Second-Order Asymptotically Optimal Statistical Classification

Lin Zhou (ECE, NUS)
Joint work with Vincent Tan and Mehul Motani

Sept. 18th, 2018
Binary Hypothesis Testing (BHT): System Model

Known dist. $P_1$ $P_2$

Test Seq. $Y^n$

Classifier $H_1 : Y^n \sim P^n_1$

$H_2 : Y^n \sim P^n_2$

Task: design a classifier $X^n \rightarrow f(X^n) = H_1; g(X^n) = H_2$

Acceptance region: $A(n) = \{ y^n : y^n = H_1 \}$

Error probabilities:

$1(n) = \Pr_{f(n)}(Y^n = H_2 | H_1) = P^n_1 f A(n)$

$2(n) = \Pr_{f(n)}(Y^n = H_1 | H_2) = P^n_2 f A(n)$

Neyman-Pearson Lemma: likelihood ratio test is optimal.
Binary Hypothesis Testing (BHT): System Model

Task: design a classifier $\delta_n$

$$
\delta_n : \mathcal{X}^n \rightarrow \{H_1, H_2\}.
$$
**Binary Hypothesis Testing (BHT): System Model**

- **Test Seq.** $Y^n$
- Known dist. $P_1$ $P_2$
- Classifier
- $H_1 : Y^n \sim P_1^n$
- $H_2 : Y^n \sim P_2^n$

- **Task:** design a classifier $\delta_n$

\[
\delta_n : \mathcal{X}^n \rightarrow \{H_1, H_2\}
\]

- **Acceptance region:** $\mathcal{A}(\delta_n) := \{y^n : \delta_n(y^n) = H_1\}$.
Binary Hypothesis Testing (BHT): System Model

Known dist. $P_1$ $P_2$

Test Seq. $Y^n$ $\rightarrow$ Classifier $\rightarrow$

$H_1: Y^n \sim P^n_1$

$H_2: Y^n \sim P^n_2$

- Task: design a classifier $\delta_n$

$$\delta_n: \mathcal{X}^n \rightarrow \{H_1, H_2\}.$$ 

- Acceptance region: $\mathcal{A}(\delta_n) := \{y^n : \delta_n(y^n) = H_1\}.$

- Error probabilities:

$$\beta_1(\delta_n) := \Pr\{\delta_n(Y^n) = H_2 | H_1\} = P^n_1\{\mathcal{A}^c(\delta_n)\},$$
Binary Hypothesis Testing (BHT): System Model

- Known dist. $P_1$, $P_2$
- Test Seq. $Y^n$
- Classifier
- $H_1: Y^n \sim P_1^n$
- $H_2: Y^n \sim P_2^n$

**Task:** design a classifier $\delta_n$

$\delta_n: \mathcal{X}^n \rightarrow \{H_1, H_2\}$.

**Acceptance region:** $\mathcal{A}(\delta_n) := \{y^n: \delta_n(y^n) = H_1\}$.

**Error probabilities:**

$\beta_1(\delta_n) := \Pr\{\delta_n(Y^n) = H_2 | H_1\} = P_1^n\{A^c(\delta_n)\}$,

$\beta_2(\delta_n) := \Pr\{\delta_n(Y^n) = H_1 | H_2\} = P_2^n\{A(\delta_n)\}$.
Binary Hypothesis Testing (BHT): System Model

- **Task**: design a classifier $\delta_n$

  $$\delta_n : \mathcal{X}^n \rightarrow \{H_1, H_2\}.$$ 

- **Acceptance region**: $\mathcal{A}(\delta_n) := \{y^n : \delta_n(y^n) = H_1\}$. 

- **Error probabilities**:

  $$\beta_1(\delta_n) := \Pr\{\delta_n(Y^n) = H_2|H_1\} = P_1^n\{\mathcal{A}^c(\delta_n)\},$$

  $$\beta_2(\delta_n) := \Pr\{\delta_n(Y^n) = H_1|H_2\} = P_2^n\{\mathcal{A}(\delta_n)\}.$$ 

- **Neyman-Pearson Lemma**: likelihood ratio test is optimal.
BHT: Bayesian and Non-Bayesian Settings

- Bayesian setting: under prior distribution of hypotheses:

\[ P_{e,n}^* := \inf_{\delta_n} \{ \Pr\{H_1\} \beta_1(\delta_n) + \Pr\{H_2\} \beta_2(\delta_n) \}. \]
BHT: Bayesian and Non-Bayesian Settings

- **Bayesian setting:** under prior distribution of hypotheses:
  \[ P_{e,n}^* := \inf_{\delta_n} \{ \Pr\{H_1\} \beta_1(\delta_n) + \Pr\{H_2\} \beta_2(\delta_n) \}. \]

- **Asymptotics for Bayesian setting (Chernoff Information):**
  \[
  \lim_{n \to \infty} -\frac{1}{n} \log P_{e,n}^* = C(P_1, P_2).
  \]
BHT: Bayesian and Non-Bayesian Settings

- Bayesian setting: under prior distribution of hypotheses:
  \[ P_{e,n}^* := \inf_{\delta_n} \{ \Pr\{H_1\} \beta_1(\delta_n) + \Pr\{H_2\} \beta_2(\delta_n) \} . \]

- Asymptotics for Bayesian setting (Chernoff Information):
  \[ \lim_{n \to \infty} -\frac{1}{n} \log P_{e,n}^* = C(P_1, P_2). \]

- Non-Bayesian setting: for any \( \varepsilon \in (0, 1) \),
  \[ \beta_{1,n}^*(\varepsilon) := \inf_{\delta_n: \beta_2(\delta_n) \leq \varepsilon} \beta_1(\delta_n). \]
BHT: Bayesian and Non-Bayesian Settings

- Bayesian setting: under prior distribution of hypotheses:
  \[ P_{e,n}^* := \inf_{\delta_n} \{ \Pr\{H_1\} \beta_1(\delta_n) + \Pr\{H_2\} \beta_2(\delta_n) \} \]

- Asymptotics for Bayesian setting (Chernoff Information):
  \[ \lim_{n \to \infty} - \frac{1}{n} \log P_{e,n}^* = C(P_1, P_2) \]

- Non-Bayesian setting: for any \( \varepsilon \in (0, 1) \),
  \[ \beta_{1,n}(\varepsilon) := \inf_{\delta_n: \beta_2(\delta_n) \leq \varepsilon} \beta_1(\delta_n) \]

- Exponential type-I error \( \varepsilon = \exp(-nE) \) (Blahut 1974)
  \[ \lim_{n \to \infty} - \frac{1}{n} \log \beta_{1,n}(\varepsilon) = \min_{Q: D(Q||P_1) \leq E} D(Q||P_2) \]
Constant type-I error (Chernoff-Stein Lemma)

\[ \lim_{n \to \infty} - \frac{1}{n} \log \beta_{1,n}^*(\varepsilon) = D(P_2 \parallel P_1), \quad \forall \varepsilon \in (0, 1). \]
Constant type-I error (Chernoff-Stein Lemma)

\[
\lim_{n \to \infty} \frac{1}{n} \log \beta_{1,n}^*(\varepsilon) = D(P_2\|P_1), \quad \forall \varepsilon \in (0, 1).
\]

