Fundamental Limits of Heterogeneous Distributed Detection: Price of Anonymity

Available at ***arXiv:1805.03554***

Wei-Ning Chen
wnchen@ntu.edu.tw

Ho-Chun Chen
r04942078@ntu.edu.tw

I-Hsiang Wang
ihwang@ntu.edu.tw

National Taiwan University

Anonymous Heterogeneous Detection

Optimal Decision Rules, Error Exponents, and the Price of Anonymity

Available at ***arXiv:1805.03554***

Wei-Ning Chen
wnchen@ntu.edu.tw

Ho-Chun Chen
r04942078@ntu.edu.tw

I-Hsiang Wang
ihwang@ntu.edu.tw

National Taiwan University

Distributed Detection

- <u>Each Sensor</u>

  1) Observes a sample Y_i

  2) Processes it, and

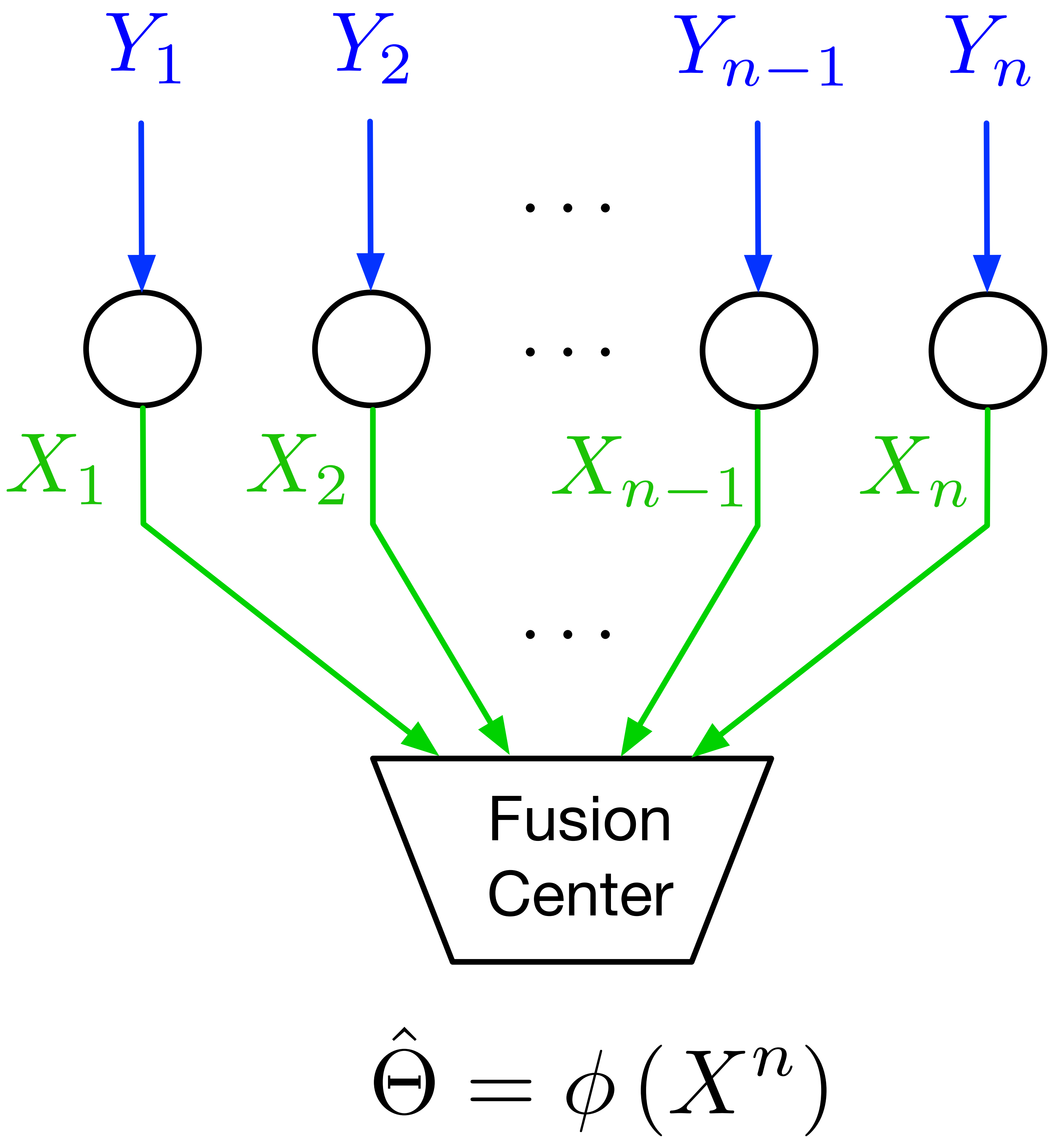
  3) Send a message X_i to FC

  (Local Decision Function: $X_i = \gamma_i\left(Y_i\right)$)[1]

- <u>FC detects hypothesis based on X^n</u>

- <u>Simplification</u>:

  absorb local decision functions γ_i into distributions



[1] J. N. Tsitsiklis , "Decentralized detection," in Advances in Statistical Signal Processing, 1990

