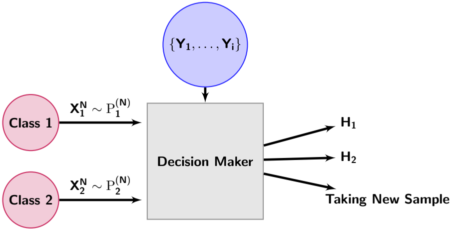
Statistical Classification with Empirically Observed Statistics

Mahdi Haghifam⋆
Joint work with
Vincent Y. F. Tan† and Ashish Khisti⋆
⋆Department of Electrical and Computer Engineering, University of Toronto.
†Department of Electrical and Computer Engineering & Department of Mathematics, National University of Singapore.

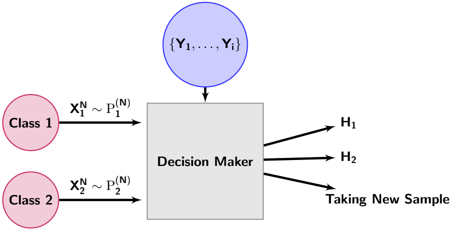
Outline

Problem Definition



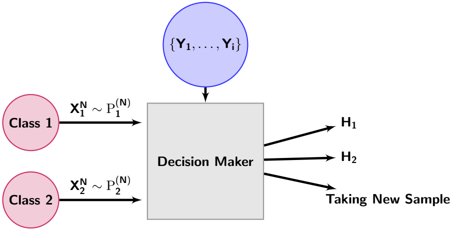
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.

Problem Definition



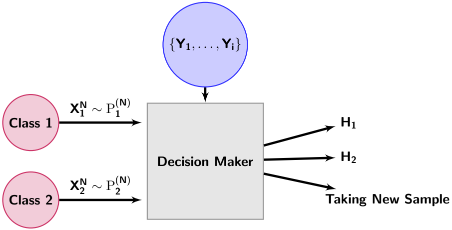
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**

Problem Definition



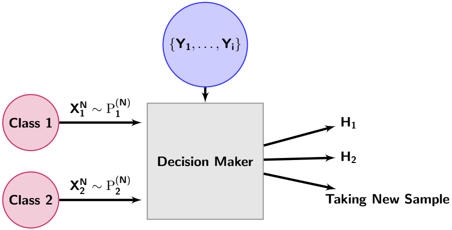
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.

Problem Definition



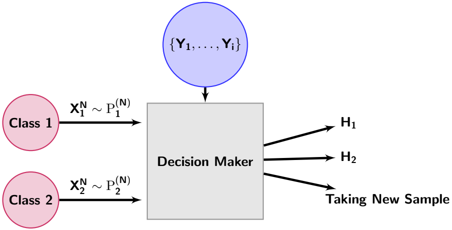
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- Two hypotheses:
  - $\mathbf{H_1}$: The test sequence belongs to Class 1.
  - $\mathbf{H_2}$: The test sequence belongs to Class 2.

Problem Definition



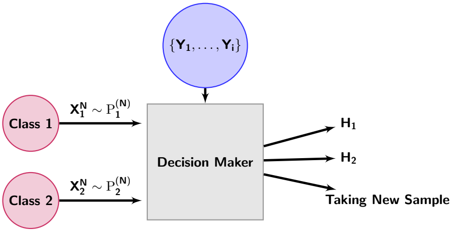
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- Two hypotheses:
  - $\mathbf{H}_1$: The test sequence belongs to Class 1.
  - $\mathbf{H}_2$: The test sequence belongs to Class 2.
- Decision maker's actions at each time $n \in \mathbb{N}$:

Problem Definition



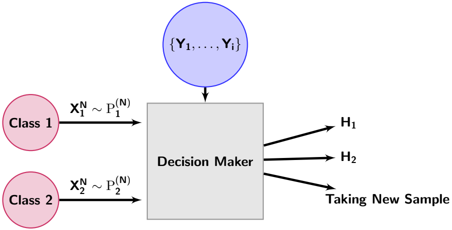
- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- Two hypotheses:
  - $\mathbf{H}_1$: The test sequence belongs to Class 1.
  - $\mathbf{H}_2$: The test sequence belongs to Class 2.
- Decision maker's actions at each time $n \in \mathbb{N}$:
  1. Declare $\mathbf{H}_1$

Problem Definition



- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- Two hypotheses:
  - $\mathbf{H}_1$: The test sequence belongs to Class 1.
  - $\mathbf{H}_2$: The test sequence belongs to Class 2.
- Decision maker's actions at each time $n \in \mathbb{N}$:
  1. Declare $\mathbf{H}_1$
  2. Declare $\mathbf{H}_2$

Problem Definition



- Two training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}$ and $X_2^N \sim \mathrm{P}_2^{(N)}$.
- **P_1 and P_2 are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- Two hypotheses:
  - $\mathbf{H}_1$: The test sequence belongs to Class 1.
  - $\mathbf{H}_2$: The test sequence belongs to Class 2.
- Decision maker's actions at each time $n \in \mathbb{N}$:
  1. Declare $\mathbf{H}_1$
  2. Declare $\mathbf{H}_2$
  3. Continue drawing a new test sample.