Second-order asymptotics (Strassen 1962): for any \(\varepsilon \in (0, 1)\),

\[
\frac{1}{n} \log \beta_{1,n}^*(\varepsilon) = D(P_2\|P_1) + \sqrt{\frac{V(P_2\|P_1)}{n}} \Phi^{-1}(\varepsilon) + \frac{1}{2} \frac{\log n}{n} + O(1),
\]

where

\[
V(P_2\|P_1) = \text{Var}_{P_2} \left[ \log \frac{P_2(X)}{P_1(X)} \right].
\]
From BHT to Binary Classification

 Known dist. $P_1$ $P_2$

 Test Seq. $Y^n$

 Classifier

 H$_1$: $Y^n \sim P_1^n$

 H$_2$: $Y^n \sim P_2^n$

Training sequences $X^N_i$ for unknown distributions $(P_1, P_2)$.

Task: Design a test $\phi_n$: $X^2_n + N! f$ $H_1: Y^n \sim P_1^n$ and $X^N_1$ are generated according to the same distribution.

H$_2$: $Y^n \sim P_2^n$ and $X^N_2$ are generated according to the same distribution.

Assumption: $N = \lceil n \rceil$ for some $2 \in \mathbb{R}^+$. 

Zhou-Tan-Motani (NUS) Statistical Classification Group Meeting 5/23
From BHT to Binary Classification

$$(P_1, P_2) \text{ unknown}$$ in practical machine learning applications: image classification, junk mail identification, etc.
From BHT to Binary Classification

- $(P_1, P_2)$ unknown in practical machine learning applications: image classification, junk mail identification, etc.
- Training sequences $X_i^N \sim P_i^N$ for unknown distributions $(P_1, P_2)$.

---

(\(P_1, P_2\)) unknown in practical machine learning applications: image classification, junk mail identification, etc.

Training sequences \(X_i^N \sim P_i^N\) for unknown distributions \((P_1, P_2)\).

Task: Design a test \(\phi_n: \mathcal{X}^{2n+N} \rightarrow \{H_1, H_2\}\)

- \(H_1: Y^n\) and \(X_1^N\) are generated according to same distribution.
- \(H_2: Y^n\) and \(X_2^N\) are generated according to same distribution.

---

From BHT to Binary Classification

- $(P_1, P_2)$ unknown in practical machine learning applications: image classification, junk mail identification, etc.
- Training sequences $X_i^N \sim P_i^N$ for unknown distributions $(P_1, P_2)$.
- Task: Design a test $\phi_n : X^{2n+N} \rightarrow \{H_1, H_2\}$
  - $H_1$: $Y^n$ and $X_1^N$ are generated according to same distribution.
  - $H_2$: $Y^n$ and $X_2^N$ are generated according to same distribution.
- Assumption: $N = \lceil \alpha n \rceil$ for some $\alpha \in \mathbb{R}_+$.

---

Definitions and Gutman’s Test

Given $\phi_n$ and a pair of distributions $(P_1, P_2)$, define

$$\beta_1(\phi_n | P_1, P_2) := \Pr\{\phi_n(X_1^N, X_2^N, Y^n) = H_2 | H_1\},$$
Definitions and Gutman’s Test

Given $\phi_n$ and a pair of distributions $(P_1, P_2)$, define

$$\beta_1(\phi_n|P_1, P_2) := \Pr \{ \phi_n(X_1^N, X_2^N, Y^n) = H_2|H_1 \},$$
$$\beta_2(\phi_n|P_1, P_2) := \Pr \{ \phi_n(X_1^N, X_2^N, Y^n) = H_1|H_2 \}.$$
Definitions and Gutman’s Test

- Given $\phi_n$ and a pair of distributions $(P_1, P_2)$, define
  
  \[ \beta_1(\phi_n|P_1, P_2) := \Pr\{\phi_n(X_1^N, X_2^N, Y^n) = H_2|H_1\}, \]
  \[ \beta_2(\phi_n|P_1, P_2) := \Pr\{\phi_n(X_1^N, X_2^N, Y^n) = H_1|H_2\}. \]

- Generalized Jensen-Shannon divergence
  
  \[ \text{GJS}(P_1, P_2, \alpha) := \alpha D\left(P_1 \parallel \frac{\alpha P_1 + P_2}{1 + \alpha}\right) + D\left(P_2 \parallel \frac{\alpha P_1 + P_2}{1 + \alpha}\right). \]
Definitions and Gutman’s Test

- Given $\phi_n$ and a pair of distributions $(P_1, P_2)$, define
  
  \[ \beta_1(\phi_n|P_1, P_2) := \Pr\{\phi_n(X_1^N, X_2^N, Y^n) = H_2|H_1]\}, \]
  
  \[ \beta_2(\phi_n|P_1, P_2) := \Pr\{\phi_n(X_1^N, X_2^N, Y^n) = H_1|H_2\}. \]

- Generalized Jensen-Shannon divergence
  
  \[ \text{GJS}(P_1, P_2, \alpha) := \alpha D\left( P_1 \parallel \frac{\alpha P_1 + P_2}{1 + \alpha} \right) + D\left( P_2 \parallel \frac{\alpha P_1 + P_2}{1 + \alpha} \right). \]

- Gutman’s test (TIT 1989): fix a positive $\lambda$
  
  \[ \phi_{\text{Gut}}^n(x_1^N, x_2^N, y^n) = \begin{cases} 
  H_1 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) \leq \lambda, \\
  H_2 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) > \lambda. 
\end{cases} \]
Asymptotic performance: For any \((P_1, P_2) \in \mathcal{P}^2(\mathcal{X})\),

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n^{\text{Gut}} | P_1, P_2) \geq \lambda,
\]

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_2(\phi_n^{\text{Gut}} | P_1, P_2) = F(P_1, P_2, \alpha, \lambda).
\]
Asymptotic performances of Gutman’s Test

- Asymptotic performance: For any \((P_1, P_2) \in \mathcal{P}^2(X)\),
  \[
  \liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n^{\text{Gut}} | P_1, P_2) \geq \lambda, \\
  \liminf_{n \to \infty} -\frac{1}{n} \log \beta_2(\phi_n^{\text{Gut}} | P_1, P_2) = F(P_1, P_2, \alpha, \lambda).
  \]

- \(F(P_1, P_2, \alpha, \lambda) > 0\) if \(\lambda < \text{GJS}(P_1, P_2, \alpha)\).
Asymptotic performance: For any \((P_1, P_2) \in \mathcal{P}^2(\mathcal{X})\),

\[
\liminf_{n \to \infty} \frac{1}{n} \log \beta_1(\phi_n^{\text{Gut}} | P_1, P_2) \geq \lambda,
\]

\[
\liminf_{n \to \infty} \frac{1}{n} \log \beta_2(\phi_n^{\text{Gut}} | P_1, P_2) = F(P_1, P_2, \alpha, \lambda).
\]

\(F(P_1, P_2, \alpha, \lambda) > 0\) if \(\lambda < \text{GJS}(P_1, P_2, \alpha)\).

