An information theoretic proof of a hypercontractive inequality

Ehud Friedgut

Weizmann Institute of Science

April 30, 2015

Playing in a theatre near you

Overview

1. The Beckner-Bonami-Gross inequality
   - Propoganda
   - The Fourier-Walsh basis
   - Definitions of the operator
     - The tensor definition
     - The spectral definition
     - The noise/averaging definition
   - The inequality, and the dual version
2. An entropy proof of the dual (w/ Rödl)
3. An information theoretic proof of the primal version

Applications

Quote from [BT]:

Applications

Quote from [BT]:

First introduced into theoretical computer science by the celebrated work of Kahn, Kalai, and Linial [15], the Hypercontractive Inequality has seen utility in a surprisingly wide variety of areas, spanning distributed computing, random graphs, k-SAT, social choice, inapproximability, learning theory, metric spaces, statistical physics, convex relaxation hierarchies, etc. [2, 6, 22, 8, 9, 10, 11, 5, 18, 17, 21, 13, 19, 1]. In almost every one of these results there are no known alternate proofs that do not require the use of hypercontractivity.

The Fourier-Walsh Basis

Consider the coordinates X_i of a vector X as functions on $\{-1,1\}^n$

The Fourier-Walsh Basis

Consider the coordinates X_i of a vector X as functions on $\{-1,1\}^n$
For every $I \subseteq \{1,\ldots,n\}$ define

$$X_I := \prod_{i\in I} X_i$$

The Fourier-Walsh Basis

Consider the coordinates X_i of a vector X as functions on $\{-1,1\}^n$
For every $I \subseteq \{1,\dots,n\}$ define

$$X_I := \prod_{i \in I} X_i$$

This set, of 2^n monomials, forms an orthonormal basis of the space of real functions on $\{-1,1\}^n$.

The Fourier-Walsh Basis

Consider the coordinates X_i of a vector X as functions on $\{-1,1\}^n$
For every $I \subseteq \{1,\ldots,n\}$ define

$$X_I := \prod_{i \in I} X_i$$

This set, of 2^n monomials, forms an orthonormal basis of the space of real functions on $\{-1,1\}^n$.
Every real function on $\{-1,1\}^n$ has a unique expansion

$$f = \sum \widehat{f}(I) X_I$$

The tensor definition of the operator

Let $\epsilon \in [0,1]$.

Let $f : \{-1,1\} \to \mathbf{R}$, $f(X) = aX + b$.

The tensor definition of the operator

Let $\epsilon \in [0, 1]$.

Let $f : \{-1, 1\} \to \mathbf{R}$, $f(X) = aX + b$.

Define $T_\epsilon(f)(X) := \epsilon aX + b$.

The tensor definition of the operator

Let $\epsilon \in [0, 1]$.

Let $f : \{-1, 1\} \to \mathbf{R}$, $f(X) = aX + b$.

Define $T_\epsilon(f)(X) := \epsilon aX + b$.

Then $T_\epsilon := T_\epsilon^{\otimes n}$ is a linear operator acting on real functions on $\{-1, 1\}^n$

The spectral definition of the operator

Let $\epsilon \in [0, 1]$. Define

$$T_\epsilon X_i = \epsilon X_i$$

The spectral definition of the operator

Let $\epsilon \in [0,1]$. Define

$$T_\epsilon X_i = \epsilon X_i$$

$$T_\epsilon X_I = \prod_{\{i \in I\}} T_\epsilon X_i = \epsilon^{|I|} X_I$$

The spectral definition of the operator

Let $\epsilon \in [0,1]$. Define

$$T_\epsilon X_i = \epsilon X_i$$

$$T_\epsilon X_I = \prod_{\{i \in I\}} T_\epsilon X_i = \epsilon^{|I|} X_I$$

$$T_\epsilon f = \sum \widehat{f}(I) \epsilon^{|I|} X_I$$

The operator as a noise/averaging operator

Let $\epsilon \in [0,1]$. Let X be chosen from any distribution on $\{-1,1\}^n$.