Problem Definition (cont'd)

Definition (Test)

A test is a pair $\Phi = (T, d)$ where

- $T \in \mathbb{N}$ is a stopping time.
- $d : \left(X_1^N, X_2^N, Y^T\right) \to \{1, 2\}$ is the terminal decision rule.

Problem Definition (cont'd)

Definition (Test)

A test is a pair $\Phi = (T, d)$ where

- $T \in \mathbb{N}$ is a stopping time.
- $d : \left(X_1^N, X_2^N, Y^T\right) \to \{1, 2\}$ is the terminal decision rule.

Definition (Error Probabilities)

For a test Φ, the error probability under the hypothesis H_i is defined as $\mathbb{P}_\Phi\left(\mathcal{E} \middle| H_i\right) = \mathbb{P}_\Phi\left(d \neq i \middle| H_i\right)$. Note that $\mathcal{E}$ and T are random variables depend on N i.e., the length of the training sequence.

Problem Definition (cont'd)

Definition (Test)

A test is a pair $\Phi = (T, d)$ where

- $T \in \mathbb{N}$ is a stopping time.
- $d : \left(X_1^N, X_2^N, Y^T\right) \to \{1, 2\}$ is the terminal decision rule.

Definition (Error Probabilities)

For a test Φ, the error probability under the hypothesis H_i is defined as $\mathbb{P}_\Phi\left(\mathcal{E}\middle|H_i\right) = \mathbb{P}_\Phi\left(d \neq i\middle|H_i\right)$. Note that $\mathcal{E}$ and T are random variables depend on N i.e., the length of the training sequence.

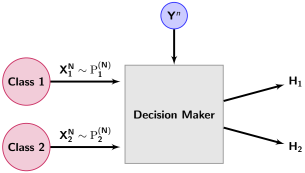
Definition (Error Exponents)

For a test Φ such that $\mathbb{P}_\Phi\left(\mathcal{E}|H_i\right) \to 0$ as $N \to \infty$, we define the error exponents as

$$E_i(\Phi) = \liminf_{N\to\infty} \frac{-\log \mathbb{P}_\Phi\left(\mathcal{E}|H_i\right)}{\mathbb{E}_\Phi\left[T|H_i\right]}$$

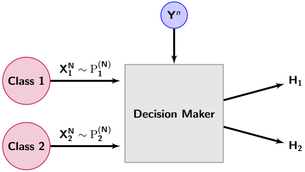
Prior Works

- Gutman (TIT 1989) setup:
  - two training sequences: $X_1^N \sim \mathrm{P}_1^N$, $X_2^N \sim \mathrm{P}_2^N$
  - test sequence of **Fixed-Length** n: Y^n

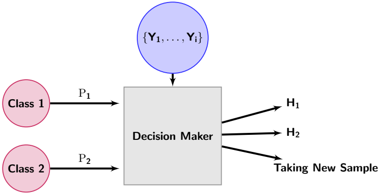


Prior Works

- Gutman (TIT 1989) setup:
  - two training sequences: $X_1^N \sim \mathrm{P}_1^N$, $X_2^N \sim \mathrm{P}_2^N$
  - test sequence of **Fixed-Length** n: Y^n



- Sequential Hypothesis Testing (Wald 1945) :
  - Distributions $(\mathrm{P}_1, \mathrm{P}_2)$ are assumed to be **Known**.



Prior Works: More on Gutman's Results

- For any $\alpha \in \mathbb{R}_+$, define Generalized Jensen-Shannon Divergence as

$$\mathrm{GJS}\left(\mathrm{P}_1, \mathrm{P}_2, \alpha\right) \triangleq \alpha \mathrm{KL}\left(\mathrm{P}_1 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right) + \mathrm{KL}\left(\mathrm{P}_2 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right)$$

Prior Works: More on Gutman's Results

- For any $\alpha \in \mathbb{R}_+$, define Generalized Jensen-Shannon Divergence as

$$\mathrm{GJS}\left(\mathrm{P}_1, \mathrm{P}_2, \alpha\right) \triangleq \alpha \mathrm{KL}\left(\mathrm{P}_1 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right) + \mathrm{KL}\left(\mathrm{P}_2 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right)$$

- $\alpha \triangleq \frac{N}{n} \rightarrow$ Number of training samples per one test sample.

