Data Driven Detection Methods

GEORGE V. MOUSTAKIDES
UNIVERSITY OF PATRAS, GREECE

Outline

- Hypothesis Testing
  - Mathematical Formulation
  - Data Driven Approach
  - The Consistency Property

- Likelihood Ratio Estimation
  - Optimization Problems with Consistent Solutions
  - Data Driven Implementation

- Detection in Time Series
  - I.i.d. processes
  - Markov processes

- What is Missing from the Analysis ?

Hypothesis Testing

Mathematical Formulation

For a random vector X we assume the following two hypotheses

$$\begin{aligned} \mathsf{H}_0: &\quad X \sim \mathsf{f}_0(X),\ \ \mathbb{P}(\mathsf{H}_0) \\ \mathsf{H}_1: &\quad X \sim \mathsf{f}_1(X),\ \ \mathbb{P}(\mathsf{H}_1) \end{aligned}$$

For every X need to decide if it comes from H_0 or H_1

Decide using a *Decision Function* $\mathsf{D}(X) \in \{0, 1\}$

Would like to optimize $\mathsf{D}(X)$

Plethora of applications in diverse scientific fields!!!

Bayesian Approach

Minimize decision error probability

$$\min_{\mathsf{D}}\Big\{\mathbb{P}(\mathsf{D}=1|\mathsf{H}_0)\mathbb{P}(\mathsf{H}_0)+\mathbb{P}(\mathsf{D}=0|\mathsf{H}_1)\mathbb{P}(\mathsf{H}_1)\Big\}$$

$$\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \frac{\mathbb{P}(\mathsf{H}_0)}{\mathbb{P}(\mathsf{H}_1)} \quad \equiv \quad \frac{\mathsf{f}_1(X)\mathbb{P}(\mathsf{H}_1)}{\mathsf{f}_0(X)\mathbb{P}(\mathsf{H}_0)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 1$$

For $\omega(\mathsf{r})$ strictly increasing

$$\mathsf{r}(X) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 1 \;\equiv\; \omega\big(\mathsf{r}(X)\big) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \omega(1), \qquad \mathsf{r}(X)=\frac{\mathsf{f}_1(X)\mathbb{P}(\mathsf{H}_1)}{\mathsf{f}_0(X)\mathbb{P}(\mathsf{H}_0)}$$

Neyman-Pearson Approach

$$\mathsf{H}_0: \quad X \sim \mathsf{f}_0(X), \quad \mathbb{P}(\mathsf{H}_0)$$
$$\mathsf{H}_1: \quad X \sim \mathsf{f}_1(X), \quad \mathbb{P}(\mathsf{H}_1)$$

Maximize detection probability $\mathbb{P}(\mathsf{D}=1|\mathsf{H}_1)$

subject to false alarm probability constraint $\mathbb{P}(\mathsf{D}=1|\mathsf{H}_0) \le \alpha$

$$\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \lambda, \quad \mathbb{P}\left(\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)} \ge \lambda \Big| \mathsf{H}_0\right) = \alpha$$

For $\omega(\mathsf{r})$ strictly increasing

$$\omega\big(\mathsf{r}(X)\big) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \eta, \quad \mathbb{P}\Big(\omega\big(\mathsf{r}(X)\big) \ge \eta \Big| \mathsf{H}_0\Big) = \alpha, \qquad \mathsf{r}(X) = \frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}$$

Data Driven Approach

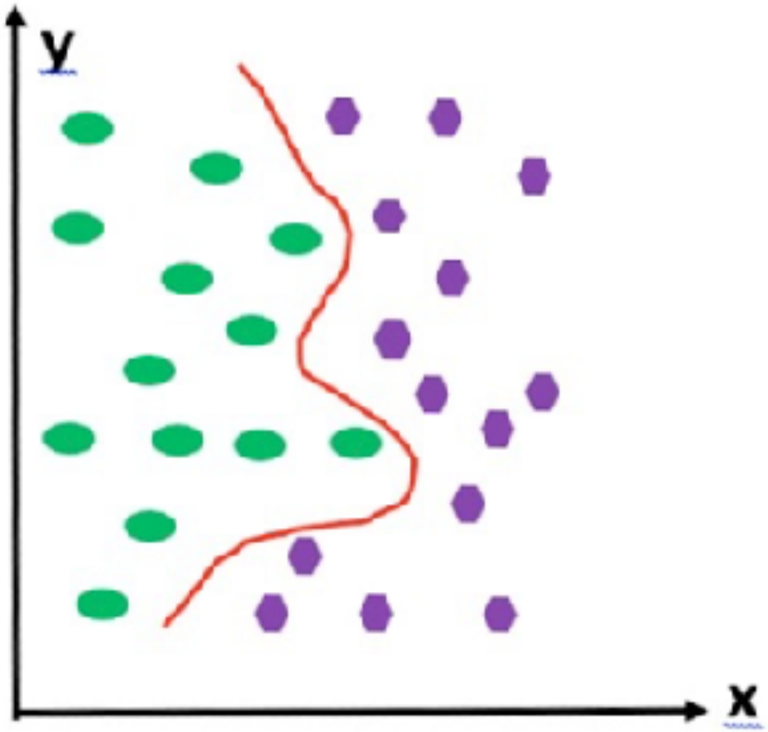
$\mathsf{H}_0: \quad X \sim \mathsf{f}_0(X), \quad \mathbb{P}(\mathsf{H}_0)$

