Hypercontractivity parameters using information measures

Chandra Nair, Yan Nan Wang

The Chinese University of Hong Kong

Jul 11, 2016

Definitions and equivalent characterizations

$(X, Y) \sim \mu_{XY}$ is said to be (λ_1, λ_2)-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \le \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \ge 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \ge 1$. (Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}$.)

Equivalent characterizations:

- A simple exercise

$$\|\mathrm{E}(g(Y)|X)\|_{\lambda_1'} \le \|g(Y)\|_{\lambda_2} \quad \forall g(Y). \quad \frac{1}{\lambda_1'} = 1 - \frac{1}{\lambda_1}.$$

DEFINITIONS AND EQUIVALENT CHARACTERIZATIONS

$(X,Y) \sim \mu_{XY}$ is said to be (λ_1, λ_2)-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \le \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \ge 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \ge 1$. (Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}$.)

Equivalent characterizations:

1. A simple exercise

$$\|\mathrm{E}(g(Y)|X)\|_{\lambda_1'} \le \|g(Y)\|_{\lambda_2} \quad \forall\, g(Y). \quad \frac{1}{\lambda_1'} = 1 - \frac{1}{\lambda_1}.$$

2. Using relative entropies (Carlen – Cordero-Erasquin ’09, N ’14, Friedgut ’15)

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y) \le D(\nu_{XY} \| \mu_{XY}) \quad \forall \nu_{XY} \ll \mu_{XY}.$$

3. Using mutual information and auxiliary variables (N ’14)

$$\frac{1}{\lambda_1} I(U;X) + \frac{1}{\lambda_2} I(U;Y) \le I(U;XY) \quad \forall \mu_{U|XY}.$$

DEFINITIONS AND EQUIVALENT CHARACTERIZATIONS

$(X, Y) \sim \mu_{XY}$ is said to be (λ_1, λ_2)-hypercontractive if

$$\mathrm{E}(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2} \quad \forall f(X), g(Y).$$

Interested in $\lambda_1, \lambda_2 \geq 1$ and $\frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$. (Notation: $\|Z\|_\lambda = \mathrm{E}(|Z|^\lambda)^{1/\lambda}$.)

Equivalent characterizations:

1. A simple exercise

$$\|\mathrm{E}(g(Y)|X)\|_{\lambda_1'} \leq \|g(Y)\|_{\lambda_2} \quad \forall g(Y). \quad \frac{1}{\lambda_1'} = 1 - \frac{1}{\lambda_1}.$$

2. Using relative entropies (Carlen – Cordero-Erasquin '09, N '14, Friedgut '15)

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y) \leq D(\nu_{XY} \| \mu_{XY}) \quad \forall \nu_{XY} \ll \mu_{XY}.$$

3. Using mutual information and auxiliary variables (N'14)

$$\frac{1}{\lambda_1} I(U; X) + \frac{1}{\lambda_2} I(U; Y) \leq I(U; XY) \quad \forall \mu_{U|XY}.$$

4. Using concave envelopes of a functional (N'14)

$$\mathfrak{C}\left[H(XY) - \frac{1}{\lambda_1}H(X) - \frac{1}{\lambda_2}H(Y)\right]_{\mu_{XY}} = H_\mu(XY) - \frac{1}{\lambda_1}H_\mu(X) - \frac{1}{\lambda_2}H_\mu(Y).$$

Question 1: Why is hypercontractivity useful for information theory?

- Hungarian school ('73-): interested in characterizing "image sizes" over noisy channels
    - An input set $B \subseteq \mathcal{X}^n$
    - Image: union of typical outputs of each point in B when passed through a noisy channel
- Wanted to understand feasible communication rates
- Strong converses to coding theorems

4. Using concave envelopes of a functional (N'14)

$$\mathfrak{C}\left[H(XY) - \frac{1}{\lambda_1}H(X) - \frac{1}{\lambda_2}H(Y)\right]_{\mu_{XY}} = H_\mu(XY) - \frac{1}{\lambda_1}H_\mu(X) - \frac{1}{\lambda_2}H_\mu(Y).$$

Question 1: Why is hypercontractivity useful (in information theory)?

- Hungarian school ('73-): interested in characterizing "image sizes" over noisy channels
  - An input set $B \subseteq \mathcal{X}^n$
  - Image: union of typical outputs of each point in B when passed through a noisy channel
- Wanted to understand feasible communication rates
- Strong converses to coding theorems

EQUIVALENT CHARACTERIZATIONS: CTD ...