The operator as a noise/averaging operator

Let $\epsilon \in [0,1]$. Let X be chosen from any distribution on $\{-1,1\}^n$.

Let Y be such that for every $1 \leq i \leq n$, the coordinate Y_i is chosen independently so that $PR[Y_i = X_i] = \frac{1+\epsilon}{2}$, or, in other words, $E[X_i Y_i] = \epsilon$.

The operator as a noise/averaging operator

Let $\epsilon \in [0, 1]$. Let X be chosen from any distribution on $\{-1, 1\}^n$.

Let Y be such that for every $1 \leq i \leq n$, the coordinate Y_i is chosen independently so that $PR[Y_i = X_i] = \frac{1+\epsilon}{2}$, or, in other words, $E[X_i Y_i] = \epsilon$.

X and Y are called an ϵ-correlated pair.

The operator as a noise/averaging operator

Let $\epsilon \in [0,1]$. Let X be chosen from any distribution on $\{-1,1\}^n$.

Let Y be such that for every $1 \leq i \leq n$, the coordinate Y_i is chosen independently so that $PR[Y_i = X_i] = \frac{1+\epsilon}{2}$, or, in other words, $E[X_i Y_i] = \epsilon$.

X and Y are called an ϵ-correlated pair.

Define for any f and fixed X

$$T_\epsilon(f)(X) = E[f(Y)],$$

where X and Y are ϵ correlated.

The inequality and its dual

Bonami[68,70],Gross[75],Beckner[75]

Let $f : \{-1,1\}^n \to \mathbf{R}$, and $\epsilon \in [0,1]$. Then

$$|T_\epsilon f|_2 \leq |f|_{1+\epsilon^2}$$

The inequality and its dual

Bonami[68,70],Gross[75],Beckner[75]

Let $f : \{-1,1\}^n \to \mathbf{R}$, and $\epsilon \in [0,1]$. Then

$$|T_\epsilon f|_2 \leq |f|_{1+\epsilon^2}$$

Dual version

Let $f : \{-1,1\}^n \to \mathbf{R}$ be a polynomial of degree m, and $q \geq 2$. Then

$$|f|_q \leq \left(\sqrt{q-1}\right)^m |f|_2$$

Theorems

F, Rödl, 2001

Theorems

F, Rödl, 2001

We might have written in the introduction:

Theorems

F, Rödl, 2001

We might have written in the introduction: "Bonami and others proved a similar result in the seventies, however our result is more recent, and less general"

Theorems

F, Rödl, 2001

We might have written in the introduction: "Bonami and others proved a similar result in the seventies, however our result is more recent, and less general" (...but it uses entropy.)

Theorems

F, Rödl, 2001

We might have written in the introduction: "Bonami and others proved a similar result in the seventies, however our result is more recent, and less general" (...but it uses entropy.)

Theorem: Let $f : \{-1,1\}^n$ be a polynomial of degree m. Then

$$|f|_4 \leq (\sqrt[4]{28})^m |f|_2$$

Theorems

F, Rödl, 2001

We might have written in the introduction: "Bonami and others proved a similar result in the seventies, however our result is more recent, and less general" (...but it uses entropy.)

Theorem: Let $f : \{-1,1\}^n$ be a polynomial of degree m. Then

$$|f|_4 \leq (\sqrt[4]{28})^m |f|_2$$

Blais, Tan, 2013

Let $f : \{-1,1\}^n$ be a polynomial of degree m, and q an even positive integer. Then

$$|f|_q \leq \left(\sqrt{q-1}\right)^m |f|_2$$

Note that this is the optimal constant.

