IEEE International Symposium on Information Theory
TAIPEI 2023

Information Spectrum Converse for Minimum Entropy Couplings and Functional Representations

Anuj Yadav
LINX
EPFL

Joint work with :

Yanina Y. Shkel
EPFL

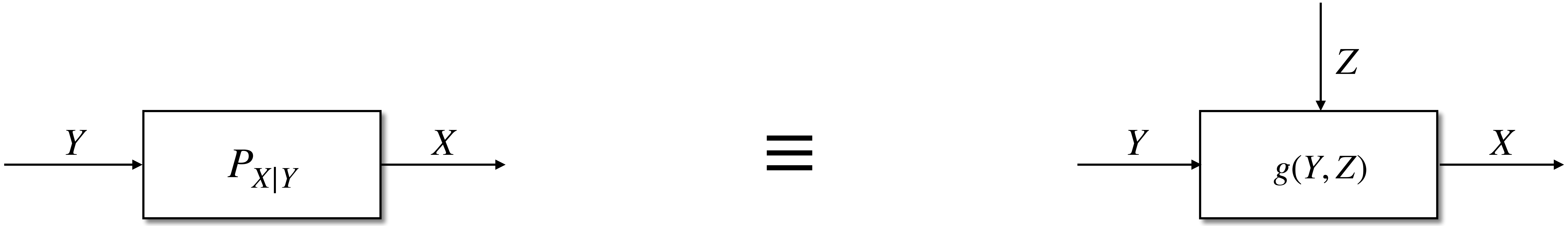
EPFL

EPFL Information Processing Group

Functional Representation Lemma

Given $(X, Y) \sim P_{XY}$, there exists Z $(Z \perp Y)$ and a function $g(\cdot\,,\cdot)$ such that $X = g(Y, Z)$ i.e.,

$$H(X \mid Y, Z) = 0$$
$$I(Y; Z) = 0$$



El Gamal, A., & Kim, Y. (2011). Network Information Theory. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139030687

Minimum Entropy Functional Representations

Given: $(X, Y) \sim P_{XY}$

Find: Z (or $P_{Z|XY}$)

Such that: $Y \perp Z,$
$X = g(Y, Z)$

Minimize: $H_\alpha(Z) \quad (\forall \alpha \geq 0)$

Minimum Entropy Couplings

Given: m marginal distributions i.e.,

$$\{P_1, P_2, \cdots, P_m\}$$

Find: coupling $(X_1, X_2, \cdots, X_m)$

Such that: $X_i \sim P_i$

$\forall i \in \{1,2,\cdots,m\}$

Minimize: $H_\alpha(X_1, X_2, \cdots, X_m)$

$$(\forall \alpha \geq 0)$$

The two problems are ONE!

Minimum Rényi entropy of Z in FRL,

i.e., $H_\alpha(Z)$

$\equiv$

Rényi entropy of the Minimum Entropy Coupling for $\{P_{X|Y=y}\}_{y \in \mathcal{Y}}$

Kocaoglu, M., Dimakis, A., Vishwanath, S., & Hassibi, B. (2017). Entropic Causal Inference. Proceedings of the AAAI Conference on Artificial Intelligence, 31(1).

However……

- Computing $H_\alpha(Z)$ or $H_\alpha(X_1, X_2, \cdots, X_m)$ is NP-hard.

- **Lower bounds** — Converse type results
- **Upper bounds** — Achievability type results

However......

- Computing $H_\alpha(Z)$ or $H_\alpha(X_1, X_2, \cdots, X_m)$ is NP-hard.

- **Lower bounds** — Converse type results
- **Upper bounds** — Achievability type results

- Concerned with Lower bounds here !!

Entropy and Information

Factsheet

Information :

$$\imath_X(x) = \log \frac{1}{P_X(x)}$$

Information spectrum of X :

$$\mathbb{F}_{\imath_X}(t) = \mathbb{P}\left[\imath_X(X) \leq t\right]$$

$\forall t \in [0, \infty)$

Entropy and Information

Factsheet

Information :

$$\imath_X(x) = \log \frac{1}{P_X(x)}$$

Information spectrum of X :

$$\mathbb{F}_{\imath_X}(t) = \mathbb{P}\left[\imath_X(X) \leq t\right]$$

$\forall t \in [0,\infty)$

Shannon entropy :

$$H(X) = \mathbb{E}\left[\imath_X(X)\right]$$

$$= \int_0^\infty \left(1 - \mathbb{F}_{\imath_X}(t)\right) dt$$

Rényi entropy :

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \left(\mathbb{E}\left[2^{(1-\alpha)\imath_X(X)}\right]\right)$$

$\alpha \in [0,1) \cup (1,\infty)$

Majorization ($\preceq_m$)

Factsheet

Definition :

given $Q = (q_1, q_2, q_3, \dots), \quad q_1 \geq q_2 \geq q_3 \dots$

$P = (p_1, p_2, p_3, \dots), \quad p_1 \geq p_2 \geq p_3 \dots$

we say $Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k = 1, 2, \dots$$

Majorization ($\preceq_m$)

Factsheet

Definition :

given $Q = (q_1, q_2, q_3, \dots), \quad q_1 \geq q_2 \geq q_3 \dots$

$P = (p_1, p_2, p_3, \dots), \quad p_1 \geq p_2 \geq p_3 \dots$

we say $Q \preceq_m P$

if $\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k = 1, 2, \dots$

Greatest lower bound w.r.t Majorization :

$$\bigwedge_{i=1}^{m} P_i \preceq_m P_i, \quad i \in \{1, \dots, m\}$$

$$Q \preceq_m P_i \implies Q \preceq_m \bigwedge_{i=1}^{m} P_i$$

Majorization ($\preceq_m$)

Factsheet

Definition :

given $Q = (q_1, q_2, q_3, \ldots), \quad q_1 \geq q_2 \geq q_3 \ldots$

$P = (p_1, p_2, p_3, \ldots), \quad p_1 \geq p_2 \geq p_3 \ldots$

we say $Q \preceq_m P$

if $$\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i, \quad \forall k = 1, 2, \ldots$$

Greatest lower bound w.r.t Majorization :

$$\bigwedge_{i=1}^{m} P_i \preceq_m P_i, \quad i \in \{1, \ldots, m\}$$

$$Q \preceq_m P_i \implies Q \preceq_m \bigwedge_{i=1}^{m} P_i$$

Schur concavity :

$$Q \preceq_m P \implies H_\alpha(Q) \geq H_\alpha(P)$$

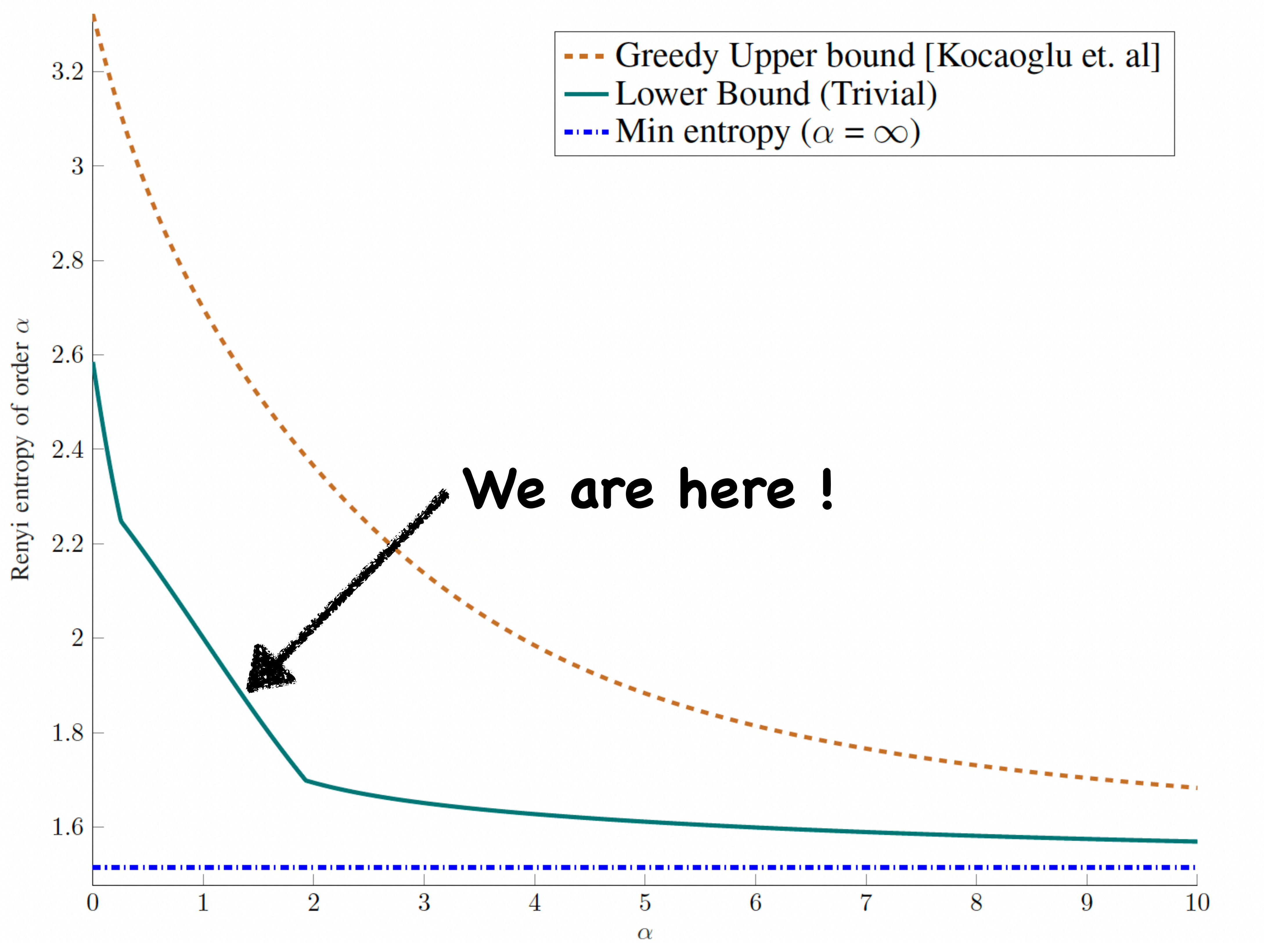
Existing Lower bounds

A very basic one

$$H_\alpha(Z) \geq \max_{y \in \mathcal{Y}} H_\alpha(X \mid Y = y)$$

or

$$H_\alpha(X_1, X_2, \cdots, X_m) \geq \max_{i \in \{1,2,\cdots,m\}} H_\alpha(X_i)$$



Existing Lower bounds

A very basic one

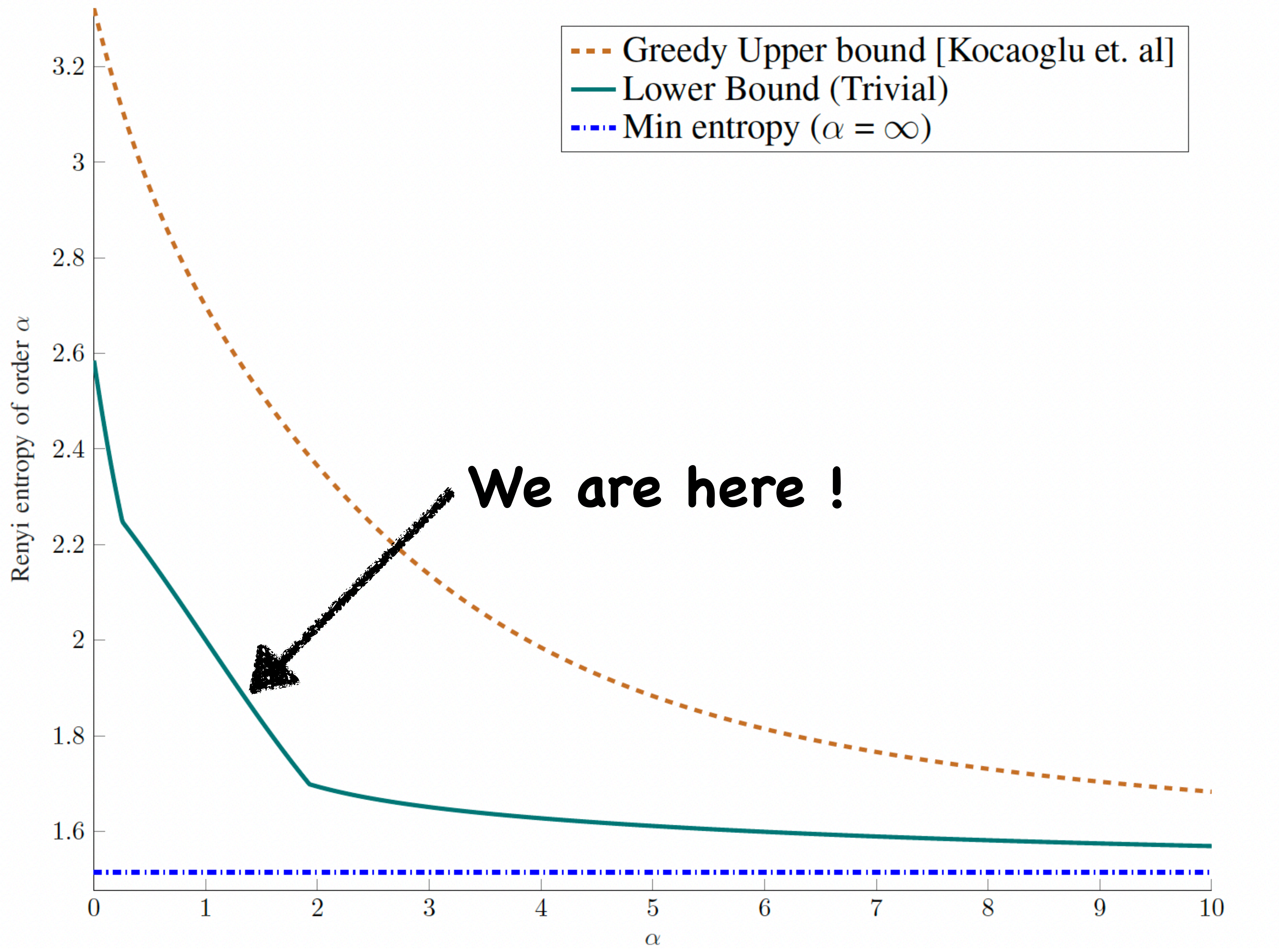
- $P_{\{X_1,\cdots,X_m\}} \sqsubseteq P_{X_i} \; (\forall i \in \{1,\cdots,m\})$
- $P_{\{X_1,\cdots,X_m\}} \preceq_m P_{X_i} \; (\forall i \in \{1,\cdots,m\})$

