The Sample Complexity of Simple Binary Hypothesis Testing

Conference on Learning Theory, 2024

Ankit Pensia

Po-Ling Loh

Varun Jog

Introducing simple binary hypothesis testing

Let p and q be two known distributions on $\mathcal{X}$.

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$ such that, w.h.p., $\widehat{\theta} = \theta$

- A foundational problem in statistics
  - Well-understood in two extreme regimes

Introducing simple binary hypothesis testing

Let p and q be two known distributions on $\mathcal{X}$.

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$ such that, w.h.p., $\widehat{\theta} = \theta$

- A foundational problem in statistics
  - Well-understood in two extreme regimes
- (Regime I) $n = 1$: Single-sample regime

Introducing simple binary hypothesis testing

Let p and q be two known distributions on $\mathcal{X}$.

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$ such that, w.h.p., $\widehat{\theta} = \theta$

- A foundational problem in statistics
  - Well-understood in two extreme regimes
- (Regime I) $n = 1$: Single-sample regime
- (Regime II) $n \to \infty$: Asymptotic regime

Introducing simple binary hypothesis testing

Let p and q be two known distributions on $\mathcal{X}$.

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$ such that, w.h.p., $\widehat{\theta} = \theta$

- A foundational problem in statistics
  - Well-understood in two extreme regimes
- (Regime I) $n = 1$: Single-sample regime
- (Regime II) $n \to \infty$: Asymptotic regime

Can we develop a **non-asymptotic** understanding?

Introducing Bayesian formulation

Let p, q be two known distributions.

Input: $\triangleright$ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output:

Introducing Bayesian formulation

Let p, q be two known distributions. Let the prior $\alpha \in (0, 0.5]$ and failure probability $\delta \in (0, \alpha/4]$

Input: $\triangleright$ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output:

Introducing Bayesian formulation

Let $\boldsymbol{p}$, $\boldsymbol{q}$ be two known distributions. Let the prior $\boldsymbol{\alpha} \in (0, 0.5]$ and failure probability $\boldsymbol{\delta} \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = \boldsymbol{p}) = \boldsymbol{\alpha}$

Input: ▷ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output:

Introducing Bayesian formulation

Let p, q be two known distributions. Let the prior $\alpha \in (0, 0.5]$ and failure probability $\delta \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = p) = \alpha$

▷ Conditioned on $\Theta = \theta$,

Input: ▷ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output:

Introducing Bayesian formulation

Let p, q be two known distributions. Let the prior $\alpha \in (0, 0.5]$ and failure probability $\delta \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = p) = \alpha$

▷ Conditioned on $\Theta = \theta$,

Input: ▷ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output: $\widehat{\theta}(X_{1:n})$ such that $\mathbb{P}\left(\widehat{\theta}(X_{1:n}) \neq \Theta\right) \leq \delta$

Introducing Bayesian formulation

Let p, q be two known distributions. Let the prior $\alpha \in (0, 0.5]$ and failure probability $\delta \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = p) = \alpha$

▷ Conditioned on $\Theta = \theta$,

Input: ▷ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output: $\widehat{\theta}(X_{1:n})$ such that $\mathbb{P}\left(\widehat{\theta}(X_{1:n}) \neq \Theta\right) \leq \delta$

- Sample complexity: necessary & sufficient samples

Introducing Bayesian formulation

Let $\boldsymbol{p}$, $\boldsymbol{q}$ be two known distributions. Let the prior $\boldsymbol{\alpha} \in (0, 0.5]$ and failure probability $\boldsymbol{\delta} \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = \boldsymbol{p}) = \boldsymbol{\alpha}$

▷ Conditioned on $\Theta = \theta$,

Input: ▷ $X_1, \ldots, X_n$ are sampled i.i.d. from θ

Output: $\widehat{\theta}(X_{1:n})$ such that $\mathbb{P}\left(\widehat{\theta}(X_{1:n}) \neq \Theta\right) \leq \boldsymbol{\delta}$

- Sample complexity: necessary & sufficient samples $\boldsymbol{n}^\star(p, q, \alpha, \delta)$

Introducing Bayesian formulation

Let p, q be two known distributions. Let the prior $\alpha \in (0, 0.5]$ and failure probability $\delta \in (0, \alpha/4]$