$\mathsf{H}_1: \quad X \sim \mathsf{f}_1(X), \quad \mathbb{P}(\mathsf{H}_1)$

$X_1^0 \; X_2^0 \; \ldots \; X_{n_0}^0$ — Sampled from f_0

$X_1^1 \; X_2^1 \; \ldots \; X_{n_1}^1$ — Sampled from f_1

$$\mathbb{P}(\mathsf{H}_i) \approx \frac{n_i}{n_0 + n_1}$$



Design border to separate the two datasets

What is the best border ?

$$\text{All } X: \; \frac{\mathsf{f}_1(X)\mathbb{P}(\mathsf{H}_1)}{\mathsf{f}_0(X)\mathbb{P}(\mathsf{H}_0)} = 1$$

Instead of a "border", design a decision like function $\mathsf{v}(X)$

$$\mathsf{v}(X) = \begin{cases} -1 & \text{when } X \text{ from } \mathsf{H}_0 \\ 1 & \text{when } X \text{ from } \mathsf{H}_1. \end{cases}$$

Use parametric family of functions $\mathsf{u}(X,\theta)$ and optimize θ solving

$$\mathsf{J}(\theta) = \frac{1}{n_0 + n_1} \left\{ \sum_{i=1}^{n_0} \Big(-1 - \mathsf{u}(X_i^0, \theta) \Big)^2 + \sum_{j=1}^{n_1} \Big(1 - \mathsf{u}(X_j^1, \theta) \Big)^2 \right\}$$

$$\min_{\theta} \mathsf{J}(\theta) \;\Rightarrow\; \theta_{\mathsf{o}} \;\Rightarrow\; \mathsf{u}(X, \theta_{\mathsf{o}})$$

For every X to test decide as follows: $\mathsf{u}(X, \theta_{\mathsf{o}}) \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} 0$

Works “well”!! Why??

Understanding using Asymptotic Analysis

$$n_0, n_1 \to \infty, \qquad \mathsf{u}(X,\theta) \to \mathsf{v}(X)$$

$$\mathsf{J}(\theta) = \frac{n_0}{n_0+n_1} \frac{1}{n_0} \sum_{i=1}^{n_0} \Big(1 + \mathsf{u}(X_i^0,\theta)\Big)^2 + \frac{n_1}{n_0+n_1} \frac{1}{n_1} \sum_{j=1}^{n_1} \Big(1 - \mathsf{u}(X_j^1,\theta)\Big)^2$$

$$\mathsf{J}(\mathsf{v}) = \mathbb{P}(\mathsf{H}_0)\mathbb{E}_0\left[\Big(1+\mathsf{v}(X)\Big)^2\right] + \mathbb{P}(\mathsf{H}_1)\mathbb{E}_1\left[\Big(1-\mathsf{v}(X)\Big)^2\right]$$

$$\min_\theta \mathsf{J}(\theta) \to \min_\mathsf{v} \mathsf{J}(\mathsf{v})$$

$$\theta_o \Rightarrow \mathsf{u}(X,\theta_o) \approx \mathsf{v}_o(X)$$

$$\mathbb{E}_1\left[\Big(1-\mathsf{v}(X)\Big)^2\right]=\mathbb{E}_0\left[\Big(1-\mathsf{v}(X)\Big)^2\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}\right]$$

$$\mathsf{J}(\mathsf{v})=\mathbb{P}(\mathsf{H}_0)\mathbb{E}_0\left[\Big(1+\mathsf{v}(X)\Big)^2+\mathsf{r}(X)\Big(1-\mathsf{v}(X)\Big)^2\right] \qquad \mathsf{r}(X)=\frac{\mathsf{f}_1(X)\mathbb{P}(\mathsf{H}_1)}{\mathsf{f}_0(X)\mathbb{P}(\mathsf{H}_0)}$$

minimize for each X

$$\mathsf{v}_o(X)=\frac{\mathsf{r}(X)-1}{\mathsf{r}(X)+1}=\omega\big(\mathsf{r}(X)\big), \quad \text{where } \omega(\mathsf{r})=\frac{\mathsf{r}-1}{\mathsf{r}+1} \text{ strictly increasing}$$

Test equivalent to Bayes: $\mathsf{v}_o(X)=\omega\big(\mathsf{r}(X)\big)\underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}}\omega(1)=0 \quad\Rightarrow\quad \mathsf{u}(X,\theta_o)\underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}}0$