Prior Works: More on Gutman's Results

- For any $\alpha \in \mathbb{R}_+$, define Generalized Jensen-Shannon Divergence as

$$\mathrm{GJS}\left(\mathrm{P}_1, \mathrm{P}_2, \alpha\right) \triangleq \alpha \mathrm{KL}\left(\mathrm{P}_1 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right) + \mathrm{KL}\left(\mathrm{P}_2 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right)$$

- $\alpha \triangleq \frac{N}{n} \rightarrow$ Number of training samples per one test sample.
- Gutman's test $\phi_n^{(\mathrm{Gut})}$ is

$$\phi_n^{(\mathrm{Gut})} \triangleq \begin{cases} 2 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \geq \lambda \\ 1 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \leq \lambda \end{cases}.$$

Prior Works: More on Gutman's Results

- For any $\alpha \in \mathbb{R}_+$, define Generalized Jensen-Shannon Divergence as

$$\mathrm{GJS}\left(\mathrm{P}_1, \mathrm{P}_2, \alpha\right) \triangleq \alpha \mathrm{KL}\left(\mathrm{P}_1 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right) + \mathrm{KL}\left(\mathrm{P}_2 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right)$$

- $\alpha \triangleq \frac{N}{n} \rightarrow$ Number of training samples per one test sample.
- Gutman's test $\phi_n^{(\mathrm{Gut})}$ is

$$\phi_n^{(\mathrm{Gut})} \triangleq \begin{cases} 2 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \geq \lambda \\ 1 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \leq \lambda \end{cases}.$$

- **Asymptotic Optimality:** The type-II error probability of Gutman't test is the least among all tests with equal type-I error exponent.

Prior Works: More on Gutman's Results

- For any $\alpha \in \mathbb{R}_+$, define Generalized Jensen-Shannon Divergence as

$$\mathrm{GJS}\left(\mathrm{P}_1, \mathrm{P}_2, \alpha\right) \triangleq \alpha \mathrm{KL}\left(\mathrm{P}_1 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right) + \mathrm{KL}\left(\mathrm{P}_2 \| \frac{\alpha \mathrm{P}_1 + \mathrm{P}_2}{1+\alpha}\right)$$

- $\alpha \triangleq \frac{N}{n} \rightarrow$ Number of training samples per one test sample.
- Gutman's test $\phi_n^{(\mathrm{Gut})}$ is

$$\phi_n^{(\mathrm{Gut})} \triangleq \begin{cases} 2 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \geq \lambda \\ 1 & \text{if } \mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^n}, \alpha\right) \leq \lambda \end{cases}.$$

- **Asymptotic Optimality:** The type-II error probability of Gutman't test is the least among all tests with equal type-I error exponent.

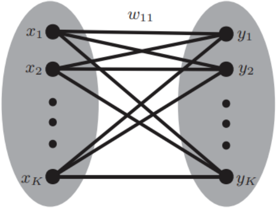
Our goal in this work is to propose a **sequential** version of Gutman's framework.

Other Related Works

- Second-order analysis [Zhou et al, 2018]: What happened if n and N in Gutman's setup are finite?

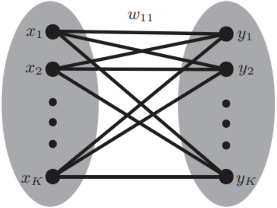
Other Related Works

- Second-order analysis [Zhou et al, 2018]: What happened if *n* and *N* in Gutman's setup are finite?
- Multiple Test Sequences [Unnikrishnan 2017]



Other Related Works

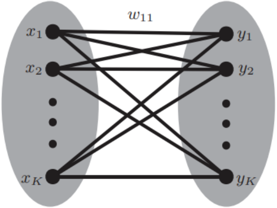
- Second-order analysis [Zhou et al, 2018]: What happened if n and N in Gutman's setup are finite?
- Multiple Test Sequences [Unnikrishnan 2017]



- Alphabet Size [Kelly et al 2013]: Assume the scenario that the alphabet size grows with length of training and test sequences $\Rightarrow$ model for natural language $\Rightarrow$ what is the largest possible growth rate?

Other Related Works

- Second-order analysis [Zhou et al, 2018]: What happened if *n* and *N* in Gutman's setup are finite?
- Multiple Test Sequences [Unnikrishnan 2017]



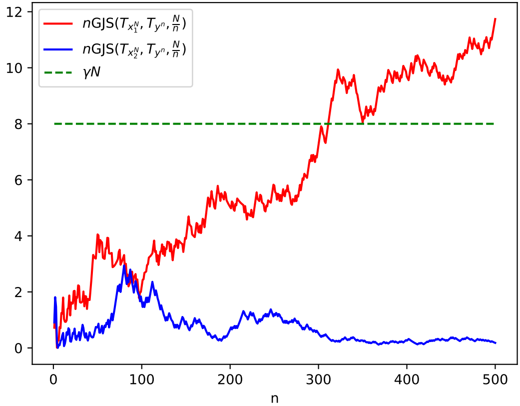
- Alphabet Size [Kelly et al 2013]: Assume the scenario that the alphabet size grows with length of training and test sequences ⇒ model for natural language ⇒ what is the largest possible growth rate?
- Closeness Testing [Acharya et al, 2012]. Given two sequences, we want to determine whether they are from the same distribution or not.

Main Results: Motivation

- Design $\{f_t^{(i)}\}_{t\in\mathbb{N}}$ for $i \in \{1, 2\}$.
- $f_t^{(i)} : \mathcal{X}^N \times \mathcal{X}^t \to \mathbb{R}$.
- $f_t^{(i)} \to$ measures whether the test sequence and training sequence i generated according to the same distribution or not.