4. Using concave envelopes of a functional (N'14)

$$\mathfrak{C}\left[H(XY) - \frac{1}{\lambda_1}H(X) - \frac{1}{\lambda_2}H(Y)\right]_{\mu_{XY}} = H_\mu(XY) - \frac{1}{\lambda_1}H_\mu(X) - \frac{1}{\lambda_2}H_\mu(Y).$$

Question 1: Why is hypercontractivity useful (in information theory)?

- Hungarian school ('73-): interested in characterizing "image sizes" over noisy channels
  - An input set $\mathcal{B} \subseteq \mathcal{X}^n$
  - Image: union of typical outputs of each point in $\mathcal{B}$ when passed through a noisy channel
- Wanted to understand feasible communication rates
- Strong converses to coding theorems

4. Using concave envelopes of a functional (N'14)

$$\mathfrak{C}\left[H(XY) - \frac{1}{\lambda_1}H(X) - \frac{1}{\lambda_2}H(Y)\right]_{\mu_{XY}} = H_\mu(XY) - \frac{1}{\lambda_1}H_\mu(X) - \frac{1}{\lambda_2}H_\mu(Y).$$

Question 1: Why is hypercontractivity useful (in information theory)?

- Hungarian school ('73-): interested in characterizing "image sizes" over noisy channels
  - An input set $\mathcal{B} \subseteq \mathcal{X}^n$
  - Image: union of typical outputs of each point in $\mathcal{B}$ when passed through a noisy channel
- Wanted to understand feasible communication rates
- Strong converses to coding theorems

Some relevant publications:

- Rudolf Ahlswede and Peter Gács. "Spreading of sets in product spaces and hypercontraction of the Markov operator". In: The Annals of Probability (1976), pp. 925–939
- J Körner and K Marton. "Images of a set via two channels and their role in Mult-User Communication". In: IEEE Trans. Info. Theory IT-23 (Nov, 1977), pp. 751–761

Question 2: Why are these equivalent characterizations useful?

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality. (Concave envelope characterization)
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath-N '15) (mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization) (this talk).

Hypercontractivity and Information Theory

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality.
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath N '15) (mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization (this talk)).

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
    - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result
    - Establish a stronger extremal inequality.
- Establish the strong data-processing inequality between sums of i.i.d. random variables
    - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality. (Concave envelope characterization)
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath-N '15)(mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization) (this talk).

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality. (Concave envelope characterization)
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath-N '15) (mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization) (this talk).

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality. (Concave envelope characterization)
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath-N '15) (mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization) (this talk).

Question 2: Why are these equivalent characterizations useful?

Evaluation of hypercontractivity parameters has been very useful

- Measure concentration, quantum physics, theoretical computer science
- Ties up evaluation of hypercontractivity to computation of regions in multiuser information theory (auxiliary variables).
  - Borrowing and lending of techniques for *evaluation* and *single-letterization*
- Method for proving optimality of Gaussians to recover Gross's result (N '14)
  - Establish a stronger extremal inequality. (Concave envelope characterization)
- Establish the strong data-processing inequality between sums of i.i.d. random variables (Kamath-N '15) (mutual information characterization)
  - Implies (and strengthens) a result by Dembo et. al. on maximal correlations
- Establish hypercontractivity for the binary erasure channel (relative entropy characterization) (this talk).

Hypercontractivity region for the binary symmetric channel

- Bonami's inequality (famous result)
- This paper: techniques recover yet another proof of this result.

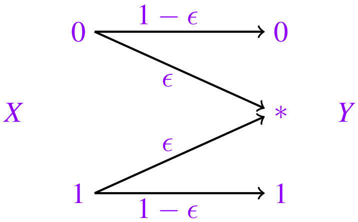

Figure: Binary erasure channel

Goal: Determine $\lambda_1, \lambda_2 \geq 1, \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$ such that

$$E(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2}.$$

Binary erasure channel with uniform inputs

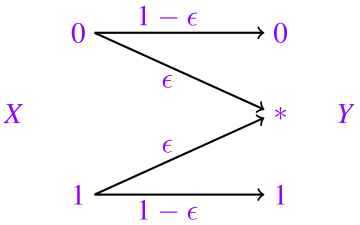

Figure: Binary erasure channel

Goal: Determine $\lambda_1, \lambda_2 \geq 1, \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$ such that

$$E(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2}.$$

Correlation lower bound ['70s] (Necessary condition)