Equivalently,

$$P_Z \sqsubseteq P_{X|Y=y}$$

and $\quad \forall y \in \mathcal{Y}$

$$P_Z \preceq P_{X|Y=y}$$



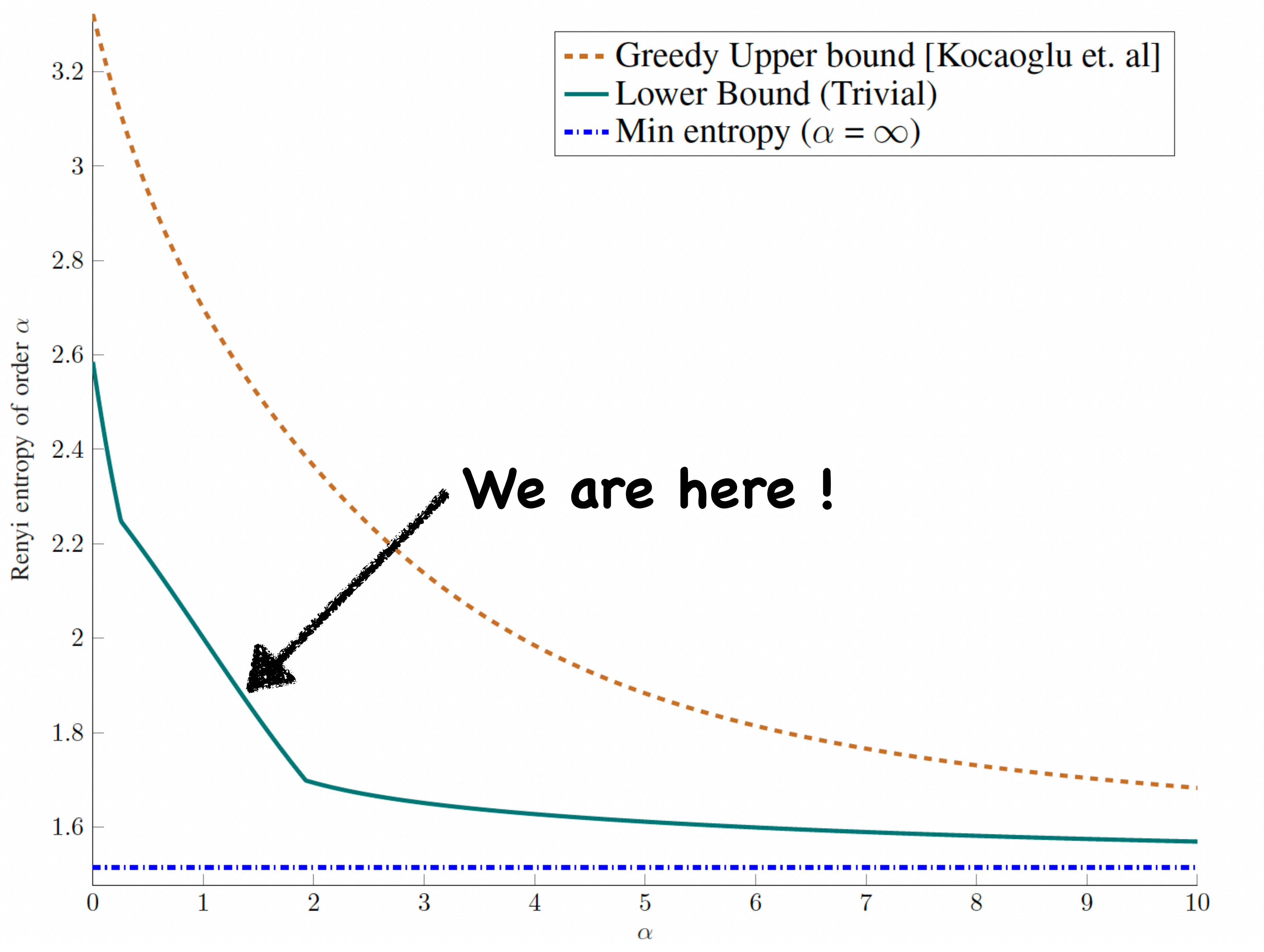
Existing Lower bounds

A very basic one

- $P_{\{X_1,\cdots,X_m\}} \sqsubseteq P_{X_i} \ (\forall i \in \{1,\cdots,m\})$
- $P_{\{X_1,\cdots,X_m\}} \preceq_m P_{X_i} \ (\forall i \in \{1,\cdots,m\})$

$$\Downarrow \text{ [Schur Concavity]}$$

$$H_\alpha(X_1, X_2, \cdots, X_m) \geq \max_{i\in\{1,2,\cdots,m\}} H_\alpha(X_i)$$



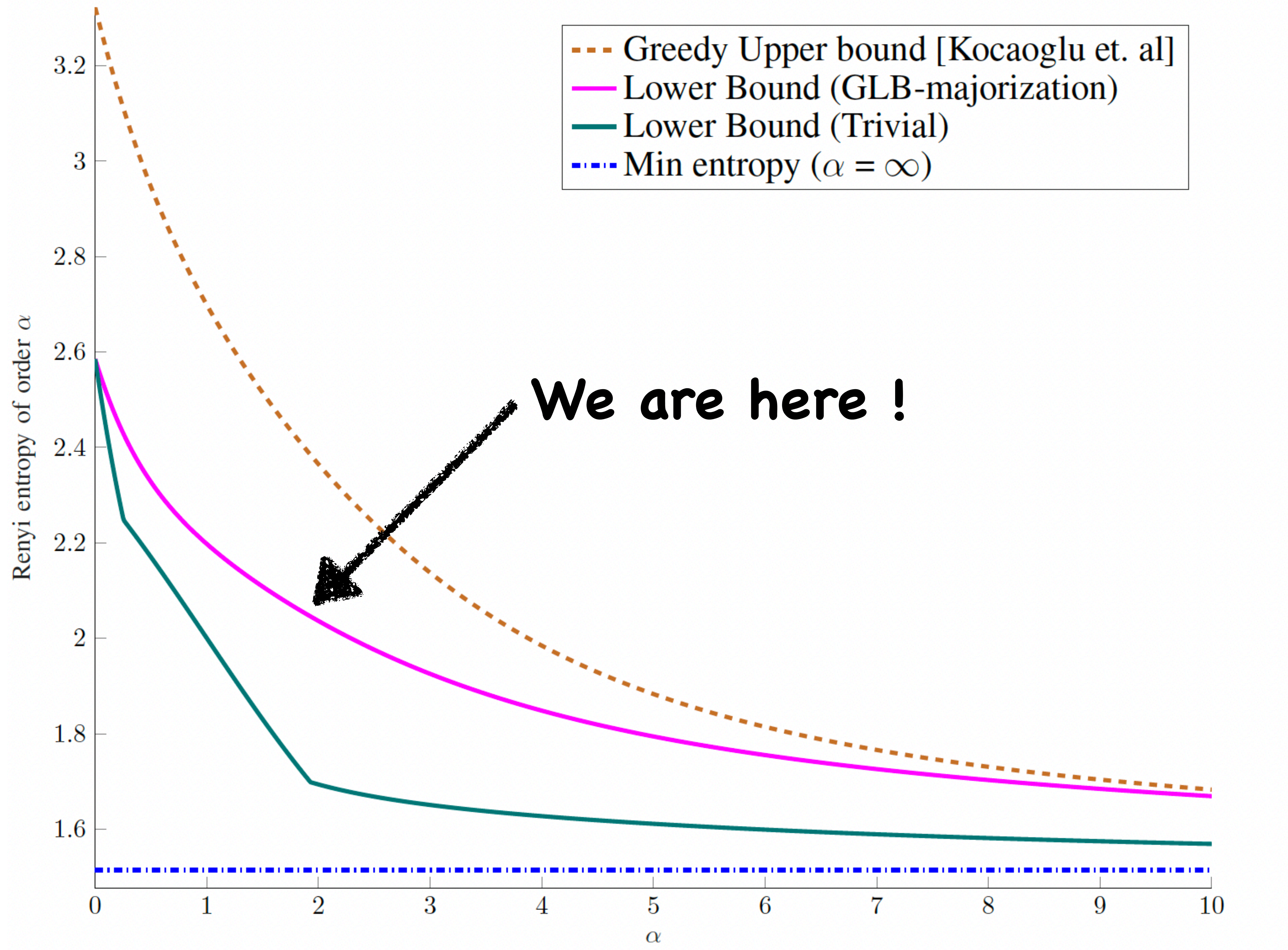
Existing Lower bounds

Based on Majorization ($\preceq_m$)

$$H_\alpha(Z) \geq H_\alpha\Big(\bigwedge_{y \in \mathcal{Y}} P_{X|Y=y}\Big)$$

or

$$H_\alpha(X_1, \cdots, X_m) \geq H_\alpha\Big(\bigwedge_{i=1}^{m} P_i\Big)$$

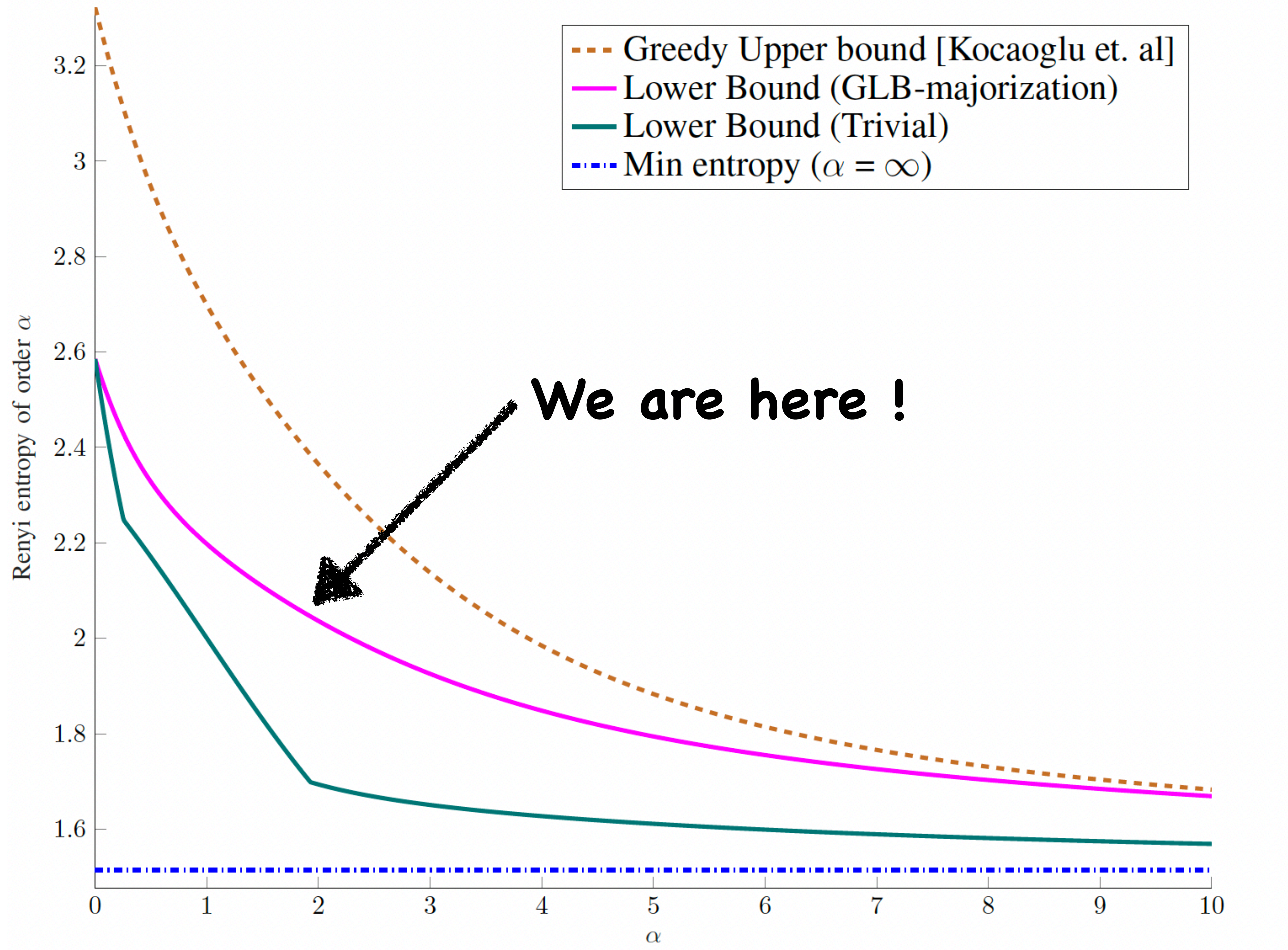


F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Existing Lower bounds

Based on Majorization ($\preceq_m$)

- $P_{\{X_1,\cdots,X_m\}} \preceq_m P_i \ (\forall i \in \{1,\cdots,m\})$
- Recall, $\preceq_m$ is a partial order and complete lattice.