Main Results: Motivation

- Design $\{f_t^{(i)}\}_{t\in\mathbb{N}}$ for $i \in \{1,2\}$.
- $f_t^{(i)} : \mathcal{X}^N \times \mathcal{X}^t \to \mathbb{R}$.
- $f_t^{(i)} \to$ measures whether the test sequence and training sequence i generated according to the same distribution or not.



Main Results: Proposed Test

Proposed Test for the Sequential Setting

Let $\Phi_{\mathsf{seq}} = (T_{\mathsf{seq}}, d_{\mathsf{seq}})$. Our proposed test consists of

$$T_{\mathsf{seq}} = \inf\left\{n \geq 1 : \exists i \in \{1,2\} \text{ such that } n\mathrm{GJS}\left(\mathcal{T}_{X_i^N}, \mathcal{T}_{Y^n}, \frac{N}{n}\right) \geq \gamma N\right\},$$

Main Results: Proposed Test

Proposed Test for the Sequential Setting

Let $\Phi_{\text{seq}} = (T_{\text{seq}}, d_{\text{seq}})$. Our proposed test consists of

$$T_{\text{seq}} = \inf\left\{ n \geq 1 : \exists i \in \{1,2\} \text{ such that } n\mathrm{GJS}\left(\mathcal{T}_{X_i^N}, \mathcal{T}_{Y^n}, \frac{N}{n}\right) \geq \gamma N \right\},$$

$$d_{\text{seq}} = \begin{cases} 1 & \text{if } T_{\text{seq}}\mathrm{GJS}\left(\mathcal{T}_{X_2^N}, \mathcal{T}_{Y^{T_{\text{seq}}}}, \frac{N}{T_{\text{seq}}}\right) \geq \gamma N, \\ 2 & \text{if } T_{\text{seq}}\mathrm{GJS}\left(\mathcal{T}_{X_1^N}, \mathcal{T}_{Y^{T_{\text{seq}}}}, \frac{N}{T_{\text{seq}}}\right) \geq \gamma N \end{cases}.$$

Main Results: Achievable Error Exponent

Theorem(Achievable Error Exponent)

Fix pair $(\mathrm{P}_1, \mathrm{P}_2) \in \mathcal{P}(\mathcal{X})^2$. Then, the proposed test achieves

$$E_1(\Phi_{\mathsf{seq}}) = \liminf_{N \to \infty} \frac{-\log \mathbb{P}_{\Phi_{\mathsf{seq}}}(\mathcal{E}|H_1)}{\mathbb{E}_{\Phi_{\mathsf{seq}}}[T_{\mathsf{seq}}|H_1]} \geq \mathrm{GJS}(\mathrm{P}_2, \mathrm{P}_1, \beta^\star), \quad (1)$$

$$E_2(\Phi_{\mathsf{seq}}) = \liminf_{N \to \infty} \frac{-\log \mathbb{P}_{\Phi_{\mathsf{seq}}}(\mathcal{E}|H_2)}{\mathbb{E}_{\Phi_{\mathsf{seq}}}[T_{\mathsf{seq}}|H_2]} \geq \mathrm{GJS}(\mathrm{P}_1, \mathrm{P}_2, \theta^\star). \quad (2)$$

Here, in (17), $\beta^\star$ is the solution of $\mathrm{GJS}(\mathrm{P}_2, \mathrm{P}_1, \beta^\star) = \gamma\beta^\star$. Similarly, $\theta^\star$ in (2) is the solution of $\mathrm{GJS}(\mathrm{P}_1, \mathrm{P}_2, \theta^\star) = \gamma\theta^\star$.

Main Results: Achievable Error Exponent

Theorem(Achievable Error Exponent)

Fix pair $(\mathrm{P}_1, \mathrm{P}_2) \in \mathcal{P}(\mathcal{X})^2$. Then, the proposed test achieves

$$E_1(\Phi_{\text{seq}}) = \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi_{\text{seq}}}(\mathcal{E}|H_1)}{\mathbb{E}_{\Phi_{\text{seq}}}[T_{\text{seq}}|H_1]} \geq \mathrm{GJS}(\mathrm{P}_2, \mathrm{P}_1, \beta^\star), \quad (1)$$

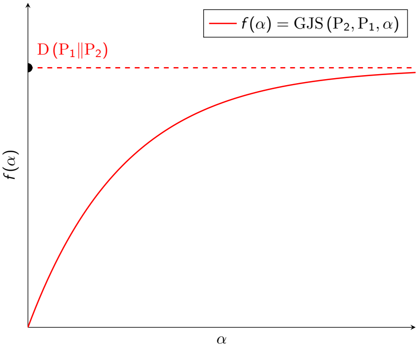
$$E_2(\Phi_{\text{seq}}) = \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi_{\text{seq}}}(\mathcal{E}|H_2)}{\mathbb{E}_{\Phi_{\text{seq}}}[T_{\text{seq}}|H_2]} \geq \mathrm{GJS}(\mathrm{P}_1, \mathrm{P}_2, \theta^\star). \quad (2)$$

Here, in (17), $\beta^\star$ is the solution of $\mathrm{GJS}(\mathrm{P}_2, \mathrm{P}_1, \beta^\star) = \gamma\beta^\star$. Similarly, $\theta^\star$ in (2) is the solution of $\mathrm{GJS}(\mathrm{P}_1, \mathrm{P}_2, \theta^\star) = \gamma\theta^\star$.