*Universally* Exponentially consistent for \(P_1 \neq P_2\).
Asymptotic Performances of Gutman’s Test

Asymptotic performance: For any \((P_1, P_2) \in \mathcal{P}^2(\mathcal{X})\),

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n^{\text{Gut}} | P_1, P_2) \geq \lambda,
\]

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_2(\phi_n^{\text{Gut}} | P_1, P_2) = F(P_1, P_2, \alpha, \lambda).
\]

\(F(P_1, P_2, \alpha, \lambda) > 0\) if \(\lambda < \text{GJS}(P_1, P_2, \alpha)\).

Universally Exponentially consistent for \(P_1 \neq P_2\).

Analogous to Blahut’s result for BHT.
Asymptotic Optimality of Gutman’s Test

Asymptotic Optimality: Fix any sequence of tests \( \{\phi_n\}_{n=1}^{\infty} \) s.t.

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n|\tilde{P}_1, \tilde{P}_2) \geq \lambda, \ \forall \ \tilde{P}_1 \neq \tilde{P}_2
\]
Asymptotic Optimality: Fix any sequence of tests \( \{\phi_n\}_{n=1}^{\infty} \) s.t.

\[
\liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n|\tilde{P}_1, \tilde{P}_2) \geq \lambda, \ \forall \ \tilde{P}_1 \neq \tilde{P}_2
\]

then for any \((P_1, P_2)\),

\[
\beta_2(\phi_n|P_1, P_2) \geq \beta_2(\phi_n^{\text{Gut}}|P_1, P_2).
\]
Asymptotic Optimality of Gutman’s Test

- Asymptotic Optimality: Fix any sequence of tests \( \{ \phi_n \}_{n=1}^{\infty} \) s.t.
  \[
  \liminf_{n \to \infty} -\frac{1}{n} \log \beta_1(\phi_n|\tilde{P}_1, \tilde{P}_2) \geq \lambda, \ \forall \ \tilde{P}_1 \neq \tilde{P}_2
  \]
  then for any \((P_1, P_2)\),
  \[
  \beta_2(\phi_n|P_1, P_2) \geq \beta_2(\phi_n^{\text{Gut}}|P_1, P_2).
  \]

- Analogous to Neyman-Pearson lemma for BHT.
Motivation and Problem Formulation

- Motivation: finite sample length
Motivation and Problem Formulation

- Motivation: finite sample length → Non-asymptotic analysis
Motivation and Problem Formulation

- **Motivation:** finite sample length → Non-asymptotic analysis
- **Attempt One:** Chernoff-Stein setting

\[
E^*(n, \alpha, \varepsilon | P_1, P_2) := \sup \left\{ E \in \mathbb{R}_+ : \exists \phi_n \text{ s.t.} \right. \\
\beta_1(\phi_n | P_1, P_2) \leq \exp(-nE), \text{ and } \beta_2(\phi_n | P_1, P_2) \leq \varepsilon .
\]
Motivation and Problem Formulation

- Motivation: finite sample length → Non-asymptotic analysis
- Attempt One: Chernoff-Stein setting non-universal

\[ E^*(n, \alpha, \varepsilon|P_1, P_2) := \sup \left\{ E \in \mathbb{R}_+ : \exists \phi_n \text{ s.t.} \right. \]

\[ \beta_1(\phi_n|P_1, P_2) \leq \exp(-nE), \text{ and } \beta_2(\phi_n|P_1, P_2) \leq \varepsilon \right\} . \]
Motivation and Problem Formulation

- **Motivation:** finite sample length $\rightarrow$ Non-asymptotic analysis
- **Attempt One:** Chernoff-Stein setting non-universal

$$E^*(n, \alpha, \varepsilon| P_1, P_2) := \sup \left\{ E \in \mathbb{R}_+ : \exists \phi_n \text{ s.t.} \right.$$

$$\beta_1(\phi_n| P_1, P_2) \leq \exp(-nE), \text{ and } \beta_2(\phi_n| P_1, P_2) \leq \varepsilon \right\}.$$

- **Attempt Two:** Fully Universal setting

$$\lambda^*(n, \alpha, \varepsilon) := \sup \left\{ \lambda \in \mathbb{R}_+ : \exists \phi_n \text{ s.t. } \forall \tilde{P}_1 \neq \tilde{P}_2, \right.$$

$$\beta_1(\phi_n| \tilde{P}_1, \tilde{P}_2) \leq \exp(-n\lambda), \text{ and } \beta_2(\phi_n| \tilde{P}_1, \tilde{P}_2) \leq \varepsilon \right\}.$$
Motivation and Problem Formulation

- Motivation: finite sample length → Non-asymptotic analysis
- Attempt One: Chernoff-Stein setting non-universal

\[ E^*(n, \alpha, \varepsilon | P_1, P_2) := \sup \left\{ E \in \mathbb{R}_+ : \exists \phi_n \text{ s.t.} \right. \]
\[ \left. \beta_1(\phi_n | P_1, P_2) \leq \exp(-nE), \text{ and } \beta_2(\phi_n | P_1, P_2) \leq \varepsilon \right\}. \]

- Attempt Two: Fully Universal setting pessimistic

\[ \lambda^*(n, \alpha, \varepsilon) := \sup \left\{ \lambda \in \mathbb{R}_+ : \exists \phi_n \text{ s.t.} \forall \tilde{P}_1 \neq \tilde{P}_2, \right. \]
\[ \left. \beta_1(\phi_n | \tilde{P}_1, \tilde{P}_2) \leq \exp(-n\lambda), \text{ and } \beta_2(\phi_n | \tilde{P}_1, \tilde{P}_2) \leq \varepsilon \right\}. \]
Problem Formulation Continued

- **Partially Universal** setting: For any $\varepsilon \in (0, 1)$ and any $P_1 \neq P_2$

  $$\lambda^*(n, \alpha, \varepsilon|P_1, P_2) := \sup \left\{ \lambda \in \mathbb{R}^+_0 : \exists \phi_n \text{ s.t. } \beta_2(\phi_n|P_1, P_2) \leq \varepsilon, \text{ and } \right.$$

  $$\forall \tilde{P}_1 \neq \tilde{P}_2, \beta_1(\phi_n|\tilde{P}_1, \tilde{P}_2) \leq \exp(-n\lambda) \right\}.$$
Problem Formulation Continued

- **Partially Universal** setting: For any $\varepsilon \in (0, 1)$ and any $P_1 \neq P_2$

$$\lambda^*(n, \alpha, \varepsilon|P_1, P_2) := \sup \left\{ \lambda \in \mathbb{R}_+ : \exists \phi_n \text{ s.t. } \beta_2(\phi_n|P_1, P_2) \leq \varepsilon, \text{ and} \right.$$ \[
\forall \tilde{P}_1 \neq \tilde{P}_2, \beta_1(\phi_n|\tilde{P}_1, \tilde{P}_2) \leq \exp(-n\lambda) \right\}.