Let $f = \sum \widehat{f}(I) X_I$, where every X_I is monomial of degree m, $X_I = \prod_{i \in I} X_i$. Then

Let $f = \sum \widehat{f}(I)X_I$, where every X_I is monomial of degree m, $X_I = \prod_{i \in I} X_i$. Then

$$|f|_2^2 = E(f^2) = \sum \widehat{f}(I)^2$$

Let $f = \sum \widehat{f}(I)X_I$, where every X_I is monomial of degree m, $X_I = \prod_{i \in I} X_i$. Then

$$|f|_2^2 = E(f^2) = \sum \widehat{f}(I)^2$$

and

$$|f|_4^4 = E(f^4) = \sum_{I \Delta J \Delta K \Delta L = \emptyset} \widehat{f}(I)\widehat{f}(J)\widehat{f}(K)\widehat{f}(L).$$

Plan of proof

Let $\mathcal{I}\Delta\mathcal{J}\Delta\mathcal{K}\Delta\mathcal{L}=\emptyset$. Fix a partition P of $\{1,\ldots,n\}$ with parts corresponding to the Venn diagram of $(\mathcal{I},\mathcal{J},\mathcal{K},\mathcal{L})$ and prove that

$$\left(\sum_{I\Delta J\Delta K\Delta L=\emptyset}\widehat{f}(I)\widehat{f}(J)\widehat{f}(K)\widehat{f}(L)\right)^2\leq$$

$$\left(\sum_I\widehat{f}(I)^2\right)\left(\sum_J\widehat{f}(J)^2\right)\left(\sum_K\widehat{f}(K)^2\right)\left(\sum_L\widehat{f}(L)^2\right)$$

where all sums are only over quadruples of sets, that are consistent with P.

Plan of proof

Let $\mathcal{I}\Delta\mathcal{J}\Delta\mathcal{K}\Delta\mathcal{L}=\emptyset$. Fix a partition P of $\{1,\ldots,n\}$ with parts corresponding to the Venn diagram of $(\mathcal{I},\mathcal{J},\mathcal{K},\mathcal{L})$ and prove that

$$\left(\sum_{I\Delta J\Delta K\Delta L=\emptyset}\widehat{f}(I)\widehat{f}(J)\widehat{f}(K)\widehat{f}(L)\right)^2 \leq$$

$$\left(\sum_I \widehat{f}(I)^2\right)\left(\sum_J \widehat{f}(J)^2\right)\left(\sum_K \widehat{f}(K)^2\right)\left(\sum_L \widehat{f}(L)^2\right)$$

where all sums are only over quadruples of sets, that are consistent with P.

To this end invoke a *fractional version* of Shearer's lemma.

Plan of proof

Let $\mathcal{I}\Delta\mathcal{J}\Delta\mathcal{K}\Delta\mathcal{L} = \emptyset$. Fix a partition P of $\{1,\ldots,n\}$ with parts corresponding to the Venn diagram of $(\mathcal{I},\mathcal{J},\mathcal{K},\mathcal{L})$ and prove that

$$\left(\sum_{I\Delta J\Delta K\Delta L=\emptyset} \widehat{f}(I)\widehat{f}(J)\widehat{f}(K)\widehat{f}(L)\right)^2 \leq$$

$$\left(\sum_I \widehat{f}(I)^2\right)\left(\sum_J \widehat{f}(J)^2\right)\left(\sum_K \widehat{f}(K)^2\right)\left(\sum_L \widehat{f}(L)^2\right)$$

where all sums are only over quadruples of sets, that are consistent with P.

To this end invoke a *fractional version* of Shearer's lemma.

Show that the above expressions for a *random partition* reflect $|f|_4$ and $|f|_2$ fairly well. (This already introduces a loss of a multiplicative constant).

Theorems

Shearer's Lemma['86]

Let t be a positive integer. Let $E \subseteq P(V)$, and let $F_1 \ldots F_r \subseteq V$ such that every vertex in V belongs to at least t of the sets F_i. Let $E_i = \{e \cap F_i : e \in E\}$. Then

$$|E|^t \leq \prod |E_i|.$$

Theorems

Shearer's Lemma['86]

Let t be a positive integer. Let $E \subseteq P(V)$, and let $F_1 \ldots F_r \subseteq V$ such that every vertex in V belongs to at least t of the sets F_i. Let $E_i = \{e \cap F_i : e \in E\}$. Then

$$|E|^t \leq \prod |E_i|.$$

Fractional version [F, 2004]

Let $e_i := e \cap F_i$. Let every edge $e_i \in E_i$ be endowed with a nonnegative weight $w_i(e_i)$. Then

$$\left(\sum_{e \in E} \prod_{i=1}^{r} w_i(e_i)\right)^t \leq \prod_i \sum_{e_i \in E_i} w_i(e_i)^t.$$

So.