▷ Generate $\Theta \in \{p, q\}$ randomly with $\mathbb{P}(\Theta = p) = \alpha$

▷ Conditioned on $\Theta = \theta$,

Input: ▷ $X_1, \dots, X_n$ are sampled i.i.d. from θ

Output: $\widehat{\theta}(X_{1:n})$ such that $\mathbb{P}\left(\widehat{\theta}(X_{1:n}) \neq \Theta\right) \leq \delta$

- Sample complexity: necessary & sufficient samples $n^\star(p, q, \alpha, \delta)$

Our goal: characterizing sample complexity

Existing results

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics: well-understood

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics: well-understood
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n\cdot \mathrm{hel}^2(p,q))}$

Note that the Bayesian error exponent does not depend on the actual value of π_1 and π_2, as long as they are nonzero. Essentially, the effect of the prior is washed out for large sample sizes. The optimum decision rule

Cover and Thomas, Elements of Information Theory

Existing results

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

- Asymptotics: well-understood
  - failure probability $\overset{n\to\infty}{\to} e^{-\Theta(n\cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics: well-understood
  - failure probability $\overset{n \to \infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- **Non-asymptotic**

Existing results

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- Non-asymptotic
  - **Prediction** is true for the uniform prior ☺

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- Non-asymptotic
  - Prediction is true for the uniform prior ☺
  - More generally, $\frac{\log(\alpha/\delta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- Non-asymptotic
  - Prediction is true for the uniform prior ☺
  - More generally, $\frac{\log(\alpha/\delta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Exists cases where inequalities are sharp

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Existing results

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - **Predicts** $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- Non-asymptotic
  - Prediction is true for the uniform prior ☺
  - More generally, $\frac{\log(\alpha/\delta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Exists cases where inequalities are sharp
    - Hence, prediction is incorrect in general ☹

Existing results

Candidate distributions: p, q

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

- Asymptotics
  - failure probability $\overset{n\to\infty}{\longrightarrow} e^{-\Theta(n \cdot \mathrm{hel}^2(p,q))}$
  - Predicts $n^\star(p, q, \alpha, \delta) \overset{??}{\approx} \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Independent of prior, symmetric in p and q
- Non-asymptotic
  - Prediction is true for the uniform prior ☺
  - More generally, $\frac{\log(\alpha/\delta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)}$
    - Exists cases where inequalities are sharp
    - Hence, prediction is incorrect in general ☹

Can we tightly characterize the sample complexity?

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p, q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p,q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

$$n^\star(p, q, \alpha, \delta) \asymp \begin{cases} \\ \\ \\ \end{cases}$$

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p, q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

$$n^\star(p, q, \alpha, \delta) \asymp \begin{cases} \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta; X_1)}, & \delta \in [\frac{\alpha}{100}, \frac{\alpha}{4}) \end{cases}$$

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p, q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

$$n^\star(p,q,\alpha,\delta) \asymp \begin{cases} \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta;X_1)}, & \delta \in [\frac{\alpha}{100}, \frac{\alpha}{4}) \\ \log(\alpha/\delta) \cdot n^\star(p,q,\alpha',\alpha'/8) & \delta \in (\alpha^2, \frac{\alpha}{100}] \end{cases}$$

$$\alpha' := \alpha^{\frac{1}{\log(\alpha/\delta)}}$$

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p, q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

$$n^\star(p, q, \alpha, \delta) \asymp \begin{cases} \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta; X_1)}, & \delta \in [\frac{\alpha}{100}, \frac{\alpha}{4}) \\ \log(\alpha/\delta) \cdot n^\star(p, q, \alpha', \alpha'/8) & \delta \in (\alpha^2, \frac{\alpha}{100}] \\ \frac{\log(1/\delta)}{\mathrm{hel}^2(p, q)} & \delta \leq \alpha^2 \end{cases}$$

Prior: $\mathbb{P}(\Theta = p) = \alpha$

Conditionally iid samples $X_1, \ldots, X_n$

Desired failure prob. $\mathbb{P}(\widehat{\theta} \neq \Theta) \leq \delta$

Sample complexity: $n^*(p, q, \alpha, \delta)$

Our results

Theorem: [PJL24]

