parameters before and after change.

### Example

**Exact statistics and maximum likelihood:**

\[
\frac{1}{2} \hat{\Lambda}^i(\bar{x}) = i D_{KL} \left( P_{\hat{\theta}^i_{0 \text{ml}}(\bar{x})} \parallel P_{\theta^i_{0 \text{ml}}(\bar{x})} \right) + (n - i) D_{KL} \left( P_{\hat{\theta}^i_{1 \text{ml}}(\bar{x})} \parallel P_{\theta^i_{0 \text{ml}}(\bar{x})} \right).
\]

**Example**

**Approximate statistics:**

\[
\frac{1}{2} \hat{\Lambda}^i(\bar{x}) = (n - i) D_{KL} \left( P_{\hat{\theta}^i_{1 \text{ml}}(\bar{x})} \parallel P_{\hat{\theta}^0(\bar{x})} \right).
\]
Proposition

For an exponential family, the generalized likelihood ratio satisfies:

\[ \frac{1}{2} \hat{\Lambda}^i(\bar{x}) = i \phi(\hat{\eta}_0^i(\bar{x})) + (n - i) \phi(\hat{\eta}_1^i(\bar{x})) - n \phi(\hat{\eta}_0(\bar{x})) + \Delta^i_{ml}(\bar{x}) . \]

where the corrective term \( \Delta^i_{ml} \) compared to maximum likelihood estimation equals:

\[
\Delta^i_{ml}(\bar{x}) = i(\hat{\eta}_0^i_{ml}(\bar{x}) - \hat{\eta}_0^i(\bar{x}))^\top \nabla \phi(\hat{\eta}_0^i(\bar{x})) + (n - i)(\hat{\eta}_1^i_{ml}(\bar{x}) - \hat{\eta}_1^i(\bar{x}))^\top \nabla \phi(\hat{\eta}_1^i(\bar{x}))
\]
\[
- n(\hat{\eta}_0^i_{ml}(\bar{x}) - \hat{\eta}_0(\bar{x}))^\top \nabla \phi(\hat{\eta}_0(\bar{x})) .
\]
Revisiting through convex duality

Proposition

For an exponential family, the generalized likelihood ratio satisfies:

\[
\frac{1}{2} \hat{\lambda}^i(x) = i \phi(\hat{\eta}_0^i(x)) + (n - i) \phi(\hat{\eta}_1^i(x)) - n \phi(\hat{\eta}_0(x)) + \Delta_{ml}^i(x).
\]

where the corrective term \( \Delta_{ml}^i \) compared to maximum likelihood estimation equals:

\[
\Delta_{ml}^i(x) = i (\hat{\eta}_0^i(x) - \hat{\eta}_0^i(x))^\top \nabla \phi(\hat{\eta}_0(x)) + (n - i) (\hat{\eta}_1^i(x) - \hat{\eta}_1^i(x))^\top \nabla \phi(\hat{\eta}_1(x)) - n (\hat{\eta}_0^i(x) - \hat{\eta}_0(x))^\top \nabla \phi(\hat{\eta}_0(x)).
\]

Example

The exact generalized likelihood ratio for unknown parameters and maximum likelihood estimation verifies:

\[
\frac{1}{2} \hat{\lambda}^i(x) = i \phi(\hat{\eta}_0^i_{ml}(x)) + (n - i) \phi(\hat{\eta}_1^i_{ml}(x)) - n \phi(\hat{\eta}_0_{ml}(x)).
\]
Figure: Segmentation of well-log data.
Figure: Segmentation of the daily log-return of the Dow Jones.
Discussion

Summary:
- Standard non-Bayesian approach to sequential change detection.
- Dually flat information geometry of exponential families.
- Generalized likelihood ratios with arbitrary estimators.
- Attractive scheme for exact inference when unknown parameters.
Discussion

- **Summary:**
  - Standard non-Bayesian approach to sequential change detection.
  - Dually flat information geometry of exponential families.
  - Generalized likelihood ratios with arbitrary estimators.
  - Attractive scheme for exact inference when unknown parameters.

- **Perspectives:**
  - **Direct extensions:**
    - Non-steep or curved exponential families.
    - Maximum a posteriori estimators.
  - **Asymptotic properties:**
    - Distribution of the test statistics.
    - Optimality formulation and analysis.
  - **Statistical dependence:**
    - Autoregressive models.
    - Non-linear systems and particle filtering.
  - **Alternative test statistics:**
    - Reversing the problem and starting from geometric considerations.
    - Information divergences and more robust estimators.
Outline

1. Preliminaries on Information Geometry
2. Sequential Change Detection with Exponential Families
3. Real-Time Audio Segmentation
   - Context
   - Proposed approach
   - Experimental results
   - Discussion
Principle:

- Determine time boundaries that partition a sound into homogeneous and continuous temporal segments, such that adjacent segments exhibit inhomogeneities.
- Define a criterion to quantify the homogeneity of the segments.
Background

- **Principle:**
  - Determine time boundaries that partition a sound into homogeneous and continuous temporal segments, such that adjacent segments exhibit inhomogeneities.
  - Define a criterion to quantify the homogeneity of the segments.

