# On the Price of Source Anonymity in Heterogeneous Parametric Point Estimation

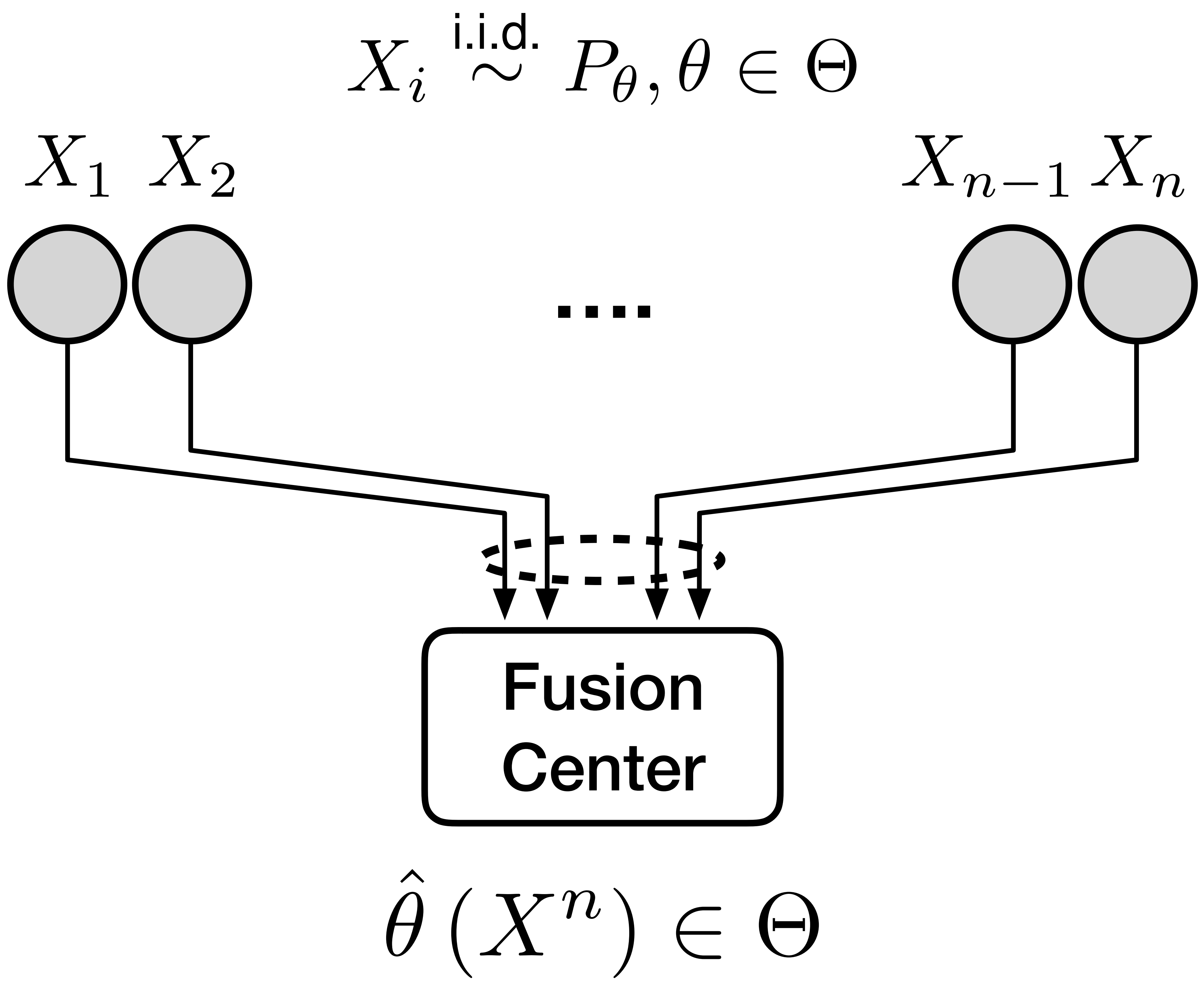

Wei-Ning Chen
wnchen@ntu.edu.tw

I-Hsiang Wang
ihwang@ntu.edu.tw

National Taiwan University

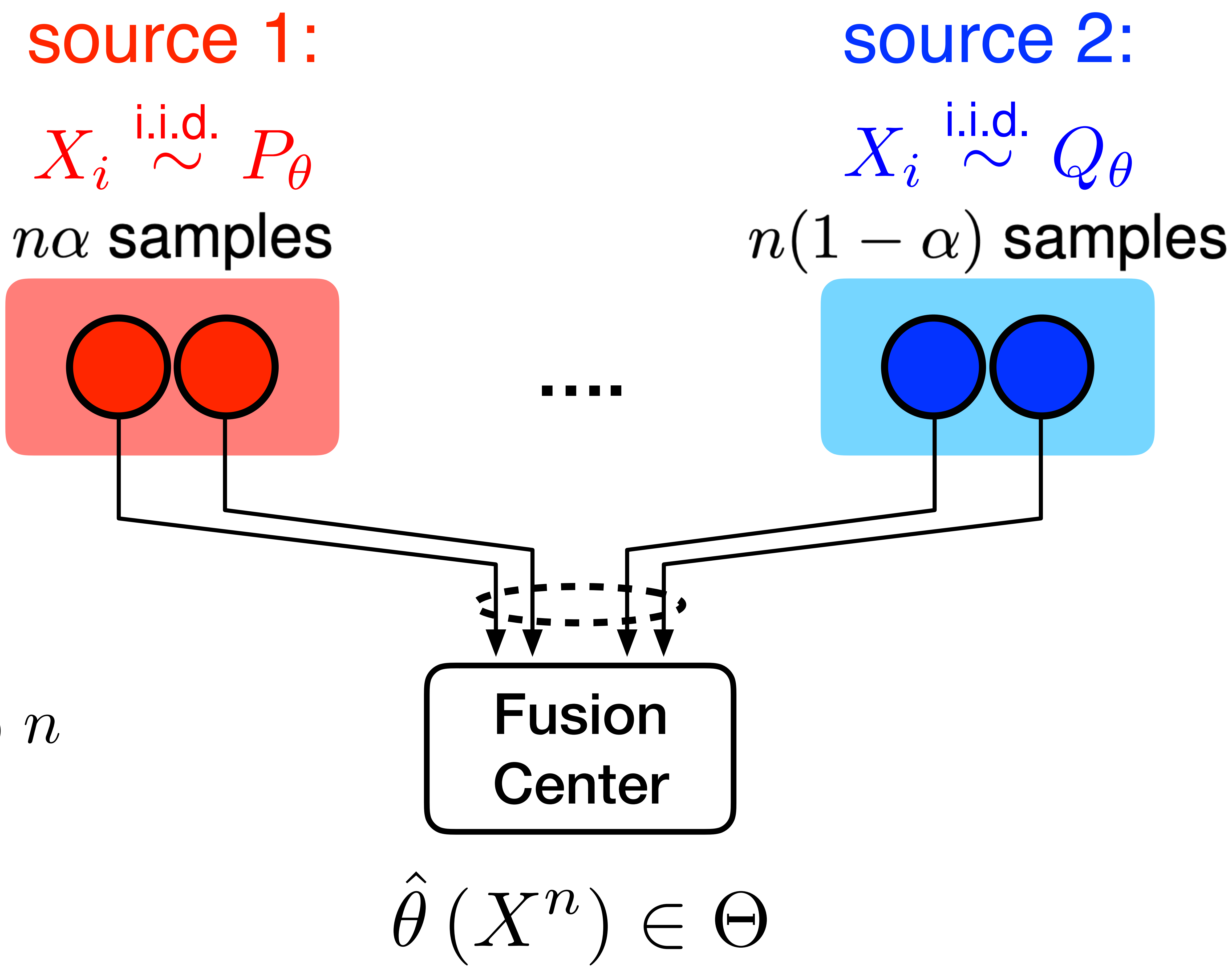
# Estimation From Heterogeneous and Shuffled Data

- <u>Distributed Estimation</u>
  - Fusion center collects $X^n \sim P_\theta^{\otimes n}$
  - Estimate $\theta \in \Theta$ from $X^n$



# Estimation From Heterogeneous and Shuffled Data

- <u>Distributed Estimation</u>
  - Fusion center collects $X^n \sim P_\theta^{\otimes n}$
  - Estimate $\theta \in \Theta$ from $X^n$
- <u>Heterogeneity</u>
  - Samples are drawn from different sources
  - # of samples from each source proportional to $n$



# Estimation From Heterogeneous and Shuffled Data

- Distributed Estimation
  - Fusion center collects $X^n \sim P_\theta^{\otimes n}$
  - Estimate $\theta \in \Theta$ from $X^n$
- Heterogeneity
  - Samples are drawn from different sources
  - # of samples from each source proportional to $n$

source 1:

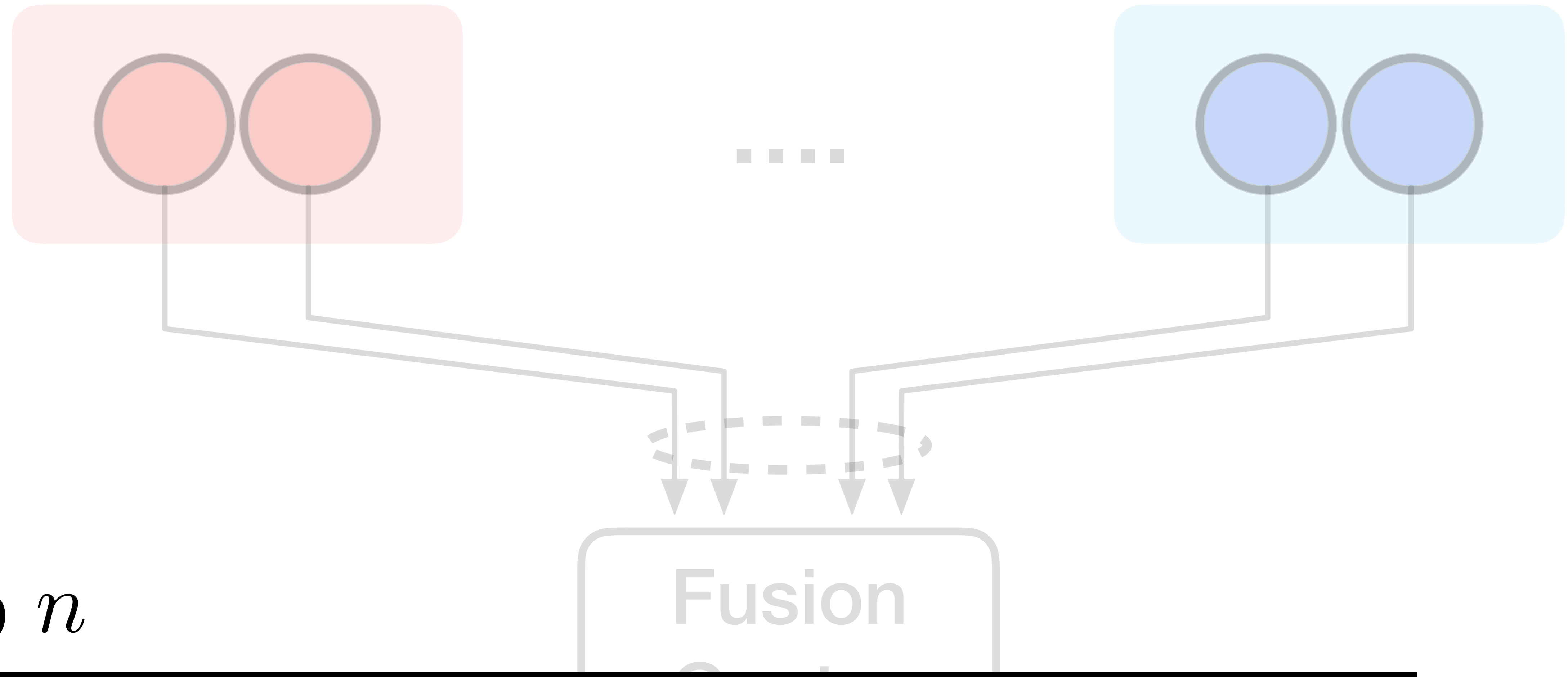
$X_i \overset{\text{i.i.d.}}{\sim} P_\theta$

$n\alpha$ samples

source 2:

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

$n(1-\alpha)$ samples




**Probabilistic (Bayesian) setting:**

→ i.i.d. on mixture distribution

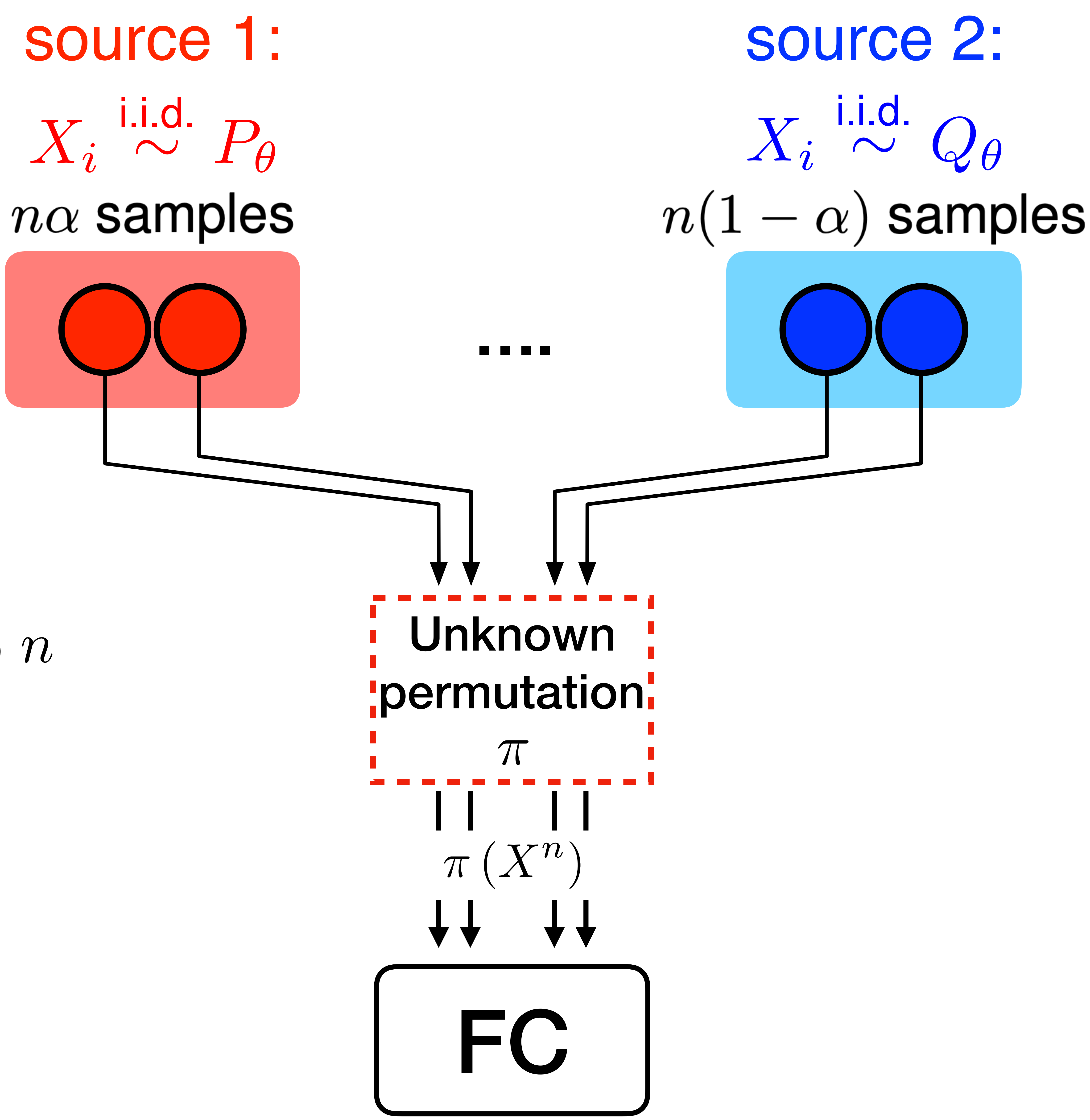
$X_i$ comes from source 1 with probability $\alpha$

**Combinatorial setting:**

Exact $\alpha$ fraction of samples come from source 1

# Estimation From Heterogeneous and Shuffled Data

- <u>Distributed Estimation</u>
  - Fusion center collects $X^n \sim P_\theta^{\otimes n}$
  - Estimate $\theta \in \Theta$ from $X^n$
- <u>Heterogeneity</u>
  - Samples are drawn from different sources
  - # of samples from each source proportional to $n$
- <u>Anonymity</u>
  - FC only observed *shuffled* data