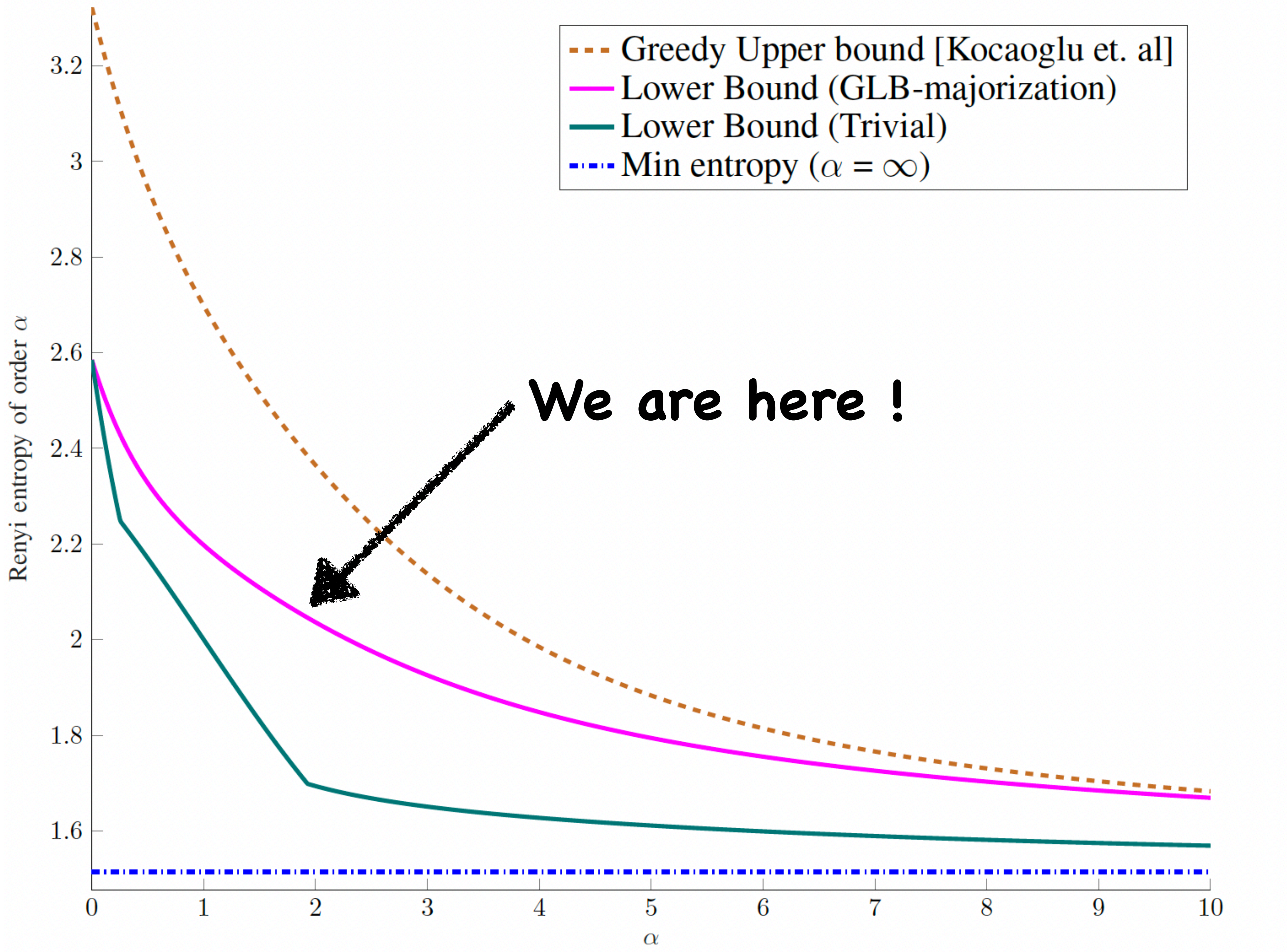


F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Existing Lower bounds

Based on Majorization ($\preceq_m$)

- $P_{\{X_1,\cdots,X_m\}} \preceq_m P_i \ (\forall i \in \{1,\cdots,m\})$
- Recall, $\preceq_m$ is a partial order and complete lattice.
- $P_{\{X_1,\cdots,X_m\}} \preceq_m \left(\bigwedge_{i=1}^{m} P_i\right)$



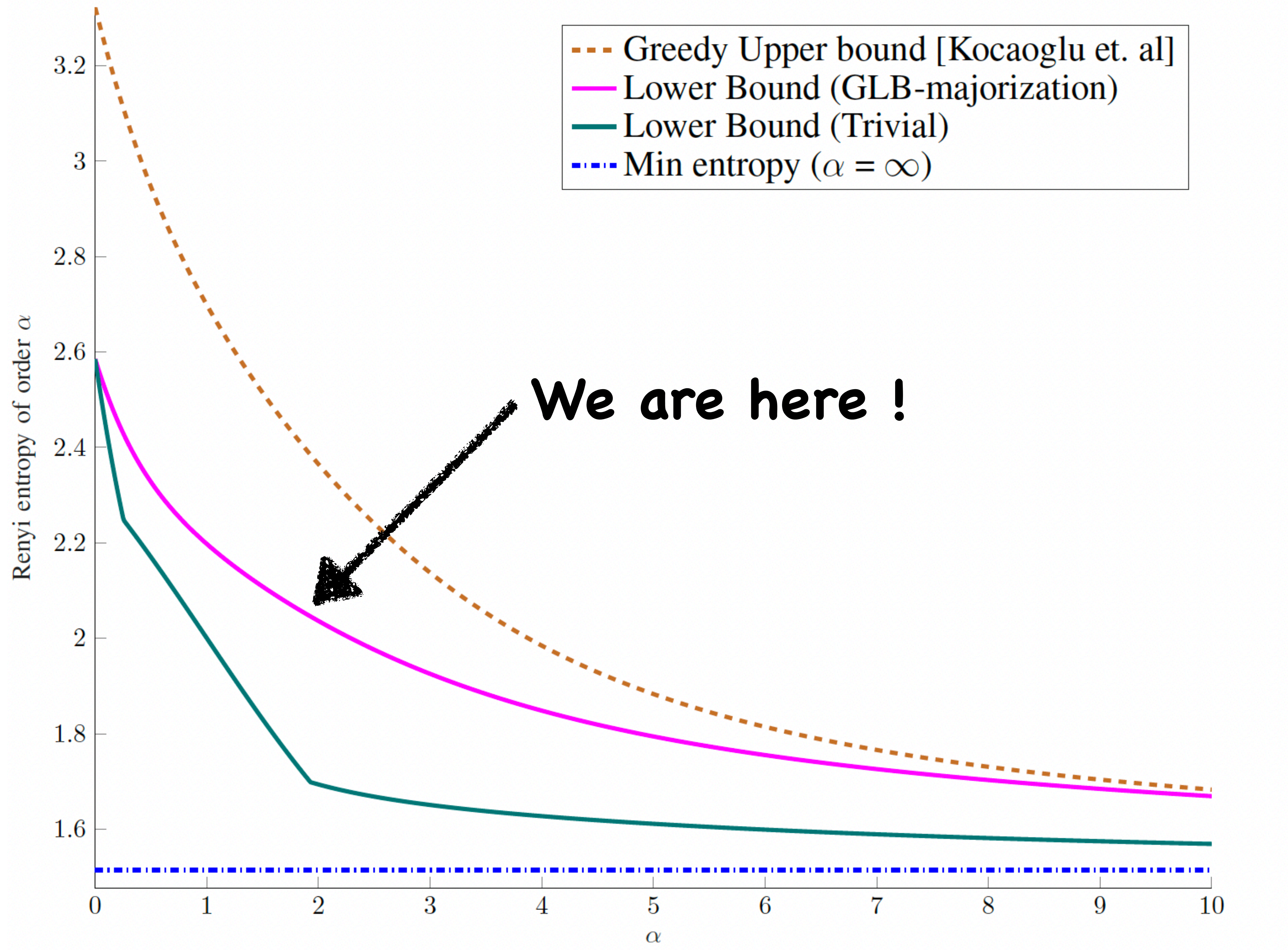
F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Existing Lower bounds

Based on Majorization ($\preceq_m$)

- $P_{\{X_1,\cdots,X_m\}} \preceq_m P_i \ (\forall i \in \{1,\cdots,m\})$
- Recall, $\preceq_m$ is a partial order and complete lattice.
- $P_{\{X_1,\cdots,X_m\}} \preceq_m \left(\bigwedge_{i=1}^{m} P_i\right)$

$$H_\alpha(X_1, \cdots, X_m) \geq H_\alpha\left(\bigwedge_{i=1}^{m} P_i\right)$$



F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," IEEE Transactions on Information Theory, vol. 65, no. 6, pp. 3436–3451, 2019.

Information-spectrum based Lower bound

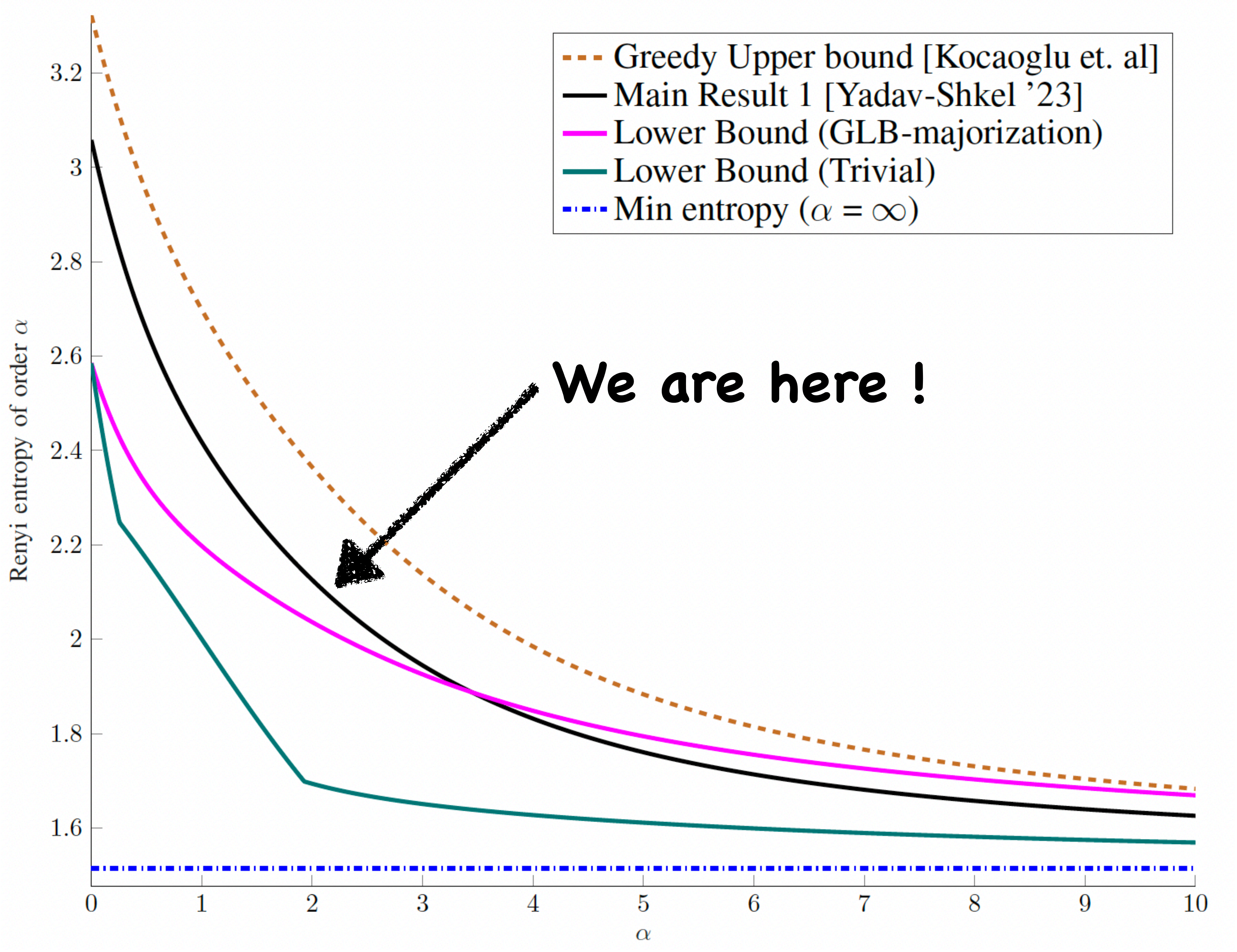
Main Result I

Theorem:

$$\mathbb{P}[\imath_Z(Z) > t] \geq \max_{y \in \mathcal{Y}} \mathbb{P}[\imath_{X|Y}(X|Y) > t \,|\, Y = y]$$

or

$$\mathbb{F}_{\imath_{X|Y=y}}(t) \geq \mathbb{F}_{\imath_Z}(t) \quad \forall y \in \mathcal{Y}$$