Heterogeneous Distributed Detection

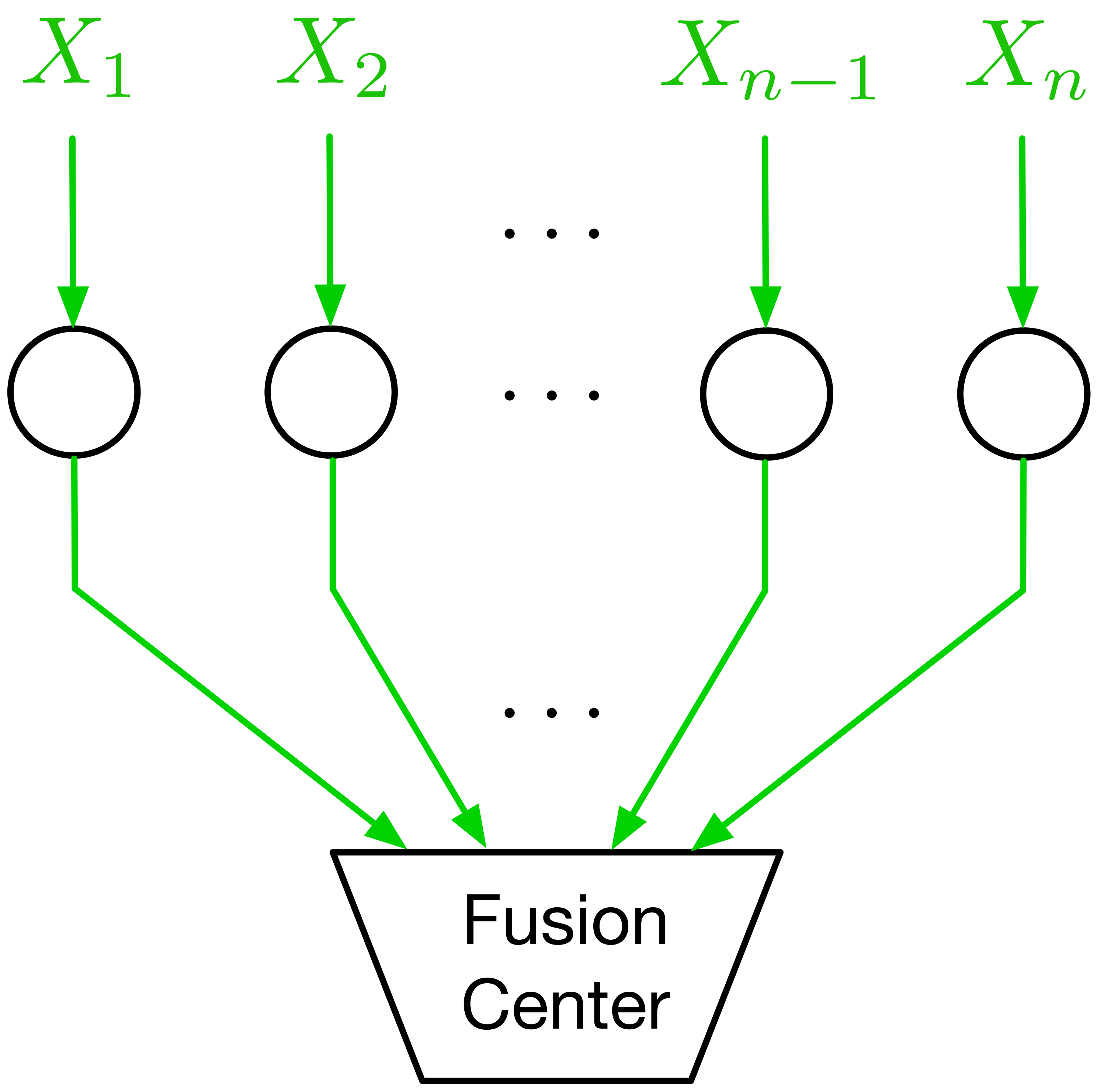
- Heterogeneity happens when *observations are not identically distributed*, or *local decision functions are not identical*.

- Heterogeneity: K group of sensors

$$X_i \overset{\text{i.i.d.}}{\sim} P_{\theta;k}, \text{ for } i \in \mathcal{I}_k$$

  - Sensors in group $\mathcal{I}_k$ follows distribution $P_{\theta;k}$
  - The k-th group has $n\alpha_k$ sensors, $\sum_{k=1}^{K} \alpha_k = 1$

- Neyman-Pearson setting: $\theta \in \{0, 1\}$
  - Minimize Type-II error prob. while keeping type-I error prob. small $(\leq \epsilon)$
  - Minimum Type-II error probability: $\beta^{(n)}(\epsilon, \alpha_1, ..., \alpha_K)$
  - Error exponent: $E(\epsilon, \boldsymbol{\alpha}) \triangleq \lim_{n\to\infty} \left\{-\frac{1}{n} \log_2 \beta^{(n)}(\epsilon, \boldsymbol{\alpha})\right\}$, if it exists



$$\hat{\Theta} = \phi(X^n)$$

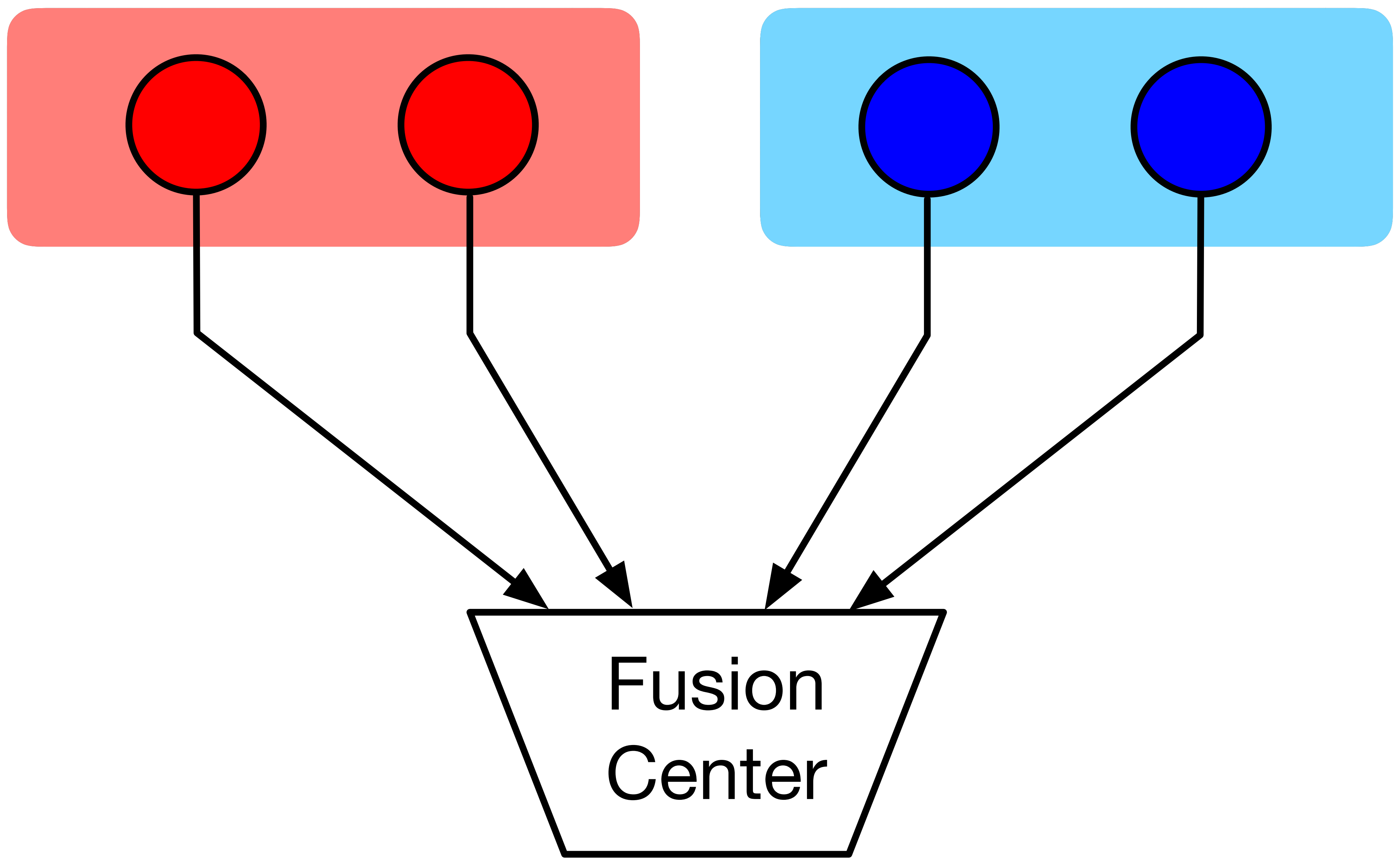
Effect of Heterogeneity without Anonymity