Shearer → Beckner

Fractional Shearer

$$\left(\sum_{e\in E}\prod_{i=1}^{r} w_i(e_i)\right)^t \leq \prod_i \sum_{e_i\in E_i} w_i(e_i)^t.$$

Shearer → Beckner

Fractional Shearer

$$\left(\sum_{e\in E}\prod_{i=1}^{r} w_i(e_i)\right)^t \leq \prod_i \sum_{e_i\in E_i} w_i(e_i)^t.$$

Comparison of norms

$$\left(\sum_{I\Delta J\Delta K\Delta L=\emptyset} \widehat{f}(I)\widehat{f}(J)\widehat{f}(K)\widehat{f}(L)\right)^2 \leq$$

$$\left(\sum_I \widehat{f}(I)^2\right)\left(\sum_J \widehat{f}(J)^2\right)\left(\sum_K \widehat{f}(K)^2\right)\left(\sum_L \widehat{f}(L)^2\right).$$

Theorems

The Boolean case

Let $\epsilon \in (0,1)$, and let $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ be nonempty. Let X be uniformly distributed on $\{0,1\}^n$, and let Y be such that for each $1 \leq i \leq n$ independently $Pr[X_i = Y_i] = \frac{1+\epsilon}{2}$. Then

$$E[\mathbf{1}_{\mathcal{X}}(X)\mathbf{1}_{\mathcal{Y}}(Y)] \leq (\mu(\mathcal{X})\mu(\mathcal{Y}))^{\frac{1}{1+\epsilon}},$$

with equality iff $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$

Theorems

The Boolean case

Let $\epsilon \in (0,1)$, and let $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$ be nonempty. Let X be uniformly distributed on $\{0,1\}^n$, and let Y be such that for each $1 \leq i \leq n$ independently $Pr[X_i = Y_i] = \frac{1+\epsilon}{2}$. Then

$$E[\mathbf{1}_{\mathcal{X}}(X)\mathbf{1}_{\mathcal{Y}}(Y)] \leq (\mu(\mathcal{X})\mu(\mathcal{Y}))^{\frac{1}{1+\epsilon}},$$

with equality iff $\mathcal{X} = \mathcal{Y} = \{0,1\}^n$

The general case

Let $\epsilon \in (0,1)$,and $X, Y \in \{0,1\}^n$ as above, and let $f, g : \{0,1\}^n \to \mathbf{R}^{\geq 0}$. Then

$$E[f(X)g(Y)] \leq |f|_{1+\epsilon}|g|_{1+\epsilon}.$$

The Boolean case

We want to prove

$$E[\mathbf{1}_{\mathcal{X}}(X)\mathbf{1}_{\mathcal{Y}}(Y)] \leq (\mu(\mathcal{X})\mu(\mathcal{Y}))^{\frac{1}{1+\epsilon}}.$$

The Boolean case

We want to prove

$$E[\mathbf{1}_{\mathcal{X}}(X)\mathbf{1}_{\mathcal{Y}}(Y)] \leq (\mu(\mathcal{X})\mu(\mathcal{Y}))^{\frac{1}{1+\epsilon}}.$$

This easily translates to

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}}(1+\epsilon)^{a(X,Y)}(1-\epsilon)^{d(X,Y)}\right)^* \leq$$

$$\frac{1}{1+\epsilon}\left(2\epsilon n+\log(|\mathcal{X}|)+\log(|\mathcal{Y}|)\right)$$

*Where a stands for "agree" and d for "disagree".