$(\lambda_1 - 1)(\lambda_2 - 1) \geq \rho^2(X;Y); \quad \rho(X,Y) = \sup_{\substack{f,g:\mathrm{E}(f)=\mathrm{E}(g)=0 \\ \mathrm{E}(f^2)=\mathrm{E}(g^2)=1}} \mathrm{E}(f(X)g(Y)).$

Binary erasure channel with uniform inputs

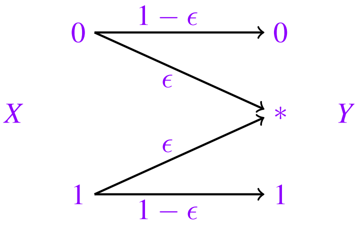

Figure: Binary erasure channel

Goal: Determine $\lambda_1, \lambda_2 \geq 1, \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$ such that

$$E(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2}.$$

Correlation lower bound ['70s] (Necessary condition) - BEC
$(\lambda_1 - 1)(\lambda_2 - 1) \geq 1 - \epsilon.$

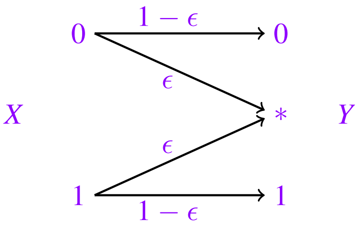

Figure: Binary erasure channel

Goal: Determine $\lambda_1, \lambda_2 \geq 1, \frac{1}{\lambda_1} + \frac{1}{\lambda_2} \geq 1$ such that

$$E(f(X)g(Y)) \leq \|f(X)\|_{\lambda_1} \|g(Y)\|_{\lambda_2}.$$

Correlation lower bound ['70s] (Necessary condition) - BSC
$(\lambda_1 - 1)(\lambda_2 - 1) \geq (1 - 2p)^2.$
Sufficiency: Bonami's inequality

MAIN RESULT

Question (to me) by V. Guruswami and J. Radhakrishnan:

- Is correlation bound tight for BEC?

Theorem (Tightness of correlation lower bound)

For BEC the correlation bound is tight, i.e. (X, Y) is (λ_1, λ_2)-hypercontractive for $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$, if and only if the following condition is satisfied:

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1).$$

Remarks:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all ϵ if $\lambda_2 \geq \frac{4}{3}$.

Idea of the proof:

- $\lambda_2 \geq 2$: Mimic Janson's proof technique for BSC (i.e. Bonami-Beckner)
- Remainder of regime
  - Use the relative entropy characterization
  - Inspired by Friedgut's proof for BSC for case $\lambda_1 = \lambda_2 = \rho$.

Main result

Question (to me) by V. Guruswami and J. Radhakrishnan:

- Is correlation bound tight for BEC?

Theorem (Tightness of correlation lower bound)

For BEC the correlation bound is tight, i.e. (X, Y) *is* (λ_1, λ_2)*-hypercontractive for* $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$, *if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1).$$

Remarks:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all ϵ if $\lambda_2 \geq \frac{4}{3}$.

Idea of the proof:

- $\lambda_2 \geq 2$: Mimic Janson's proof technique for BSC (i.e. Bonami-Beckner)
- Remainder of regime
  - Use the relative entropy characterization
  - Inspired by Friedgut's proof for BSC for case $\lambda_1 = \lambda_2 = \rho$.

MAIN RESULT

Question (to me) by V. Guruswami and J. Radhakrishnan:

- Is correlation bound tight for BEC?

Theorem (Tightness of correlation lower bound)

For BEC the correlation bound is tight, i.e. (X, Y) *is* (λ_1, λ_2)*-hypercontractive for* $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$*, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1).$$

Remarks:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all ϵ if $\lambda_2 \geq \frac{4}{3}$.

Idea of the proof:

- $\lambda_2 \geq 2$: Mimic Janson's proof technique for BSC (i.e. Bonami-Beckner)
- Remainder of regime
  - Use the relative entropy characterization
  - Inspired by Friedgut's proof for BSC for case $\lambda_1 = \lambda_2 = \mu$.

Main result

Question (to me) by V. Guruswami and J. Radhakrishnan:

- Is correlation bound tight for BEC?

