Local continuity bounds for entropies of finite probability distributions

Eric P. Hanson[†], Nilanjana Datta[†],

[†]Univ. Cambridge, DAMPT

UNIVERSITY OF CAMBRIDGE

CCIMI CANTAB CAPITAL INSTITUTE FOR THE MATHEMATICS OF INFORMATION

Motivation

Goal: local continuity bound for the entropy H. Why?

- Entropy is a fundamental quantity in information theory
  - Shannon source coding. Given i.i.d. data distributed with probability distribution p, the asymptotic optimal lossless data-compression rate is $H(p)$.
- Related to other quantities (e.g. conditional entropy when you don't condition on anything)
- The von Neumman entropy of a d-dimensional quantum state ρ is the entropy of its eigenvalues, which form a probability distribution.

Continuity bounds (CB) and local CB (LCB), by example

Consider $f : [0,1] \subset \mathbb{R} \to \mathbb{R}, \quad f(x) = x^2$. Then

$$|f(x) - f(y)| = |x^2 - y^2| = (x + y)|x - y|.$$

Continuity bound: suppose we don't know x, y but know

- $x, y \in [0,1]$
- $|x - y| \leq \varepsilon$.

Continuity bound:

$$|f(x) - f(y)| \leq 2\varepsilon$$

Local continuity bound: Additionally, know the value x. LCB:

$$|f(x) - f(y)| \leq (x + 1)\varepsilon.$$

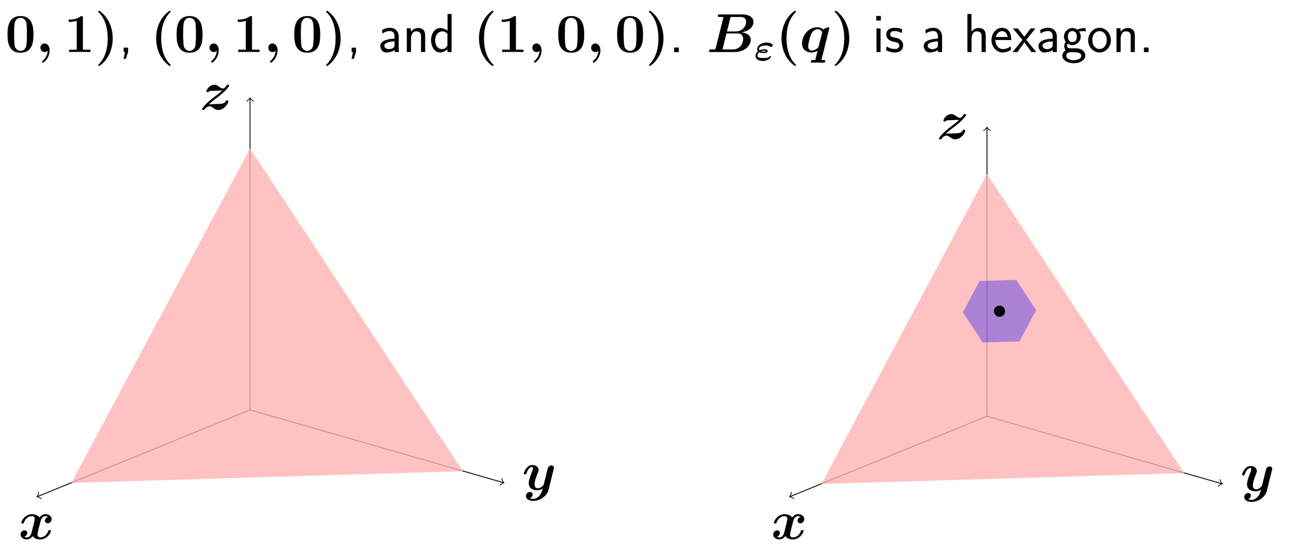
Tighter bound for any $x < 1$.

Geometry of probability distributions

- Finite set $\Omega = \{1, 2, \ldots, d\}$ on d elements.
- Let Δ be the set of probability distributions on Ω.
  - $p \in \Delta$ means $p : \Omega \to \mathbb{R}$, $p(i) \geq 0$, $\sum_{i=1}^d p(i) = 1$.
  - We think of $p \in \Delta$ as a vector in $\mathbb{R}^d$: $p(1)$ is the first coordinate, $p(i)$ the ith coordinate, etc.
- Locality: we fix $q \in \Delta$ and $\varepsilon \in [0,1]$. Set of distributions near q:

$$B_\varepsilon(q) := \{p \in \Delta : \frac{1}{2}\|p - q\|_1 \leq \varepsilon\}.$$

- In 3 dimensions: $\Omega = \{x, y, z\}$. Δ is the triangle with vertices $(0,0,1)$, $(0,1,0)$, and $(1,0,0)$. $B_\varepsilon(q)$ is a hexagon.



Left: the set Δ. Right: q is the black dot, $B_\varepsilon(q)$ the purple hexagon.