The Boolean case

Letting $s \leq r$ be positive integers so that $\frac{1+\epsilon}{2} = \frac{r}{s+r}$ and $\frac{1-\epsilon}{2} = \frac{s}{r+s}$ this gives

The Boolean case

Letting $s \leq r$ be positive integers so that $\frac{1+\epsilon}{2} = \frac{r}{s+r}$ and $\frac{1-\epsilon}{2} = \frac{s}{r+s}$ this gives

$$\log \left(\sum_{X \in \mathcal{X}} \sum_{Y \in \mathcal{Y}} r^{a(X,Y)} s^{d(X,Y)} \right) \leq$$

$$n(\log(r+s) - s/r) + \frac{r+s}{2r} (\log(|\mathcal{X}|) + \log(|\mathcal{Y}|))$$

The Boolean case

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} r^{a(X,Y)}s^{d(X,Y)}\right) \leq$$

$$n(\log(r+s) - s/r) + \frac{r+s}{2r}(\log(|\mathcal{X}|) + \log(|\mathcal{Y}|))$$

The Boolean case

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} r^{a(X,Y)}s^{d(X,Y)}\right) \leq$$

$$n(\log(r+s)-s/r)+\frac{r+s}{2r}(\log(|\mathcal{X}|)+\log(|\mathcal{Y}|))$$

$$H(Z) \leq n(\log(r+s)-s/r)+\frac{r+s}{2r}(H(X)+H(Y))$$

The Boolean case

$$H(Z) \leq n(\log(r+s) - s/r) + \frac{r+s}{2r}(H(X) + H(Y))$$

The Boolean case

$$H(Z) \leq n(\log(r+s) - s/r) + \frac{r+s}{2r}(H(X) + H(Y))$$

Using the chain rule this will follow from

$$H(Z_i|Past) \leq (\log(r+s) - s/r) + \frac{r+s}{2r}(H(X_i|Past) + H(Y_i|Past)).$$

The Boolean case

$$H(Z_i|Past) \leq (\log(r+s) - s/r) + \frac{r+s}{2r}(H(X_i|Past) + H(Y_i|Past)).$$

The Boolean case

$$H(Z_i|Past) \leq (\log(r+s) - s/r) + \frac{r+s}{2r}(H(X_i|Past) + H(Y_i|Past)).$$

Using $H(Z_i) = H(X_i, Y_i) + H(Z_i|X_i, Y_i)$ this is equivalent to

$$\frac{r+s}{2r}\left(H(X_i) + H(Y_i)\right) - H(X_i, Y_i) - (Pr[X_i = Y_i]) \log r$$

$$-(Pr[X_i \neq Y_i]) \log s + \log(r+s) - \frac{s}{r} \geq 0$$

The Boolean case

$$H(Z_i|Past) \leq (\log(r+s) - s/r) + \frac{r+s}{2r}(H(X_i|Past) + H(Y_i|Past)).$$

Using $H(Z_i) = H(X_i, Y_i) + H(Z_i|X_i, Y_i)$ this is equivalent to

$$\frac{r+s}{2r}(H(X_i) + H(Y_i)) - H(X_i, Y_i) - (Pr[X_i = Y_i]) \log r$$

$$-(Pr[X_i \neq Y_i]) \log s + \log(r+s) - \frac{s}{r} \geq 0$$

Note this is invariant if s and r are multiplied by a positive constant, so set $r = 1, s = \delta$, $0 \leq \delta \leq 1$.

The Boolean case

So now we have an elementary calculus problem. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a distribution of (X, Y) on $\{0,1\}^2$. Prove

The Boolean case

So now we have an elementary calculus problem. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a distribution of (X, Y) on $\{0,1\}^2$. Prove

$$F_\delta \begin{pmatrix} a & b \\ c & d \end{pmatrix} := (1+\delta)\left[H(X) + H(Y)\right] - H(X,Y)$$

$$-(Pr[X \neq Y]) \log \delta + \log(1+\delta) - \delta \geq 0$$

The Boolean case

So now we have an elementary calculus problem. Let $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ be a distribution of (X, Y) on $\{0,1\}^2$. Prove

$$F_\delta \begin{pmatrix} a & b \\ c & d \end{pmatrix} := (1+\delta)\,[H(X) + H(Y)] - H(X,Y)$$

$$-(Pr[X \neq Y])\log\delta + \log(1+\delta) - \delta \geq 0$$

Show $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{\delta}{2+2\delta} & \frac{1}{2+2\delta} \\ \frac{1}{2+2\delta} & \frac{\delta}{2+2\delta} \end{pmatrix}$ is a unique minimum.