- Relationship with fully universal setting

$$\lambda^*(n, \alpha, \varepsilon) = \min_{P_1 \neq P_2} \lambda^*(n, \alpha, \varepsilon|P_1, P_2).$$
Problem Formulation Continued

- **Partially Universal** setting: For any \( \varepsilon \in (0, 1) \) and any \( P_1 \neq P_2 \)

\[
\lambda^*(n, \alpha, \varepsilon | P_1, P_2) := \sup \left\{ \lambda \in \mathbb{R}_+ : \exists \phi_n \text{ s.t. } \beta_2(\phi_n| P_1, P_2) \leq \varepsilon, \text{ and} \right. \\
\left. \forall \bar{P}_1 \neq \bar{P}_2, \beta_1(\phi_n| \bar{P}_1, \bar{P}_2) \leq \exp(-n \lambda) \right\}.
\]

- **Relationship with fully universal setting**

\[
\lambda^*(n, \alpha, \varepsilon) = \min_{P_1 \neq P_2} \lambda^*(n, \alpha, \varepsilon | P_1, P_2).
\]

- **Gutman’s asymptotic result** implies that

\[
\liminf_{n \to \infty} \lambda^*(n, \alpha, \varepsilon | P_1, P_2) \geq \text{GJS}(P_1, P_2, \alpha).
\]
Second-Order Asymptotics for Binary Classification

Given \((P_1, P_2)\), define information densities

\[
\nu_i(x|P_1, P_2, \alpha) := \log \left( \frac{(1 + \alpha)P_i(x)}{\alpha P_1(x) + P_2(x)} \right), \quad i \in [2].
\]
Second-Order Asymptotics for Binary Classification

- Given \((P_1, P_2)\), define information densities

\[ \nu_i(x|P_1, P_2, \alpha) := \log \frac{(1 + \alpha)P_i(x)}{\alpha P_1(x) + P_2(x)}, \quad i \in [2]. \]

- Dispersion function

\[ V(P_1, P_2, \alpha) = \alpha \text{Var}_{P_1}[\nu_1(X|P_1, P_2, \alpha)] + \text{Var}_{P_2}[\nu_2(X|P_1, P_2, \alpha)]. \]
Second-Order Asymptotics for Binary Classification

- Given $(P_1, P_2)$, define information densities

$$\nu_i(x|P_1, P_2, \alpha) := \log \frac{(1 + \alpha)P_i(x)}{\alpha P_1(x) + P_2(x)}, \quad i \in [2].$$

- Dispersion function

$$V(P_1, P_2, \alpha) = \alpha \text{Var}_{P_1} [\nu_1(X|P_1, P_2, \alpha)] + \text{Var}_{P_2} [\nu_2(X|P_1, P_2, \alpha)].$$

---

**Theorem 1**

For any $\varepsilon \in (0, 1)$, any $\alpha \in \mathbb{R}_+$ and any $(P_1, P_2) \in \mathcal{P}(\mathcal{X})^2$, we have

$$\lambda^*(n, \alpha, \varepsilon|P_1, P_2) = \text{GJS}(P_1, P_2, \alpha) + \sqrt{\frac{V(P_1, P_2, \alpha)}{n}} \Phi^{-1}(\varepsilon) + \Theta \left( \frac{\log n}{n} \right).$$
Remarks for Second-Order Asymptotics

\[
\begin{align*}
\lambda^*(n, \alpha, \varepsilon | P_1, P_2) &= GJS(P_1, P_2, \alpha) + \sqrt{\frac{V(P_1, P_2, \alpha)}{n}} \phi^{-1}(\varepsilon) + \Theta \left( \frac{\log n}{n} \right).
\end{align*}
\]

- Second-order optimality of Gutman’s test.
Remarks for Second-Order Asymptotics

\[ \lambda^*(n, \alpha, \varepsilon|P_1, P_2) = GJS(P_1, P_2, \alpha) + \sqrt{\frac{V(P_1, P_2, \alpha)}{n}} \phi^{-1}(\varepsilon) + \Theta \left( \frac{\log n}{n} \right). \]

- Second-order optimality of Gutman’s test.
- Influence of \( \alpha \): \( GJS(P_1, P_2, \alpha) \) is increasing in \( \alpha \);
Remarks for Second-Order Asymptotics

\[
\lambda^*(n, \alpha, \varepsilon | P_1, P_2) = GJS(P_1, P_2, \alpha) + \sqrt{\frac{V(P_1, P_2, \alpha)}{n}} \Phi^{-1}(\varepsilon) + \Theta \left( \frac{\log n}{n} \right). 
\]

- Second-order optimality of Gutman's test.
- Influence of \( \alpha \): \( GJS(P_1, P_2, \alpha) \) is increasing in \( \alpha \);

\[
\lim_{\alpha \to 0} GJS(P_1, P_2, \alpha) = 0, 
\]
Remarks for Second-Order Asymptotics

\[ \lambda^*(n, \alpha, \varepsilon | P_1, P_2) = \text{GJS}(P_1, P_2, \alpha) + \sqrt{\frac{\text{V}(P_1, P_2, \alpha)}{n}} \Phi^{-1}(\varepsilon) + \Theta \left(\frac{\log n}{n}\right). \]

- Second-order optimality of Gutman’s test.
- Influence of \( \alpha \): \( \text{GJS}(P_1, P_2, \alpha) \) is increasing in \( \alpha \);

\[
\lim_{\alpha \to 0} \text{GJS}(P_1, P_2, \alpha) = 0, \quad \lim_{\alpha \to \infty} \text{GJS}(P_1, P_2, \alpha) = D(P_2 \parallel P_1).
\]
Remarks for Second-Order Asymptotics

\[ \lambda^*(n, \alpha, \varepsilon \mid P_1, P_2) \]

\[ = \text{GJS}(P_1, P_2, \alpha) + \sqrt{\frac{\text{V}(P_1, P_2, \alpha)}{n}} \phi^{-1}(\varepsilon) + \Theta\left(\frac{\log n}{n}\right). \]

- Second-order optimality of Gutman’s test.
- Influence of \( \alpha \): \( \text{GJS}(P_1, P_2, \alpha) \) is increasing in \( \alpha \);
  \[ \lim_{\alpha \to 0} \text{GJS}(P_1, P_2, \alpha) = 0, \quad \lim_{\alpha \to \infty} \text{GJS}(P_1, P_2, \alpha) = D(P_2 \parallel P_1). \]
- Recover second-order asymptotics for BHT by Strassen
  \[ \lim_{\alpha \to \infty} \text{V}(P_1, P_2, \alpha) = \text{Var}_{P_2}[\log(P_2(X)/P_1(X))] = \text{V}(P_2 \parallel P_1). \]
A special case: two sample homogeneity testing problem
A special case: two sample homogeneity testing problem

Input: Test sequence $Y^n$ and training sequence $X_1^N$. 
Remarks Continued

- A special case: two sample homogeneity testing problem
  - Input: Test sequence $Y^n$ and training sequence $X_1^N$.
  - Task: design a test $\phi_n$ to classify two Hypotheses:
    - $H_1$: $X_1^N$ and $Y^n$ are generated according to the same distribution;
    - $H_2$: $X_1^N$ and $Y^n$ are generated according to different distributions.