Let p and q be two distributions with $\mathrm{TV}(p, q) \leq 0.5$.
Let the prior be $\alpha \in (0, 0.5]$ and failure probability $\delta \leq \alpha/4$. Then

$$n^\star(p, q, \alpha, \delta) \asymp \begin{cases} \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta; X_1)}, & \delta \in [\frac{\alpha}{100}, \frac{\alpha}{4}) \\ \log(\alpha/\delta) \cdot n^\star(p, q, \alpha', \alpha'/8) & \delta \in (\alpha^2, \frac{\alpha}{100}] \\ \frac{\log(1/\delta)}{\mathrm{hel}^2(p,q)} & \delta \leq \alpha^2 \end{cases}$$

A complete characterization of sample complexity

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{\text{Initial uncertainty}}{\text{Information per sample}} = \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta;X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)

$$\textsc{FailureProb} = \sum_{x_1,\ldots,x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha)q\left(x_{1:n}\right)\Big)$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p, q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$\text{FAILUREPROB} = \sum_{x_1, \ldots, x_n} \min\left(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\right)$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p, q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$\textsc{FailureProb} = \sum_{x_1,\ldots,x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\Big)$$
$$\leq \alpha^{1-\lambda} \textstyle\sum_{x_1,\ldots,x_n} p(x_{1:n})^{1-\lambda} q(x_{1:n})^\lambda$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta;X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p,q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$\begin{aligned}\textsc{FailureProb} &= \sum_{x_1,\ldots,x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\Big) \\ &\leq \alpha^{1-\lambda} \textstyle\sum_{x_1,\ldots,x_n} p(x_{1:n})^{1-\lambda} q(x_{1:n})^\lambda \\ &= \alpha^{1-\lambda}(1 - \mathcal{H}_\lambda(p^{\otimes n}, q^{\otimes n}))\end{aligned}$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p, q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$
\begin{aligned}
\textsc{FailureProb} &= \sum_{x_1, \ldots, x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\Big) \\
&\leq \alpha^{1-\lambda} \textstyle\sum_{x_1, \ldots, x_n} p(x_{1:n})^{1-\lambda} q(x_{1:n})^\lambda \\
&= \alpha^{1-\lambda}(1 - \mathcal{H}_\lambda(p^{\otimes n}, q^{\otimes n})) \\
&= \alpha^{1-\lambda}\left(1 - \mathcal{H}_\lambda(p, q)\right)^n
\end{aligned}
$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta;X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p,q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$
\begin{aligned}
\textsc{FailureProb} &= \sum_{x_1,\ldots,x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\Big) \\
&\leq \alpha^{1-\lambda} \textstyle\sum_{x_1,\ldots,x_n} p(x_{1:n})^{1-\lambda} q(x_{1:n})^\lambda \\
&= \alpha^{1-\lambda}(1 - \mathcal{H}_\lambda(p^{\otimes n}, q^{\otimes n})) \\
&= \alpha^{1-\lambda}\left(1 - \mathcal{H}_\lambda(p,q)\right)^n \quad \overset{???}{\leq} \alpha/4
\end{aligned}
$$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p, q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,

$$
\begin{aligned}
\textsc{FailureProb} &= \sum_{x_1,\dots,x_n} \min\Big(\alpha p\left(x_{1:n}\right), (1-\alpha) q\left(x_{1:n}\right)\Big) \\
&\leq \alpha^{1-\lambda} \textstyle\sum_{x_1,\dots,x_n} p(x_{1:n})^{1-\lambda} q(x_{1:n})^\lambda \\
&= \alpha^{1-\lambda}(1 - \mathcal{H}_\lambda(p^{\otimes n}, q^{\otimes n})) \\
&= \alpha^{1-\lambda}\left(1 - \mathcal{H}_\lambda(p, q)\right)^n \quad \overset{???}{\leq} \ \alpha/4
\end{aligned}
$$

Suffices to take $n = \frac{\lambda \log(1/\alpha)}{\mathcal{H}_\lambda(p,q)}$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta;X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p,q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,
  - $n^\star \leq \frac{\lambda \log(1/\alpha)}{\mathcal{H}_\lambda(p,q)}$

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\mathrm{bin}}(\alpha)}{I(\Theta;X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p,q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,
  - $n^\star \leq \frac{\lambda \log(1/\alpha)}{\mathcal{H}_\lambda(p,q)}$
- **Incomparable** bounds ☹