The Boolean case

Using Lagrange multipliers, a necessary condition for a local extramum is

The Boolean case

Using Lagrange multipliers, a necessary condition for a local extramum is

$$\frac{\delta}{a}\left[(a+b)(a+c)\right]^{(1+\delta)/2} = \quad (1)$$

$$\frac{1}{b}\left[(a+b)(b+d)\right]^{(1+\delta)/2} = \quad (2)$$

$$\frac{1}{c}\left[(a+c)(c+d)\right]^{(1+\delta)/2} = \quad (3)$$

$$\frac{\delta}{d}\left[(c+d)(b+d)\right]^{(1+\delta)/2}. \quad (4)$$

The Boolean case

This implies (without much effort...) $a = d$ and $ad = \delta^2 bc$.

The Boolean case

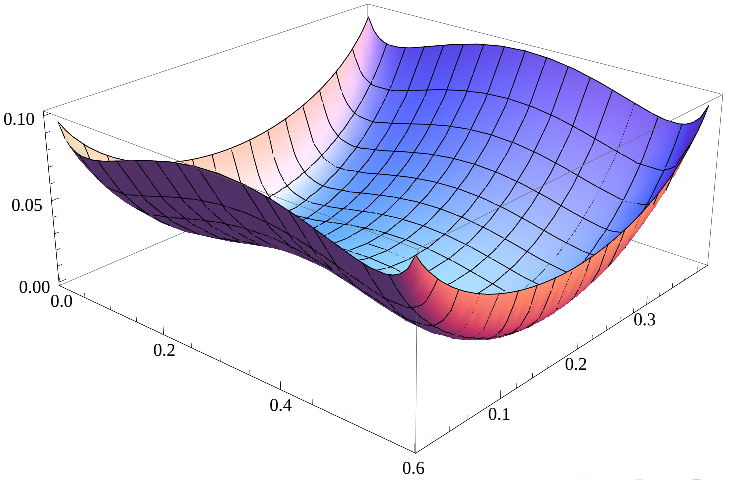
This implies (without much effort...) $a = d$ and $ad = \delta^2 bc$. All that is missing is $b = c$.

The Boolean case

This implies (without much effort...) $a = d$ and $ad = \delta^2 bc$. All that is missing is $b = c$. Sadly, for some values of δ there are other local minima.

The Boolean case

This implies (without much effort...) $a = d$ and $ad = \delta^2 bc$. All that is missing is $b = c$. Sadly, for some values of δ there are other local minima.



The Boolean case

However...

The Boolean case

However... we haven't really used the facts that

$$H(Z_i|X_i = 0, Y_i = 1) = H(Z_i|X_i = 1, Y_i = 0)$$

and

$$H(Z_i|X_i = 1, Y_i = 1) = H(Z_i|X_i = 0, Y_i = 0).$$

The Boolean case

However... we haven't really used the facts that

$$H(Z_i|X_i = 0, Y_i = 1) = H(Z_i|X_i = 1, Y_i = 0)$$

and

$$H(Z_i|X_i = 1, Y_i = 1) = H(Z_i|X_i = 0, Y_i = 0).$$

So instead of using

$$H(Z_i) = H(X_i, Y_i) + H(Z_i|X_i, Y_i)$$

The Boolean case

However... we haven't really used the facts that

$$H(Z_i|X_i = 0, Y_i = 1) = H(Z_i|X_i = 1, Y_i = 0)$$

and

$$H(Z_i|X_i = 1, Y_i = 1) = H(Z_i|X_i = 0, Y_i = 0).$$

So instead of using

$$H(Z_i) = H(X_i, Y_i) + H(Z_i|X_i, Y_i)$$

we let W indicate whether $X_i = Y_i$ or not, and use

$$H(Z_i) = H(W) + H(Z_i|W).$$

The Boolean case

This leads to a slightly different expression, with the Lagrange multipliers now yielding $a + d = \delta(b + c)$.