Example: Two Group (K=2)

Group 1:

$X_i \overset{\text{i.i.d.}}{\sim} P_\theta$

$n(1-\alpha)$ sensors



Group 2:

$X_i \overset{\text{i.i.d.}}{\sim} Q_\theta$

$n\alpha$ sensors

When FC is informed of the group that each sensor belongs to:

$$\Rightarrow E_{\text{informed}}(\epsilon, \alpha) = (1-\alpha)D\left(P_0 \| P_1\right) + \alpha D\left(Q_0 \| Q_1\right)$$

weighted combination of ‘resolvability’ of different groups!

Sensor Anonymity

- Why consider anonymity?
  - Privacy : some sensors might not be willing to reveal their groups
  - Communication cost : identifying its group requires sending extra $\log_2 K$ bits from each sensor.

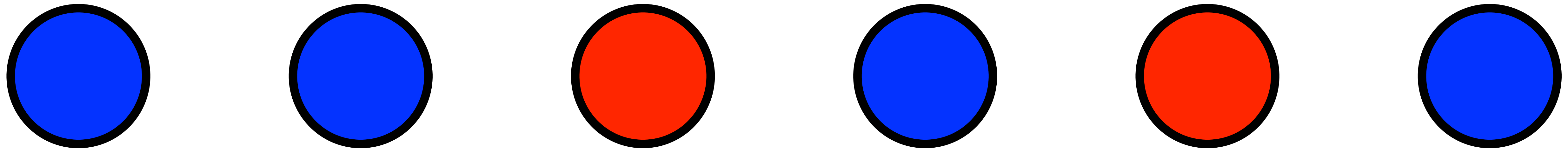
- What is the price of anonymity?

  In the homogeneous setting, no price at all.

- How about the heterogeneous setting?

Composite Hypothesis Testing

- Not sure about which group each sensor belongs to?
  $\Rightarrow$ design algo. with performance guarantee **for all possible scenarios**



sensor ID i	1	2	3	4	5	6
group assignment $\sigma(i)$	2	2	1	2	1	2

- Formally speaking:

$$\begin{cases} \mathcal{H}_0 : X^n \sim \mathbb{P}_{0;\sigma} \triangleq \prod_{i=1}^{n} P_{0;\sigma(i)}, & \text{for some } \sigma \\ \mathcal{H}_1 : X^n \sim \mathbb{P}_{1;\sigma} \triangleq \prod_{i=1}^{n} P_{1;\sigma(i)}, & \text{for some } \sigma \end{cases}$$

$$\sigma : [n] \to [K],\ s.t.\ |\{i : \sigma(i) = k\}| = n\alpha_k$$

$$\boldsymbol{P}_\theta \triangleq \begin{bmatrix} P_{\theta;1} \\ P_{\theta;2} \\ \vdots \\ P_{\theta;K} \end{bmatrix}$$

group distributions

Composite Hypothesis Testing

$$\begin{cases} \mathcal{H}_0 : X^n \sim \mathbb{P}_{0;\sigma} \triangleq \prod_{i=1}^{n} P_{0;\sigma(i)}, & \text{for some } \sigma \\ \mathcal{H}_1 : X^n \sim \mathbb{P}_{1;\sigma} \triangleq \prod_{i=1}^{n} P_{1;\sigma(i)}, & \text{for some } \sigma \end{cases}$$

$\sigma : [n] \to [k]$, s.t. $|\{i|\sigma(i) = k\}| = n\alpha_k$

- Example: $K = 2$, $\boldsymbol{\alpha} = (\frac{1}{3}, \frac{2}{3})$ (red : blue$= 1 : 2$)

red	$X_i \overset{\text{i.i.d.}}{\sim} P_{\theta;1}$
blue	$X_i \overset{\text{i.i.d.}}{\sim} P_{\theta;2}$

σ	red blue blue	blue red blue	blue blue red
possible dist. under $\mathcal{H}_0$	$P_{0;1}P_{0;2}P_{0;2}$	$P_{0;2}P_{0;1}P_{0;2}$	$P_{0;2}P_{0;2}P_{0;1}$
possible dist. under $\mathcal{H}_1$	$P_{1;1}P_{1;2}P_{1;2}$	$P_{1;2}P_{1;1}P_{1;2}$	$P_{1;2}P_{1;2}P_{1;1}$

Minimax Neyman-Pearson Formulation

- Probability of errors:

$$\mathsf{P_F}^{(n)}(\phi) \triangleq \max_{\sigma} \mathbb{P}_{0;\sigma}\{\phi(X^n) = 1\}$$ (the worst case Type-I error probability)

$$\mathsf{P_M}^{(n)}(\phi) \triangleq \max_{\sigma} \mathbb{P}_{1;\sigma}\{\phi(X^n) = 0\}$$ (the worst case Type-II error probability)

- Neyman-Pearson Regime :

$$\beta^{(n)}(\epsilon, \boldsymbol{\alpha}) \triangleq \min_{\phi} \mathsf{P_M}^{(n)}(\phi)$$

$$\text{s.t. } \mathsf{P_F}^{(n)}(\phi) < \epsilon$$

Huber[1973], Kuznetsov[1982], Veeravalli [1994], etc.