Proof overview: linear regime of failure probability

- Standard argument using TV distance and Hellinger divergence fail
- **Lower bound** (using Fano's method)
  - $n^\star \geq \frac{h_{\text{bin}}(\alpha)}{I(\Theta; X_1)}$
- **Upper bound** (using Hellinger Tensorizarion)
  - Generalized Hellinger divergence, $\mathcal{H}_\lambda(p, q) := 1 - \sum_i p_i^{1-\lambda} q_i^\lambda$,
  - $n^\star \leq \frac{\lambda \log(1/\alpha)}{\mathcal{H}_\lambda(p,q)}$
- **Incomparable** bounds ☹
- **Our contribution**: These bounds are **equivalent** for a special λ ☺

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?
    - For non-uniform prior, may not even decrease! *When?*

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?
    - For non-uniform prior, may not even decrease! *When?*
  - Optimal constants?

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?
    - For non-uniform prior, may not even decrease! *When?*
  - Optimal constants?
  - Quantum hypothesis testing

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?
    - For non-uniform prior, may not even decrease! *When?*
  - Optimal constants?
  - Quantum hypothesis testing
  - M-ary testing (dependence on M)

Conclusion

- Characterized instance-optimal sample complexity
- **Open questions**
  - **Weak** detection: failure probability $\delta = \alpha(1-\gamma)$ and $\gamma \to 0$
    - For uniform prior, sample complexity decreases with γ. *Precise scaling*?
    - For non-uniform prior, may not even decrease! *When?*
  - Optimal constants?
  - Quantum hypothesis testing
  - M-ary testing (dependence on M)

Thank You!

Introducing prior-free formulation

Let p, q be two known distributions. Let α, β satisfy $0 \le \alpha \le \beta \ll 1$

Input: i.i.d. samples $X_1, \dots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$:

Introducing prior-free formulation

Let p, q be two known distributions. Let α, β satisfy $0 \leq \alpha \leq \beta \ll 1$

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$:

Type-I error. $\mathbb{P}_{X_{1:n} \sim \mathbf{p}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq p\right) \leq \alpha$

Type-II error. $\mathbb{P}_{X_{1:n} \sim \mathbf{q}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq q\right) \leq \beta$

Introducing prior-free formulation

Let p, q be two known distributions. Let α, β satisfy $0 \leq \alpha \leq \beta \ll 1$

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$:

Type-I error. $\mathbb{P}_{X_{1:n} \sim \mathbf{p}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq p\right) \leq \alpha$

Type-II error. $\mathbb{P}_{X_{1:n} \sim \mathbf{q}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq q\right) \leq \beta$

- Sample complexity: minimum number of samples needed

Introducing prior-free formulation

Let p, q be two known distributions. Let α, β satisfy $0 \leq \alpha \leq \beta \ll 1$

Input: i.i.d. samples $X_1, \ldots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$:

Type-I error. $\mathbb{P}_{X_{1:n} \sim \mathbf{p}^{\otimes n}}\left(\widehat{\theta}(X_{1:n}) \neq p\right) \leq \alpha$

Type-II error. $\mathbb{P}_{X_{1:n} \sim \mathbf{q}^{\otimes n}}\left(\widehat{\theta}(X_{1:n}) \neq q\right) \leq \beta$

- Sample complexity: minimum number of samples needed
- Existing results: $\frac{\log(\alpha/\beta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\beta)}{\mathrm{hel}^2(p,q)}$

Introducing prior-free formulation

Let p, q be two known distributions. Let α, β satisfy $0 \leq \alpha \leq \beta \ll 1$

Input: i.i.d. samples $X_1, \dots, X_n$ from a distribution $\theta \in \{p, q\}$

Output: $\widehat{\theta}(X_{1:n})$:

Type-I error. $\mathbb{P}_{X_{1:n} \sim \mathbf{p}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq p\right) \leq \alpha$

Type-II error. $\mathbb{P}_{X_{1:n} \sim \mathbf{q}^{\otimes n}} \left(\widehat{\theta}(X_{1:n}) \neq q\right) \leq \beta$

- Sample complexity: minimum number of samples needed
- Existing results: $\frac{\log(\alpha/\beta)}{\mathrm{hel}^2(p,q)} \leq$ sample complexity $\leq \frac{\log(1/\beta)}{\mathrm{hel}^2(p,q)}$

Can we close this gap?