Equivalence in the limit

Consistency (with respect to the Bayes test)

$$\min_{\theta} \mathsf{J}(\theta) = \min_{\theta} \left\{ \sum_{i=1}^{n_0} \Big(1 + \mathsf{u}(X_i^0, \theta)\Big)^2 + \sum_{j=1}^{n_1} \Big(1 - \mathsf{u}(X_j^1, \theta)\Big)^2 \right\}$$

$$\Rightarrow \ \theta_{\mathsf{o}} \ \Rightarrow \ \mathsf{u}(X, \theta_{\mathsf{o}})$$

$$\min_{\mathsf{v}} \mathsf{J}(\mathsf{v}) = \min_{\mathsf{v}} \left\{ \mathbb{P}(\mathsf{H}_0)\mathbb{E}_0 \left[\Big(1 + \mathsf{v}(X)\Big)^2 \right] + \mathbb{P}(\mathsf{H}_1)\mathbb{E}_1 \left[\Big(1 - \mathsf{v}(X)\Big)^2 \right] \right\}$$

$$\Rightarrow \ \mathsf{v}_{\mathsf{o}}(X) \ = \ \omega\big(\mathsf{r}(X)\big)$$

Expect: $\mathsf{u}(X, \theta_{\mathsf{o}}) \approx \omega\big(\mathsf{r}(X)\big)$

Optimum Test: $\omega\big(\mathsf{r}(X)\big) \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \omega(1)$, Close to Optimum: $\mathsf{u}(X, \theta_{\mathsf{o}}) \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \omega(1)$

Develop data driven methods for estimation of $\omega\big(\mathsf{r}(X)\big)$ for other $\omega(\mathsf{r})$

Consistent tests eventually prevail over inconsistent tests

Likelihood Ratio Estimation

$$\mathsf{r}(X) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 1 \;\equiv\; \omega_1\big(\mathsf{r}(X)\big) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \omega_1(1) \;\equiv\; \omega_2\big(\mathsf{r}(X)\big) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \omega_2(1)$$

$$\mathsf{u}(X,\theta_{\mathsf{o}}) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 1 \;\not\equiv\; \mathsf{u}_1(X,\theta_1) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \omega_1(1) \;\not\equiv\; \mathsf{u}_2(X,\theta_2) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \omega_2(1)$$

For function $\omega(\mathsf{r})$ can we define cost

$$\mathsf{J}(\mathsf{v}) = \mathbb{P}(\mathsf{H}_0)\mathbb{E}_0\left[\phi\big(\mathsf{v}(X)\big)\right] + \mathbb{P}(\mathsf{H}_1)\mathbb{E}_1\left[\psi\big(\mathsf{v}(X)\big)\right]$$

so that $\min_{\mathsf{v}} \mathsf{J}(\mathsf{v}) \;\Rightarrow\; \mathsf{v}_{\mathsf{o}}(X) = \omega\big(\mathsf{r}(X)\big)$?

THEOREM: Select strictly increasing function $\omega(\mathsf{r})$ and strictly negative function $\rho(z)$. Define

$$\psi'(z) = \rho(z), \quad \phi'(z) = -\omega^{-1}(z)\rho(z)$$

then the solution of the optimization problem

$$\min_{\mathsf{v}} \mathsf{J}(\mathsf{v}) = \min_{\mathsf{v}} \Big\{ \mathbb{P}(\mathsf{H}_0)\mathbb{E}_0\big[\phi\big(\mathsf{v}(X)\big)\big] + \mathbb{P}(\mathsf{H}_1)\mathbb{E}_1\big[\psi\big(\mathsf{v}(X)\big)\big] \Big\}$$

satisfies $\mathsf{v}_o(X) = \arg\min_{\mathsf{v}} \mathsf{J}(\mathsf{v}) = \omega\big(\mathsf{r}(X)\big)$

Same optimal solution $\omega(\mathsf{r}(X))$ for all functions $\rho(z) < 0$

Examples of functions

A: $\omega(\mathsf{r}) = \mathsf{r} \in \mathbb{R}_+$ (likelihood ratio)

$$\rho(z) = -1, z \geq 0 \;\Rightarrow\; \phi(z) = \frac{z^2}{2}, \;\; \psi(z) = -z$$

Mean Square

B: $\omega(\mathsf{r}) = \log(\mathsf{r}) \in \mathbb{R}$ (log-likelihood ratio)

$$\rho(z) = -e^{-0.5z} \;\Rightarrow\; \phi(z) = 2e^{0.5z}, \;\; \psi(z) = 2e^{-0.5z}$$

Exponential

C: $\omega(\mathsf{r}) = \frac{\mathsf{r}}{\mathsf{r}+1} \in [0,1]$ (posterior probability)

$$\rho(z) = -\frac{1}{z}, z \in [0,1] \;\Rightarrow\; \phi(z) = -\log(1-z), \;\; \psi(z) = -\log(z)$$

Cross Entropy

Data Driven Implementation

$$\mathsf{J}(\mathsf{v}) = \mathbb{P}(\mathsf{H}_0)\mathbb{E}_0\left[\phi\big(\mathsf{v}(X)\big)\right] + \mathbb{P}(\mathsf{H}_1)\mathbb{E}_1\left[\psi\big(\mathsf{v}(X)\big)\right]$$

$$\mathsf{J}(\theta) = \frac{1}{n_0+n_1}\left\{\sum_{i=1}^{n_0}\phi\big(\mathsf{u}(X_i^0,\theta)\big) + \sum_{j=1}^{n_1}\psi\big(\mathsf{u}(X_j^1,\theta)\big)\right\}$$

$$\mathsf{u}(X,\theta_{\mathsf{o}}) \approx \omega\left(\frac{\mathsf{f}_1(X)\mathbb{P}(\mathsf{H}_1)}{\mathsf{f}_0(X)\mathbb{P}(\mathsf{H}_0)}\right)$$

$$\mathsf{J}(\theta) = \frac{1}{n_0}\sum_{i=1}^{n_0}\phi\big(\mathsf{u}(X_i^0,\theta)\big) + \frac{1}{n_1}\sum_{j=1}^{n_1}\psi\big(\mathsf{u}(X_j^1,\theta)\big)$$