# Estimation From Heterogeneous and Shuffled Data

***Example: n=3***

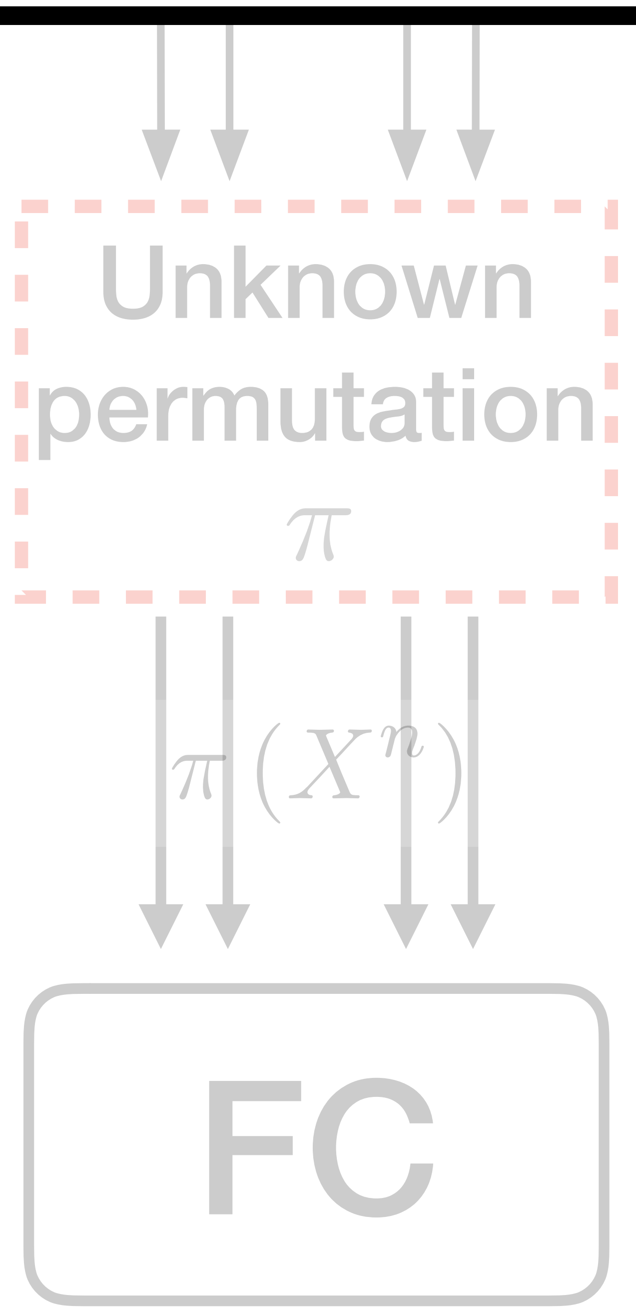
**Possible situations:**

$(X_1, X_2, X_3)$ $(X_2, X_1, X_3)$ $(X_2, X_3, X_1)$

$(X_1, X_3, X_2)$ $(X_3, X_1, X_2)$ $(X_3, X_2, X_1)$

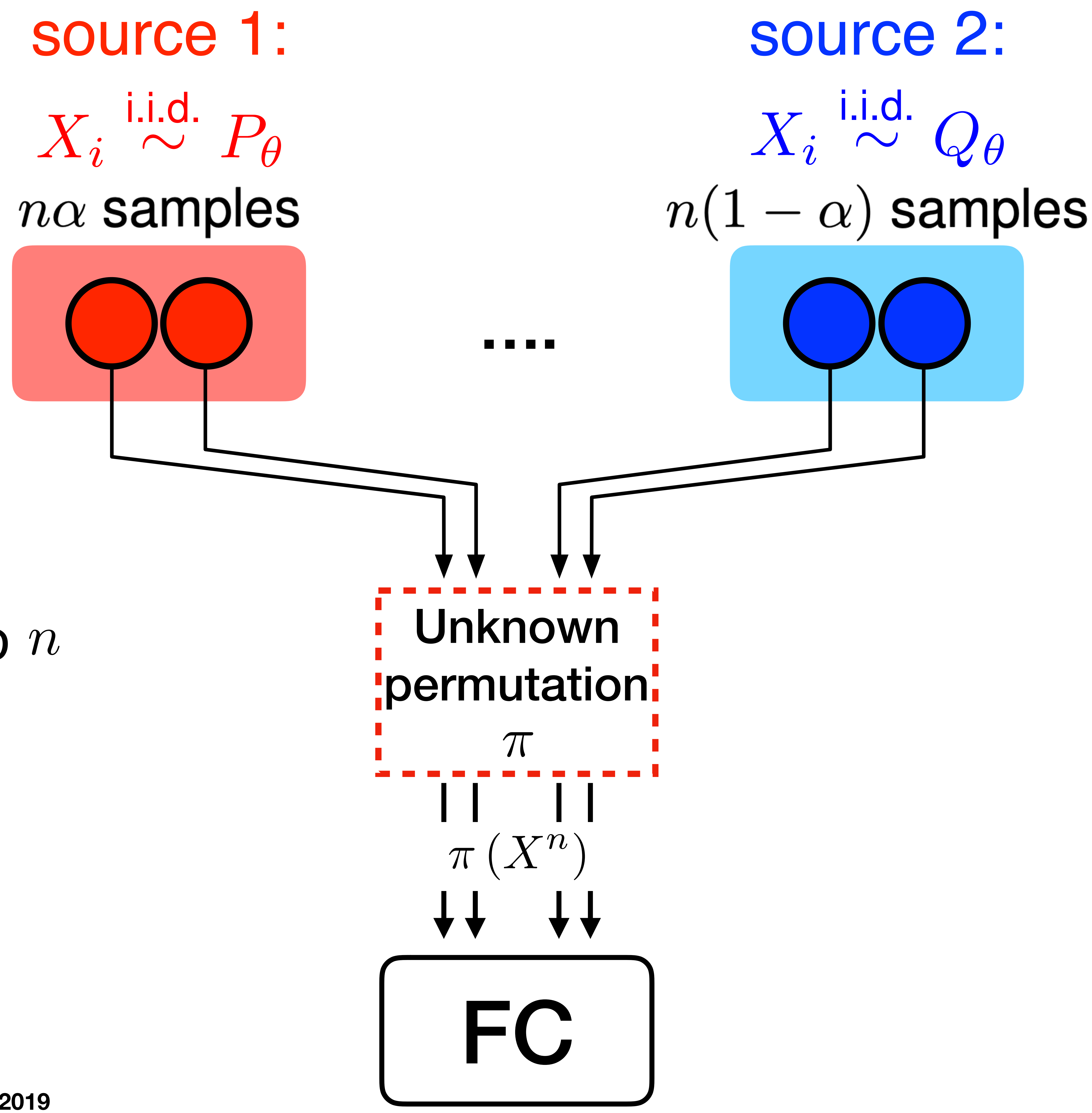
- Samples are drawn from different sources
- # of samples from each source proportional to $n$

- Anonymity
  - FC only observed *shuffled* data



# Estimation From Heterogeneous and Shuffled Data

- Distributed Estimation
  - Fusion center collects $X^n \sim P_\theta^{\otimes n}$
  - Estimate $\theta \in \Theta$ from $X^n$
- Heterogeneity
  - Samples are drawn from different sources
  - # of samples from each source proportional to $n$
- Anonymity
  - FC only observed *shuffled* data
  - Due to *privacy*[1] or *identification cost*




[1] Úlfar Erlingsson et al. , "Amplification by shuffling: From local to central differential privacy via anonymity," SODA 2019

# Effect of Heterogeneity without Anonymity

- **Performance of an Estimator**

$$\text{MSE}(\hat{\theta}) = \mathbb{E}_\theta\left[\left(\hat{\theta}(X^n) - \theta\right)^2\right]$$

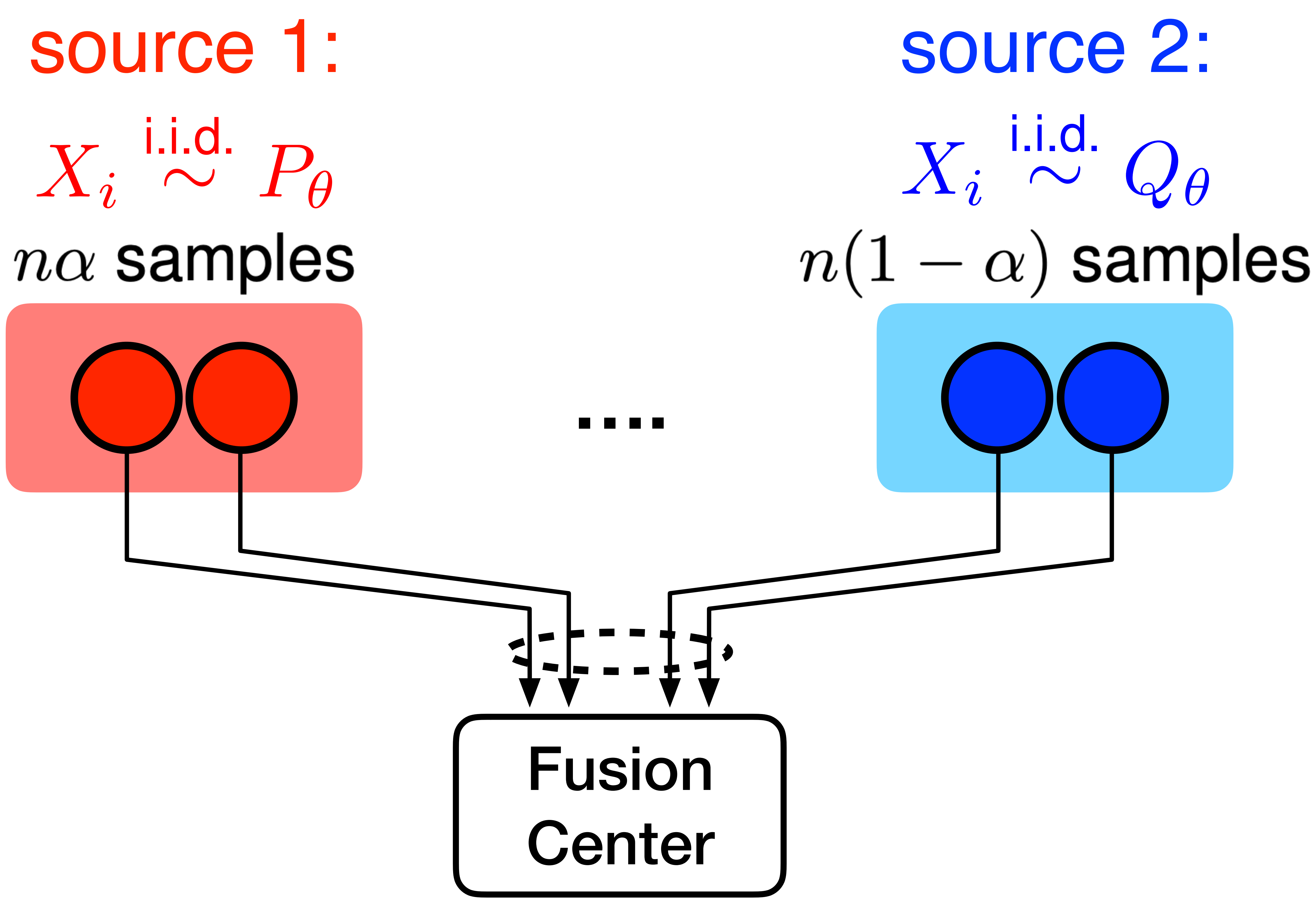
- **Cramér-Rao Lower Bound**

For any unbiased estimator $\hat{\theta}$, we have

$$\text{MSE}\left(\hat{\theta}\right) \geq \frac{1}{I_{\mathbb{P}}(\theta)}$$

- **Fisher Information**

$$I_{\mathbb{P}}(\theta) \triangleq \mathbb{E}_{\mathbb{P}_\theta}\left[\left(\frac{\partial}{\partial\theta}\log\mathbb{P}_\theta(X)\right)^2\right]$$
$$= n\left[\alpha I_P(\theta) + (1-\alpha) I_Q(\theta)\right]$$




$$\hat{\theta}\left(X^n\right) \in \Theta$$

$$X^n \sim \mathbb{P}_\theta \triangleq P_\theta^{\otimes n\alpha} Q_\theta^{\otimes n(1-\alpha)}$$