Asymmetry in Gutman's test

$\phi_{Gut}(x_{N1}; x_{N2}; y_n) = \begin{cases} 
H_1 & \text{if } \text{GJS}^T(x_{N1}; x_{N2}; y_n) \\
H_2 & \text{if } \text{GJS}^T(x_{N1}; x_{N2}; y_n) > \end{cases}$

Classification with rejection.

Zhou-Tan-Motani (NUS)  
Statistical Classification  
Group Meeting 13 / 23
A special case: two sample homogeneity testing problem

- Input: Test sequence $Y^n$ and training sequence $X_1^N$.
- Task: design a test $\phi_n$ to classify two Hypotheses:
  - $H_1$: $X_1^N$ and $Y^n$ are generated according to the same distribution;
  - $H_2$: $X_1^N$ and $Y^n$ are generated according to different distributions.

Asymmetry in Gutman’s test

$$\phi_n^{\text{Gut}}(x_1^N, x_2^N, y^n) = \begin{cases} 
H_1 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) \leq \lambda, \\
H_2 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) > \lambda.
\end{cases}$$
Remarks Continued

- A special case: two sample homogeneity testing problem
  - Input: Test sequence $Y^n$ and training sequence $X_1^N$.
  - Task: design a test $\phi_n$ to classify two Hypotheses:
    - $H_1$: $X_1^N$ and $Y^n$ are generated according to the same distribution;
    - $H_2$: $X_1^N$ and $Y^n$ are generated according to different distributions.

- Asymmetry in Gutman’s test

$$
\phi_n^{\text{Gut}}(x_1^N, x_2^N, y^n) = \begin{cases} 
H_1 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) \leq \lambda, \\
H_2 & \text{if } \text{GJS}(\hat{T}_{x_1^N}, \hat{T}_{y^n}, \alpha) > \lambda.
\end{cases}
$$

- Classification with rejection.
Binary source: \( \mathcal{X} = \{0, 1\} \).
\( \alpha = 2. \)

Target distributions
\( P_1 = \text{Bern}(0.2), \)
\( P_2 = \text{Bern}(0.4). \)

Target type-II error probability
\( \varepsilon = 0.2. \)

Threshold in Gutman’s test
\[
\hat{\lambda} := \text{GJS}(P_1, P_2, \alpha) \\
+ \sqrt{\frac{V(P_1, P_2, \alpha)}{n}} \Phi^{-1}(\varepsilon).
\]
Numerical Simulation: Type-II Error Probability

- Binary source: $\mathcal{X} = \{0, 1\}$.
- $\alpha = 2$.
- Target distributions
  $P_1 = \text{Bern}(0.2)$,
  $P_2 = \text{Bern}(0.4)$.
- Target type-II error probability $\varepsilon = 0.2$.
- Threshold in Gutman’s test
  \[
  \hat{\lambda} := \text{GJS}(P_1, P_2, \alpha) + \sqrt{V(P_1, P_2, \alpha) \frac{n}{\alpha}} \Phi^{-1}(\varepsilon).
  \]
Numerical Simulation: Type-I Error Probability

- Target distributions $P_1 = \text{Bern}(0.2)$ and $P_2 = \text{Bern}(0.228)$
Numerical Simulation: Type-I Error Probability

- Target distributions $P_1 = \text{Bern}(0.2)$ and $P_2 = \text{Bern}(0.228)$
Classification of Multiple Hypotheses with Rejection

Training Seq. $X_1^N \ X_2^N \ \cdots \ \ X_M^N$

Test Seq. $Y^n$

Classifier

Hypotheses: $\{H_r, H_1, \ldots, H_M\}$
Classification of Multiple Hypotheses with Rejection

- Training Seq. $X_1^N, X_2^N, \ldots, X_M^N$
- Test Seq. $Y^n$
- Classifier
- Output: $\{H_r, H_1, \ldots, H_M\}$

- $X_i^N \sim P_i^N$ for a tuple of unknown distributions $\{P_1, \ldots, P_M\}$.
Classification of Multiple Hypotheses with Rejection

- $X_i^N \sim P_i^N$ for a tuple of unknown distributions $\{P_1, \ldots, P_M\}$.
- Task: design a test $\psi_n$ to classify the following hypotheses
  - $H_j$, $j \in [M]$: $Y^n$ and $X_j^N$ are generated according to the same distribution;

![Diagram](image_url)
Classification of Multiple Hypotheses with Rejection

- $X_i^N \sim P_i^N$ for a tuple of unknown distributions $\{P_1, \ldots, P_M\}$.
- Task: design a test $\psi_n$ to classify the following hypotheses
  - $H_j$, $j \in [M]$: $Y^n$ and $X_j^N$ are generated according to the same distribution;
  - $H_r$: none of $(X_1^N, \ldots, X_M^N)$ is generated according to the same distribution as $Y^n$. 

\[ \begin{align*}
\text{Training Seq.} & \quad X_1^N \quad X_2^N \quad \cdots \quad X_M^N \\
\text{Test Seq.} & \quad Y^n \quad \rightarrow \quad \text{Classifier} \quad \{H_r, H_1, \ldots, H_M\}
\end{align*} \]
Classification of Multiple Hypotheses with Rejection

**Task:** design a test \( \psi_n \) to classify the following hypotheses

- \( H_j, j \in [M] \): \( Y^n \) and \( X_j^N \) are generated according to the same distribution;
- \( H_r \): none of \( (X_1^N, \ldots, X_M^N) \) is generated according to the same distribution as \( Y^n \).