Theorem (Tightness of correlation lower bound)

For BEC the correlation bound is tight, i.e. (X, Y) *is* (λ_1, λ_2)*-hypercontractive for* $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$*, if and only if the following condition is satisfied:*

$$\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1).$$

Remarks:

- Always holds when $\epsilon \leq \frac{1}{2}$
- Holds for all ϵ if $\lambda_2 \geq \frac{4}{3}$.

Idea of the proof:

- $\lambda_2 \geq 2$: Mimic Janson's proof technique for BSC (i.e. Bonami-Beckner)
- Remainder of regime
  - Use the relative entropy characterization
  - Inspired by Friedgut's proof for BSC for case $\lambda_1 = \lambda_2 = \rho$.

Case 1: $\lambda_2 \geq 2$

We have $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. Need to show that

$$\|\mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1} \quad \forall f(X),$$

where $\lambda_2' = \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate.

Let $f(0) = 1 - \delta, f(1) = 1 + \delta$. Need to show

$$\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}$$

CASE 1: $\lambda_2 \geq 2$

We have $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. Need to show that

$$\|\mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1} \quad \forall f(X),$$

where $\lambda_2' = \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate.

Let $f(0) = 1 - \delta, f(1) = 1 + \delta$. Need to show

$$\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}.$$

Case 1: $\lambda_2 \geq 2$

We have $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. Need to show that

$$\|\mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1} \quad \forall f(X),$$

where $\lambda_2' = \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate.

Let $f(0) = 1 - \delta, f(1) = 1 + \delta$. Need to show

$$\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}.$$

Rewriting

$$1 + (1-\epsilon)\sum_{k=1}^{\infty}\binom{\lambda_2'}{2k}\delta^{2k} \leq \left(1 + \sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k}\right)^{\frac{\lambda_2'}{\lambda_1}}$$

CASE 1: $\lambda_2 \geq 2$

We have $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. Need to show that

$$\|\mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1} \quad \forall f(X),$$

where $\lambda_2' = \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate.

Let $f(0) = 1 - \delta, f(1) = 1 + \delta$. Need to show

$$\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}.$$

Suffices to show that (since $(1+x)^a \geq (1+ax), a > 1, x > 0$)

$$1 + (1-\epsilon)\sum_{k=1}^{\infty}\binom{\lambda_2'}{2k}\delta^{2k} \leq 1 + \frac{\lambda_2'}{\lambda_1}\sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k} \leq \left(1 + \sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k}\right)^{\frac{\lambda_2'}{\lambda_1}}$$

Case 1: $\lambda_2 \geq 2$

We have $(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$. Need to show that

$$\|\mathrm{E}(f(X)|Y)\|_{\lambda_2'} \leq \|f(X)\|_{\lambda_1} \quad \forall f(X),$$

where $\lambda_2' = \frac{\lambda_2}{\lambda_2 - 1}$, the Hölder conjugate.

Let $f(0) = 1 - \delta, f(1) = 1 + \delta$. Need to show

$$\left[\frac{1-\epsilon}{2}(1-\delta)^{\lambda_2'} + \frac{1-\epsilon}{2}(1+\delta)^{\lambda_2'} + \epsilon\right]^{\frac{1}{\lambda_2'}} \leq \left[\frac{1}{2}(1-\delta)^{\lambda_1} + \frac{1}{2}(1+\delta)^{\lambda_1}\right]^{\frac{1}{\lambda_1}}.$$

Suffices to show that (since $(1+x)^a \geq (1+ax), a > 1, x > 0$)

$$1 + (1-\epsilon)\sum_{k=1}^{\infty}\binom{\lambda_2'}{2k}\delta^{2k} \leq 1 + \frac{\lambda_2'}{\lambda_1}\sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k} \leq \left(1 + \sum_{k=1}^{\infty}\binom{\lambda_1}{2k}\delta^{2k}\right)^{\frac{\lambda_2'}{\lambda_1}}$$

For $k \geq 1$, since $1 \leq \lambda_1 \leq \lambda_2' \leq 2$ immediate that

$$(1-\epsilon)\binom{\lambda_2'}{2k} \leq \frac{\lambda_2'}{\lambda_1}\binom{\lambda_1}{2k}.$$

Remarks

The previous argument

- Essentially due to Janson
- Works for any BMS (binary memoryless symmetric)
  - $W(Y = i|X = 1) = W(Y = -i|X = -1) = \mu_i, \ \ i = 1, .., K$
  - Necessary & Sufficient: $(\lambda_1 - 1)(\lambda_2 - 1) = \sum_{i=1}^{K} \frac{(\mu_i - \mu_{-i})^2}{\mu_i + \mu_{-i}}, \quad \lambda_2 \geq 2.$
- Similar reasoning can be extended for $\lambda_2 \geq \frac{3}{2}$.
- Could not (does not) extend to $\lambda_2 \geq \frac{4}{3}$.