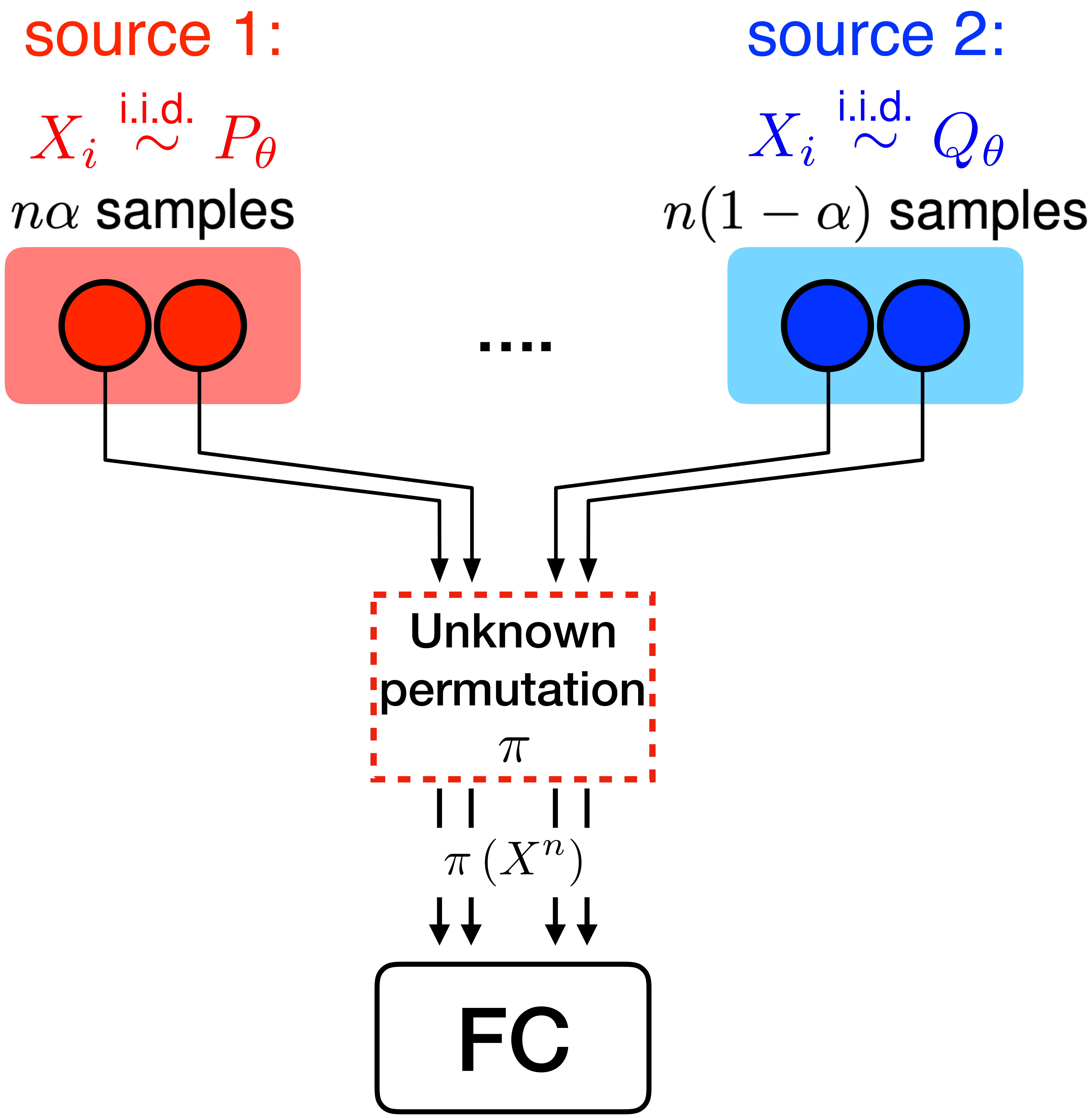
# Effect of Anonymity

- **Performance evaluation**
  - MSE depends on the shuffling
  - Study the *worst-case permutation*

$$\text{MSE}\left(\hat{\theta}\right) \triangleq \max_{\pi \in \mathcal{S}_n} \mathbb{E}_\theta \left[ \left( \hat{\theta}\left(\pi\left(X^n\right)\right) - \theta \right)^2 \right]$$

- **How to address anonymity?**

- **What is the price of anonymity?**
  - *In the homogeneous setting, no price at all!*



# Main Result I — Sufficiency of Type

- The *type* of the samples $\Pi_{X^n}$ is sufficient to estimate $\theta$
- Consider a convex loss function $\ell : \Theta \times \Theta \to [0, \infty),$ and define the *worst-case risk* as

$$R^*_\theta(\phi) \triangleq \max_{\pi \in \mathcal{S}_n} \mathbb{E}_\theta \left[\ell\left(\theta, \phi\left(\pi\left(X^n\right)\right)\right)\right]$$

**<u>Theorem</u>** *(Sufficiency of Type)*

For any estimator $\hat{\phi}\left(X^n\right)$, there exists an estimator $\hat{\theta}\left(X^n\right)$ which depends only on $\Pi_{X^n}$, and

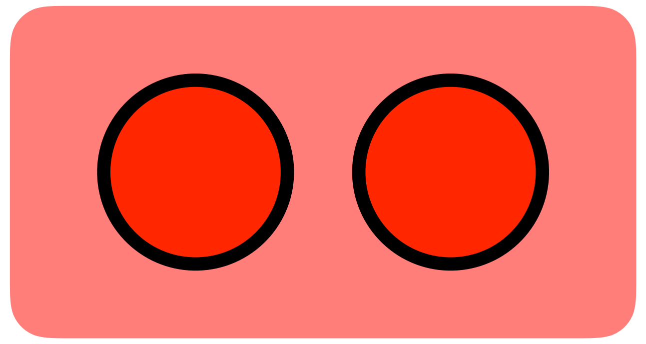
$$R^*_\theta\left(\hat{\theta}\right) \le R^*_\theta\left(\hat{\phi}\right).$$

# Main Result II — Lower Bound on MSE

- A lower bound on worst-case MSE:

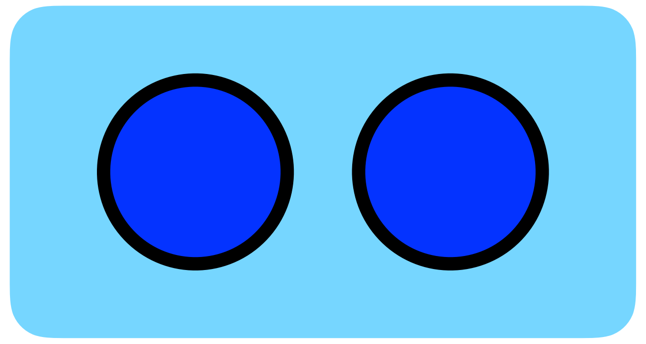
> **Theorem** *(Lower Bound on MSE)*
>
> $$\forall \text{ unbiased } \hat{\theta}, \ \text{MSE}\left(\hat{\theta}\right) \geq \frac{1}{n I_M(\theta)} + o(1/n),$$
>
> where $M_\theta \triangleq \alpha P_\theta + (1-\alpha) Q_\theta$ is the mixture distribution.

# Main Result II — Lower Bound on MSE



$n\alpha$ samples

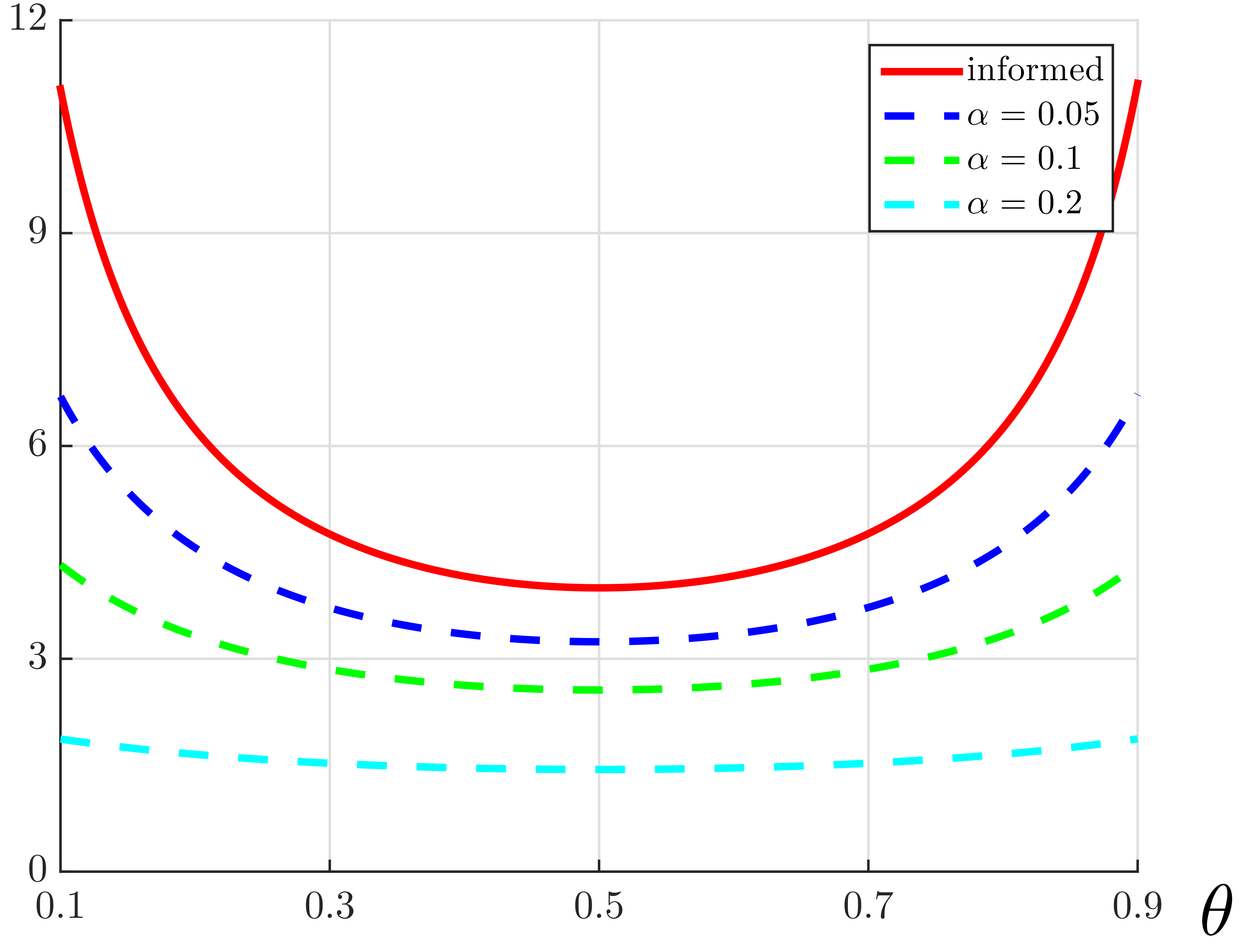
$X_i \overset{\text{i.i.d.}}{\sim} P_\theta = \text{Ber}(\theta)$