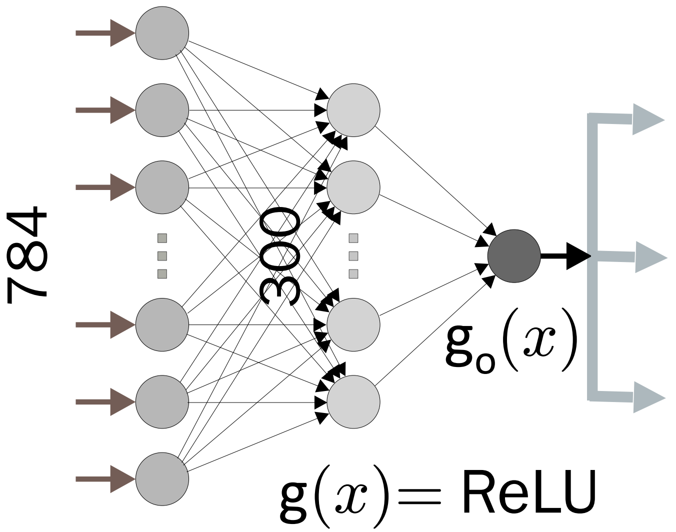
$$\mathsf{u}(X,\theta_{\mathsf{o}}) \approx \omega\left(\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}\right)$$

Example: Classification Problem

From dataset MNIST isolate handwritten numerals 4 and 9

Gray scale images 28 X 28 = 784 pixels. Design classifier using training data. Examine performance using testing data.

Neural network 784 X 300 X 1



Mean square $g_o(x)=$ ReLU, $u_1(X,\theta_1)$

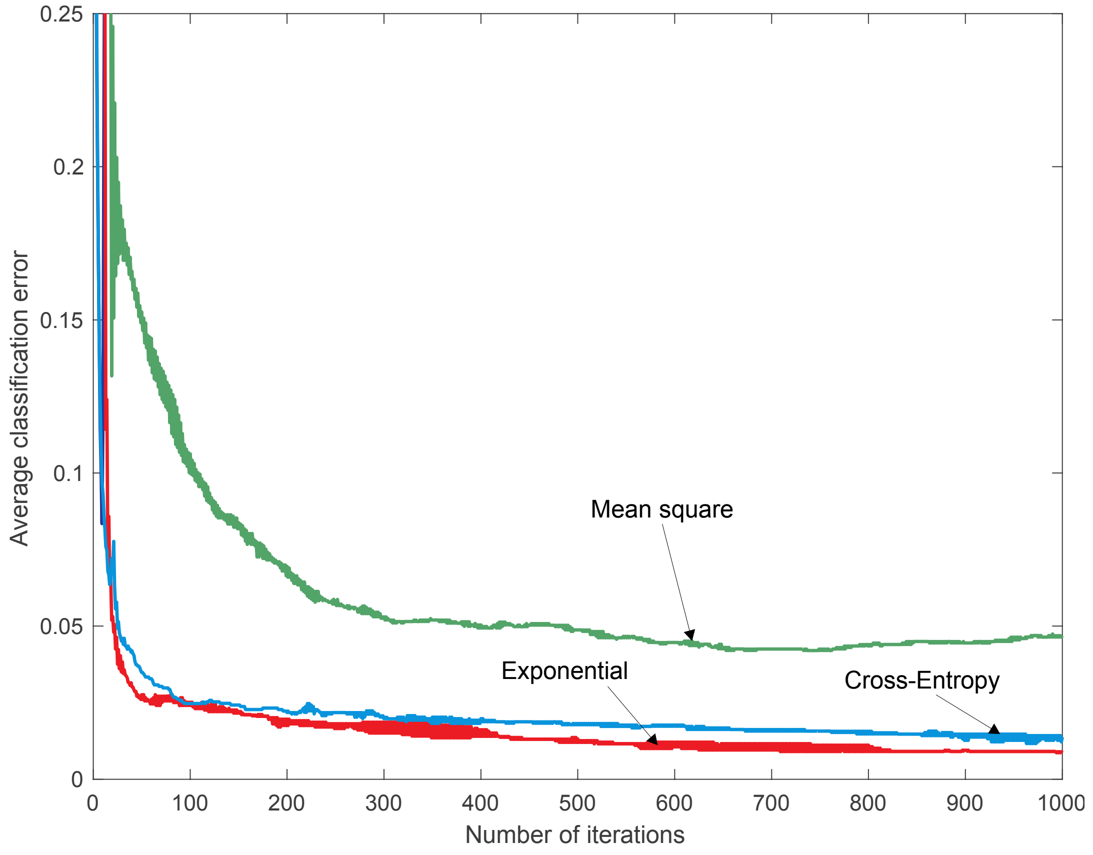
Exponential $g_o(x)= x$, $u_2(X,\theta_2)$