Information-spectrum based Lower bound

Main Result I

$$\mathbb{F}_{\imath_{X|Y=y}}(t) \geq \mathbb{F}_{\imath_Z}(t) \quad \forall y \in \mathcal{Y}$$

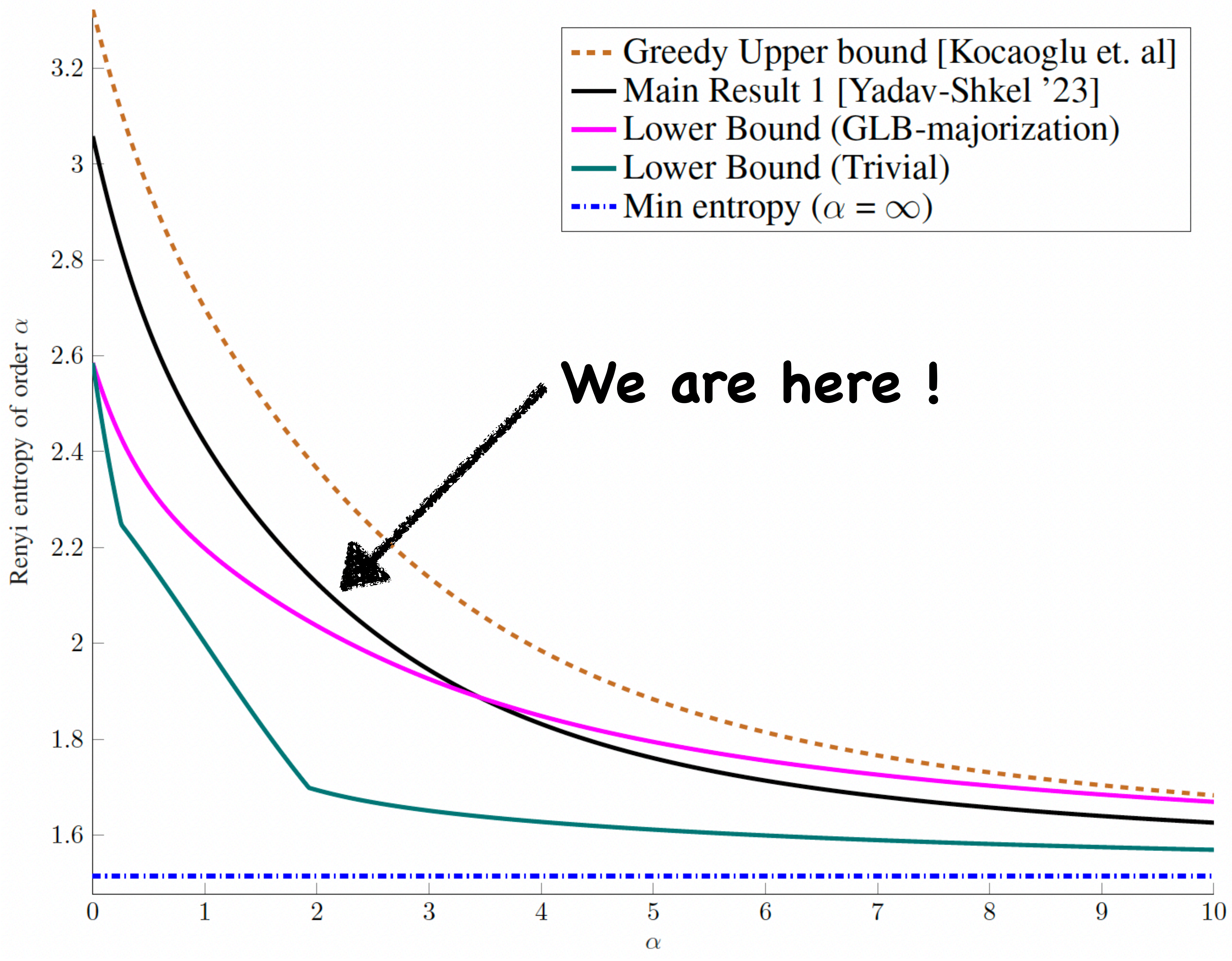
$$\Downarrow$$

$$H(Z) = \mathbb{E}[\imath_Z(Z)]$$

$$= \int_0^\infty \left(1 - \mathbb{F}_{\imath_Z}(t)\right) dt$$

$$\geq \int_0^\infty \max_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(t)\right) dt$$

$$H(Z) \geq \int_0^\infty \max_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(t)\right) dt$$

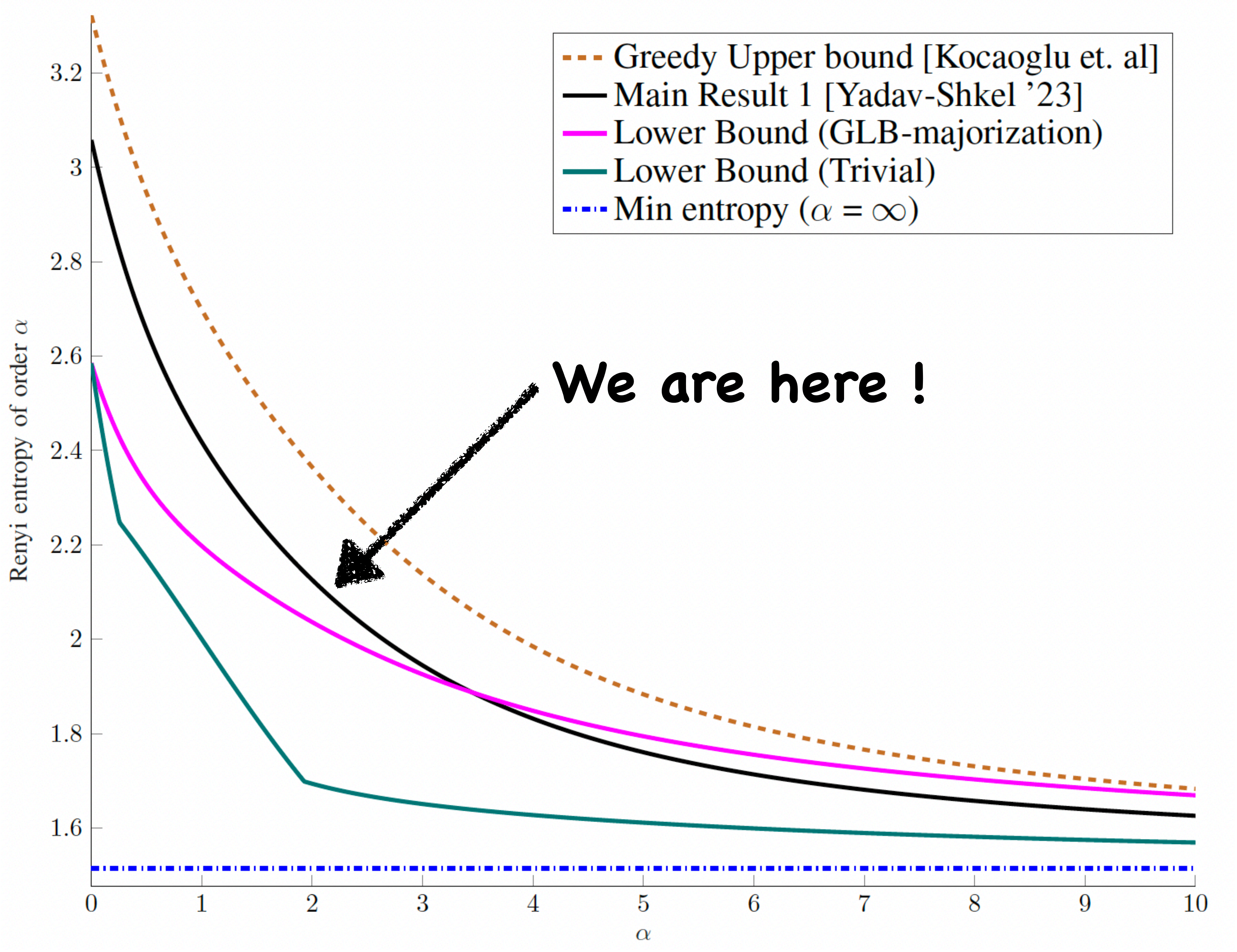


Information- spectrum based Lower bound

Main Result I

$$H(Z) \geq \int_0^{\infty} \max_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(t)\right) dt$$

- Can be similarly extended for Rényi entropy of Z (for any $\alpha \geq 0$)



‘Majorization’

Information-spectrum sense ($\preceq_\imath$)

We say : P majorizes Q in an information-spectrum sense, i.e.,

$$Q \preceq_\imath P$$

if

$$\mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0,\infty)$$

Also introduced in :

C. T. Li, "Infinite Divisibility of Information," in IEEE Transactions on Information Theory, vol. 68, no. 7, pp. 4257-4271, July 2022.

'Majorization'

Information-spectrum sense ($\preceq_\iota$)

We say :

$$Q \preceq_\iota P$$

if $\mathbb{F}_{\iota_Q}(t) \leq \mathbb{F}_{\iota_P}(t), \quad \forall t \in [0,\infty)$

Recall that :

$$Q \preceq_m P$$

if $\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i \quad \forall k \in \{1,\cdots,m\}$

- **Lemma 1:** $Q \preceq_\iota P \implies Q \preceq_m P$

Information-spectrum based Lower bound

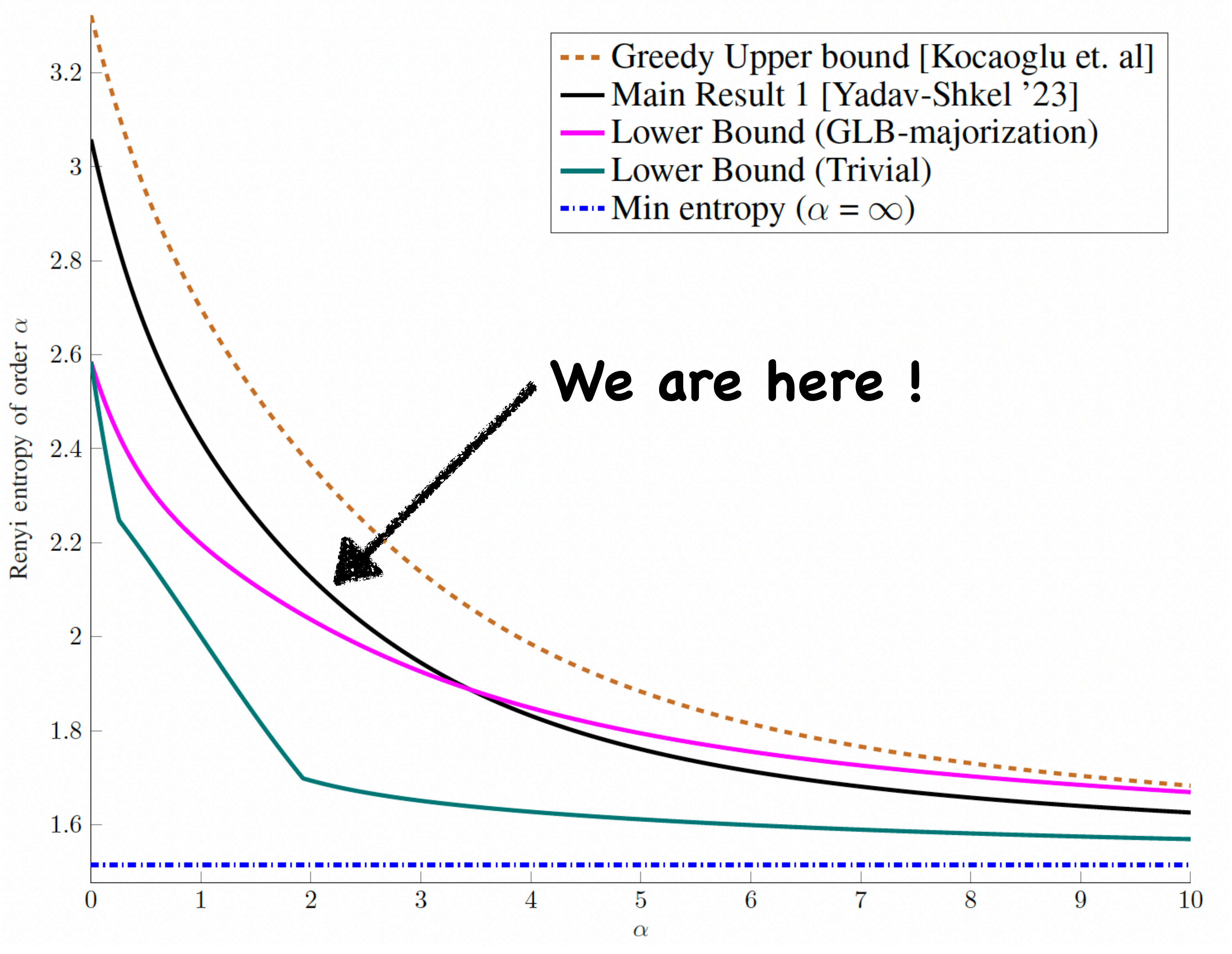
Recall Main Result I

Theorem:

$$\mathbb{P}[\imath_Z(Z) > t] \geq \max_{y \in \mathcal{Y}} \mathbb{P}[\imath_{X|Y}(X|Y) > t \,|\, Y = y]$$

or

$$\mathbb{F}_{\imath_{X|Y=y}}(t) \geq \mathbb{F}_{\imath_Z}(t) \quad \forall y \in \mathcal{Y}$$