$n(1-\alpha)$ samples

$X_i \overset{\text{i.i.d.}}{\sim} Q_\theta = \text{Ber}(1-\theta)$

**IR** $\alpha I_P + (1-\alpha) I_Q$ **v.s.** $I_{\alpha P + (1-\alpha) Q}$




**Informed**

$$\alpha I_P(\theta) + (1-\alpha) I_Q(\theta) = \frac{1}{\theta(1-\theta)}$$

**Anonymous**

$$I_M(\theta), \text{ with } M_\theta = \alpha P_\theta + (1-\alpha) Q_\theta$$

**Theorem** *(Sufficiency of Type)*

For any estimator $\hat{\phi}\left(X^n\right)$, there exists an estimator $\hat{\theta}\left(X^n\right)$ which depends only on $\Pi_{X^n}$, and

$$R_\theta^*\left(\hat{\theta}\right) \le R_\theta^*\left(\hat{\phi}\right).$$

***proof.***

1) Construct $\hat{\theta}$ by symmetrizing $\hat{\phi}$:

$$\hat{\theta}\left(x^n\right) \triangleq \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\hat{\phi}\left(\sigma\left(x^n\right)\right)$$

2)

$$\begin{aligned}
\mathrm{R}_\theta^*\left(\hat{\theta}\right) &= \max_{\pi\in\mathcal{S}_n}\mathbb{E}_\theta\left[\ell\left(\theta,\frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\hat{\phi}\left(\sigma\circ\pi\left(X^n\right)\right)\right)\right] \\
&=\mathbb{E}_\theta\left[\ell\left(\theta,\frac{1}{n!}\sum_{\pi\in\mathcal{S}_n}\hat{\phi}\left(\pi\left(X^n\right)\right)\right)\right] \\
&\le\frac{1}{n!}\sum_{\pi\in\mathcal{S}_n}\mathbb{E}_\theta\left[\ell\left(\theta,\hat{\phi}\left(\pi\left(X^n\right)\right)\right)\right] \\
&\le \max_{\pi\in\mathcal{S}_n}\mathbb{E}_\theta\left[\ell\left(\theta,\hat{\phi}\left(\pi\left(X^n\right)\right)\right)\right] = \mathrm{R}_\theta^*\left(\hat{\phi}\right)
\end{aligned}$$

$\square$

# Proof of Result I

**Theorem** *(Sufficiency of Type)*

For any estimator $\hat{\phi}(X^n)$, there exists an estimator $\hat{\theta}(X^n)$ which depends only on $\Pi_{X^n}$, and

$$R^*_\theta\left(\hat{\theta}\right) \le R^*_\theta\left(\hat{\phi}\right).$$

***We should design an estimator based on*** $\Pi_{X^n}$

$$\hat{\theta}(x^n) \triangleq \frac{1}{n!}\sum_{\sigma\in\mathcal{S}_n}\hat{\phi}(\sigma(x^n))$$

$$\begin{aligned}
&= \mathbb{E}_\theta\left[\ell\left(\theta, \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n}\hat{\phi}(\pi(X^n))\right)\right] \\
&\le \frac{1}{n!}\sum_{\pi\in\mathcal{S}_n}\mathbb{E}_\theta\left[\ell\left(\theta, \hat{\phi}(\pi(X^n))\right)\right] \\
&\le \max_{\pi\in\mathcal{S}_n}\mathbb{E}_\theta\left[\ell\left(\theta, \hat{\phi}(\pi(X^n))\right)\right] = \mathsf{R}^*_\theta\left(\hat{\phi}\right)
\end{aligned}$$

□

# Fisher Information of the Type

- For estimators based on $\Pi_{X^n}$, the MSE no longer depend on $\sigma$.
- It suffices to compute the FI based on $\Pi_{X^n}$.
- <u>Notation</u>
  - $\tilde{\mathbb{P}}_\theta$ denotes the distribution of $\Pi_{X^n}$ (defined on $\mathcal{P}_\mathcal{X}$)
  - $I_{\tilde{\mathbb{P}}}(\theta)$ denotes the FI of $\tilde{\mathbb{P}}_\theta$
- <u>Goal:</u>

$$I_{\tilde{\mathbb{P}}}(\theta) = nI_M(\theta) + o(n), \text{ where } M_\theta(x) = \alpha P_\theta(x) + (1-\alpha)Q_\theta(x)$$

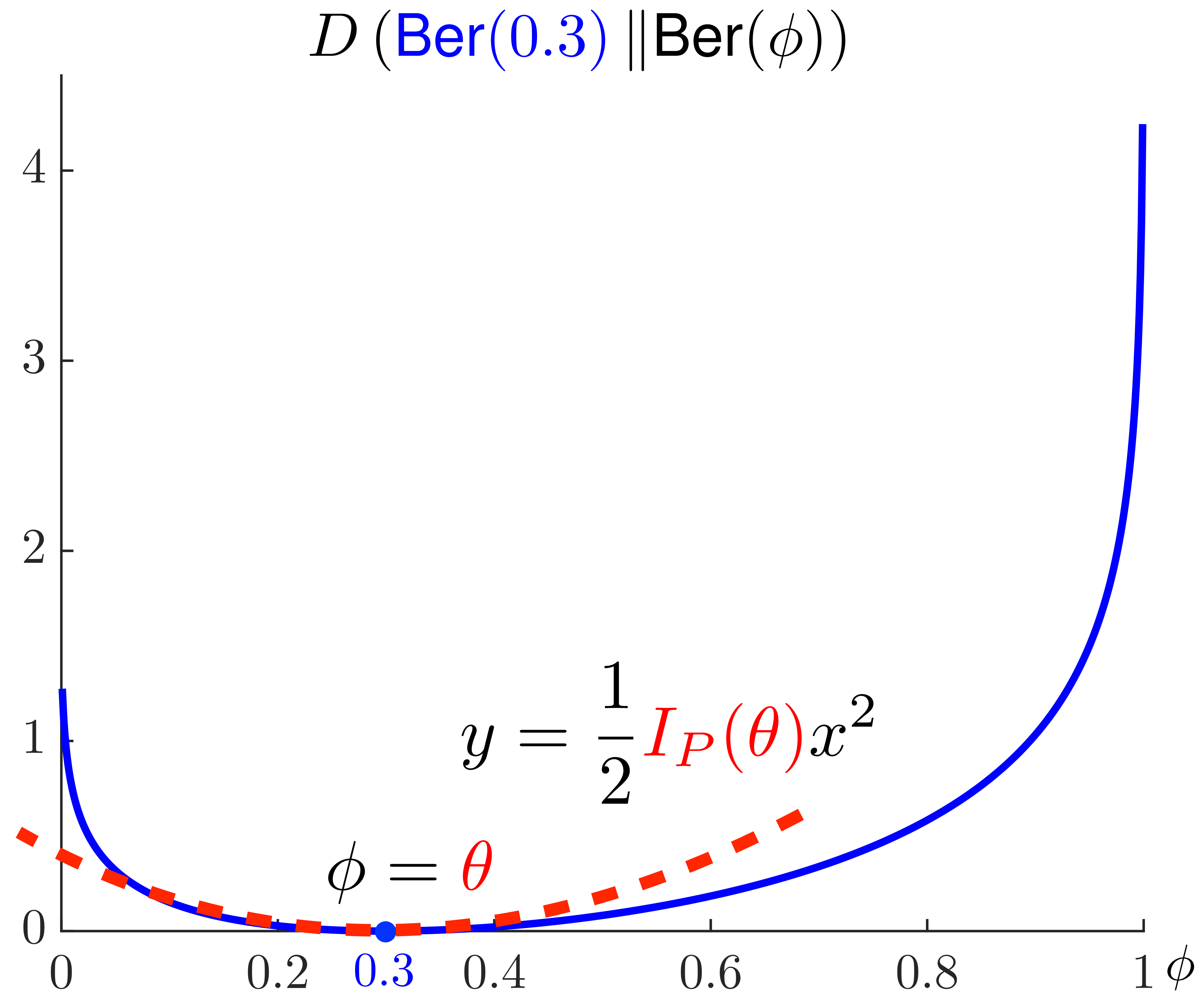
# From Divergence to Fisher Information (1)