Strategy for LCB

$$H : \Delta \to \mathbb{R}, \qquad H(p) = -\sum_{i=1}^{d} p(i) \log p(i).$$

Fix $q \in \Delta$ and $\varepsilon \in [0,1]$. We always have

$$H(p) - H(q) \leq \max_{p \in B_\varepsilon(q)} H(p) - H(q)$$

and similarly $H(q) - H(p) \leq H(q) - \min_{p \in B_\varepsilon(q)} H(p)$, so

$$|H(p) - H(q)| \leq \max\{\max_{p \in B_\varepsilon(q)} H(p) - H(q), H(q) - \min_{p \in B_\varepsilon(q)} H(p)\}$$

which is a function of q and ε only.

Idea: Understand

$$p_\varepsilon^*(q) := \operatorname{argmax}_{p \in B_\varepsilon(q)} H(p), \quad \text{and} \quad p_{*,\varepsilon} := \operatorname{argmin}_{p \in B_\varepsilon(q)}(p).$$

Finding the minimizer (I)

- Wish to find the minimizer of H over $B_\varepsilon(q)$,
- H is concave, $B_\varepsilon(q)$ is convex, so minimizer is an extreme point
- $B_\varepsilon(q)$ is a polytope; has finitely many extreme points.

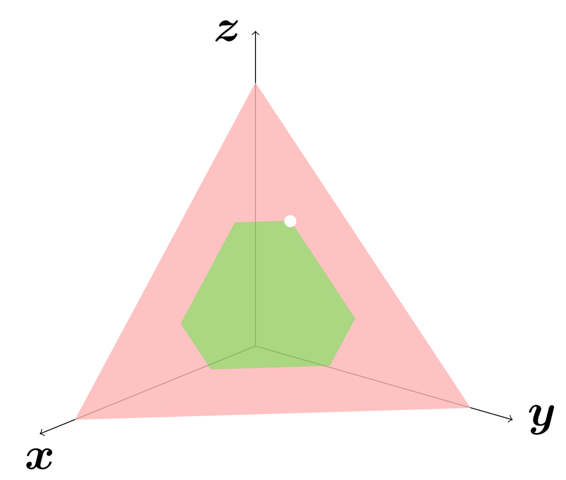
Interlude: Majorization

- Majorization is a *partial order* on Δ.
- For $p \in \Delta$, define $p^\downarrow$ as the perm. of p which is in descending order.
- p majorizes p', written $p \succ p'$, if $\sum_{i=1}^k p^\downarrow(i) \geq \sum_{i=1}^k p'^\downarrow(i)$ for each $k = 1, 2, \ldots, d$.

Thm. Set of p' majorized by p is the convex hull of the permutations of p:

$$\{p' \in \Delta : p \succ p'\} = \operatorname{conv}\{\pi(p) : \pi \in S_d\}.$$

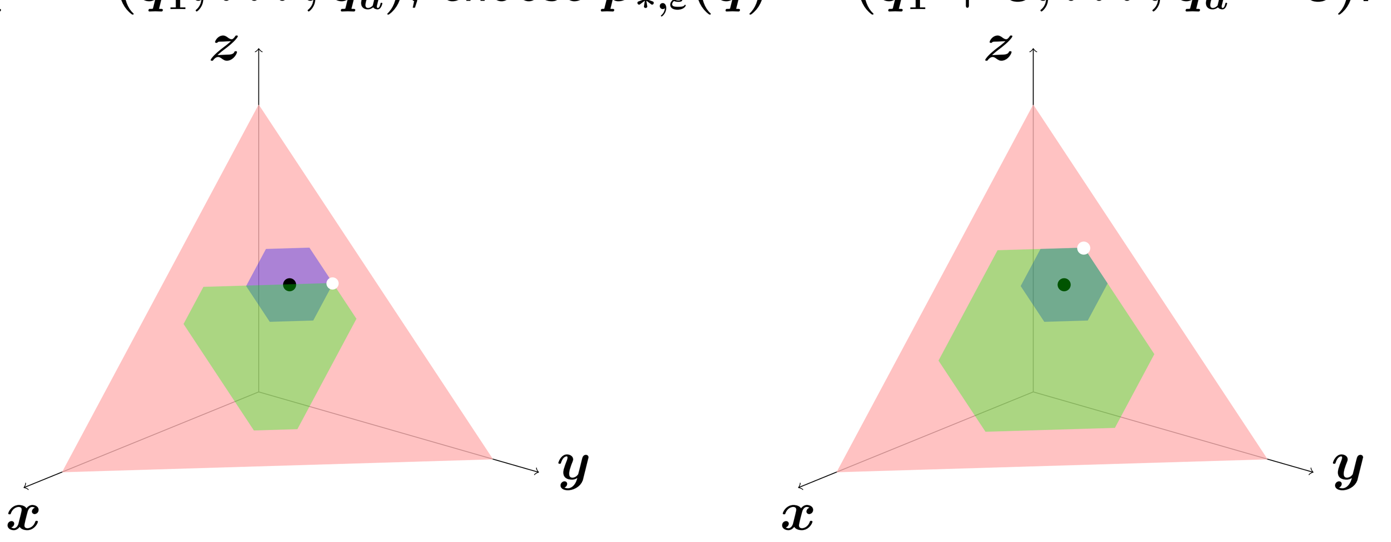
- For example, if p is the white dot, the green polygon is the set of probability distributions majorized by p.