- Type-II error exponent:

$$E(\epsilon, \boldsymbol{\alpha}) \triangleq \lim_{n\to\infty} \left\{ -\frac{1}{n} \log_2 \beta^{(n)}(\epsilon, \boldsymbol{\alpha}) \right\}$$

Main Contribution : Optimal Test

- An intuitive test : first *estimate the group assignment* σ, then do LRT

  $\Rightarrow$ Generalized likelihood ratio test

$$\phi(x^n) = \begin{cases} 1, & \text{if } \ell(x^n) < \tau \\ \gamma, & \text{if } \ell(x^n) = \tau \\ 0, & \text{if } \ell(x^n) > \tau \end{cases}$$

is this optimal ?

~~$$\ell_{\mathsf{GLRT}}(x^n) \triangleq \frac{\sup_\sigma \mathbb{P}_{0;\sigma}(x^n)}{\sup_\sigma \mathbb{P}_{1;\sigma}(x^n)}$$~~

- Optimal Decision Rule :

mixture likelihood ratio test

$$\ell(x^n) \triangleq \frac{\sum_\sigma \mathbb{P}_{0;\sigma}(x^n)}{\sum_\sigma \mathbb{P}_{1;\sigma}(x^n)}$$

likelihood ratio between uniform mixture under $\mathcal{H}_0$ to $\mathcal{H}_1$

Main Contribution : Type-II Error Exponent

- A generalized 'divergence' :

$$D_{\boldsymbol{\alpha}}(\boldsymbol{P},\boldsymbol{Q}) \triangleq \min_{\boldsymbol{U}\in(\mathcal{P}_{\mathcal{X}})^K} \sum_{k=1}^{K} \alpha_k D\left(U_k \,\|\, Q_k\right)$$

$$\text{s.t. } \boldsymbol{\alpha}^\intercal \boldsymbol{U} = \boldsymbol{\alpha}^\intercal \boldsymbol{P}$$

  - ▸ Plays a similar role as KL divergence in simple hypothesis testing

recall

$$\boldsymbol{P} = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_K \end{bmatrix}$$

- Type-II error exponent :

$$E(\epsilon,\boldsymbol{\alpha}) = D_{\boldsymbol{\alpha}}(\boldsymbol{P}_0,\boldsymbol{P}_1)$$

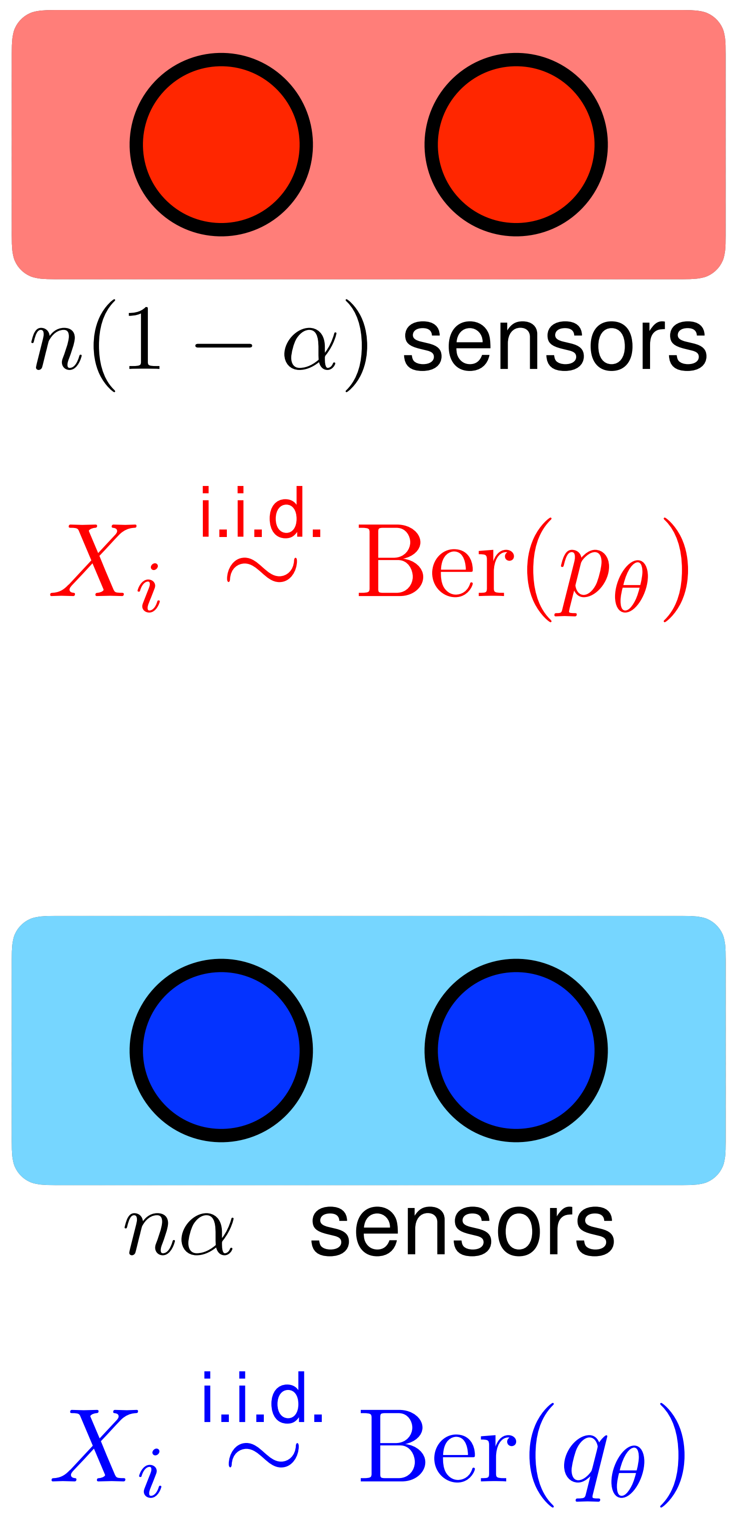
  - ▸ Independent of ϵ, convex in $\boldsymbol{\alpha}$