- One can start with direct analysis on $I_{\tilde{\mathbb{P}}}(\theta)$
  - Hard to analyze due to the messy form of $\tilde{\mathbb{P}}_\theta$

- Relation between KL divergence and FI

**Fact**

$$I_P(\theta) = \frac{\partial^2}{\partial \phi^2} D\left(P_\theta \,\|P_\phi\right)\Big|_{\phi=\theta}$$

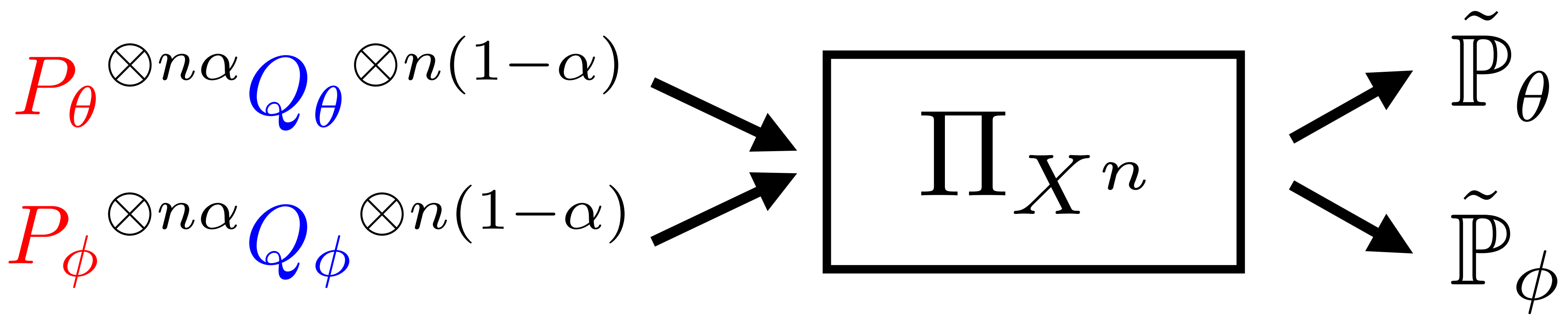


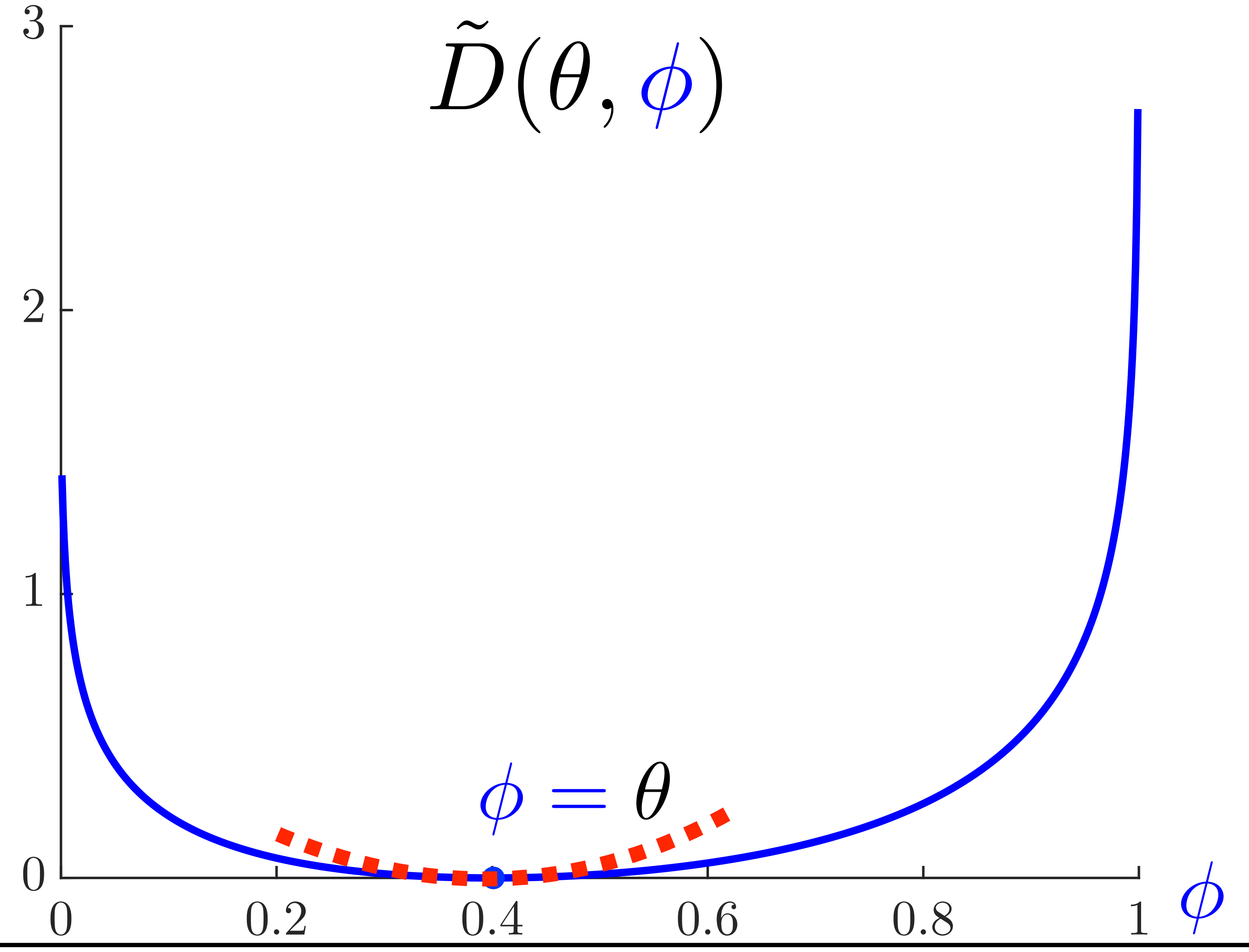

**Fisher information is the *curvature of KL divergence***

# From Divergence to Fisher Information (2)

First compute $D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right)$, and then extend to $I_{\tilde{\mathbb{P}}}(\theta)$!

**Previous Works on *Hypothesis Testing***




## Asymptotic Divergence

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \asymp \begin{array}{ll} \min\limits_{V_0, V_1} & \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \\ \text{s.t.} & \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta \end{array}$$

$$\triangleq \tilde{D}(\theta, \phi)$$

## Asymptotic Info. Rate $\frac{1}{n} I_{\tilde{\mathbb{P}}}(\theta)$

$$\frac{\partial^2}{\partial \phi^2} \tilde{D}(\theta, \phi)\Big|_{\phi=\theta} = \lim_{\Delta\theta \to 0} \frac{\tilde{D}(\theta, \theta + \Delta\theta)}{\Delta\theta^2}$$

approx. by a quadratic problem

$$= I_M(\theta)$$

[2] W.-N. Chen and I.-H. Wang, "Anonymous Heterogeneous Distributed Detection: Optimal Decision Rules, Error Exponents, and the Price of Anonymity," TIT 2019