- Connection to entropy: if $\varphi : \Delta \to \mathbb{R}$ is symmetric and concave, then $p \succ p'$ implies $\varphi(p) \leq \varphi(p')$. This is called *Schur concavity*.

Finding the minimizer (II)

- If $q^\downarrow = (q_1, \ldots, q_d)$, choose $p_{*,\varepsilon}(q) = (q_1 + \varepsilon, \ldots, q_d - \varepsilon)$.

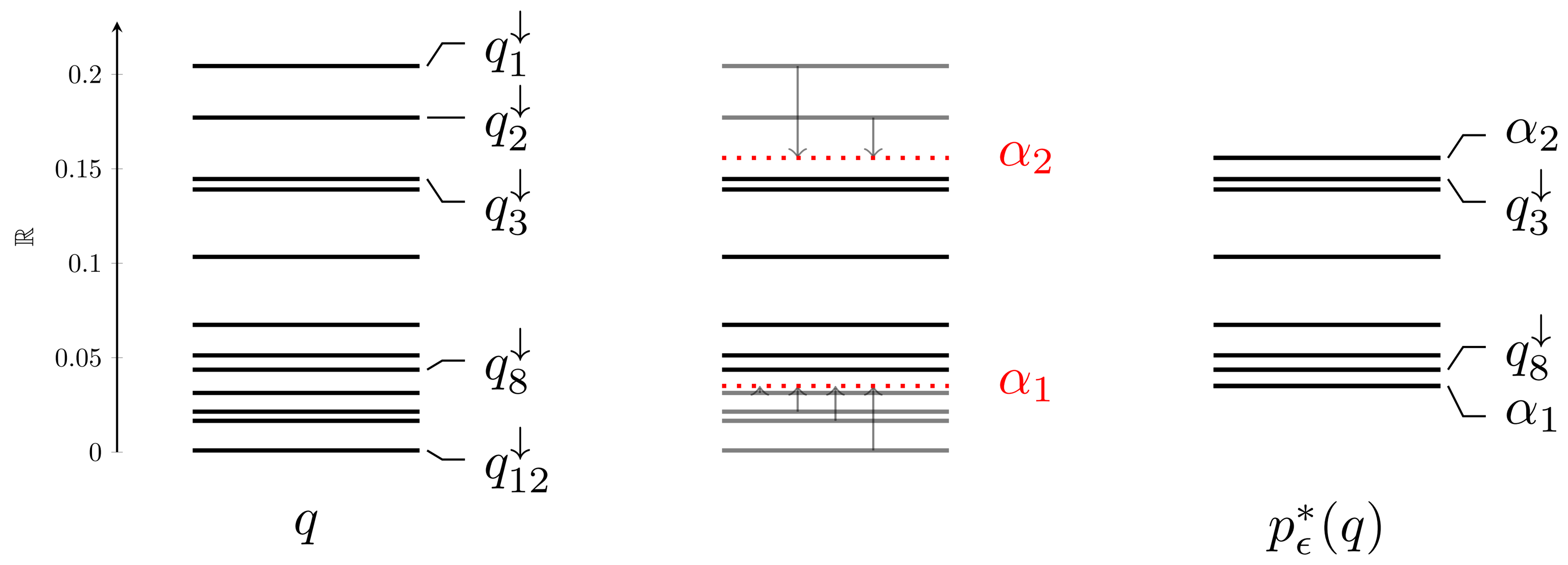


Left: Set of points majorized by a vertex; does not contain $B_\varepsilon(q)$.
Right: Set of points majorized by the vertex $p_{*,\varepsilon}(q)$. Thus $p_{*,\varepsilon}(q)$ majorizes any $p' \in B_\varepsilon(q)$, so minimizes *any* Schur concave function!

A hint from convex optimization

- Maximize differential concave function subject to 1-norm constraint (over quantum states or probability distributions).
- Commuting case: more tractable. The ith entry of the maximizer is either $q(i)$ or one of two values, α_1 or α_2.

The maximizer, by example ($d = 12$)



In fact, $p_\varepsilon^*(q)$ is majorized by every $p' \in B_\varepsilon(q)$! Maximizes *any* Schur concave function.

Performance of resulting LCB for the entropy

500 uniformly random choices of $q \in \Delta$ and $\varepsilon \in [0,1]$. For each q and ε, constructed $p_{*,\varepsilon}(q)$ and $p_\varepsilon^*(q)$, calculated the LCB, and plotted versus ε, in red. Below blue line is good!