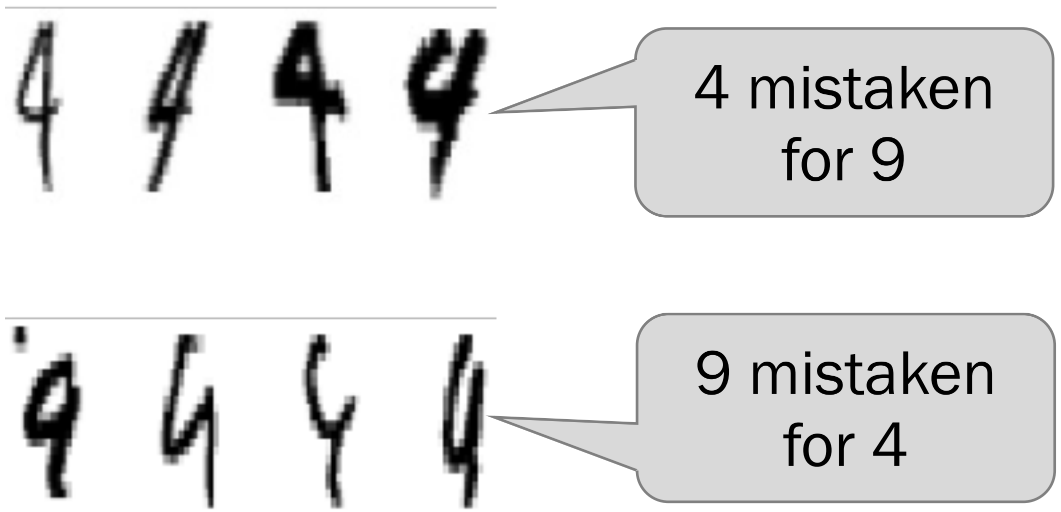
Cross entropy $g_o(x)=$ sigmoid, $u_3(X,\theta_3)$

# of parameters 235801

Training set: 5500 “4” and 5500 “9”. Testing set: 982 “4” and 1009 “9”

$$\mathsf{u}_1(X,\theta_1) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 1, \quad \mathsf{u}_2(X,\theta_2) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} 0, \quad \mathsf{u}_3(X,\theta_3) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \frac{1}{2}$$





Detection in Time Series

More practically interesting case: Testing of time series $\{ X_1 , X_2 ,...., X_n \}$

The whole set of measurements under H_0 or H_1

For testing we need likelihood ratio

$$\frac{\mathsf{f}_1(X_n,\ldots,X_1)}{\mathsf{f}_0(X_n,\ldots,X_1)} = \frac{\mathsf{f}_1(X_n|X_{n-1},\ldots,X_1)}{\mathsf{f}_0(X_n|X_{n-1},\ldots,X_1)} \frac{\mathsf{f}_1(X_{n-1}|X_{n-2},\ldots,X_1)}{\mathsf{f}_0(X_{n-1}|x_{X-2},\ldots,X_1)} \cdots \frac{\mathsf{f}_1(X_1)}{\mathsf{f}_0(X_1)}$$

When i.i.d. under each hypothesis

$$\frac{\mathsf{f}_1(X_n,\ldots,X_1)}{\mathsf{f}_0(X_n,\ldots,X_1)} = \frac{\mathsf{f}_1(X_n)}{\mathsf{f}_0(X_n)} \cdots \frac{\mathsf{f}_1(X_1)}{\mathsf{f}_0(X_1)}$$

Test to be used

$$\sum_{i=1}^{n} \log\left(\frac{\mathsf{f}_1(X_i)}{\mathsf{f}_0(X_i)}\right) \overset{\mathsf{H}_1}{\underset{\mathsf{H}_0}{\gtreqless}} \eta$$

Interested in estimating $\omega\big(\mathsf{r}(X)\big) = \omega\left(\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}\right)$

We are given training data:

$\{X_1^0, \ldots, X_{n_0}^0\}$ following H_0

$\{X_1^1, \ldots, X_{n_1}^1\}$ following H_1

For each $\omega(\mathsf{r})$ of interest, minimize corresponding $\mathsf{J}(\theta)$

$$\mathsf{J}(\theta) = \frac{1}{n_0}\sum_{i=1}^{n_0}\phi\big(\mathsf{u}(X_i^0,\theta)\big) + \frac{1}{n_1}\sum_{j=1}^{n_1}\psi\big(\mathsf{u}(X_j^1,\theta)\big)$$

$$\mathsf{u}(X,\theta_{\mathsf{o}}) \approx \omega\left(\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}\right)$$

$$\text{For } \omega(\mathsf{r}) = \mathsf{r}, \quad \mathsf{u}_1(X,\theta_1) \approx \frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}, \quad \text{use} \quad \log\big(\mathsf{u}_1(X,\theta_1)\big) \quad \text{(Mean Square)}$$

$$\text{For } \omega(\mathsf{r}) = \log(\mathsf{r}), \quad \mathsf{u}_2(X,\theta_2) \approx \log\left(\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}\right), \quad \text{use} \quad \mathsf{u}_2(X,\theta_2) \quad \text{(Exponential)}$$

$$\text{For } \omega(\mathsf{r}) = \frac{\mathsf{r}}{\mathsf{r}+1}, \quad \mathsf{u}_3(X,\theta_3) \approx \frac{\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}}{\frac{\mathsf{f}_1(X)}{\mathsf{f}_0(X)}+1}, \quad \text{use} \quad \log\left(\frac{\mathsf{u}_3(X,\theta_3)}{1-\mathsf{u}_3(X,\theta_3)}\right)$$