Information-spectrum based Lower bound

Recall Main Result I

<u>Theorem:</u>

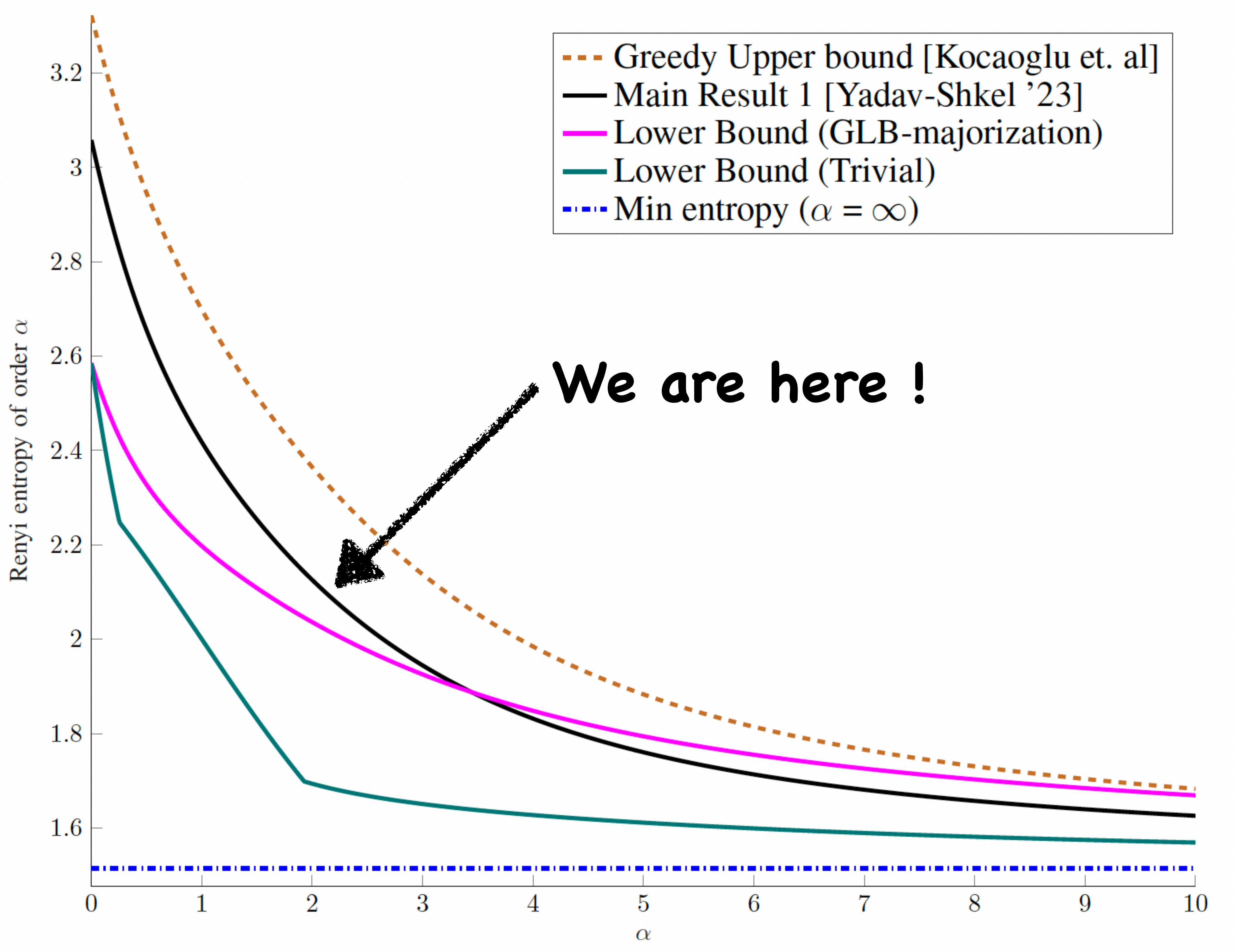
$$\mathbb{P}[\imath_Z(Z) > t] \geq \max_{y \in \mathcal{Y}} \mathbb{P}[\imath_{X|Y}(X|Y) > t \,|\, Y = y]$$

or

$$\mathbb{F}_{\imath_{X|Y=y}}(t) \geq \mathbb{F}_{\imath_Z}(t) \quad \forall y \in \mathcal{Y}$$

$$\Downarrow$$

$$P_Z \preceq_\imath P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$



Information-spectrum based Lower bound

Towards Main Result II

Recall :

$$P_Z \preceq_m P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

$$\Downarrow$$

$$H_\alpha(Z) \geq H_\alpha\Big(\bigwedge_{y\in\mathcal{Y}} P_{X|Y=y}\Big)$$

Similarly :

$$P_Z \preceq_\iota P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

$$\Downarrow \quad ???$$

$$H_\alpha(Z) \geq H_\alpha(Q^*)$$

Greatest lower bound w.r.t information spectrum?

Information-spectrum based Lower bound

Towards Main Result II

Recall :

$$P_Z \preceq_m P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

$$\Downarrow$$

$$H_\alpha(Z) \geq H_\alpha\Big(\bigwedge_{y \in \mathcal{Y}} P_{X|Y=y}\Big)$$

Similarly :

$$P_Z \preceq_\imath P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

$$\Downarrow \quad ???$$

$$H_\alpha(Z) \geq H_\alpha(Q^*)$$

~~Greatest lower bound w.r.t information spectrum?~~

'Majorization'

Information-spectrum sense ($\preceq_\imath$)

We say :

$$Q \preceq_\imath P$$

if $\mathbb{F}_{\imath_Q}(t) \leq \mathbb{F}_{\imath_P}(t), \quad \forall t \in [0,\infty)$

Recall that :

$$Q \preceq_m P$$

if $\sum_{i=1}^{k} q_i \leq \sum_{i=1}^{k} p_i \quad \forall k \in \{1,\cdots,m\}$

- **Lemma 1:** $Q \preceq_\imath P \implies Q \preceq_m P$
- **Lemma 2:** $\mathcal{F} = \{Q \colon Q \preceq_\imath P_i \quad \forall i \in \{1,\ldots,m\}\}$

$$\exists Q^* \in \mathcal{F} \text{ s.t. } Q \preceq_m Q^* \quad \forall Q \in \mathcal{F}$$

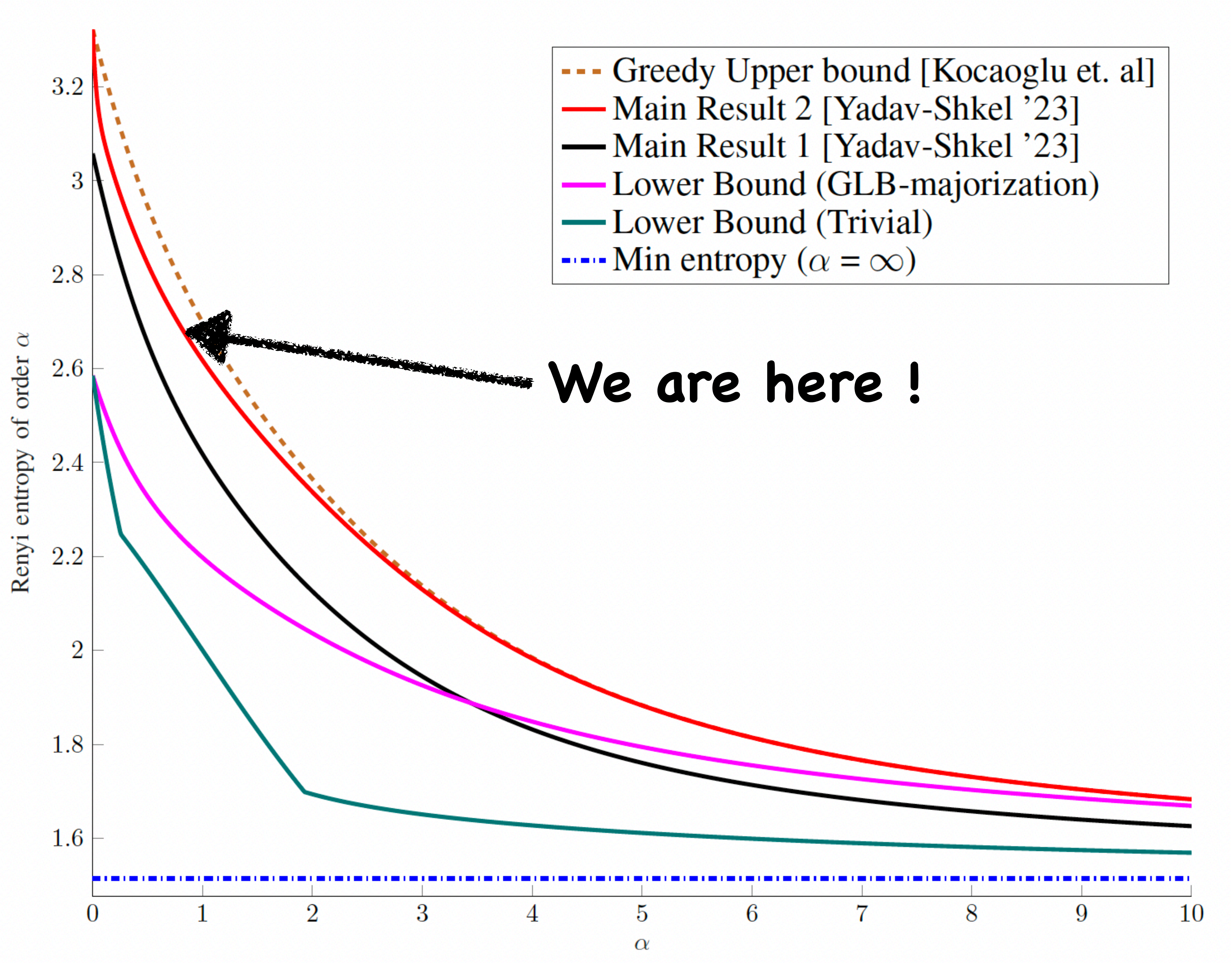
Information-spectrum based Lower bound

Main Result II

$$P_Z \preceq_\imath P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

<u>Define:</u>

- $\mathcal{S} = \{Q : Q \preceq_\imath P_{X|Y=y} \ \forall y \in \mathcal{Y}\}$



Information-spectrum based Lower bound

Main Result II

$$P_Z \preceq_\imath P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

<u>Define:</u>

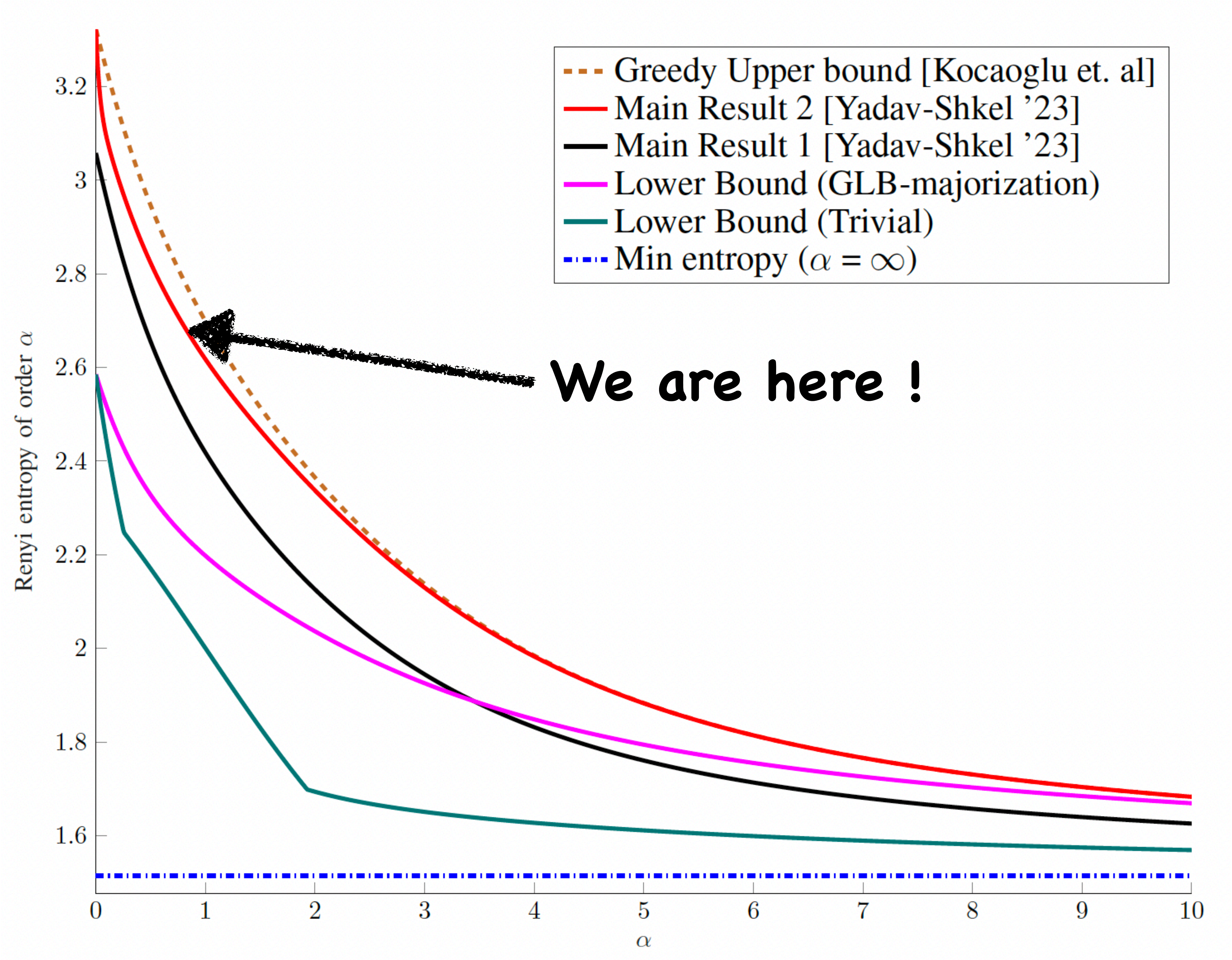
- $\mathcal{S} = \{Q : Q \preceq_\imath P_{X|Y=y} \ \forall y \in \mathcal{Y}\}$