REMARKS

The previous argument

- Essentially due to Janson
- Works for any BMS (binary memoryless symmetric)
  - $W(Y=i|X=1) = W(Y=-i|X=-1) = \mu_i, \;\; i = 1, .., K$
  - Necessary & Sufficient: $(\lambda_1 - 1)(\lambda_2 - 1) = \sum_{i=1}^{K} \frac{(\mu_i - \mu_{-i})^2}{\mu_i + \mu_{-i}}, \quad \lambda_2 \geq 2.$
- Similar reasoning can be extended for $\lambda_2 \geq \frac{3}{2}$.
- Could not (does not) extend to $\lambda_2 \geq \frac{4}{3}$.

Remarks

The previous argument

- Essentially due to Janson
- Works for any BMS (binary memoryless symmetric)
  - $W(Y = i|X = 1) = W(Y = -i|X = -1) = \mu_i, \ \ i = 1, .., K$
  - Necessary & Sufficient: $(\lambda_1 - 1)(\lambda_2 - 1) = \sum_{i=1}^{K} \frac{(\mu_i - \mu_{-i})^2}{\mu_i + \mu_{-i}}, \quad \lambda_2 \geq 2.$
- Similar reasoning can be extended for $\lambda_2 \geq \frac{3}{2}$.
- Could not (does not) extend to $\lambda_2 \geq \frac{4}{3}$.

CASE 2: $\lambda_2 < 2$ AND $\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1)$

$(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon.$

We wish to show that $\max \nu_{XY} (\ll \mu_{XY}^{BEC})$

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC}) = \begin{cases} 0 & \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & \text{o.w.} \end{cases}$$

$(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon$.

We wish to show that $\max \nu_{XY} (\ll \mu_{XY}^{BEC})$

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC}) = \begin{cases} 0 & \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & \text{o.w.} \end{cases}$$

Easy: maximum will occur in the interior.

Stationarity conditions yield (Friedgut)

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{0*}) - \frac{1}{\lambda_2'} \ln \frac{\nu_{00}}{1 - \epsilon}$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{11} + \nu_{1*}) - \frac{1}{\lambda_2'} \ln \frac{\nu_{11}}{1 - \epsilon}$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{0*}) + \frac{1}{\lambda_2} \ln(\nu_{0*} + \nu_{1*}) - \frac{1}{\lambda_2} \ln 2 - \ln \nu_{0*} + \frac{1}{\lambda_2'} \ln \epsilon$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{11} + \nu_{1*}) + \frac{1}{\lambda_2} \ln(\nu_{0*} + \nu_{1*}) - \frac{1}{\lambda_2} \ln 2 - \ln \nu_{1*} + \frac{1}{\lambda_2'} \ln \epsilon$$

Case 2: $\lambda_2 < 2$ and $\epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1)$

$(\lambda_1 - 1)(\lambda_2 - 1) = 1 - \epsilon.$

We wish to show that $\max \nu_{XY} (\ll \mu_{XY}^{BEC})$

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC}) = \begin{cases} 0 & \epsilon - \frac{1}{2} \leq \frac{3}{2}(\lambda_2 - 1) \\ > 0 & \text{o.w.} \end{cases}$$

Easy: maximum will occur in the interior.

Stationarity conditions yield (Friedgut)

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{0*}) - \frac{1}{\lambda_2'} \ln \frac{\nu_{00}}{1 - \epsilon}$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{11} + \nu_{1*}) - \frac{1}{\lambda_2'} \ln \frac{\nu_{11}}{1 - \epsilon}$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{00} + \nu_{0*}) + \frac{1}{\lambda_2} \ln(\nu_{0*} + \nu_{1*}) - \frac{1}{\lambda_2} \ln 2 - \ln \nu_{0*} + \frac{1}{\lambda_2'} \ln \epsilon$$

$$k = \frac{1}{\lambda_1} \ln(\nu_{11} + \nu_{1*}) + \frac{1}{\lambda_2} \ln(\nu_{0*} + \nu_{1*}) - \frac{1}{\lambda_2} \ln 2 - \ln \nu_{1*} + \frac{1}{\lambda_2'} \ln \epsilon$$