- **Approaches:**
  - Supervised: high-level classes and automatic classification.
  - Unsupervised: statistical and distance-based approaches:
    - Musical onset detection [Bello et al., 2005, Dixon, 2006].
    - Speaker segmentation [Kemp et al., 2000, Kotti et al., 2008].
Motivations and contributions

- Issues of unsupervised approaches to audio segmentation:
  - Often tailored to particular types of signal and homogeneity criterion.
  - Specific distance functions or models.
  - Some are offline.
  - Others approximate the exact statistics.
Motivations and contributions

- Issues of unsupervised approaches to audio segmentation:
  - Often tailored to particular types of signal and homogeneity criterion.
  - Specific distance functions or models.
  - Some are offline.
  - Others approximate the exact statistics.

- Goals towards a unifying framework for audio segmentation:
  - Arbitrary types of signals homogeneity criteria.
  - Large choice of distance functions or models.
  - Real-time constraints.
  - Exact online inference.
Motivations and contributions

- **Issues of unsupervised approaches to audio segmentation:**
  - Often tailored to particular types of signal and homogeneity criterion.
  - Specific distance functions or models.
  - Some are offline.
  - Others approximate the exact statistics.

- **Goals towards a unifying framework for audio segmentation:**
  - Arbitrary types of signals homogeneity criteria.
  - Large choice of distance functions or models.
  - Real-time constraints.
  - Exact online inference.

- **Contributions in this context:**
  - Generic framework for real-time audio segmentation.
  - Unification of several standard approaches.
  - Online change detection with exponential families.
  - Exact generalized likelihood ratios and maximum likelihood.
System architecture

- **Segmentation scheme:**
  1. Represent frames with a short-time sound description.
  2. Model the observations with probability distributions.
  3. Detect sequentially changes in the distribution parameters.

Audio segmentation (online)

Auditory scene

Short-time sound representation

\( x_j \)

Statistical modeling

\( P_{\xi_j} \)

Change detection
System architecture

- Segmentation scheme:
  1. Represent frames with a short-time sound description.
  2. Model the observations with probability distributions.
  3. Detect sequentially changes in the distribution parameters.

- Short-time sound representation:
  - Energy for information on loudness.
  - Fourier transform for information on spectral content.
  - Mel-frequency cepstral coefficients for information on timbre.
  - Many other possibilities.

![Diagram of audio segmentation process]

Auditory scene → Short-time sound representation → Statistical modeling → Change detection

arlud.f.dessein@gmail.com August 28th 2013
System architecture

- Segmentation scheme:
  1. Represent frames with a short-time sound description.
  2. Model the observations with probability distributions.
  3. Detect sequentially changes in the distribution parameters.

- Short-time sound representation:
  - Energy for information on loudness.
  - Fourier transform for information on spectral content.
  - Mel-frequency cepstral coefficients for information on timbre.
  - Many other possibilities.

- Statistical modeling:
  - Exponential families and generalized likelihood ratios.
  - Unknown parameters and maximum likelihood.

Audio segmentation (online)

Auditory scene

Short-time sound representation

Statistical modeling

Change detection
Clarification of the relations between statistical approaches:

- **Likelihood statistics:** \(-2 \log(p(\bar{x}|H_0)/p(\bar{x}|H_1^i)) > \lambda\).
  
  - **Exact GLR:** \(\hat{n}_0 ml(\bar{x}) = \frac{1}{n} \sum_{j=1}^{n} x_j\), \(\hat{n}_0 ml^i(\bar{x}) = \frac{1}{i} \sum_{j=1}^{i} x_j\), \(\hat{n}_1 ml(\bar{x}) = \frac{1}{n-i} \sum_{j=i+1}^{n} x_j\).
  
  - **Approximate GLR on the whole window:** \(\hat{n}_0 ml(\bar{x}) \approx \hat{n}_0 ml(\bar{x}) = \frac{1}{n} \sum_{j=1}^{n} x_j\).
  
  - **Approximate GLR in a dead region:** \(\hat{n}_0 ml(\bar{x}) \approx \hat{n}_0 ml(\bar{x}) \approx \frac{1}{n_0} \sum_{j=1}^{n_0} x_j\).

- **Model selection:** \(-2 \log(p(\bar{x}|H_0)/p(\bar{x}|H_1^i)) > \lambda\).
  