Lemma 2:

$$\mathcal{F} = \{Q : Q \preceq_\imath P_i \quad \forall i \in \{1, \dots, m\}\}$$
$$\exists Q^* \in \mathcal{F} \quad \text{s.t.} \quad Q \preceq_m Q^* \quad \forall Q \in \mathcal{F}$$

- $\exists Q^* \in \mathcal{S}$ s.t.

$$Q \preceq_m Q^* \quad ; \quad \forall Q \in \mathcal{S}$$



Information- spectrum based Lower bound

Main Result II

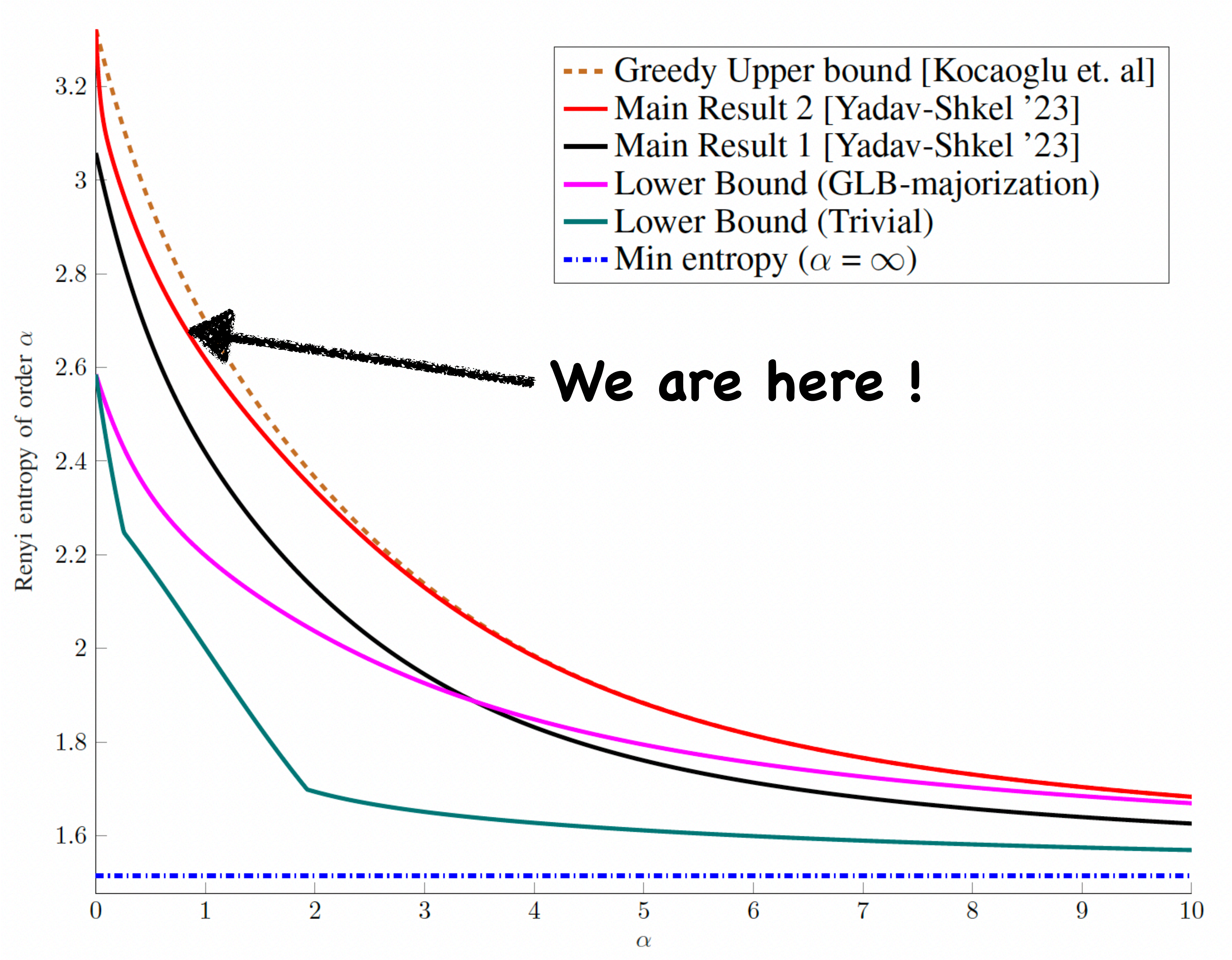
$$P_Z \preceq_\iota P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$

Define:

- $\mathcal{S} = \{Q : Q \preceq_\iota P_{X|Y=y} \; \forall y \in \mathcal{Y}\}$
- $\exists Q^* \in \mathcal{S}$ s.t.

$$Q \preceq_m Q^* \quad ; \quad \forall Q \in \mathcal{S}$$

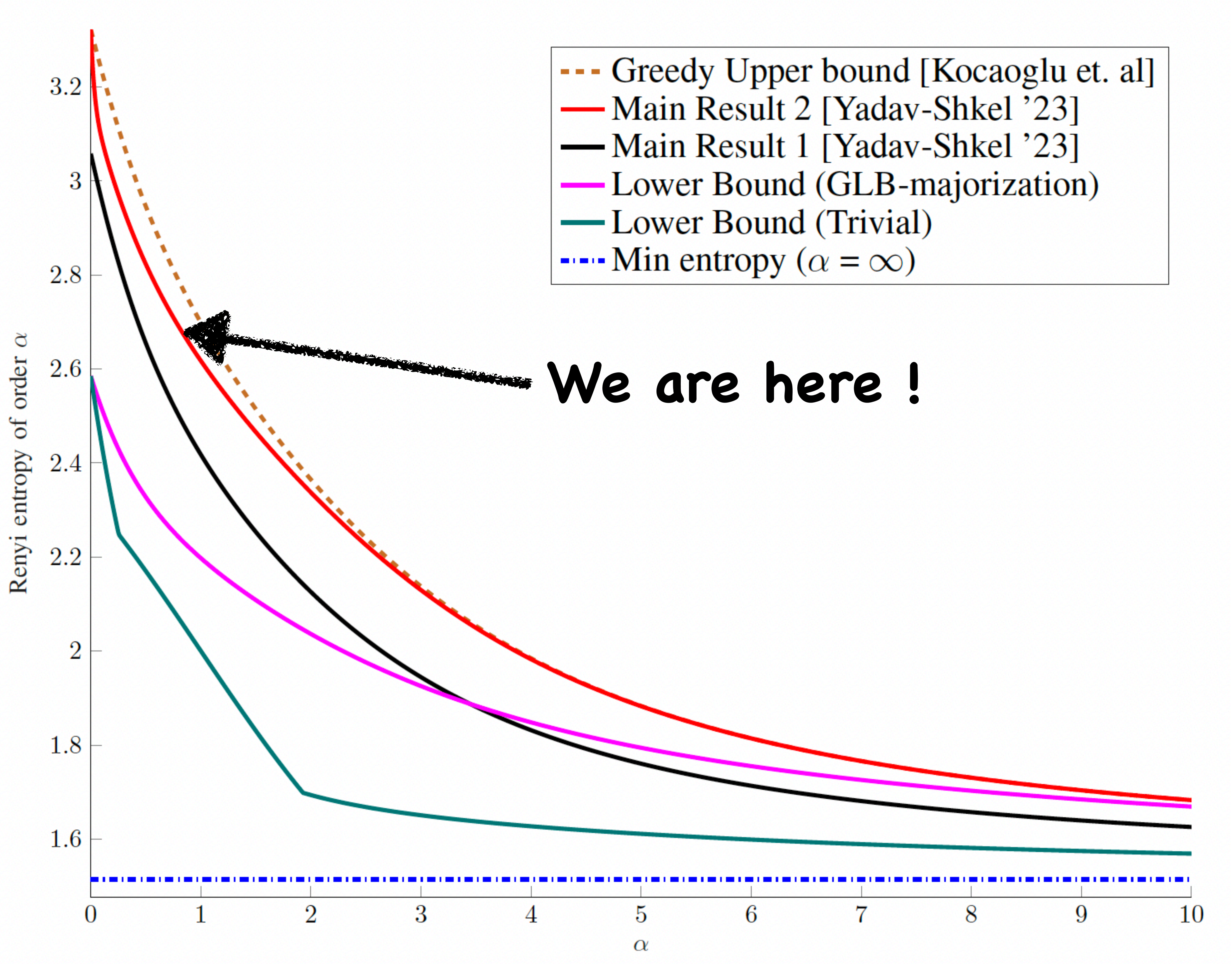
$$\implies Z \preceq_m Q^* \preceq_m P_{X|Y=y}$$



Information- spectrum based Lower bound

Main Result II

$$P_Z \preceq_\imath P_{X|Y=y} \quad \forall y \in \mathcal{Y}$$



Define:

- $\mathcal{S} = \{Q : Q \preceq_\imath P_{X|Y=y} \ \forall y \in \mathcal{Y}\}$
- $\exists Q^* \in \mathcal{S}$ s.t. $Q \preceq_m Q^*$; $\forall Q \in \mathcal{S}$