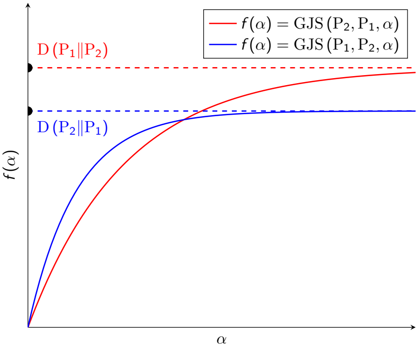
Theorem(Expected Value of the Stopping Time)

- $\mathbb{E}\left[T_{\text{seq}} \middle| H_1\right] = \frac{N}{\beta^\star}(1 + o_N(1))$
- $\mathbb{E}\left[T_{\text{seq}} \middle| H_2\right] = \frac{N}{\theta^\star}(1 + o_N(1))$

Main Results: Discussion

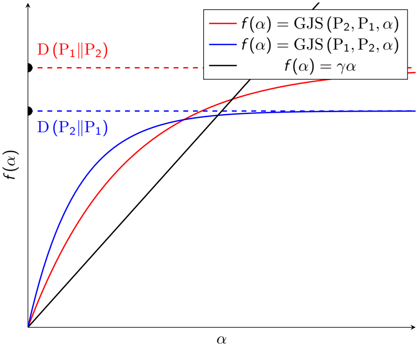


Main Results: Discussion



1. Uniqueness of the intersections $\rightarrow$ Concavity

Main Results: Discussion



1. Uniqueness of the intersections $\rightarrow$ Concavity
2. When $\beta^\star$ and $\theta^\star$ goes to infinity $\rightarrow$ Number of training samples per a test sample tends to infinity for H_1 and H_2 $\rightarrow$ $(\mathrm{D}(\mathrm{P}_1 \| \mathrm{P}_2), \mathrm{D}(\mathrm{P}_2 \| \mathrm{P}_1))$ is achievable $\rightarrow$ Converse for Sequential Hypothesis Testing:

$$\mathrm{E}_1(\Phi) \times \mathrm{E}_2(\Phi) \leq \mathrm{D}(\mathrm{P}_1 \| \mathrm{P}_2) \times \mathrm{D}(\mathrm{P}_2 \| \mathrm{P}_1)$$

Comparison with Gutman's test: Empirical Results

- The performance metric is the minimum of type-I and type-II error exponents.

Comparison with Gutman's test: Empirical Results

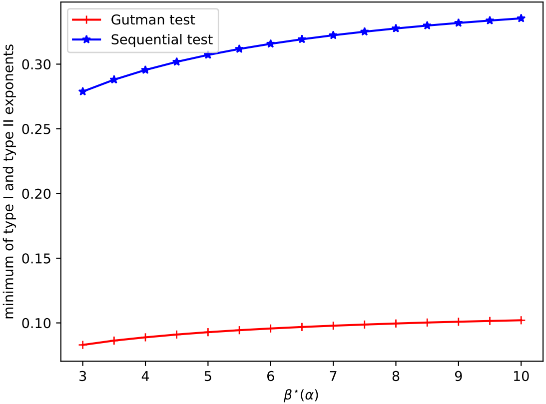
- The performance metric is the minimum of type-I and type-II error exponents.
- For Gutman's, the best trade-off is considered $\max_{\lambda \geq 0}\{\min\{E_1(\Phi_{\mathrm{Gut}}), E_2(\Phi_{\mathrm{Gut}})\}\}$.

Comparison with Gutman's test: Empirical Results

- The performance metric is the minimum of type-I and type-II error exponents.
- For Gutman's, the best trade-off is considered $\max_{\lambda \geq 0}\{\min\{E_1(\Phi_{\text{Gut}}), E_2(\Phi_{\text{Gut}})\}\}$.
- $\mathcal{X} = \{1, 2, 3, 4\}$, $\mathrm{P}_1 = [0.25, 0.25, 0.25, 0.25]$, and $\mathrm{P}_2 = [0.4, 0.5, 0.05, 0.05]$.
- $\#$ test samples under H_2 for $\Phi_{\text{seq}} = \#$ test samples for Φ_{Gut}.