(Cross Entropy)

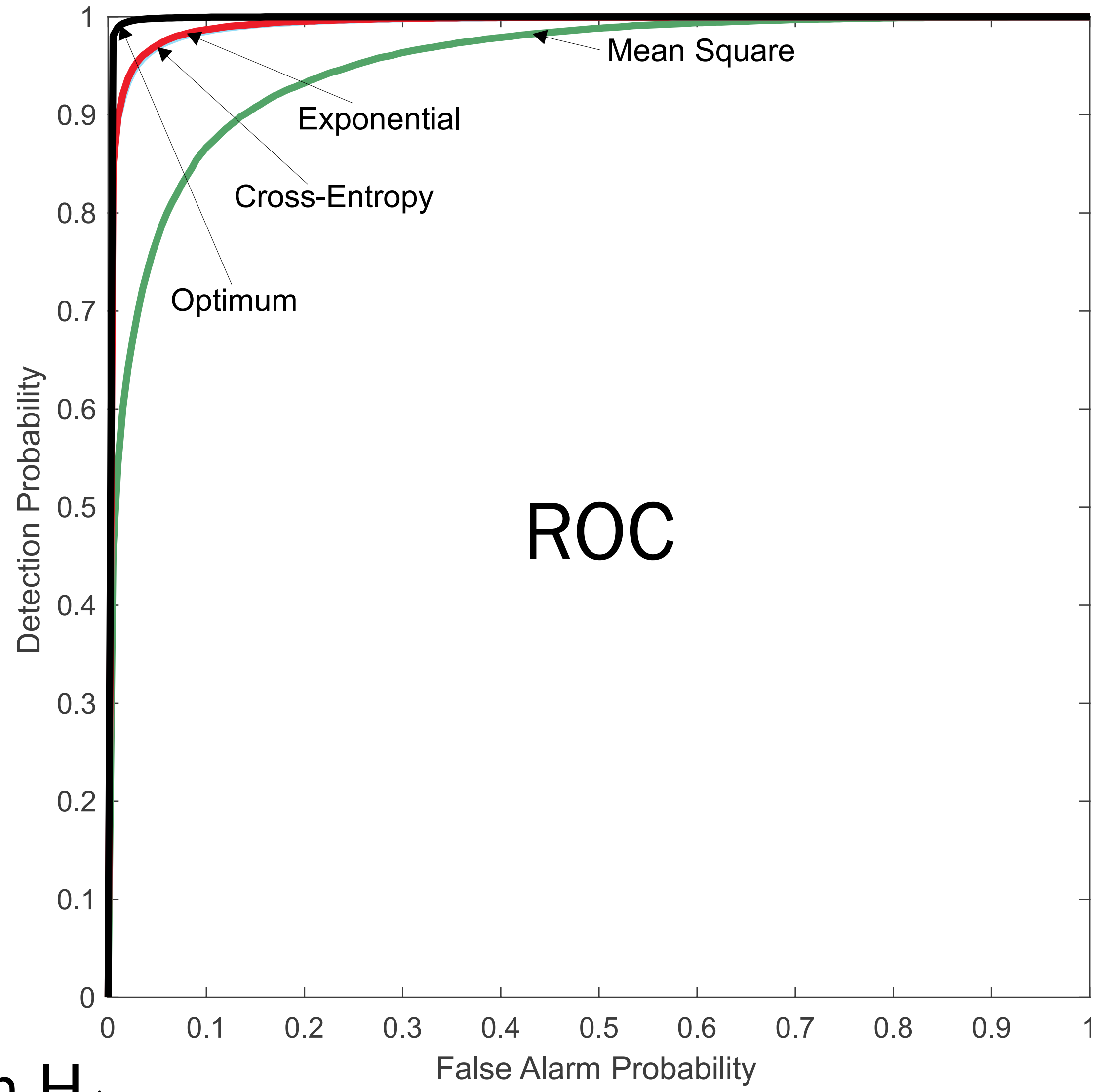
Example: Testing i.i.d. sequences

X_i length 10, $\mathsf{f}_0(X) \sim \mathcal{N}(0, I)$
$\mathsf{f}_1(X) \sim \mathcal{N}(\frac{1}{\sqrt{10}}[1 \cdots 1]^\intercal, 1.2I)$

Neural Network 10×20×1
Training data $n_0 = n_1 =$100
Produce $\mathsf{u}_1(X,\theta_1), \mathsf{u}_2(X,\theta_2), \mathsf{u}_3(X,\theta_3)$

Testing for $n =$20 samples

$$\left.\begin{array}{r} \sum_{i=1}^{20} \log\big(\mathsf{u}_1(X_i,\theta_1)\big) \\ \sum_{i=1}^{20} \mathsf{u}_2(X_i,\theta_2) \\ \sum_{i=1}^{20} \log\left(\frac{\mathsf{u}_3(X_i,\theta_3)}{1-\mathsf{u}_3(X_i,\theta_3)}\right) \end{array}\right\} \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \eta$$