**Notation:** \( X^N := (X_1^N, \ldots, X_M^N) \), \( P := (P_1, \ldots, P_M) \) and \( x^N \).

\[ X_i^N \sim P_i^N \] for a tuple of unknown distributions \( \{P_1, \ldots, P_M\} \).
Definitions for M-ary Classification with Rejection

Given any $P$ and any test $\psi_n$
Definitions for M-ary Classification with Rejection

- Given any $\mathbf{P}$ and any test $\psi_n$
  - Error probabilities:
    \[
    \beta_j(\psi_n|\mathbf{P}) := \Pr \{ \psi_n(\mathbf{X}^N, \mathbf{Y}^n) \notin \{H_j, H_r\}|H_j \}.
    \]
Definitions for M-ary Classification with Rejection

- Given any $\mathbf{P}$ and any test $\psi_n$
  - Error probabilities:
    \[
    \beta_j(\psi_n|\mathbf{P}) := \Pr \{ \psi_n(\mathbf{X}^N, Y^n) \notin \{H_j, H_r\}|H_j\}.
    \]
  - Rejection probabilities
    \[
    \zeta_j(\psi_n|\mathbf{P}) := \Pr \{ \psi_n(\mathbf{X}^N, Y^n) = H_r|H_j\}.
    \]
Definitions for M-ary Classification with Rejection

- Given any $P$ and any test $\psi_n$
  - Error probabilities:
    \[ \beta_j(\psi_n|P) := \Pr \{ \psi_n(X^N, Y^n) \notin \{H_j, H_r\}|H_j \} . \]
  - Rejection probabilities
    \[ \zeta_j(\psi_n|P) := \Pr \{ \psi_n(X^N, Y^n) = H_r|H_j \} . \]

- Fundamental limit: for any $\epsilon^M = (\epsilon_1, \ldots, \epsilon_M) \in (0, 1)^M$ and any $P$, 
Definitions for M-ary Classification with Rejection

- Given any $\mathbf{P}$ and any test $\psi_n$
  - Error probabilities:
    \[ \beta_j(\psi_n|\mathbf{P}) := \Pr \{ \psi_n(X^n, Y^n) \notin \{H_j, H_r\} | H_j \}. \]
  - Rejection probabilities
    \[ \zeta_j(\psi_n|\mathbf{P}) := \Pr \{ \psi_n(X^n, Y^n) = H_r | H_j \}. \]

- Fundamental limit: for any $\varepsilon^M = (\varepsilon_1, \ldots, \varepsilon_M) \in (0, 1)^M$ and any $\mathbf{P}$,
  \[ \lambda^*(n, \alpha, \varepsilon^M|\mathbf{P}) := \sup \left\{ \lambda \in \mathbb{R}_+ : \exists \psi_n \text{ s.t. } \forall j \in [M], \zeta_j(\psi_n|\mathbf{P}) \leq \varepsilon_j, \beta_j(\psi_n|\tilde{\mathbf{P}}) \leq \exp(-n\lambda), \forall \tilde{\mathbf{P}} \right\}. \]
Preliminaries for M-ary Classification

- Gutman’s asymptotic result for M-ary classification implies that

\[
\lim_{n \to \infty} \inf \lambda^* (n, \alpha, \varepsilon^M | \mathbf{P}) \geq \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]
Gutman’s asymptotic result for M-ary classification implies that

$$\liminf_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M | P) \geq \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha).$$

Given any $P$ and any $j \in [M]$,

$$\theta_j(P, \alpha) := \min_{i \in [M] : i \neq j} \text{GJS}(P_i, P_j, \alpha).$$
Gutman’s asymptotic result for M-ary classification implies that

\[
\liminf_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M|\mathbf{P}) \geq \min_{(i,j) \in [M]^2: i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]

Given any \( \mathbf{P} \) and any \( j \in [M] \),

\[
\theta_j(\mathbf{P}, \alpha) := \min_{i \in [M]: i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]

Assumption on \( \mathbf{P} \): uniqueness of minimizer for \( \theta_j(\mathbf{P}, \alpha) \)
Preliminaries for M-ary Classification

- Gutman’s asymptotic result for M-ary classification implies that

\[
\liminf_{n \to \infty} \lambda^*(n, \alpha, \epsilon^M|\mathbf{P}) \geq \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]

- Given any \( \mathbf{P} \) and any \( j \in [M] \),

\[
\theta_j(\mathbf{P}, \alpha) := \min_{i \in [M] : i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]

- Assumption on \( \mathbf{P} \): uniqueness of minimizer for \( \theta_j(\mathbf{P}, \alpha) \)

\[
i^*(j|\mathbf{P}, \alpha) := \arg \min_{i \in [M] : i \neq j} \text{GJS}(P_i, P_j, \alpha).
\]
Gutman’s asymptotic result for M-ary classification implies that

$$\liminf_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M|\mathbf{P}) \geq \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha).$$

Given any $\mathbf{P}$ and any $j \in [M]$,

$$\theta_j(\mathbf{P}, \alpha) := \min_{i \in [M] : i \neq j} \text{GJS}(P_i, P_j, \alpha).$$

Assumption on $\mathbf{P}$: uniqueness of minimizer for $\theta_j(\mathbf{P}, \alpha)$

$$i^*(j) := \arg \min_{i \in [M] : i \neq j} \text{GJS}(P_i, P_j, \alpha).$$
Given any $P$ and positive $\alpha$,

$$J_1(P, \alpha) := \arg \min_{j \in [M]} GJS(P_{i^*(j)}, P_j, \alpha),$$

where $GJS$ denotes the Generalized Jackknife Score.
Second-Order Asymptotics for M-ary Classification

Given any $P$ and positive $\alpha$,

$$J_1(P, \alpha) := \arg \min_{j \in [M]} GJS(P_{i*}(j), P_j, \alpha),$$

$$J_2(P, \alpha) := \arg \min_{j \in J_1(P, \alpha)} \sqrt{V(P_{i*}(j), P_j, \alpha)} \Phi^{-1}(\varepsilon_j).$$
Second-Order Asymptotics for M-ary Classification

Given any $\mathbf{P}$ and positive $\alpha$,

$$
\mathcal{J}_1(\mathbf{P}, \alpha) := \arg \min_{j \in [M]} \text{GJS}(P_{i^*(j)}, P_j, \alpha),
$$

$$
\mathcal{J}_2(\mathbf{P}, \alpha) := \arg \min_{j \in \mathcal{J}_1(\mathbf{P}, \alpha)} \sqrt{\text{V}(P_{i^*(j)}, P_j, \alpha)} \Phi^{-1}(\varepsilon_j).
$$

---

**Theorem 2**

For any $\alpha \in \mathbb{R}_+$, any $\varepsilon^M \in (0, 1)^M$ and any $\mathbf{P} \in \mathcal{P}(\mathcal{X})^M$ satisfying that the minimizer for $\theta_j(\mathbf{P}, \alpha)$ is unique for each $j \in [M]$, we have

$$
\lambda^*(n, \alpha, \varepsilon^M | \mathbf{P}) = \text{GJS}(P_{i^*(j)}, P_j, \alpha) + \sqrt{\frac{\text{V}(P_{i^*(j)}, P_j, \alpha)}{n}} \Phi^{-1}(\varepsilon_j) + O\left(\frac{\log n}{n}\right)
$$

for any $j \in \mathcal{J}_2(\mathbf{P}, \alpha)$. 
Remarks for M-ary Classification

- First-order asymptotics with strong converse: for any $\varepsilon^M \in (0, 1)^M$,
  \[
  \lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M|P) = \text{GJS}(P_{i^*}(j), P_j, \alpha)
  \]
  for any $j \in \mathcal{J}_1(P, \alpha)$. 
Remarks for M-ary Classification

- First-order asymptotics with **strong converse**: for any $\varepsilon^M \in (0, 1)^M$,

$$\lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M | P) = \text{GJS}(P_{i^*}(j), P_j, \alpha) = \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha)$$

for any $j \in J_1(P, \alpha)$. 