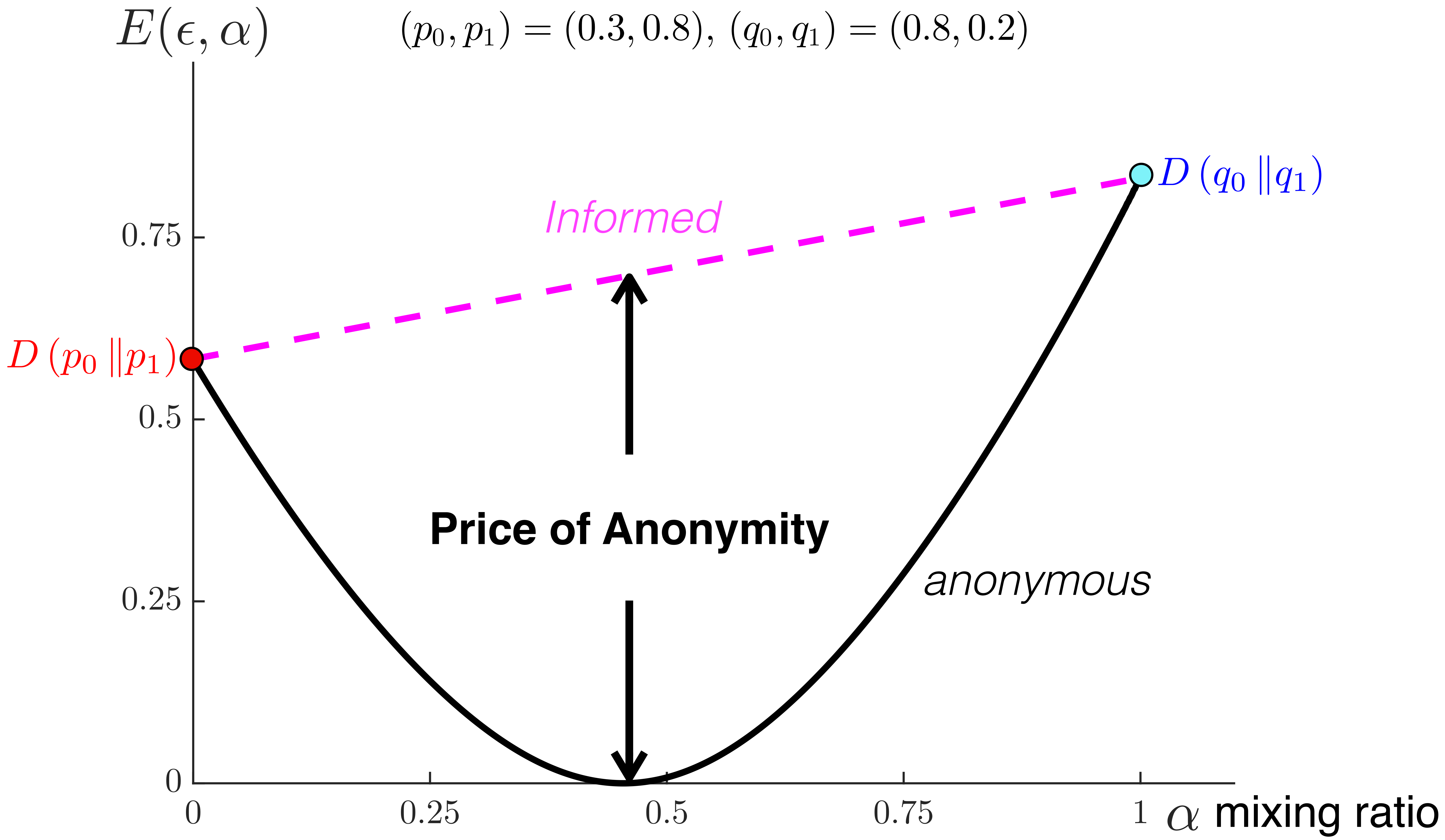
- Compared to informed case :

$$E_{\text{informed}}(\epsilon,\boldsymbol{\alpha}) = \sum_{k=1}^{K} \alpha_k D\left(P_{0;k} \,\|\, P_{1;k}\right)$$

Main Contribution : Type-II Error Exponent

Example (K=2)





Sketch of Proof : Optimal Test

- Idea :

  1) '*Symmetric test*' (tests depend only on the empirical distribution of x^n) is the best

  2) Among all symmetric tests, the *mixture likelihood ratio test (MLRT)* is optimal

Sketch of Proof : Optimal Test

- Idea :

  1) ‘*Symmetric test*’ (tests depend only on the empirical distribution of x^n) is the best

  2) Among all symmetric tests, the *mixture likelihood ratio test (MLRT)* is optimal

step 1

$\psi(\cdot)$ ⟶ symmetrization ⟶ $\phi(\cdot)$

$$\phi(x^n) = \frac{1}{n!} \sum_{\tau:\ \text{all permutations}} \psi\left(\tau(x^n)\right)$$

step 2

ϕ is better : $\mathsf{P}_\mathsf{F}(\phi) \le \mathsf{P}_\mathsf{F}(\psi)$, and $\mathsf{P}_\mathsf{M}(\phi) \le \mathsf{P}_\mathsf{M}(\psi)$

proof

$$\begin{aligned}\mathsf{P}_\mathsf{F}(\phi) &= \max_\sigma \mathsf{E}_{\mathbb{P}_{0;\sigma}}\left[\frac{1}{n!}\sum_\tau \psi\circ\tau(X^n)\right] \\ &= \max_\sigma \frac{1}{n!}\sum_\tau \mathsf{E}_{\mathbb{P}_{0;\sigma}}\left[\psi\circ\tau(X^n)\right] \\ &\le \frac{1}{n!}\sum_\tau \max_\sigma \mathsf{E}_{\mathbb{P}_{0;\sigma}}\left[\psi\circ\tau(X^n)\right] \\ &= \mathsf{P}_\mathsf{F}(\psi)\end{aligned}$$