100000×20 realizations from H_0 and from H_1

Markovian processes

Consider Markovian processes with "memory" m

$$\frac{\mathsf{f}_1(X_n|X_{n-1},\ldots,X_1)}{\mathsf{f}_0(X_n|X_{n-1},\ldots,X_1)}=\frac{\mathsf{f}_1(X_n|X_{n-1},\ldots,X_{n-m})}{\mathsf{f}_0(X_n|X_{n-1},\ldots,X_{n-m})}$$

$$\frac{\mathsf{f}_1(X_n,\ldots,X_1)}{\mathsf{f}_0(X_n,\ldots,X_1)}=\frac{\mathsf{f}_1(X_n|X_{n-1},\ldots,X_{n-m})}{\mathsf{f}_0(X_n|X_{n-1},\ldots,X_{n-m})}\cdots\frac{\mathsf{f}_1(X_{m+1}|X_m,\ldots,X_1)}{\mathsf{f}_0(X_{m+1}|X_m,\ldots,X_1)}\times\frac{\mathsf{f}_1(X_m,\ldots,X_1)}{\mathsf{f}_0(X_m,\ldots,X_1)}$$

Can we estimate likelihood ratio of conditional densities?

a) Through data dynamics (classical)

b) Directly (proposed)

Classical Approach

Most common model, Autoregressive

$$X_t = A_1^i X_{t-1} + \cdots + A_m^i X_{t-m} + W_t, \quad i = 0, 1$$

$$X_t = G_i(X_{t-1}, \ldots, X_{t-m}, \theta^i) + W_t$$

Use training data $\{X_1^0, \ldots, X_{n_0}^0\}$ and $\{X_1^1, \ldots, X_{n_1}^1\}$ to solve

$$\min_{\theta^i} \sum_{t=1}^{n_i} \left(X_t^i - G_i(X_{t-1}^i, \ldots, X_{t-m}^i, \theta^i) \right)^2 \quad \Rightarrow \quad \theta_\mathsf{o}^i$$

$$W_t^i = X_t^i - G_i(X_{t-1}^i, \ldots, X_{t-m}^i, \theta_\mathsf{o}^i), \quad \Sigma_i = \frac{1}{n_i} \sum_{t=1}^{n_i} W_t^i (W_t^i)^\intercal$$

Assume $\{W_t^i\}$ i.i.d. Gaussian $\mathcal{N}(0, \Sigma_i)$

X_t given $\{X_{t-1},\ldots,X_{t-m}\}$ under hypothesis H_i

Gaussian with mean $G_i(X_{t-1},\ldots,X_{t-m},\theta_\mathsf{o}^i)$ and covariance Σ_i

To test $\{X_1,\ldots,X_n\}$

$$W_t^i = X_t - G_i(X_{t-1},\ldots,X_{t-m},\theta_\mathsf{o}^i)$$

$$\frac{\mathsf{f}_1(X_t|X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_t|X_{t-1},\ldots,X_{t-m})} = \frac{e^{-\frac{1}{2}(W_t^1)^\intercal\Sigma_1^{-1}W_t^1}}{e^{-\frac{1}{2}(W_t^0)^\intercal\Sigma_0^{-1}W_t^0}}\sqrt{\frac{|\Sigma_0|}{|\Sigma_1|}}$$

Not purely data driven

Gaussian assumption arbitrary, not necessarily suitable for all data!

Proposed Approach

$$\log\left(\frac{\mathsf{f}_1(X_t|X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_t|X_{t-1},\ldots,X_{t-m})}\right) = \log\left(\frac{\frac{\mathsf{f}_1(X_t,X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_1(X_{t-1},\ldots,X_{t-m})}}{\frac{\mathsf{f}_0(X_t,X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_{t-1},\ldots,X_{t-m})}}\right) =$$

$$\log\left(\frac{\mathsf{f}_1(X_t,X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_t,X_{t-1},\ldots,X_{t-m})}\right) - \log\left(\frac{\mathsf{f}_1(X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_{t-1},\ldots,X_{t-m})}\right)$$

$$\mathsf{u}_{m+1}(X_t,\ldots,X_{t-m},\theta_{m+1}) \qquad \mathsf{u}_m(X_t,\ldots,X_{t-m+1},\theta_m)$$

$$\log\left(\frac{\mathsf{f}_1(X_t|X_{t-1},\ldots,X_{t-m})}{\mathsf{f}_0(X_t|X_{t-1},\ldots,X_{t-m})}\right) \approx$$

$$\mathsf{u}_{m+1}(X_t,\ldots,X_{t-m},\theta_{m+1}) - \mathsf{u}_m(X_{t-1},\ldots,X_{t-m},\theta_m)$$

Example: Testing Markov sequences (proof of concept)

Scalar observations $\{x_1, \ldots, x_n\}$

$w_t \sim \mathcal{N}(0,1)$, i.i.d.