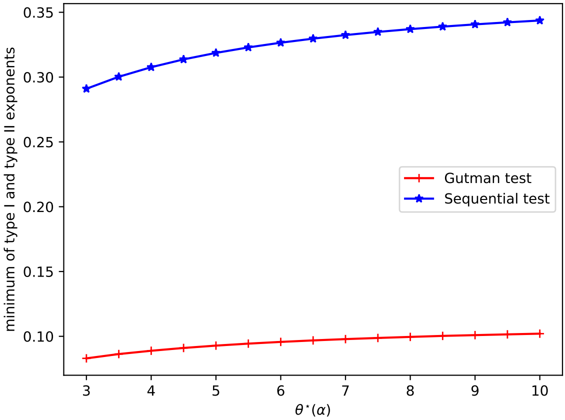
Comparison with Gutman's test: Empirical Results

- The performance metric is the minimum of type-I and type-II error exponents.
- For Gutman's, the best trade-off is considered $\max_{\lambda \geq 0}\{\min\{E_1(\Phi_{\text{Gut}}), E_2(\Phi_{\text{Gut}})\}\}$.
- $\mathcal{X} = \{1,2,3,4\}$, $\mathrm{P}_1 = [0.25, 0.25, 0.25, 0.25]$, and $\mathrm{P}_2 = [0.4, 0.5, 0.05, 0.05]$.
- $\#$ test samples under H_2 for $\Phi_{\text{seq}} = \#$ test samples for Φ_{Gut}.



Comparison with Gutman's test: Empirical Results

- \# test samples under H_1 for Φ_{seq} = # test samples for Φ_{Gut}.



Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \pi_1 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_1\right) + \pi_2 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_2\right)$$

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \pi_1 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_1\right) + \pi_2 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_2\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has two training sequence of length N, and we define the error exponent for the Bayesian scenario as

$$e_{\text{bayesian}}\left(\Phi\right) \triangleq \liminf_{N \to \infty} \frac{-\log \mathbb{P}_{\Phi}\left(\mathcal{E}\right)}{N}$$

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}(\mathcal{E}) = \pi_1 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_1\right) + \pi_2 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_2\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has two training sequence of length N, and we define the error exponent for the Bayesian scenario as

$$e_{\text{bayesian}}(\Phi) \triangleq \liminf_{N \to \infty} \frac{-\log \mathbb{P}_{\Phi}(\mathcal{E})}{N}$$

Let $\alpha \triangleq \min\{\theta^\star, \beta^\star\}$ where $\mathbb{E}\left[T_{\text{seq}} \middle| H_1\right] \simeq \frac{N}{\beta^\star}$ and $\mathbb{E}\left[T_{\text{seq}} \middle| H_2\right] \simeq \frac{N}{\theta^\star}$. *To ensure a fair comparison, we consider the case where* $\frac{N}{\alpha}$ *samples are provided for Gutman's test.*

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \pi_1 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_1\right) + \pi_2 \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_2\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has two training sequence of length N, and we define the error exponent for the Bayesian scenario as

$$e_{\text{bayesian}}\left(\Phi\right) \triangleq \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi}\left(\mathcal{E}\right)}{N}$$

Let $\alpha \triangleq \min\{\theta^\star, \beta^\star\}$ where $\mathbb{E}\left[T_{\text{seq}} \middle| H_1\right] \simeq \frac{N}{\beta^\star}$ and $\mathbb{E}\left[T_{\text{seq}} \middle| H_2\right] \simeq \frac{N}{\theta^\star}$. *To ensure a fair comparison, we consider the case where* $\frac{N}{\alpha}$ *samples are provided for Gutman's test*. Then, we have

$$e_{\text{bayesian}}\left(\Phi_{\text{seq}}\right) > e_{\text{bayesian}}\left(\Phi_{\text{Gut}}\right)$$

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **P_1, …, P_M are unknown to the decision maker.**

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **P_1, ..., P_M are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **$\mathrm{P}_1, \ldots, \mathrm{P}_M$ are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- M hypotheses:
  - $\mathbf{H_i}$: The test sequence belongs to Class i.

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **P_1, ..., P_M are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- M hypotheses:
  - $\mathbf{H_i}$: The test sequence belongs to Class i.
- Decision maker's actions at each time $n \in \mathbb{N}$:

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **P_1, ..., P_M are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- M hypotheses:
    - $\mathbf{H_i}$: The test sequence belongs to Class i.
- Decision maker's actions at each time $n \in \mathbb{N}$:
    1. Declare $\mathbf{H_i}$ for $i \in \{1, \ldots, M\}$ and Stop.

Sequential Classification with Multiple classes

- M training sequences: $X_1^N \sim \mathrm{P}_1^{(N)}, \ldots, X_M^N \sim \mathrm{P}_M^{(N)}$.
- **P_1, ..., P_M are unknown to the decision maker.**
- Y_n : *test sample* generated at each time $n \in \mathbb{N}$.
- M hypotheses:
  - **$\mathbf{H_i}$**: The test sequence belongs to Class i.
- Decision maker's actions at each time $n \in \mathbb{N}$:
  1. Declare $\mathbf{H_i}$ for $i \in \{1, \ldots, M\}$ and Stop.
  2. Continue drawing a new test sample.