  - **AIC:** \(\lambda = 2d\).
  
  - **BIC:** \(\lambda = d \log n\).
  
  - **Penalized BIC:** \(\lambda = \gamma d \log n\).
Clarification of the relations between statistical approaches:

- **Likelihood statistics:** \(-2 \log(p(\bar{x}|H_0)/p(\bar{x}|H_1^i)) > \lambda\).
  - Exact GLR: \(\hat{\eta}_{0\text{ ml}}(\bar{x}) = \frac{1}{n} \sum_{j=1}^{n} x_j\), \(\hat{\eta}_{i\text{ ml}}(\bar{x}) = \frac{1}{\bar{i}} \sum_{j=1}^{\bar{i}} x_j\), \(\hat{\eta}_{1\text{ ml}}(\bar{x}) = \frac{1}{n-\bar{i}} \sum_{j=\bar{i}+1}^{n} x_j\).
  - Approximate GLR on the whole window: \(\hat{\eta}_{0\text{ ml}}(\bar{x}) \approx \hat{\eta}_{0\text{ ml}}(\bar{x}) = \frac{1}{n} \sum_{j=1}^{n} x_j\).
  - Approximate GLR in a dead region: \(\hat{\eta}_{0\text{ ml}}(\bar{x}) \approx \hat{\eta}_{0\text{ ml}}(\bar{x}) \approx \frac{1}{n_0} \sum_{j=1}^{n_0} x_j\).

- **Model selection:** \(-2 \log(p(\bar{x}|H_0)/p(\bar{x}|H_1^i)) > \lambda\).
  - AIC: \(\lambda = 2d\).
  - BIC: \(\lambda = d \log n\).
  - Penalized BIC: \(\lambda = \gamma d \log n\).

- **Links with distance-based approaches:**
  \[
  \frac{1}{2} \hat{\Lambda}(\bar{x}) = i D_{KL} \left( P_{\hat{\theta}_{0\text{ ml}}}(\bar{x}) \bigg\| P_{\hat{\theta}_{0\text{ ml}}}(\bar{x}) \right) + \left( n - i \right) D_{KL} \left( P_{\hat{\theta}_{1\text{ ml}}}(\bar{x}) \bigg\| P_{\hat{\theta}_{0\text{ ml}}}(\bar{x}) \right).
  \]

- **Heuristics:**
  - Threshold on the observations.
  - Distance between the observations at successive frames.

- **Kernels methods:**
  - Equivalence between one-class support vector machines for novelty detection and approximate GLR statistics [Canu & Smola, 2006].
Segmentation into silence and activity

- Parameters:
  - Short-time sound representation: energy in a Mel-frequency filter bank at 11025 Hz.
  - Statistical model: Rayleigh distributions.
  - Topology: 1 dimension, continuous non-negative values.
Parameters:
- Short-time sound representation: Mel-frequency cepstral coefficients at 11025 Hz.
- Parametric statistical model: multivariate spherical normal distributions fixed variance.
- Topology: 12 dimensions, continuous real values.
Parameters:
- Short-time sound representation: Mel-frequency cepstral coefficients at 11025 Hz.
- Parametric statistical model: multivariate spherical normal distributions fixed variance.
- Topology: 12 dimensions, continuous real values.
Parameters:
- Short-time sound representation: normalized magnitude spectrum at 11025 Hz.
- Parametric statistical model: categorical distributions.
- Topology: 257 dimensions, discrete frequency histograms.
Evaluation on musical onset detection

- **Parameters:**
  - Short-time sound representation: normalized magnitude spectrum at 12600 Hz.
  - Parametric statistical model: categorical distributions.
  - Topology: 513 dimensions, discrete frequency histograms.
Evaluation on musical onset detection

- **Parameters:**
  - Short-time sound representation: normalized magnitude spectrum at 12600 Hz.
  - Parametric statistical model: categorical distributions.
  - Topology: 513 dimensions, discrete frequency histograms.

- **Evaluation of generalized likelihood ratios GLR and spectral flux SF:**
  - Difficult dataset [Leveau et al., 2004].
  - Standard methodology.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Threshold</th>
<th>$\mathcal{P}$</th>
<th>$\mathcal{R}$</th>
<th>$\mathcal{F}$</th>
<th>Distance function</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLR</td>
<td>5.00</td>
<td>60.93</td>
<td>68.55</td>
<td><strong>64.52</strong></td>
<td>Kullback-Leibler</td>
</tr>
<tr>
<td>SF</td>
<td>0.06</td>
<td>22.56</td>
<td>33.87</td>
<td>27.08</td>
<td>Euclidean</td>
</tr>
<tr>
<td>SF</td>
<td>0.10</td>
<td>34.42</td>
<td>41.26</td>
<td>37.53</td>
<td>Kullback-Leibler</td>
</tr>
<tr>
<td>SF</td>
<td>0.17</td>
<td>40.20</td>
<td>42.74</td>
<td><strong>41.43</strong></td>
<td>Half-wave rectified difference</td>
</tr>
</tbody>
</table>
Evaluation on musical onset detection

- **Parameters:**
  - Short-time sound representation: normalized magnitude spectrum at 12600 Hz.
  - Parametric statistical model: categorical distributions.
  - Topology: 513 dimensions, discrete frequency histograms.