$\mathsf{H}_0 : x_t = w_t$

$\mathsf{H}_1 : x_t = \mathsf{sign}(x_{t-1})\sqrt{|x_{t-1}|} + w_t$

$\mathsf{u}_2(x_t, x_{t-1}, \theta_2) : \ 2 \times 20 \times 1$

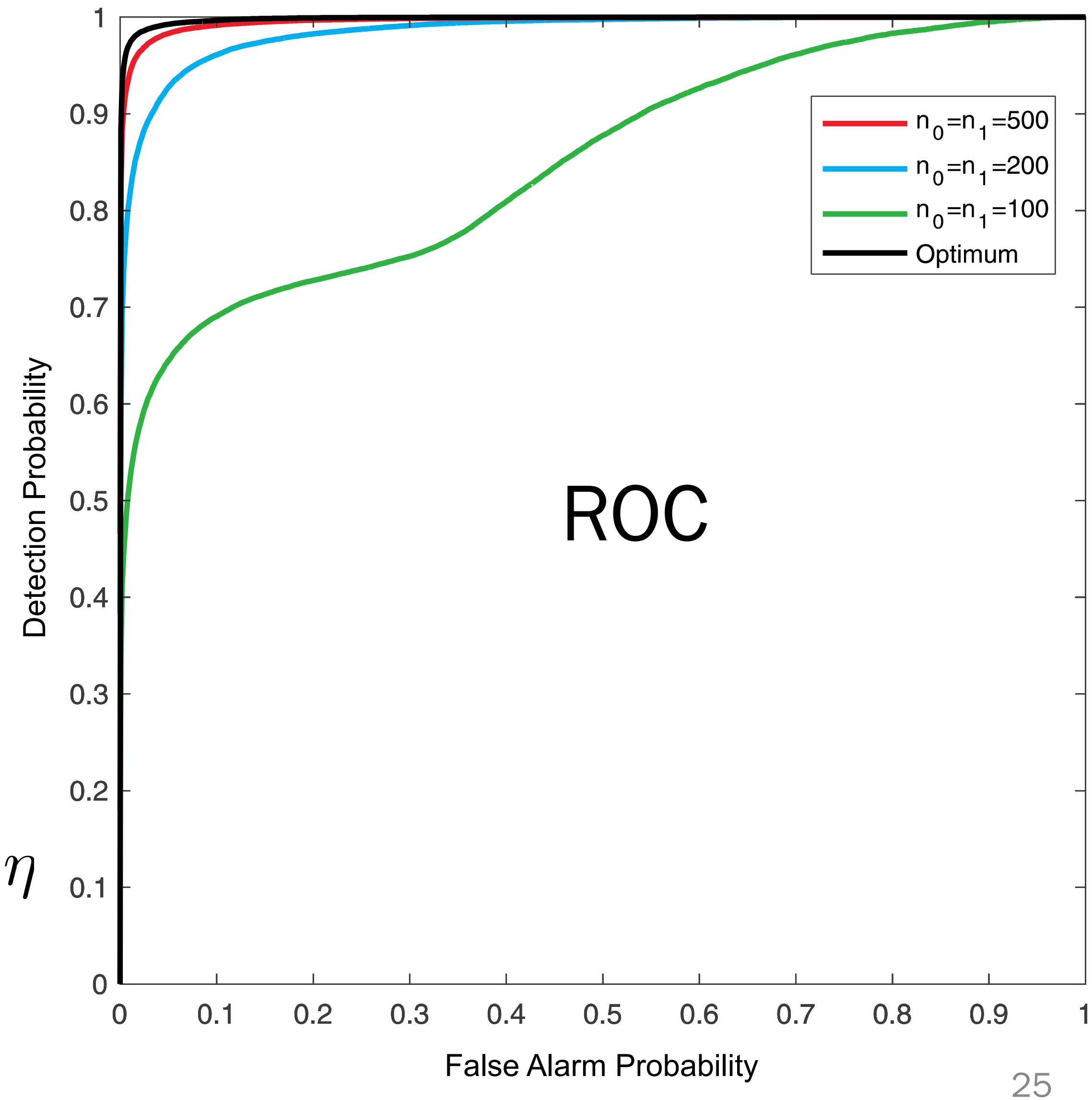
$\mathsf{u}_1(x_t, \theta_1) : \ 1 \times 10 \times 1$ (Exponential)

Training data $n_0 = n_1 = 100, 200, 500$

Testing $n = 20$, i.e. $\{x_1, \ldots, x_{20}\}$

$$\sum_{t=2}^{20} \mathsf{u}_2(x_t, x_{t-1}, \theta_2) - \sum_{t=2}^{19} \mathsf{u}_1(x_t, \theta_1) \underset{\mathsf{H}_0}{\overset{\mathsf{H}_1}{\gtreqless}} \eta$$

100000×20 samples from H_0 and H_1



What is Missing from the Analysis ?

- For fixed $\omega(\mathsf{r})$ find $\rho(z)$ that produces best estimates for $\omega(\mathsf{r}(X))$

  Requires analysis of estimation performance of neural networks for finite models (extremely challenging !!!).

- Rank functions $\omega(\mathsf{r})$ according to their approximation accuracy

  Is it possible to show that Mean Square is worse that Exponential or Cross Entropy?

Acknowledgements

Nokia Foundation

Nokia Visiting Professor Scholarship

Aalto University
School of Electrical Engineering

Department of
Signal Processing and Acoustics

Research Group of Professor
Visa Koivunen