Main Results: Proposed Test

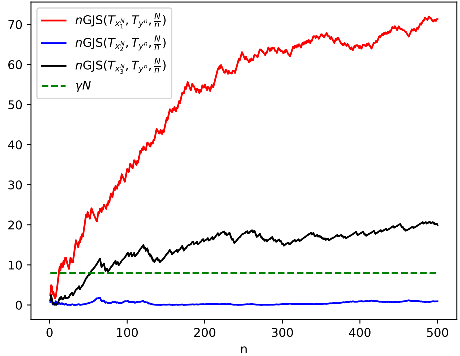
- Let
  $\Psi_n \triangleq \left\{ i \in \{1, \ldots, M\} : \exists 1 \leq k \leq n \text{ such that } k\mathrm{GJS}\left(\widehat{Q}_{X_i^N}, \widehat{Q}_{Y^k}, \frac{N}{k}\right) \geq \gamma N \right\}.$

Main Results: Proposed Test

- Let
  $\Psi_n \triangleq \left\{ i \in \{1, \ldots, M\} : \exists 1 \leq k \leq n \text{ such that } k\mathrm{GJS}\left(\widehat{Q}_{X_i^N}, \widehat{Q}_{Y^k}, \frac{N}{k}\right) \geq \gamma N \right\}.$
- The stopping time is $T_{\text{seq}}^{(M)} \triangleq \inf \left\{ n \geq 1 : \mathrm{card}\left(\Psi_n\right) \geq M - 1 \right\}.$

Main Results: Proposed Test

- Let $\Psi_n \triangleq \left\{ i \in \{1,\ldots,M\} : \exists 1 \leq k \leq n \text{ such that } k\mathrm{GJS}\left(\widehat{Q}_{X_i^N}, \widehat{Q}_{Y^k}, \frac{N}{k}\right) \geq \gamma N \right\}$.
- The stopping time is $T_{\text{seq}}^{(M)} \triangleq \inf\left\{ n \geq 1 : \mathrm{card}\left(\Psi_n\right) \geq M-1 \right\}$.
- The final decision rule is $d_{\text{seq}}^{(M)} \triangleq \{1,\ldots,M\} - \Psi_{T_{\text{seq}}^{(M)}}$



Main Results: Achievable Exponent

Theorem(Achievable Error Exponent)

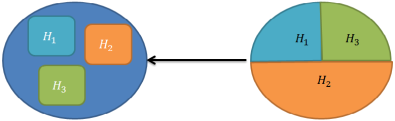
Fix $(\mathrm{P}_1, \ldots, \mathrm{P}_M) \in \mathcal{P}(\mathcal{X})^M$. Then, the proposed test achieves

$$E_i\left(\Phi_{\text{seq}}^{(M)}\right) = \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi_{\text{seq}}^{M}}(\mathcal{E}|H_i)}{\mathbb{E}_{\Phi_{\text{seq}}^{(M)}}[T_{\text{seq}}|H_i]} \geq \min_{j\in\{1,\ldots,M\}, j\neq i} \mathrm{GJS}\left(\mathrm{P}_j, \mathrm{P}_i, \beta^\star_{i(j)}\right)$$

where $\mathrm{GJS}\left(P_j, P_i, \beta^\star_{i(j)}\right) = \gamma\beta^\star_{i(j)}$ for $j \in \{1, \ldots, M\}, j \neq i$.

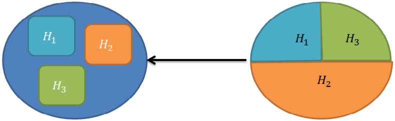
Comparison with Gutman's test: Mutliple Class

- What is the Rejection Region?



Comparison with Gutman's test: Mutliple Class

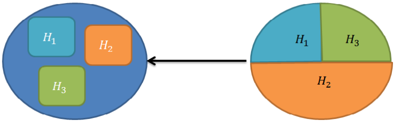
- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!

Comparison with Gutman's test: Mutliple Class

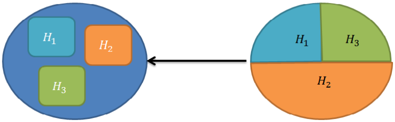
- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \dots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.

Comparison with Gutman's test: Mutliple Class

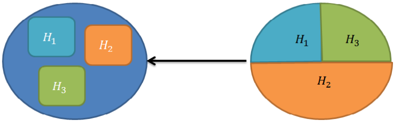
- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.
- Definition of the error probability is different!
  $\mathbb{P}\left[\mathcal{E} \middle| H_i\right] = \mathbb{P}\left[\text{final decision} \notin \{H_i, H_r\} \middle| H_i\right]$

Comparison with Gutman's test: Mutliple Class

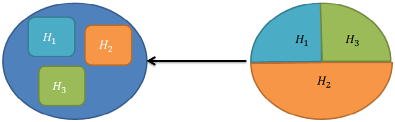
- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.
- Definition of the error probability is different!
  $\mathbb{P}\left[\mathcal{E} \middle| H_i\right] = \mathbb{P}\left[\text{final decision} \notin \{H_i, H_r\} \middle| H_i\right]$
- We also have rejection probability $\mathbb{P}\left[\mathcal{R} \middle| H_i\right] = \mathbb{P}\left[\text{final decision} = H_r \middle| H_i\right]$

Comparison with Gutman's test: Mutliple Class

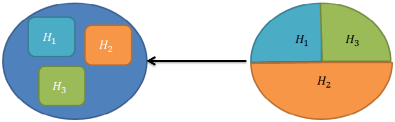
- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.
- Definition of the error probability is different!
  $\mathbb{P}\left[\mathcal{E} \middle| H_i\right] = \mathbb{P}\left[\text{final decision} \notin \{H_i, H_r\} \middle| H_i\right]$
- We also have rejection probability $\mathbb{P}\left[\mathcal{R} \middle| H_i\right] = \mathbb{P}\left[\text{final decision} = H_r \middle| H_i\right]$
- Uniform error exponent. $\liminf_{n\to\infty} \frac{-\log \mathbb{P}\left[\mathcal{E} \middle| H_i\right]}{n} \geq \lambda$ for all $i \in \{1, \ldots, M\}$.