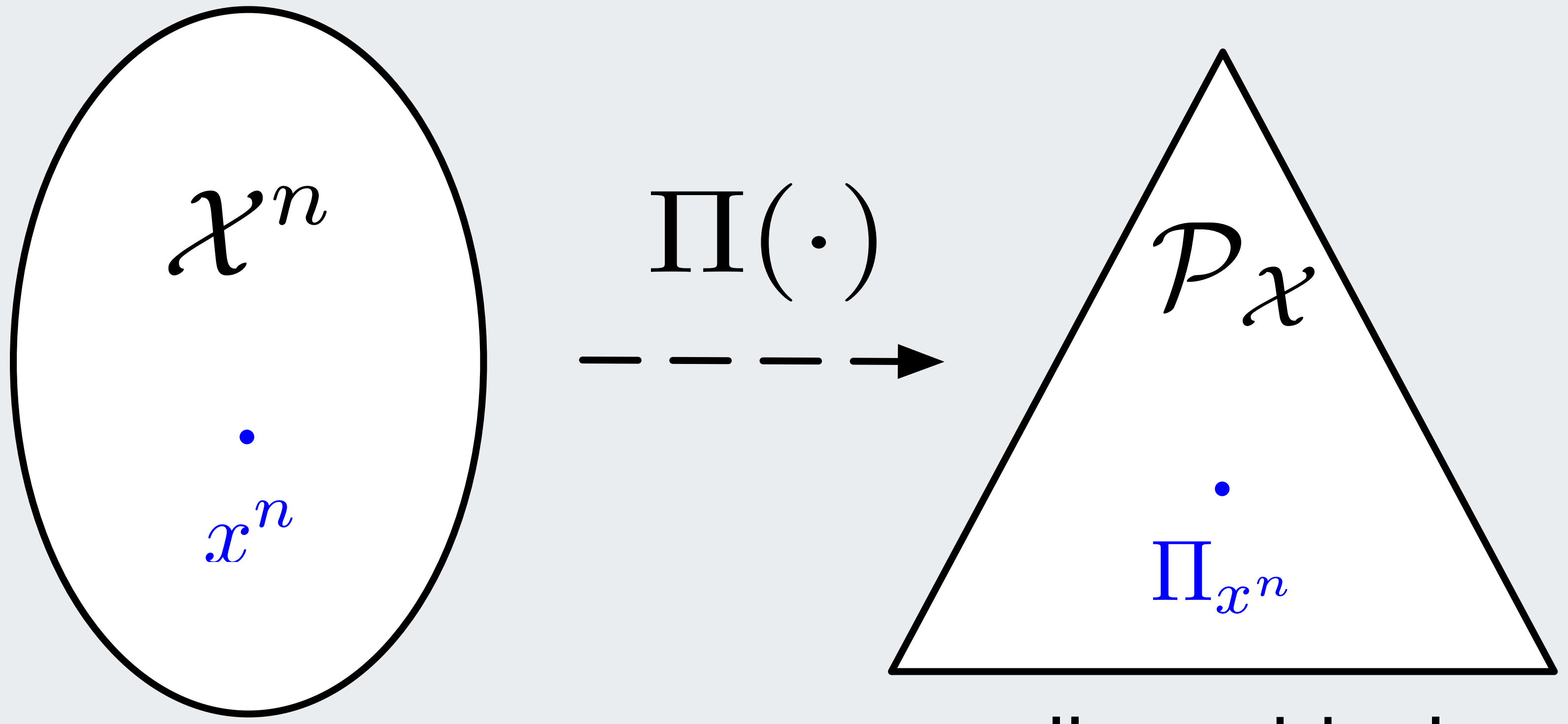
the empirical distribution contains sufficient information !

Sketch of Proof : Optimal Test

- Idea :

  1) *'Symmetric test'* (tests depend only on the empirical distribution of x^n) is the best

  2) Among all symmetric tests, the *mixture likelihood ratio test (MLRT)* is optimal



all empirical distributions

observation :

independent of σ !

$$\mathbb{P}_{\theta;\sigma}\left(T(\Pi_{x^n})\right) \triangleq \tilde{\mathbb{P}}_{\theta}(\Pi_{x^n})$$

collection of x^n with all possible orderings

Equivalent *simple* hypothesis testing on $\mathcal{P}_{\mathcal{X}}$

$$\begin{cases} \mathcal{H}_0 : \mathbb{P}_{0;\sigma}, \text{ for some } \sigma \\ \mathcal{H}_1 : \mathbb{P}_{1;\sigma}, \text{ for some } \sigma \end{cases} \Rightarrow \begin{cases} \tilde{\mathcal{H}}_0 : \tilde{\mathbb{P}}_0 \\ \tilde{\mathcal{H}}_1 : \tilde{\mathbb{P}}_1 \end{cases}$$

Neyman-Pearson lemma:

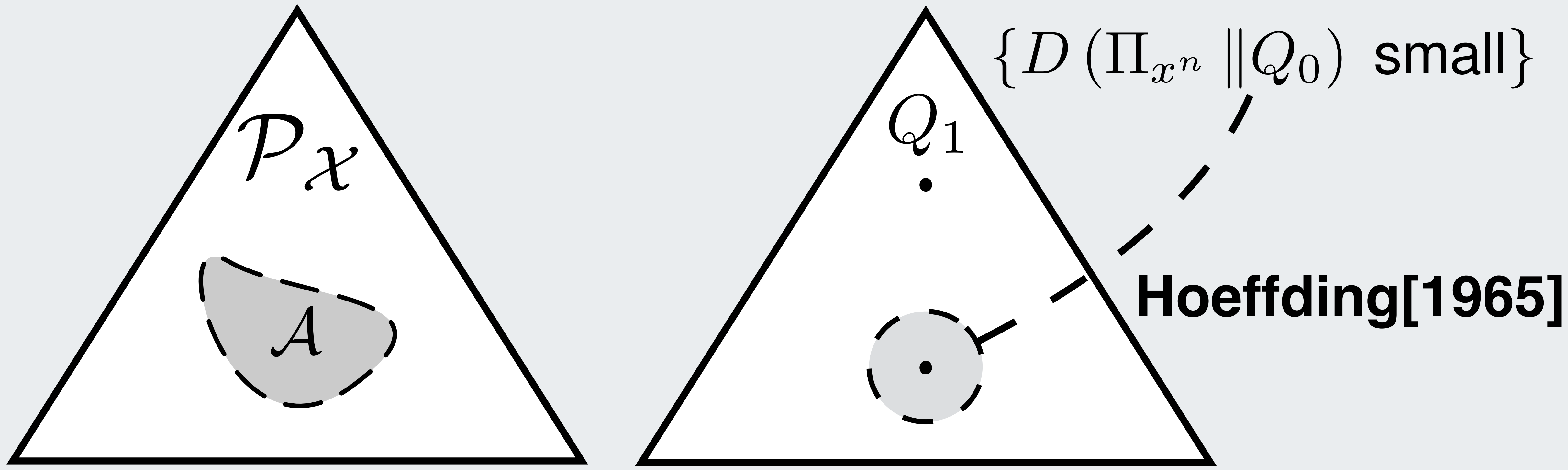
$$\ell(\boldsymbol{x}^n) = \frac{\tilde{\mathbb{P}}_0(\Pi_{\boldsymbol{x}^n})}{\tilde{\mathbb{P}}_1(\Pi_{\boldsymbol{x}^n})} = \frac{\sum_{\sigma} \mathbb{P}_{0;\sigma}(x^n)}{\sum_{\sigma} \mathbb{P}_{1;\sigma}(x^n)}$$

Asymptotic Regime : Sanov's Theorem

i.i.d simple hypothesis testing	heterogeneous anonymous testing
$\mathcal{H}_\theta : X^n \sim (Q_\theta)^{\otimes n}$	$\mathcal{H}_\theta : X^n \sim \mathbb{P}_{\theta;\sigma}$ for some σ

Sanov's Theorem

$$Q_\theta^{\otimes n}(x^n : \Pi_{x^n} \in \mathcal{A}) \approx 2^{-n\left(\min_{U \in \mathcal{A}} D(U \| Q_\theta)\right)}$$



$\Longrightarrow$ type-II error exponent : $D(Q_0 \| Q_1)$

Find exponents of large deviation events:

For any σ, we have

$$\mathbb{P}_{\theta;\sigma}(\Pi_{x^n} \in \mathcal{A}) \approx 2^{-n\left(\min_{\boldsymbol{\alpha}^\intercal \boldsymbol{U} \in \mathcal{A}} D_{\boldsymbol{\alpha}}(\boldsymbol{U}, \boldsymbol{P}_\theta)\right)},$$

with the rate function being

$$D_{\boldsymbol{\alpha}}(\boldsymbol{U}, \boldsymbol{P}_\theta) \triangleq \min_{\boldsymbol{V} \in (\mathcal{P}_\mathcal{X})^K} \sum_{k=1}^{K} \alpha_k D(V_k \| P_{\theta;k})$$

s.t. $\boldsymbol{\alpha}^\intercal \boldsymbol{V} = \boldsymbol{\alpha}^\intercal \boldsymbol{U}$

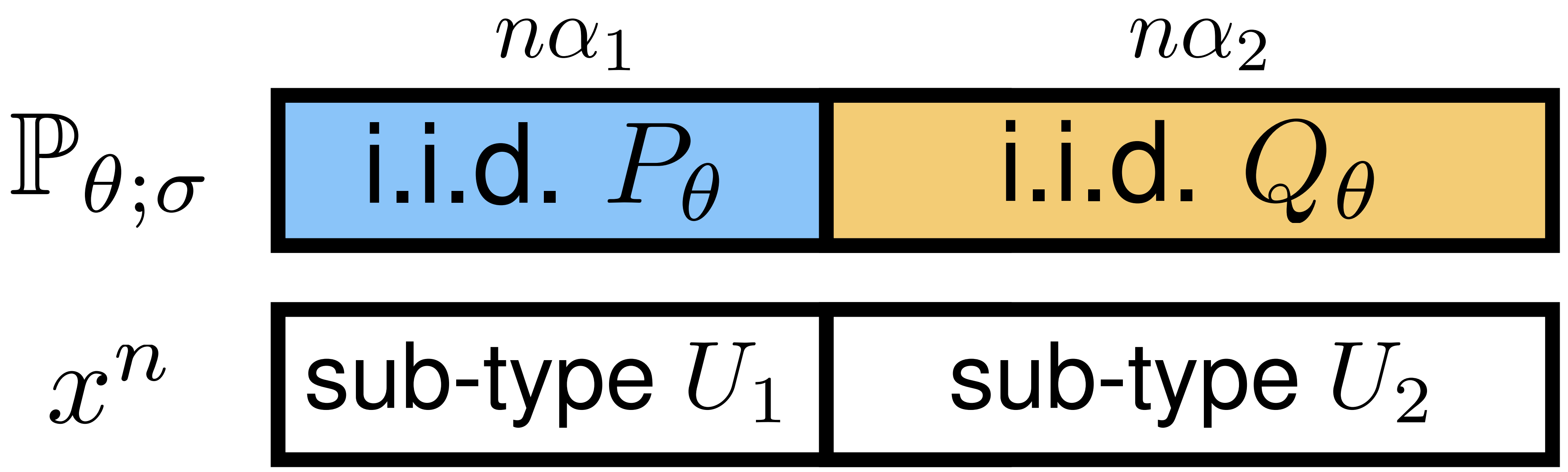
$\Longrightarrow$ type-II error exponent : $D_{\boldsymbol{\alpha}}(\boldsymbol{P}_0, \boldsymbol{P}_1)$

Key Step : non-i.i.d. Sanov's Theorem

Theorem :