$$\implies Z \preceq_m Q^* \preceq_m P_{X|Y=y}$$

$$H_\alpha(Z) \geq H_\alpha(Q^*)$$

- support size of the lower bounding distribution is enlarged

Numerical Example

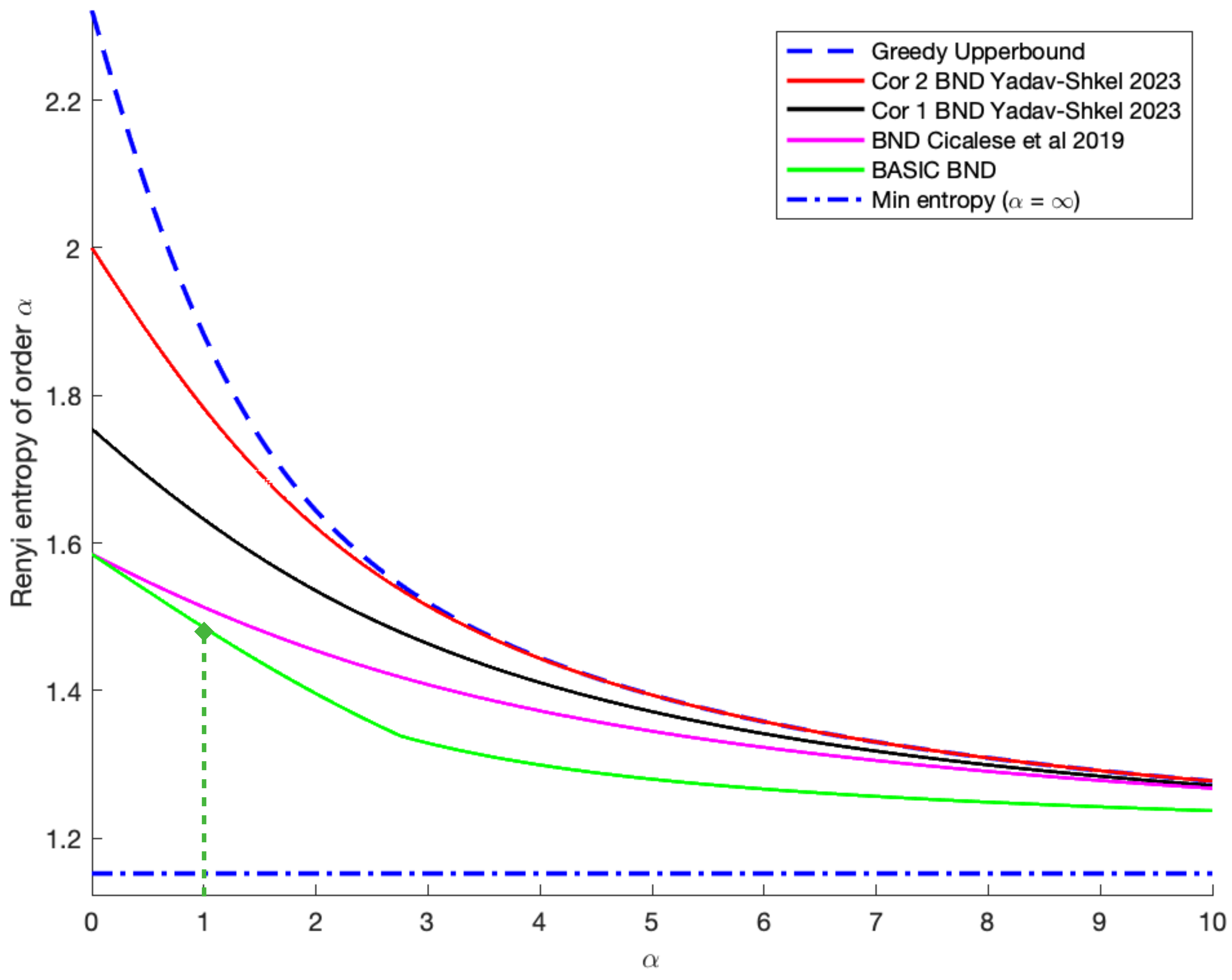
Lower bound 1

$$P_{X|Y}(\cdot\,|\,y_1) = (0.45, 0.4, 0.15)$$

$$P_{X|Y}(\cdot\,|\,y_2) = (0.5, 0.3, 0.2)$$

$$H_\alpha(Z) \geq \max_{y \in \mathcal{Y}} H_\alpha(X\,|\,Y = y)$$

$$H(Z) \geq \max\{1.4577, 1.4855\}$$
$$= 1.4855$$



Numerical Example

Lower bound 2

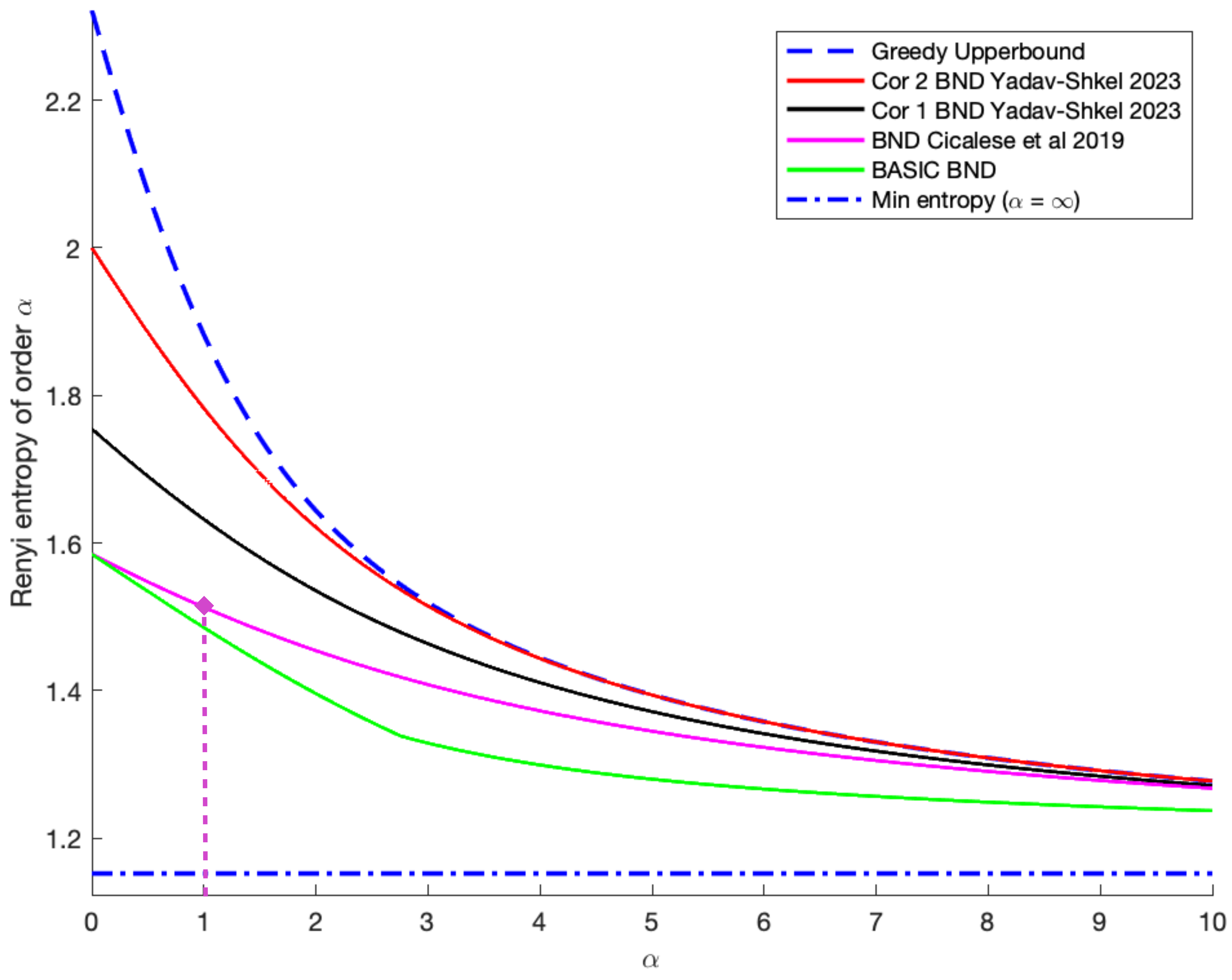
$P_{X|Y}(\cdot \mid y_1) = (0.45, 0.4, 0.15)$

$P_{X|Y}(\cdot \mid y_2) = (0.5, 0.3, 0.2)$

$$H_\alpha(Z) \geq H_\alpha\Big(\bigwedge_{y \in \mathcal{Y}} P_{X|Y=y}\Big)$$

$$\bigwedge_{y \in \mathcal{Y}} P_{X|Y=y} = (0.45, 0.35, 0.2)$$

$H(Z) \geq 1.5129$



Numerical Example

Lower bound 3 - Main Result I

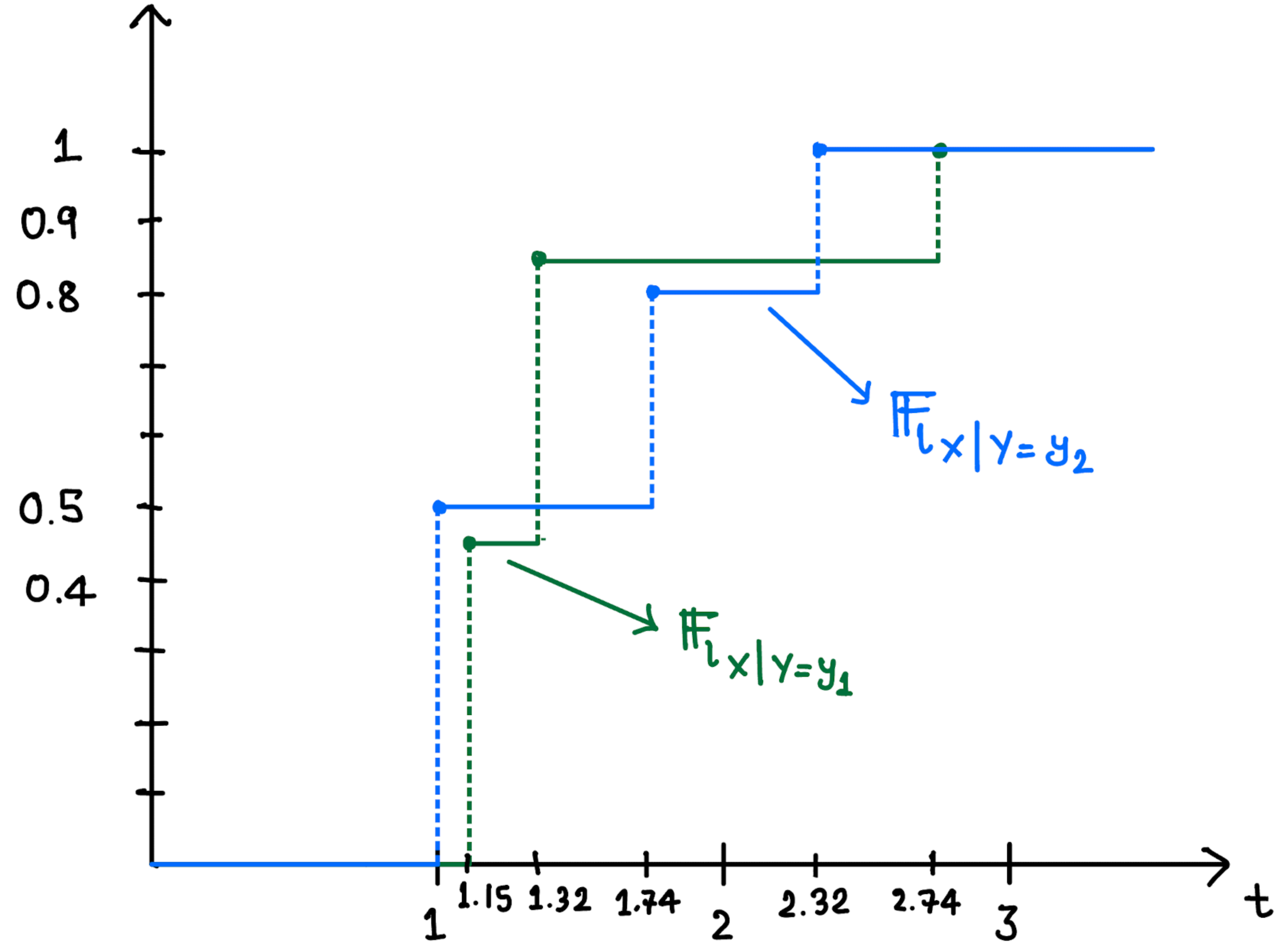
$$P_{X|Y}(\cdot \mid y_1) = (0.45, 0.4, 0.15)$$

$$P_{X|Y}(\cdot \mid y_2) = (0.5, 0.3, 0.2)$$

$$H(Z) \geq \int_0^\infty \max_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(t)\right) dt$$

$$\imath_{X|Y}(\cdot \mid y_1) \in \{1.15, 1.32, 2.74\}$$

$$\imath_{X|Y}(\cdot \mid y_2) \in \{1, 1.74, 2.32\}$$



Numerical Example

Lower bound 4 - Main Result II

$$P_{X|Y}(\cdot\,|\,y_1) = (0.45, 0.4, 0.15)$$

$$P_{X|Y}(\cdot\,|\,y_2) = (0.5, 0.3, 0.2)$$

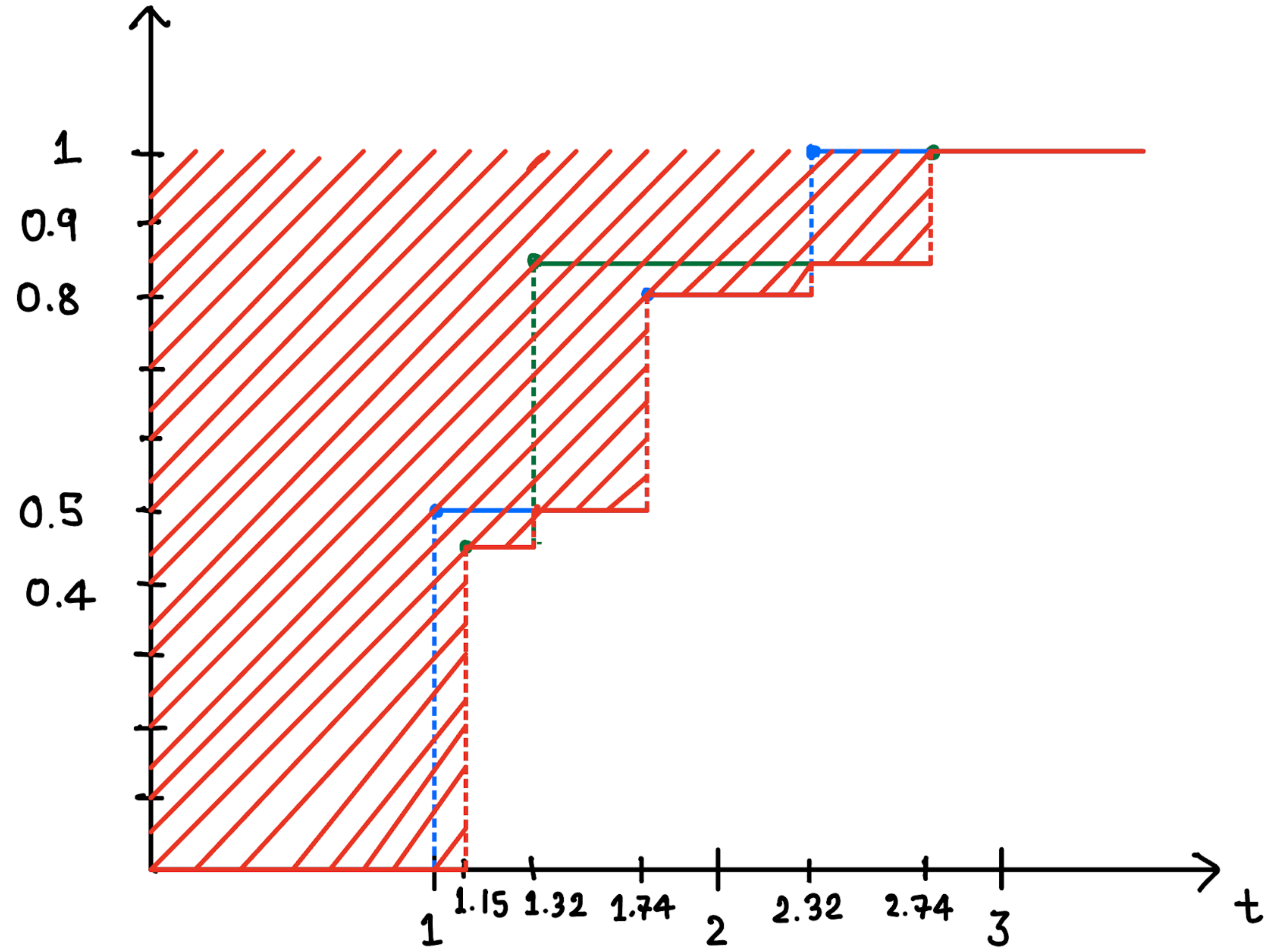
$$H(Z) \geq \int_0^\infty \max_{y \in \mathcal{Y}} \left(1 - \mathbb{F}_{\imath_{X|Y=y}}(t)\right) dt$$