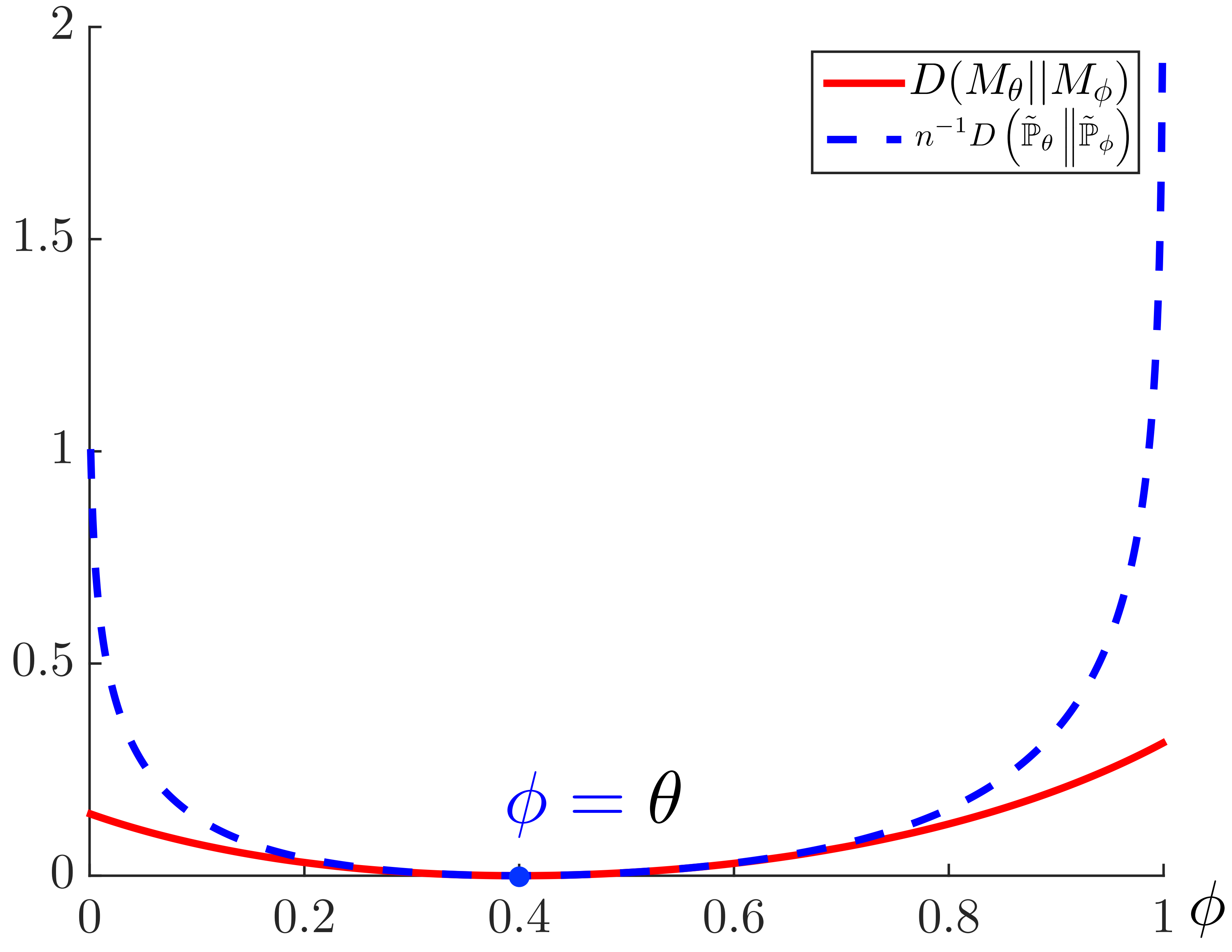
# Comparison with Hypothesis Testing

- Test $\underline{{\color{red}P_\theta}^{\otimes n\alpha}{\color{blue}Q_\theta}^{\otimes n(1-\alpha)}}$ v.s. $\underline{{\color{red}P_\phi}^{\otimes n\alpha}{\color{blue}Q_\phi}^{\otimes n(1-\alpha)}}$ from shuffled data

  <u>Worst type II error exponent</u>

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \asymp \boxed{\tilde{D}(\theta,\phi)} \geq \boxed{D\left(M_\theta \| M_\phi\right)}, \text{ with } M_\theta = \alpha P_\theta + (1-\alpha) Q_\theta$$

have same curvature at $\theta = \phi$!




- Estimate $\theta$ from $\underline{{\color{red}P_\theta}^{\otimes n\alpha}{\color{blue}Q_\theta}^{\otimes n(1-\alpha)}}$ with shuffled samples

  <u>Information Rate</u>

$$I_M(\theta), \text{ with } M_\theta = \alpha P_\theta + (1-\alpha) Q_\theta$$

[2] W.-N. Chen and I.-H. Wang, "Anonymous Heterogeneous Distributed Detection: Optimal Decision Rules, Error Exponents, and the Price of Anonymity," TIT 2019

# Summary

- Worst-case shuffling $\Rightarrow$ design the estimator based on *type*

- Asymptotic IR:

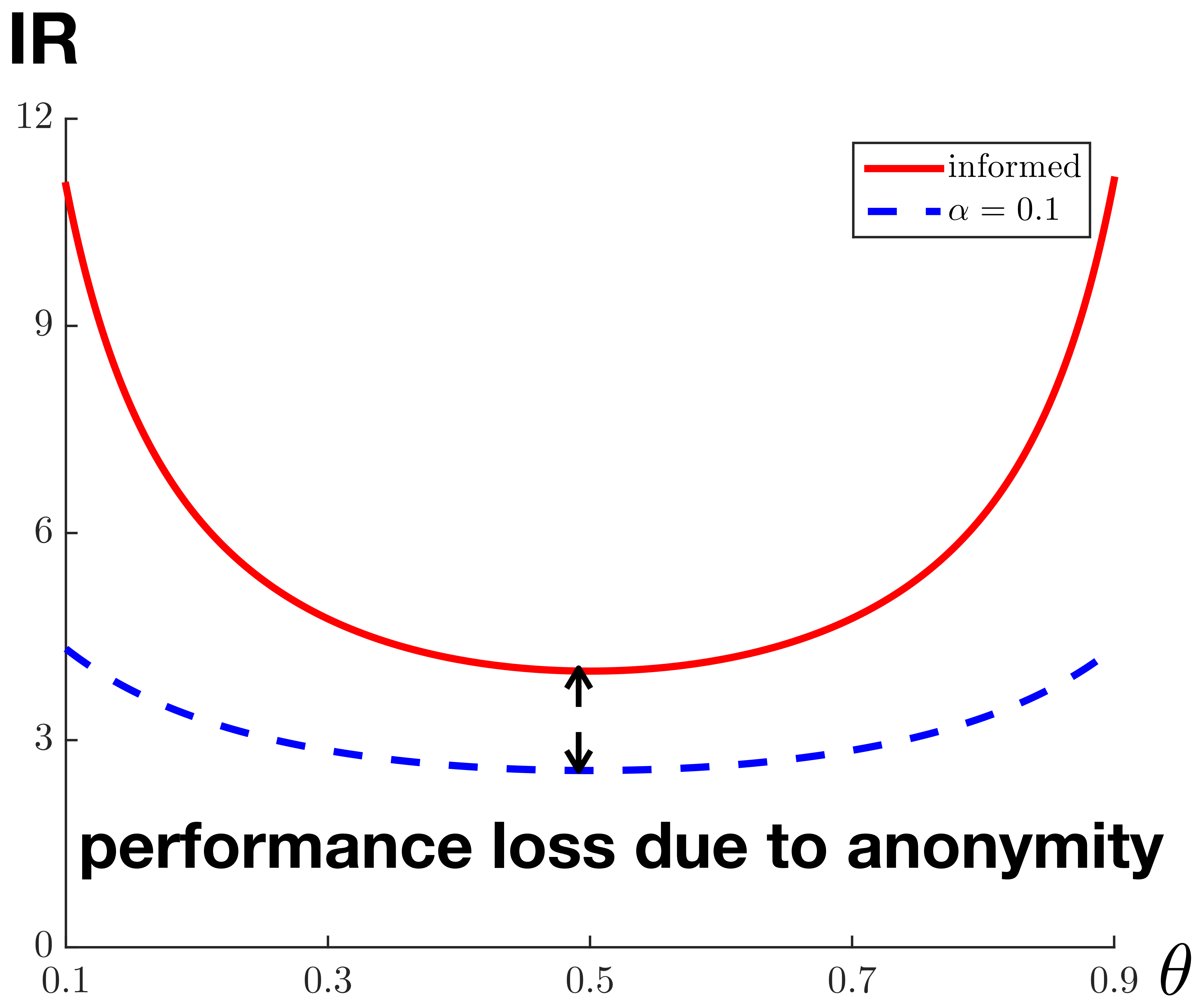
**Informed**

$$\alpha I_P(\theta) + (1-\alpha) I_Q(\theta)$$

**Anonymous**

$I_M(\theta)$, with $M_\theta = \alpha P_\theta + (1-\alpha) Q_\theta$

- Future works:
  - Upper bound on MSE
  - Relax regularity conditions



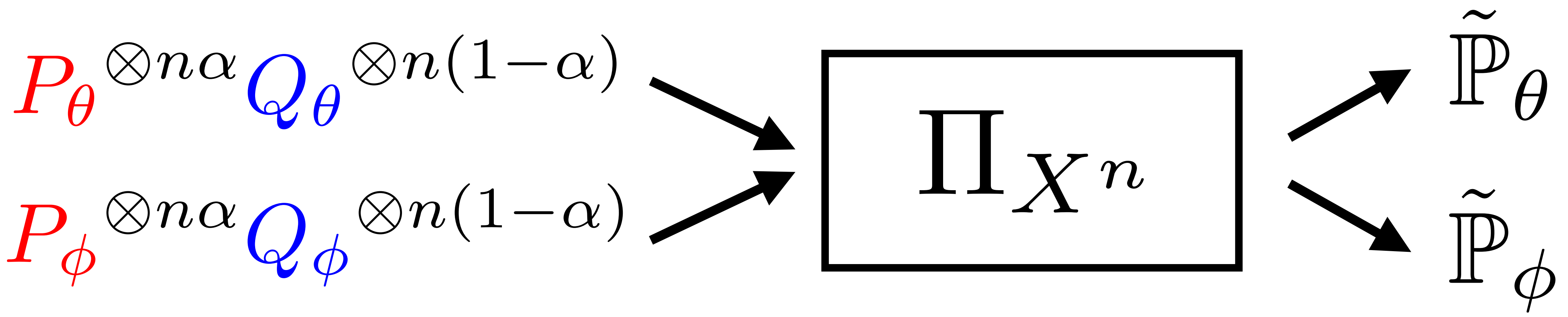
***Thanks for your attention!***

# *Back up*

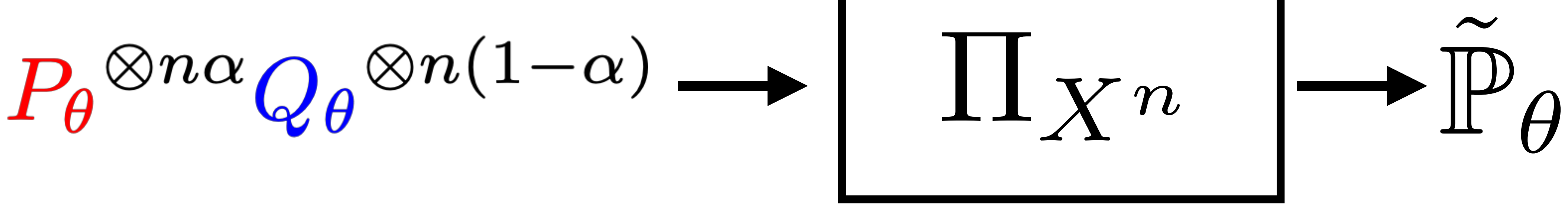
# From Divergence to Fisher Information (1)

First compute $D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right)$, and then extend to $I_{\tilde{\mathbb{P}}}(\theta)$!