Zhou-Tan-Motani (NUS)  Statistical Classification  Group Meeting  20 / 23
Remarks for M-ary Classification

- First-order asymptotics with strong converse: for any \( \varepsilon^M \in (0, 1)^M \),
  \[
  \lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M | P) = \text{GJS}(P_{i^*}(j), P_j, \alpha) = \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha)
  \]
  for any \( j \in J_1(P, \alpha) \).

- Second-order optimality of the test by Unnikrishnan (2015)
Remarks for M-ary Classification

- First-order asymptotics with strong converse: for any $\varepsilon^M \in (0, 1)^M$,

$$\lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M | \mathbf{P}) = \text{GJS}(P_{i^*}(j), P_j, \alpha) = \min_{(i,j) \in [M]^2 : i \neq j} \text{GJS}(P_i, P_j, \alpha)$$

for any $j \in J_1(\mathbf{P}, \alpha)$.

- Second-order optimality of the test by Unnikrishnan (2015)
  - Preliminaries

$$i^*(x^N, y^n) := \arg \min_{i \in [M]} \text{GJS}(\hat{T}_{x^N_i}, \hat{T}_{y^n}, \alpha),$$
Remarks for M-ary Classification

- First-order asymptotics with **strong converse**: for any $\varepsilon^M \in (0, 1)^M$,
  \[
  \lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M|\mathbf{P}) = \text{GJS}(P_{i^*}(j), P_j, \alpha) = \min_{(i,j) \in [M]^2:i \neq j} \text{GJS}(P_i, P_j, \alpha)
  \]
  for any $j \in \mathcal{J}_1(\mathbf{P}, \alpha)$.
- Second-order optimality of the test by Unnikrishnan (2015)
  - Preliminaries
    \[
    i^*(\mathbf{x}^N, y^n) := \arg \min_{i \in [M]} \text{GJS}(\hat{T}_{x_i}^N, \hat{T}_{y}^n, \alpha),
    \]
    \[
    \tilde{h}(\mathbf{x}^N, y^n) := \min_{i \in [M]:i \neq i^*(\mathbf{x}^N, y^n)} \text{GJS}(\hat{T}_{x_i}^N, \hat{T}_{y}^n, \alpha).
    \]
Remarks for M-ary Classification

- First-order asymptotics with strong converse: for any $\varepsilon^M \in (0, 1)^M$,
  \[
  \lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M | P) = \text{GJS}(P_{i^*(j)}, P_j, \alpha) = \min_{(i,j) \in [M]^2: i \neq j} \text{GJS}(P_i, P_j, \alpha)
  \]
  for any $j \in \mathcal{J}_1(P, \alpha)$.

- Second-order optimality of the test by Unnikrishnan (2015)
  - Preliminaries
    \[
    i^*(x^N, y^n) := \arg \min_{i \in [M]} \text{GJS}(\hat{T}_{x_{i^N}}, \hat{T}_{y^n}, \alpha),
    \]
    \[
    \tilde{h}(x^N, y^n) := \min_{i \in [M]: i \neq i^*(x^N, y^n)} \text{GJS}(\hat{T}_{x_i}, \hat{T}_{y^n}, \alpha).
    \]
  - Unnikrishnan’s test
    \[
    \psi_n^{\text{Unn}}(x^N, y^n) = \begin{cases} 
    H_j & \text{if } i^*(x^N, y^n) = j, \tilde{h}(x^N, y^n) \geq \tilde{\lambda} \\
    H_r & \text{if } \tilde{h}(x^N, y^n) < \tilde{\lambda}.
    \end{cases}
    \]
Remarks for M-ary Classification

- First-order asymptotics with strong converse: for any $\varepsilon^M \in (0, 1)^M$,

$$\lim_{n \to \infty} \lambda^*(n, \alpha, \varepsilon^M|\mathbf{P}) = \operatorname{GJS}(P_{i^*(j)}, P_j, \alpha) = \min_{(i,j) \in [M]^2:i \neq j} \operatorname{GJS}(P_i, P_j, \alpha)$$

for any $j \in \mathcal{J}_1(\mathbf{P}, \alpha)$.

- Second-order optimality of the test by Unnikrishnan (2015)
  
  Preliminaries

$$i^*(\mathbf{x}^N, y^n) := \arg\min_{i \in [M]} \operatorname{GJS}(\hat{T}_{x_i^N}, \hat{T}_{y^n}, \alpha),$$

$$\tilde{h}(\mathbf{x}^N, y^n) := \min_{i \in [M]:i \neq i^*(\mathbf{x}^N, y^n)} \operatorname{GJS}(\hat{T}_{x_i^N}, \hat{T}_{y^n}, \alpha).$$

- Unnikrishnan’s test

$$\psi_n^{\text{Unn}}(\mathbf{x}^N, y^n) = \begin{cases} H_j & \text{if } i^*(\mathbf{x}^N, y^n) = j, \tilde{h}(\mathbf{x}^N, y^n) \geq \tilde{\lambda} \\ H_r & \text{if } \tilde{h}(\mathbf{x}^N, y^n) < \tilde{\lambda}. \end{cases}$$
Difficulty in achievability part: identifying the index of $\tilde{h}(x^N, y^n)$
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^N, y^n)$
  - Why?
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^n, y^n)$
  - Why?
    The index of $\tilde{h}(x^n, y^n)$ changes for different $x^n, y^n$. 

Converse proof
Type-based test is optimal; Unnikrishnan's test is one of the optimal type-based tests.
Remarks Continued

- Difficulty in achievability part: identifying the index of \( \tilde{h}(x^N, y^n) \)
  - Why?
    - The index of \( \tilde{h}(x^N, y^n) \) changes for **different** \( x^N, y^n \).
  - How?
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^n, y^n)$
  - Why?
    The index of $\tilde{h}(x^n, y^n)$ changes for different $x^n, y^n$.
  - How?
    Under the assumption that the minimizer of $\theta_j(P, \alpha)$ is unique for each $j \in [M]$, we show that
    $$\lim_{n \to \infty} \Pr \left\{ \text{index of } \tilde{h}(X^n, Y^n) = i^*(j) \mid H_j \right\} = 1.$$
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^N, y^n)$
  - Why?
    The index of $\tilde{h}(x^N, y^n)$ changes for different $x^N, y^n$.
  - How?
    Under the assumption that the minimizer of $\theta_j(P, \alpha)$ is unique for each $j \in [M]$, we show that
    \[
    \lim_{n \to \infty} \Pr \left\{ \text{index of } \tilde{h}(X^N, Y^n) = i^*(j) | H_j \right\} = 1.
    \]