Comparison with Gutman's test: Mutliple Class

- What is the Rejection Region?



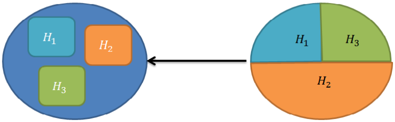
- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.
- Definition of the error probability is different!
  $\mathbb{P}\left[\mathcal{E}\middle|H_i\right] = \mathbb{P}\left[\text{final decision} \notin \{H_i, H_r\}\middle|H_i\right]$
- We also have rejection probability $\mathbb{P}\left[\mathcal{R}\middle|H_i\right] = \mathbb{P}\left[\text{final decision} = H_r\middle|H_i\right]$
- Uniform error exponent. $\liminf_{n\to\infty} \frac{-\log \mathbb{P}\left[\mathcal{E}\middle|H_i\right]}{n} \geq \lambda$ for all $i \in \{1, \ldots, M\}$.

a Fundamental Question

What is the largest possible λ such that as n goes to infinity the rejection error probability tends to zero?

Comparison with Gutman's test: Mutliple Class

- What is the Rejection Region?



- There are M+1 hypothesis:
  1. $\mathbf{H_i}$: The test sequence belongs to Class i. $\forall i \in \{1, \ldots, M\}$.
  2. $\mathbf{H_r}$: Declare Rejection!
- *fixed-length* test sequence: Y^n.
- Definition of the error probability is different!
  $\mathbb{P}\left[\mathcal{E}\middle|H_i\right] = \mathbb{P}\left[\text{final decision} \notin \{H_i, H_r\}\middle|H_i\right]$
- We also have rejection probability $\mathbb{P}\left[\mathcal{R}\middle|H_i\right] = \mathbb{P}\left[\text{final decision} = H_r\middle|H_i\right]$
- Uniform error exponent. $\liminf_{n\to\infty} \frac{-\log \mathbb{P}\left[\mathcal{E}\middle|H_i\right]}{n} \geq \lambda$ for all $i \in \{1, \ldots, M\}$.

a Fundamental Question

What is the largest possible λ such that as n goes to infinity the rejection error probability tends to zero? $\Rightarrow$ **Gutman found the answer!**

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \sum_{i=1}^{M} \pi_i \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_i\right)$$

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \sum_{i=1}^{M} \pi_i \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_i\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has M training sequence of length N, and we define the error exponent for the Bayesian scenario as $e_{\text{bayesian}}\left(\Phi\right) \triangleq \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi}(\mathcal{E})}{N}$

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \sum_{i=1}^{M} \pi_i \mathbb{P}_{\Phi}\left(\mathcal{E} \middle| H_i\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has M training sequence of length N, and we define the error exponent for the Bayesian scenario as $e_{\text{bayesian}}\left(\Phi\right) \triangleq \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi}(\mathcal{E})}{N}$ *To ensure a fair comparison, we consider the case where* $\frac{N}{\min\{\theta_1^\star,\ldots,\theta_M^\star\}}$ *samples are provided for Gutman's test.*

Comparison with Gutman's test: Theoretical Results

Bayesian Error Exponent

Define Bayesian error probability for test Φ as

$$\mathbb{P}_{\Phi}\left(\mathcal{E}\right) = \sum_{i=1}^{M} \pi_i \mathbb{P}_{\Phi}\left(\mathcal{E}\middle|H_i\right)$$

Let us assume each of both schemes, i.e., proposed and Gutman's tests, has M training sequence of length N, and we define the error exponent for the Bayesian scenario as $e_{\text{bayesian}}\left(\Phi\right) \triangleq \liminf_{N\to\infty} \frac{-\log \mathbb{P}_{\Phi}(\mathcal{E})}{N}$ *To ensure a fair comparison, we consider the case where* $\frac{N}{\min\{\theta_1^\star,\ldots,\theta_M^\star\}}$ *samples are provided for Gutman's test*. Then, we have

$$e_{\text{bayesian}}\left(\Phi_{\text{seq}}^{(M)}\right) = e_{\text{bayesian}}\left(\Phi_{\text{Gut}}^{(M)}\right)$$

Thus, our test achieves the same Bayesian error exponent without having rejection region!

Summary and Future Work

- Summary
  1. We extended the Gutman's framework for classification to the sequential setting.
  2. We proposed a test for the problem.
  3. We showed that this test surpasses the Gutman's scheme in terms of the Bayesian error exponent.
- Future Works
  1. Extension to the Markov Sources.
  2. Converse Bound.

Thank You!