**Previous Works on *Hypothesis Testing***




**Extend to *Point Estimation***




<u>Asymptotic Divergence</u>

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \asymp \begin{array}{ll} \min\limits_{V_0, V_1} & \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \\ \text{s.t.} & \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta \end{array}$$

<u>Asymptotic Fisher Information</u>

$$\frac{1}{n} I_{\tilde{\mathbb{P}}}(\theta) \asymp \lim_{n \to \infty} \frac{1}{n} \frac{\partial^2}{\partial \phi^2} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \Big|_{\phi=\theta}$$

[2] W.-N. Chen, H.-C. Chen, and I.-H. Wang, "On the Fundamental Limits of Heterogeneous Distributed Detection: Price of Anonymity," ISIT 2018

# From Divergence to Fisher Information (1)

$$\frac{1}{n} I_{\tilde{\mathbb{P}}}(\theta) \asymp \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2}{\partial \phi^2} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \Big|_{\phi=\theta}$$

$$\overset{(?)}{=} \frac{\partial^2}{\partial \phi^2} \underline{\lim_{n\to\infty} \frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right)} \Big|_{\phi=\theta}$$

**Yes if** $\left\{ \frac{1}{n(\theta-\phi)^2} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \right\}$ **is *equicontinuous*!**

<u>Asymptotic Divergence</u>

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \asymp \min_{V_0, V_1} \ \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \quad \text{s.t.} \quad \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta$$

<u>Asymptotic Fisher Information</u>

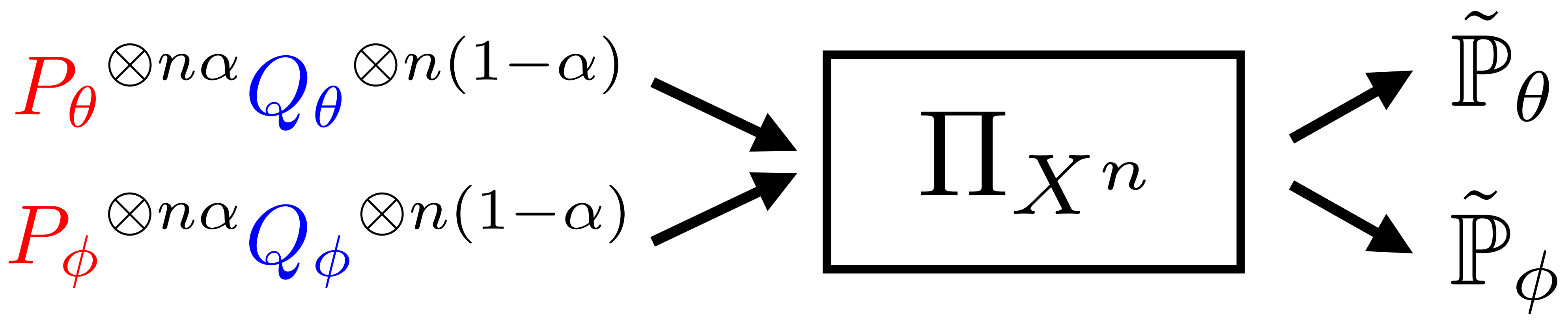
$$\frac{1}{n} I_{\tilde{\mathbb{P}}}(\theta) \asymp \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2}{\partial \phi^2} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \Big|_{\phi=\theta}$$

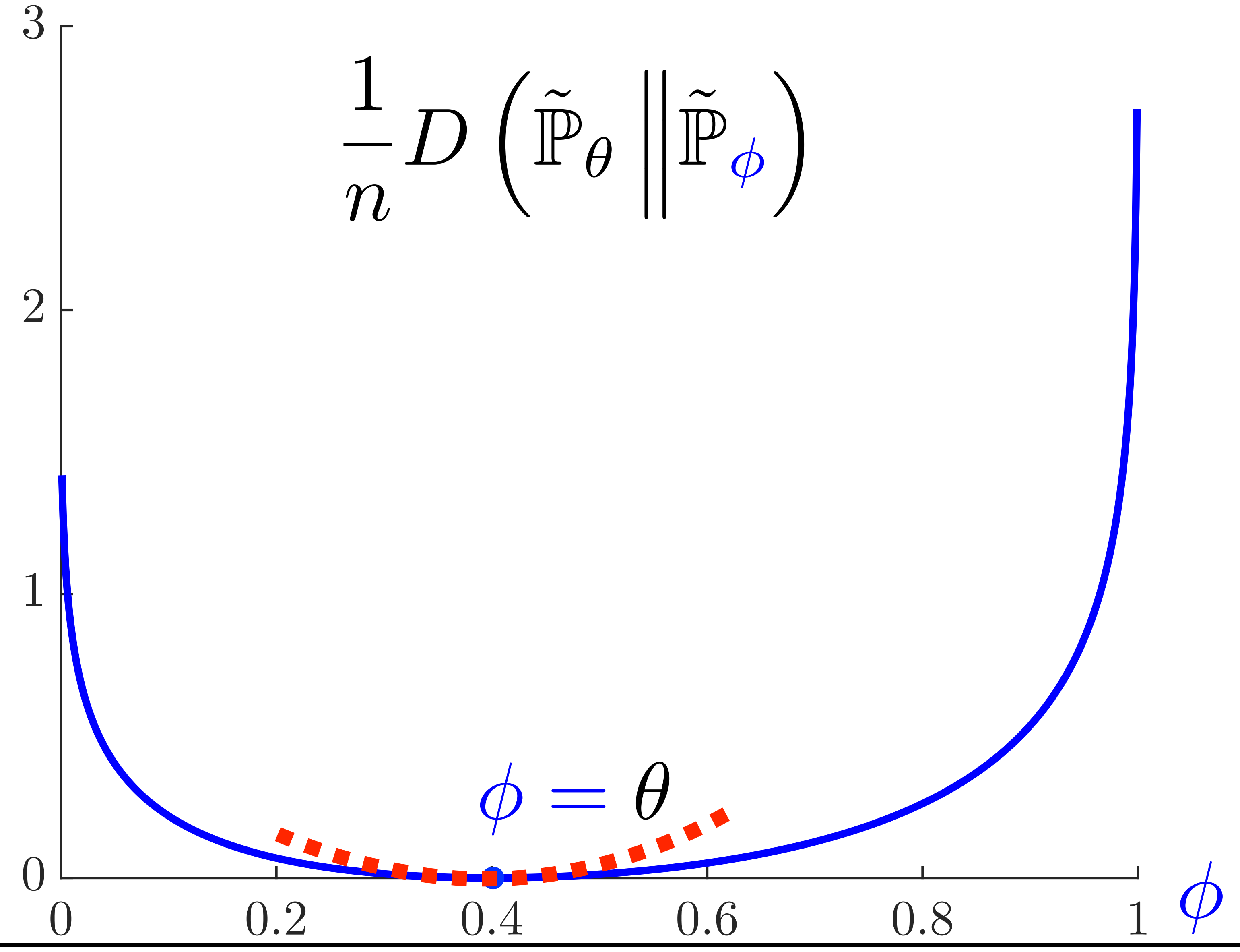
[2] W.-N. Chen, H.-C. Chen, and I.-H. Wang, "On the Fundamental Limits of Heterogeneous Distributed Detection: Price of Anonymity," ISIT 2018

# From Divergence to Fisher Information (2)

First compute $D\left(\tilde{\mathbb{P}}_\theta \,\middle\|\, \tilde{\mathbb{P}}_\phi\right)$, and then extend to $I_{\tilde{\mathbb{P}}}(\theta)$!

## Previous Works on *Hypothesis Testing*




## Asymptotic Divergence

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \,\middle\|\, \tilde{\mathbb{P}}_\phi\right) \asymp \begin{array}{ll} \min\limits_{V_0, V_1} & \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \\ \text{s.t.} & \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta \end{array}$$

## Asymptotic Info. Rate $\frac{1}{n} I_{\tilde{\mathbb{P}}}(\theta)$

$$\frac{\partial^2}{\partial \phi^2} \left( \begin{array}{ll} \min\limits_{V_0, V_1} & \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \\ \text{s.t.} & \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta \end{array} \right) \Bigg|_{\phi = \theta}$$

$$= \lim_{\Delta\theta \to 0} \frac{1}{\Delta\theta^2} \left( \begin{array}{ll} \min\limits_{V_0, V_1} & \alpha D(V_0 \| P_{\theta+\Delta\theta}) + (1-\alpha) D(V_1 \| Q_{\theta+\Delta\theta}) \\ \text{s.t.} & \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta \end{array} \right)$$

approx. by a quadratic problem

$$= I_M(\theta)$$

[2] W.-N. Chen, H.-C. Chen, and I.-H. Wang, "On the Fundamental Limits of Heterogeneous Distributed Detection: Price of Anonymity," ISIT 2018

# Equicontinuity on Divergence

Example: $\alpha = 0.3$, $\theta = 0.3$, $P_\theta = \mathrm{Ber}(\theta)$, $Q_\theta = \mathrm{Ber}(\theta)$

$$\frac{1}{n} D\left(\tilde{\mathbb{P}}_\theta \middle\| \tilde{\mathbb{P}}_\phi\right) \asymp \min_{V_0, V_1} \ \alpha D(V_0 \| P_\phi) + (1-\alpha) D(V_1 \| Q_\phi) \quad \text{s.t.} \quad \alpha V_0 + (1-\alpha) V_1 = \alpha P_\theta + (1-\alpha) Q_\theta$$

Pointwise convergence for each $\phi \neq \theta$