For any σ, $\mathbb{P}_{\theta;\sigma}\left(\Pi_{x^n} \in \mathcal{A}\right) \approx 2^{-n\left(\min\limits_{V\in\mathcal{A}} d(V,\boldsymbol{P}_\theta)\right)}$, with the rate function being

$$d(V,\boldsymbol{P}_\theta) \triangleq \min_{\substack{\boldsymbol{U}\in(\mathcal{P}_\mathcal{X})^n \\ \boldsymbol{\alpha}^\intercal\boldsymbol{U}=V}} \sum_{k=1}^{K} \alpha_k D\left(U_k \,\|\, P_{\theta;k}\right)$$



Recall : $Q^{\otimes n}\left(\Pi_{x^n}\right) \approx 2^{-nD(\Pi_{x^n}\|Q)}$

$\mathbb{P}_{\theta;\sigma}(\Pi_{x^n}) \approx 2^{-n(\alpha_1 D(U_1\|P_\theta)+\alpha_2 D(U_2\|Q_\theta))}$

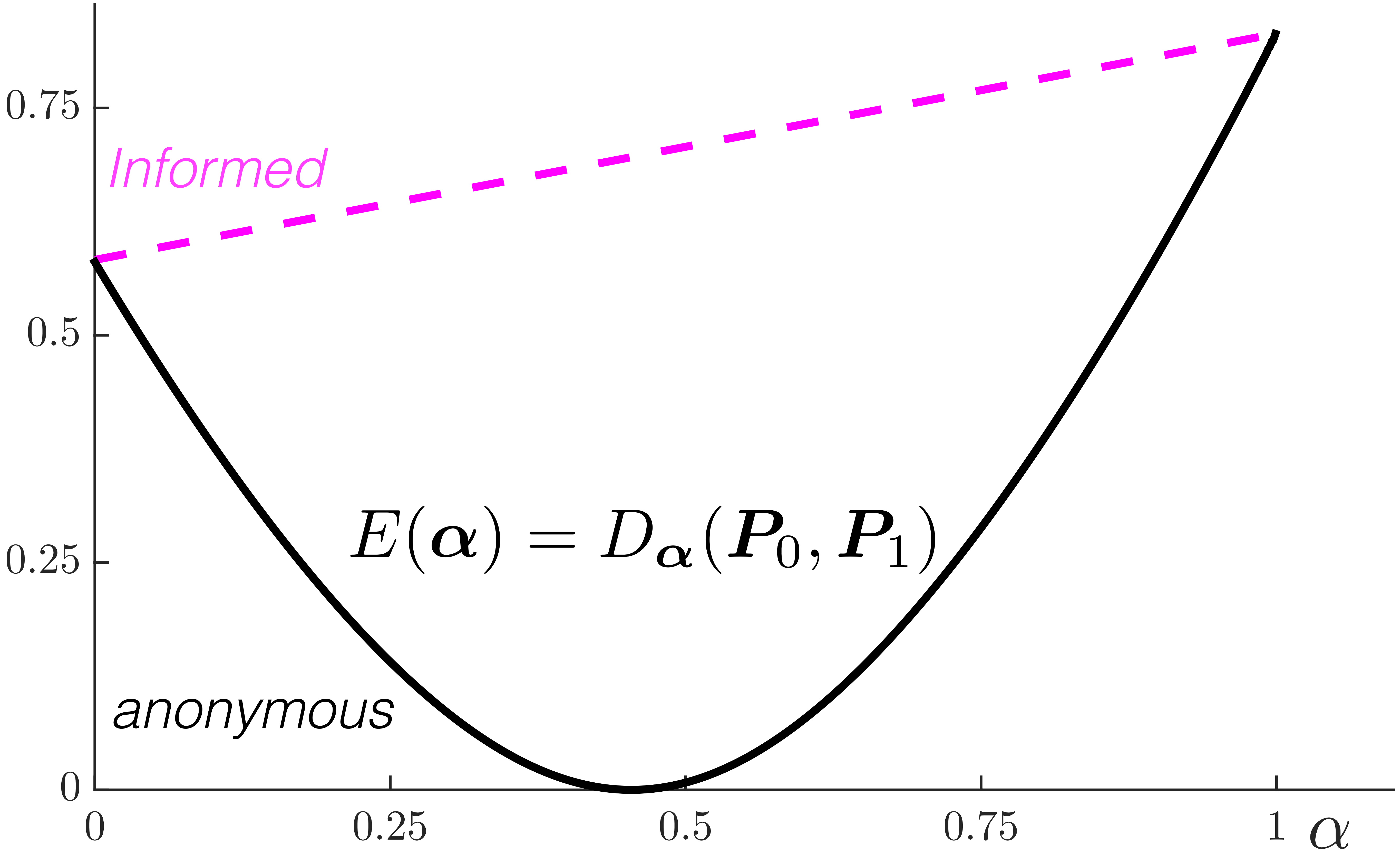
- minimize over all sub-types : $\{U_1, U_2 : \alpha_1 U_1 + \alpha_2 U_2 = V\}$
- minimize over all types : $V \in \mathcal{A}$

Regularity Conditions

- In the proof of the optimal test, we *do not* use bounds on types
  - $\mathcal{X}$ can be continuous
  - only require some assumptions on the $\sigma-$field
- For the type-II error exponent
  - similar idea in the proof of Sanov's Theorem
  - the result can be extended if $\mathcal{X}$ is a *Polish* space

Conclusions and Extension

- ***Optimal decision rule*** : mixture likelihood ratio test (MLRT) ~~GLRT~~
- ***Asymptotic*** :



Generalized divergence

$$D_{\boldsymbol{\alpha}}(\boldsymbol{U}, \boldsymbol{P}_\theta) \triangleq \min_{\boldsymbol{V} \in (\mathcal{P}_\mathcal{X})^K} \sum_{k=1}^{K} \alpha_k D\left(V_k \,\|\, P_{\theta;k}\right)$$

s.t. $\boldsymbol{\alpha}^\intercal \boldsymbol{V} = \boldsymbol{\alpha}^\intercal \boldsymbol{U}$

extended to Chernoff regime by solving information projection !

Thanks for your listening !

If interested, contact us :

ihwang@ntu.edu.tw

wnchen@ntu.edu.tw