$$\imath_{X|Y}(\cdot\,|\,y_1) \in \{1.15, 1.32, 2.74\}$$

$$\imath_{X|Y}(\cdot\,|\,y_2) \in \{1, 1.74, 2.32\}$$

$H(Z) \geq$ **Shaded Area**

$= 1.6325$



Numerical Example

Lower bound 4 - Main Result II

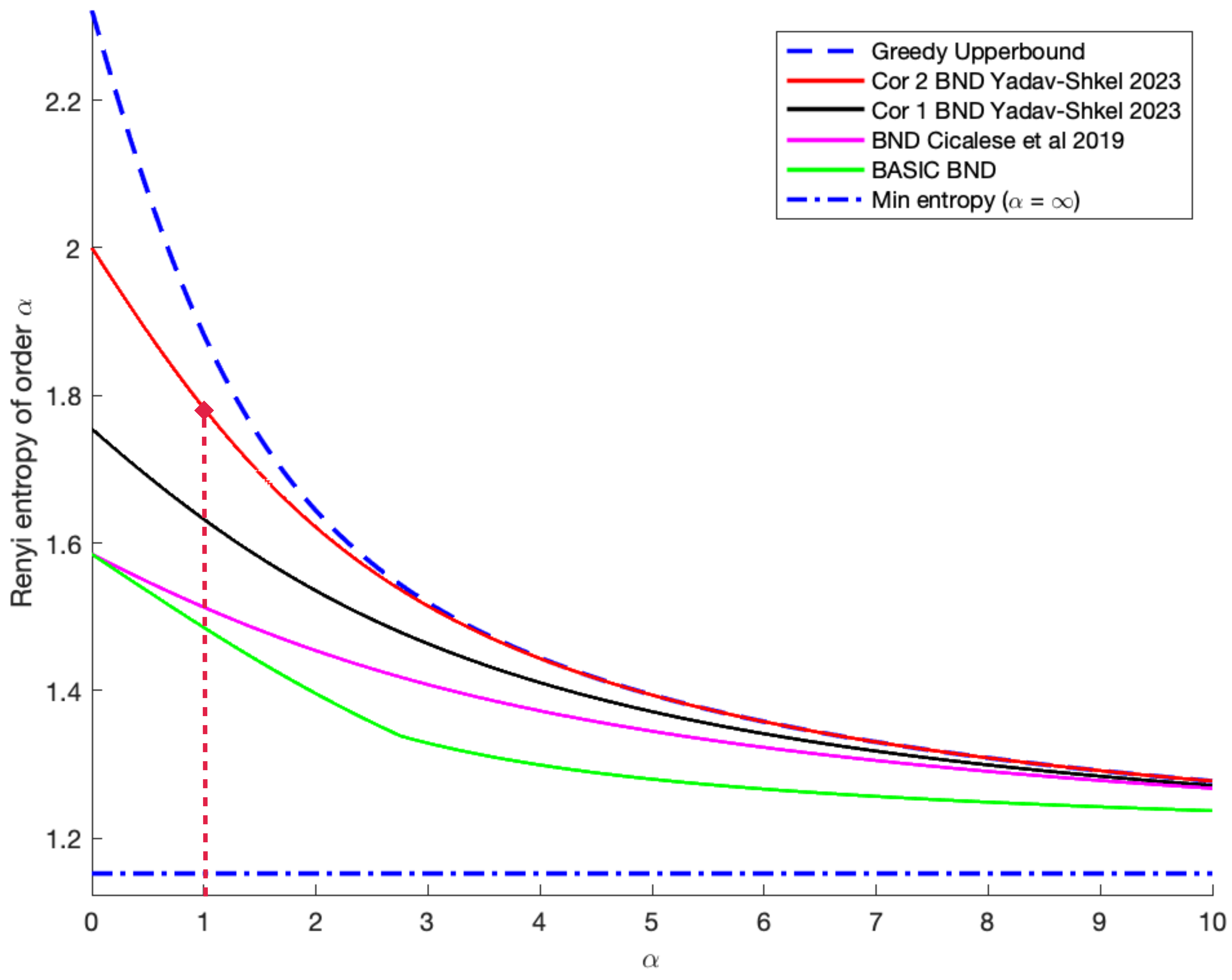
$P_{X|Y}(\cdot \mid y_1) = (0.45, 0.4, 0.15)$

$P_{X|Y}(\cdot \mid y_2) = (0.5, 0.3, 0.2)$

$$H_\alpha(Z) \geq H_\alpha(Q^*)$$

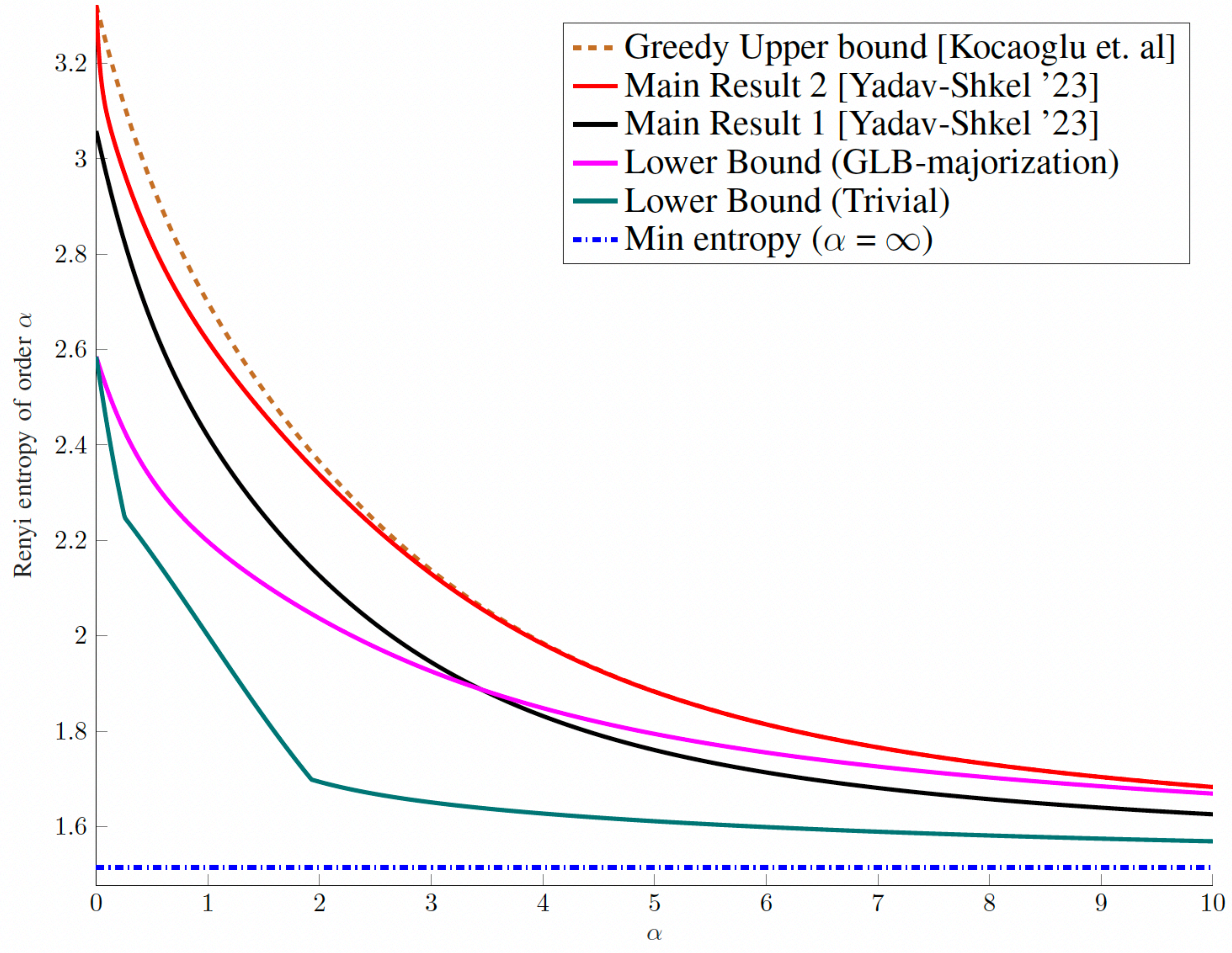
$$Q^* = (0.45, 0.3, 0.15, 0.1)$$

$$H(Z) \geq H(Q^*) = 1.7822$$



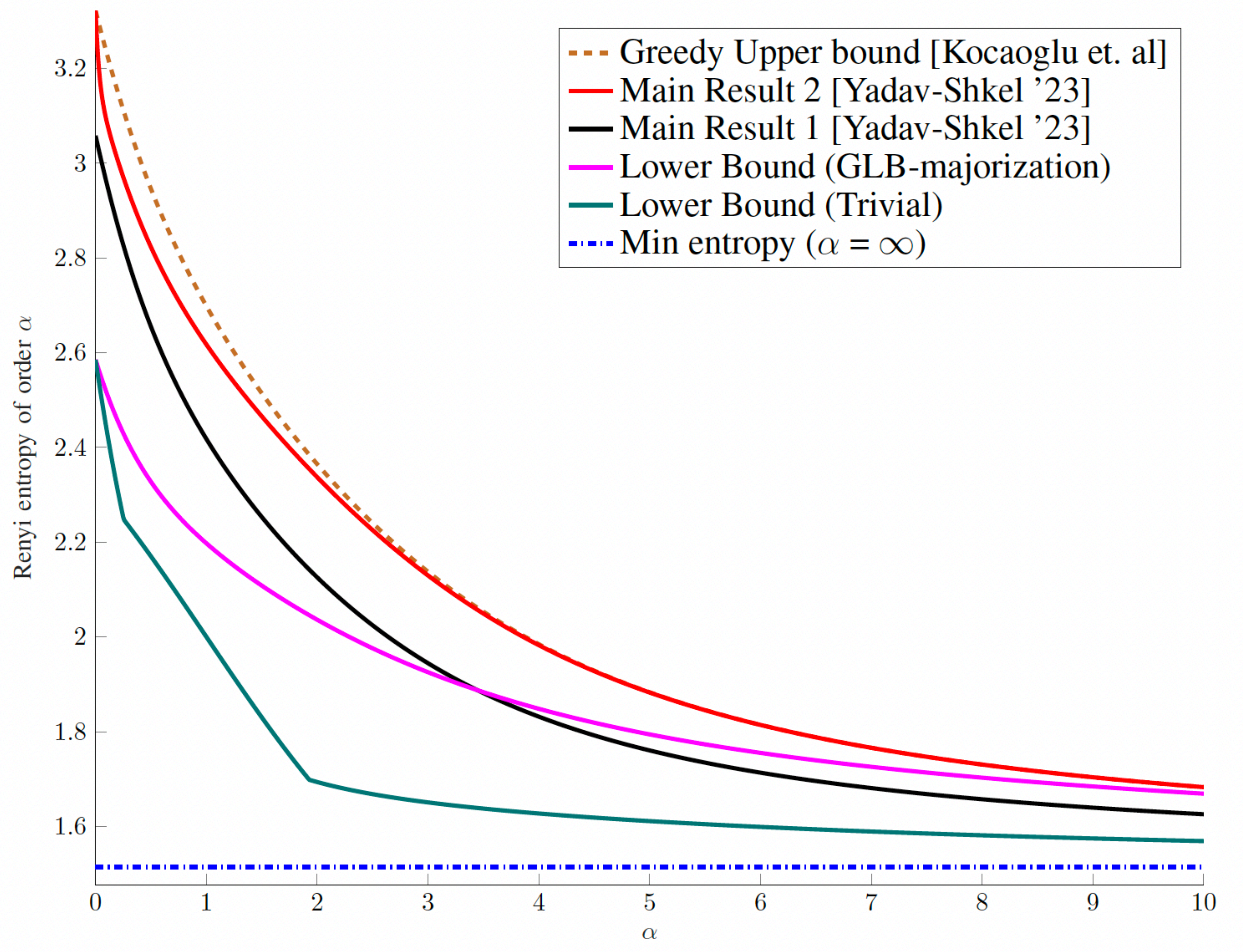
Concluding Remarks

- Converse results for two problems :
  - Functional representations
  - Minimum entropy couplings
- Tools from majorization and information spectrum.
- Improves on all the existing lower bounds.



Concluding Remarks

- Converse results for two problems :
  - Functional representations
  - Minimum entropy couplings
- Tools from majorization and information spectrum.
- Improves on all the existing lower bounds.



Concurrent Work :

Spencer Compton, Dmitriy Katz, Benjamin Qi, Kristjan Greenewald, Murat Kocaoglu, " **Minimum-Entropy Coupling Approximation Guarantees Beyond the Majorization Barrier** ", Proceedings of The 26th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 206:10445-10469, 2023.