- **Evaluation of generalized likelihood ratios GLR and spectral flux SF:**
  - Difficult dataset [Leveau et al., 2004].
  - Standard methodology.

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Threshold</th>
<th>( \mathcal{P} )</th>
<th>( \mathcal{R} )</th>
<th>( \mathcal{F} )</th>
<th>Distance function</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLR</td>
<td>5.00</td>
<td>60.93</td>
<td>68.55</td>
<td>64.52</td>
<td>Kullback-Leibler</td>
</tr>
<tr>
<td>SF</td>
<td>0.06</td>
<td>22.56</td>
<td>33.87</td>
<td>27.08</td>
<td>Euclidean</td>
</tr>
<tr>
<td>SF</td>
<td>0.10</td>
<td>34.42</td>
<td>41.26</td>
<td>37.53</td>
<td>Kullback-Leibler</td>
</tr>
<tr>
<td>SF</td>
<td>0.17</td>
<td>40.20</td>
<td>42.74</td>
<td>41.43</td>
<td>Half-wave rectified difference</td>
</tr>
</tbody>
</table>

- **Comparison to the state-of-the-art:**
  - **IG:** online symmetrized Kullback-Leibler with logarithmic frequency scale [Cont et al., 2011].
  - **LFSF:** offline spectral flux with a logarithmic frequency scale and filtering [Böck et al., 2012].
  - **TSPC:** online spectral peak classification into transients and non-transients [Röbel, 2011].
**Summary:**
- Real-system for audio segmentation.
- Various types of signals and of homogeneity criteria.
- Sequential change detection with exponential families.
- Unification and generalization of several statistical and distance-based approaches.
Discussion

Summary:
- Real-system for audio segmentation.
- Various types of signals and of homogeneity criteria.
- Sequential change detection with exponential families.
- Unification and generalization of several statistical and distance-based approaches.

Perspectives:
- Dependent observations:
  - Autoregressive models.
  - Non-linear systems and particle filtering.
- Improved robustness:
  - Post-processing by smoothing, adaptation.
  - Growing and sliding window heuristics.
- Consideration of prior information:
  - Maximum a posteriori.
  - Full Bayesian framework.
- Further applications:
  - Audio processing and music information retrieval.
  - Other domains in signal processing.
Conclusion

- **Summary of the present work:**
  - Study the application of computational information geometry to audio signal processing.
  - From sequential change detection to audio segmentation.

- **Perspectives for future work:**
  - Apply other novel or existing computational methods to audio signal processing.
  - Apply the proposed computational methods to broader applications and domains.
Conclusion

- Summary of the present work:
  - Study the application of computational information geometry to audio signal processing.
  - From sequential change detection to audio segmentation.

- Perspectives for future work:
  - Apply other novel or existing computational methods to audio signal processing.
  - Apply the proposed computational methods to broader applications and domains.

- Further readings:


- Thanks for your attention.
Bibliography I

Detection of Abrupt Changes: Theory and Application.  
Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

A tutorial on onset detection in music signals.  

Evaluating the online capabilities of onset detection methods.  
In 13th International Society for Music Information Retrieval Conference (ISMIR) (pp. 49–54). Porto, Portugal.

Kernel methods and the exponential family.  
Neurocomputing, 69(7–9), 714–720.

Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance.  

On the information geometry of audio streams with applications to similarity computing.  

An online kernel change detection algorithm.  
IEEE Transactions on Signal Processing, 53(8), 2961–2974.
Onset detection revisited.
In *9th International Conference on Digital Audio Effects (DAFx)* (pp. 133–137). Montreal, Canada.

Kernel change-point analysis.
In D. Koller, D. Schuurmans, Y. Bengio, & L. Bottou (Eds.), *Advances in Neural Information Processing Systems (NIPS)*, volume 21 (pp. 609–616). La Jolla, CA, USA: NIPS Foundation.

Catching change-points with lasso.

Multiple change-point estimation with a total variation penalty.

Strategies for automatic segmentation of audio data.

Speaker segmentation and clustering.


Using the generalized likelihood ratio statistic for sequential detection of a change-point.

Fast detection of multiple change-points shared by many signals using group LARS.