SOME CALCULATIONS

From stationarity conditions we get

$$\frac{\nu_{0*}}{\nu_{1*}} = \left(\frac{\nu_{00}+\nu_{0*}}{\nu_{11}+\nu_{1*}}\right)^{\frac{1}{\lambda_1}}, \quad \nu_{00} = \frac{\nu_{0*}^{\lambda_2'} 2^{\lambda_2'-1}}{(\nu_{0*}+\nu_{1*})^{\lambda_2'-1}} \frac{1-\epsilon}{\epsilon},$$

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2-1}} + \epsilon x^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2-1}},$$

where $x = \frac{2\nu_{0*}}{\nu_{1*}+\nu_{0*}}$.

To note: Corresponding to each solution of $x \in [0,2]$ there is a single stationary point.

Lemma

For $x \in [0,2]$, $\lambda_2 \in (1,2)$, $\epsilon \in (0,1)$ the following equation

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2-1}} + \epsilon x^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2-1}},$$

has only one root at $x = 1$ if $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$.

Proof: show that the difference between the sides decreases in $[0, 1]$.

Some calculations

From stationarity conditions we get

$$\frac{\nu_{0*}}{\nu_{1*}} = \left(\frac{\nu_{00}+\nu_{0*}}{\nu_{11}+\nu_{1*}}\right)^{\frac{1}{\lambda_1}}, \quad \nu_{00} = \frac{\nu_{0*}^{\lambda_2'} 2^{\lambda_2'-1}}{(\nu_{0*}+\nu_{1*})^{\lambda_2'-1}} \frac{1-\epsilon}{\epsilon},$$

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2-1}} + \epsilon x^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2-1}},$$

where $x = \frac{2\nu_{0*}}{\nu_{1*}+\nu_{0*}}$.
To note: Corresponding to each solution of $x \in [0,2]$ there is a single stationary point.

Lemma

For $x \in [0,2]$, $\lambda_2 \in (1,2)$, $\epsilon \in (0,1)$ the following equation

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2-1}} + \epsilon x^{\frac{\epsilon-1}{\lambda_2-1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2-1}} + \epsilon(2-x)^{\frac{\epsilon-1}{\lambda_2-1}}.$$

has only one root at $x=1$ if $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$.

Proof: show that the difference between the sides decreases in $[0,1]$.

Some calculations

From stationarity conditions we get

$$\frac{\nu_{0*}}{\nu_{1*}} = \left(\frac{\nu_{00} + \nu_{0*}}{\nu_{11} + \nu_{1*}}\right)^{\frac{1}{\lambda_1}}, \quad \nu_{00} = \frac{\nu_{0*}^{\lambda_2'} 2^{\lambda_2' - 1}}{(\nu_{0*} + \nu_{1*})^{\lambda_2' - 1}} \frac{1-\epsilon}{\epsilon},$$

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon x^{\frac{\epsilon - 1}{\lambda_2 - 1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon(2-x)^{\frac{\epsilon - 1}{\lambda_2 - 1}},$$

where $x = \frac{2\nu_{0*}}{\nu_{1*} + \nu_{0*}}$.

To note: Corresponding to each solution of $x \in [0, 2]$ there is a single stationary point.

Lemma

For $x \in [0, 2]$, $\lambda_2 \in (1, 2)$, $\epsilon \in (0, 1)$ *the following equation*

$$(1-\epsilon)x^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon x^{\frac{\epsilon - 1}{\lambda_2 - 1}} = (1-\epsilon)(2-x)^{\frac{\epsilon}{\lambda_2 - 1}} + \epsilon(2-x)^{\frac{\epsilon - 1}{\lambda_2 - 1}}.$$

has only one root at $x = 1$ *if* $(\epsilon - \frac{1}{2}) \leq \frac{3}{2}(\lambda_2 - 1)$.

Proof: show that the difference between the sides decreases in $[0, 1]$.

CASE: $(\epsilon - \frac{1}{2}) > \frac{3}{2}(\lambda_2 - 1)$

$$\nu_{XY} = \begin{bmatrix} \frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A} & \frac{\epsilon(1-\delta)}{A} & 0 \\ 0 & \frac{\epsilon(1+\delta)}{A} & \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A} \end{bmatrix}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$.