The Boolean case

This leads to a slightly different expression, with the Lagrange multipliers now yielding $a + d = \delta(b + c)$. Together with our previous information ($a = d$, $ad = \delta^2 bc$), this shows that the unique minimum on the interior of the region in question is

$$\begin{pmatrix} \frac{\delta}{2+2\delta} & \frac{1}{2+2\delta} \\ \frac{1}{2+2\delta} & \frac{\delta}{2+2\delta} \end{pmatrix}$$

as required.

The not-necessarily-Boolean case

How about non-Boolean functions?

The not-necessarily-Boolean case

How about non-Boolean functions? now we have to prove that for non-negative (w.l.o.g. positive integer-valued) f, g

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} r^{a(X,Y)} s^{d(X,Y)} f(X)g(Y)\right) \leq$$

$$n(\log(r+s) - s/r) +$$

$$\frac{r+s}{2r}\left(\log\left(\sum_{X\in 0,1^n} f(X)^{\frac{2r}{r+s}}\right) + \log\left(\sum_{Y\in 0,1^n} f(Y)^{\frac{2r}{r+s}}\right)\right).$$

The not-necessarily-Boolean case

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} r^{a(X,Y)}s^{d(X,Y)}f(X)g(Y)\right) \leq$$

$$n(\log(r+s) - s/r) +$$

$$\frac{r+s}{2r}\left(\log\left(\sum_{X\in 0,1^n} f(X)^{\frac{2r}{r+s}}\right) + \log\left(\sum_{Y\in 0,1^n} f(Y)^{\frac{2r}{r+s}}\right)\right).$$

The not-necessarily-Boolean case

$$\log\left(\sum_{X\in\mathcal{X}}\sum_{Y\in\mathcal{Y}} r^{a(X,Y)}s^{d(X,Y)}f(X)g(Y)\right)\leq$$

$$n(\log(r+s)-s/r)+$$

$$\frac{r+s}{2r}\left(\log\left(\sum_{X\in 0,1^n} f(X)^{\frac{2r}{r+s}}\right)+\log\left(\sum_{Y\in 0,1^n} f(Y)^{\frac{2r}{r+s}}\right)\right).$$

"Enhance" (Z, X, Y) of before to (Z, X, Y, a, b), where (a, b) is uniform on $\{1, \ldots, f(X)\} \times \{1, \ldots, g(Y)\}$.

Punchline

$$H(Z, a, b) = H(Z) + H((a, b)|Z) =$$

$$H(Z) + E[\log(f(X)) + \log(g(Y))]$$

Punchline

$$H(Z,a,b) = H(Z) + H((a,b)|Z) =$$

$$H(Z) + E[\log(f(X)) + \log(g(Y))]$$

Note that $H(X) + E[\log(f(X))] \leq \sum_X \log(f(X))$ and hence

Punchline

$$H(Z,a,b) = H(Z) + H((a,b)|Z) =$$
$$H(Z) + E[\log(f(X)) + \log(g(Y))]$$

Note that $H(X) + E[\log(f(X))] \leq \sum_X \log(f(X))$ and hence

$$\frac{r+s}{2r}(H(X)+H(Y)) + E[\log(f(X)) + \log(g(Y))]$$

$$\leq \frac{r+s}{2r}\left(\log\left(\sum_{X\in 0,1^n} f(X)^{\frac{2r}{r+s}}\right) + \log\left(\sum_{Y\in 0,1^n} f(Y)^{\frac{2r}{r+s}}\right)\right).$$

which is what we needed.

Thank you for your attention!