- Converse proof
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^N, y^n)$
  - Why?
    The index of $\tilde{h}(x^N, y^n)$ changes for different $x^N, y^n$.
  - How?
    Under the assumption that the minimizer of $\theta_j(P, \alpha)$ is unique for each $j \in [M]$, we show that
    \[
    \lim_{n \to \infty} \Pr \left\{ \text{index of } \tilde{h}(X^N, Y^n) = i^*(j) | H_j \right\} = 1.
    \]

- Converse proof
  - Type-based test is optimal;
Remarks Continued

- Difficulty in achievability part: identifying the index of $\tilde{h}(x^N, y^n)$
  - Why?
    The index of $\tilde{h}(x^N, y^n)$ changes for different $x^N, y^n$.
  - How?
    Under the assumption that the minimizer of $\theta_j(P, \alpha)$ is unique for each $j \in [M]$, we show that
    \[
    \lim_{n \to \infty} \Pr \left\{ \text{index of } \tilde{h}(X^N, Y^n) = i^*(j) \mid H_j \right\} = 1.
    \]

- Converse proof
  - Type-based test is optimal;
  - Unnikrishnan’s test is one of the optimal type-based tests.
For any $\varepsilon \in (0, 1)$ and $\mathbf{P}$,

$$\Lambda(n, \alpha, \varepsilon | \mathbf{P}) := \left\{ \lambda^M \in \mathbb{R}_+^M : \exists \psi_n \text{ s.t. } \sum_{j \in [M]} \zeta_j(\psi_n | \mathbf{P}) \leq \varepsilon \right\}$$

and $\forall j \in [M], \beta_j(\psi_n | \tilde{\mathbf{P}}) \leq \exp(-n\lambda_j), \forall \tilde{\mathbf{P}}$. 

Inhomogeneous Constraints on Error Probabilities

Zhou-Tan-Motani (NUS) Statistical Classification Group Meeting 22 / 23
Inhomogeneous Constraints on Error Probabilities

- For any $\varepsilon \in (0, 1)$ and $\mathbf{P}$,

$$\Lambda(n, \alpha, \varepsilon | \mathbf{P}) := \left\{ \lambda^M \in \mathbb{R}_+^M : \exists \psi_n \text{ s.t. } \sum_{j \in [M]} \zeta_j(\psi_n | \mathbf{P}) \leq \varepsilon \right\}$$

and $\forall j \in [M], \beta_j(\psi_n | \tilde{\mathbf{P}}) \leq \exp(-n \lambda_j), \forall \tilde{\mathbf{P}}$.

- In general, very hard for $M \geq 3$. 

Zhou-Tan-Motani (NUS)
Inhomogeneous Constraints on Error Probabilities

- For any $\varepsilon \in (0, 1)$ and $\mathbf{P}$,

$$\Lambda(n, \alpha, \varepsilon|\mathbf{P}) := \left\{ \lambda^M \in \mathbb{R}_+^M : \exists \psi_n \text{ s.t. } \sum_{j \in [M]} \zeta_j(\psi_n|\mathbf{P}) \leq \varepsilon \right\},$$

and $\forall j \in [M]$, $\beta_j(\psi_n|\tilde{\mathbf{P}}) \leq \exp(-n\lambda_j)$, $\forall \tilde{\mathbf{P}}$.

- In general, very hard for $M \geq 3$.
- Solvable when $M = 2$
Inhomogeneous Constraints on Error Probabilities

- For any $\varepsilon \in (0, 1)$ and $\mathbf{P}$,

$$\Lambda(n, \alpha, \varepsilon|\mathbf{P}) := \left\{ \lambda^M \in \mathbb{R}_+^M : \exists \psi_n \text{ s.t. } \sum_{j \in [M]} \zeta_j(\psi_n|\mathbf{P}) \leq \varepsilon \right\}$$

and $\forall j \in [M]$, $\beta_j(\psi_n|\tilde{\mathbf{P}}) \leq \exp(-n\lambda_j)$, $\forall \tilde{\mathbf{P}}$.

- In general, very hard for $M \geq 3$.
- Solvable when $M = 2$ using the following Gutman's test

$$\psi^G_n(x^N_1, x^N_2, y^n) := \begin{cases} 
H_1 & \text{if } \text{GJS}(\hat{T}_{x^N_2}, \hat{T}_{y^n}, \alpha) - \tilde{\lambda}_2 \geq 0, \\
H_2 & \text{if } \text{GJS}(\hat{T}_{x^N_1}, \hat{T}_{y^n}, \alpha) - \tilde{\lambda}_1 \geq 0 \\
& \text{and } \text{GJS}(\hat{T}_{x^N_2}, \hat{T}_{y^n}, \alpha) - \tilde{\lambda}_2 < 0, \\
H_i & \text{if } \text{GJS}(\hat{T}_{x^N_i}, \hat{T}_{y^n}, \alpha) - \tilde{\lambda}_i < 0, \; i \in [2].
\end{cases}$$
Summary

- Binary classification
Summary

- Binary classification
  - Non-asymptotic fundamental limit
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test

Future works

- Bayesian setting of binary classification
- Statistical version of other hypothesis testing problems, e.g. distributed detection

Full text available at arXiv 1806.00739
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance

Full text available at arXiv 1806.00739

Future works

Bayesian setting of binary classification

Statistical version of other hypothesis testing problems, e.g. distributed detection

Zhou-Tan-Motani (NUS)

Statistical Classification

Group Meeting
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance
- Generalization to classification of multiple hypotheses with rejection

Future works

- Bayesian setting of binary classification
- Statistical version of other hypothesis testing problems, e.g. distributed detection
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance
- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test

Future works
- Bayesian setting of binary classification
- Statistical version of other hypothesis testing problems, e.g. distributed detection

Full text available at arXiv 1806.00739
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance

- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test
  - Optimality of Generalization of Gutman’s test in binary classification with rejection for an Inhomogeneous setting

Full text available at arXiv 1806.00739

Future works

- Bayesian setting of binary classification
- Statistical version of other hypothesis testing problems, e.g. distributed detection

Zhou-Tan-Motani (NUS)
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance
- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test
  - Optimality of Generalization of Gutman’s test in binary classification with rejection for an Inhomogeneous setting
- Full text available at arXiv 1806.00739
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance

- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test
  - Optimality of Generalization of Gutman’s test in binary classification with rejection for an Inhomogeneous setting

- Full text available at arXiv 1806.00739

- Future works
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance

- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test
  - Optimality of Generalization of Gutman’s test in binary classification with rejection for an Inhomogeneous setting

- Full text available at arXiv 1806.00739

- Future works
  - Bayesian setting of binary classification
Summary

- Binary classification
  - Non-asymptotic fundamental limit
  - Second-order asymptotics and optimality of Gutman’s test
  - Approximation to finite blocklength performance

- Generalization to classification of multiple hypotheses with rejection
  - Second-order optimality of Unnikrishnan’s test
  - Optimality of Generalization of Gutman’s test in binary classification with rejection for an Inhomogeneous setting

- Full text available at arXiv 1806.00739

- Future works
  - Bayesian setting of binary classification
  - Statistical version of other hypothesis testing problems, e.g. distributed detection