Taylor series expansion of the term

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC})$$

around $\delta = 0$ yields an expansion

$$\frac{1}{24}\epsilon(1-\epsilon)(\lambda_2'-1)^2((2\epsilon-1)(\lambda_2'-1)-3)\delta^4 + O(\delta^6).$$

Q.E.D.

CASE: $(\epsilon - \frac{1}{2}) > \frac{3}{2}(\lambda_2 - 1)$

$$\nu_{XY} = \begin{bmatrix} \frac{(1-\delta)^{\lambda_2'}(1-\epsilon)}{A} & \frac{\epsilon(1-\delta)}{A} & 0 \\ 0 & \frac{\epsilon(1+\delta)}{A} & \frac{(1-\epsilon)(1+\delta)^{\lambda_2'}}{A} \end{bmatrix}$$

where $A = 2\epsilon + (1-\epsilon)[(1+\delta)^{\lambda_2'} + (1-\delta)^{\lambda_2'}]$.

Taylor series expansion of the term

$$\frac{1}{\lambda_1} D(\nu_X \| \mu_X^{BEC}) + \frac{1}{\lambda_2} D(\nu_Y \| \mu_Y^{BEC}) - D(\nu_{XY} \| \mu_{XY}^{BEC})$$

around $\delta = 0$ yields an expansion

$$\frac{1}{24}\epsilon(1-\epsilon)(\lambda_2' - 1)^2((2\epsilon - 1)(\lambda_2' - 1) - 3)\delta^4 + O(\delta^6).$$

Q.E.D.

Take $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$. As before look at stationarity conditions:

There is a one-to-one mapping between stationary points and roots of the equation

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t\theta + (\theta + x)^t}{(\theta + x)^t\theta + (1+\theta x)^t},$$

where $\theta = \frac{1-\rho}{1+\rho}$.

Lemma

For any $t \in (0,\infty)$ and $\theta \in (0,1) \cup (1,\infty)$ the equation

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t\theta + (\theta + x)^t}{(\theta + x)^t\theta + (1+\theta x)^t}$$

has only one root at $x = 1$ for $x \in (0,\infty)$.

This implies that the relative entropy characterization is satisfied when $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$, proving the tightness of correlation lower bound.

Take $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$. As before look at stationarity conditions:

There is a one-to-one mapping between stationary points and roots of the equation

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t \theta + (\theta + x)^t}{(\theta + x)^t \theta + (1+\theta x)^t},$$

where $\theta = \frac{1-\rho}{1+\rho}$.

Lemma

For any $t \in (0, \infty)$ and $\theta \in (0,1) \cup (1, \infty)$ the equation

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t \theta + (\theta + x)^t}{(\theta + x)^t \theta + (1+\theta x)^t}$$

has only one root at $x = 1$ for $x \in (0, \infty)$.

This implies that the relative entropy characterization is satisfied when $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$, proving the tightness of correlation lower bound.

Take $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$. As before look at stationarity conditions:

There is a one-to-one mapping between stationary points and roots of the equation

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t\theta + (\theta + x)^t}{(\theta + x)^t\theta + (1+\theta x)^t},$$

where $\theta = \frac{1-\rho}{1+\rho}$.

Lemma

For any $t \in (0, \infty)$ *and* $\theta \in (0,1) \cup (1, \infty)$ *the equation*

$$x^{t\left(\frac{1-\theta}{1+\theta}\right)^2} = \frac{(1+\theta x)^t\theta + (\theta + x)^t}{(\theta + x)^t\theta + (1+\theta x)^t}$$

has only one root at $x = 1$ *for* $x \in (0, \infty)$.

This implies that the relative entropy characterization is satisfied when $(\lambda_1 - 1)(\lambda_2 - 1) = \rho^2$, proving the tightness of correlation lower bound.

SUMMARY

What we saw

- Equivalent characterizations can be very useful for precise characterization or bounds
- Local optimality test were sufficient for BEC

SUMMARY

What we saw

- Equivalent characterizations can be very useful for precise characterization or bounds
- Local optimality test were sufficient for BEC
  - Not an isolated observation

SUMMARY

What we saw

- Equivalent characterizations can be very useful for precise characterization or bounds
- Local optimality test were sufficient for BEC
  - Not an isolated observation

Tomorrow: more on such inequalities and generalizations.

Thank You