Constrained Functional Value under General Convexity Conditions with Applications to Distributed Simulation

Yanjun Han (Stanford EE)

ISIT 2020

A motivating example

Log-concave function

A non-negative function f on $\mathbb{R}^n$ is called log-concave iff

$$f(x) = \exp(-g(x))$$

for some convex function g.

A motivating example

Log-concave function

A non-negative function f on $\mathbb{R}^n$ is called log-concave iff

$$f(x) = \exp(-g(x))$$

for some convex function g.

Theorem (Bobkov and Madiman 2011)

For any log-concave density f on $\mathbb{R}^n$, its differential entropy $h(f) = \int_{\mathbb{R}^n} -f(x) \log f(x) dx$ satisfies

$$\log\left(\frac{1}{f_{\max}}\right) \leq h(f) \leq \log\left(\frac{1}{f_{\max}}\right) + n,$$

where $f_{\max} = \sup_{x \in \mathbb{R}^n} f(x)$ is the sup-norm of the density f.

Extension to β-concave densities

β-concave function

A non-negative function f on $\mathbb{R}^n$ is called β-concave iff

$$f(x) = g(x)^{-\beta}$$

for some convex function g.

Extension to β-concave densities

β-concave function

A non-negative function f on $\mathbb{R}^n$ is called β-concave iff

$$f(x) = g(x)^{-\beta}$$

for some convex function g.

Theorem (Bobkov and Madiman 2011, Fradelizi et al. 2020)

For any β-concave density f on $\mathbb{R}^n$ with $\beta > n$, it holds that

$$\log\left(\frac{1}{f_{\max}}\right) \leq h(f) \leq \log\left(\frac{1}{f_{\max}}\right) + \sum_{i=1}^{n} \frac{\beta}{\beta - i},$$

Extension to β-concave densities

β-concave function

A non-negative function f on $\mathbb{R}^n$ is called β-concave iff

$$f(x) = g(x)^{-\beta}$$

for some convex function g.

Theorem (Bobkov and Madiman 2011, Fradelizi et al. 2020)

For any β-concave density f on $\mathbb{R}^n$ with $\beta > n$, it holds that

$$\log\left(\frac{1}{f_{\max}}\right) \leq h(f) \leq \log\left(\frac{1}{f_{\max}}\right) + \sum_{i=1}^{n} \frac{\beta}{\beta - i},$$

Intuition

The value of the density functional is constrained in a small range when the density satisfies certain convexity conditions.

A general setup

Question

Is there a general phenomenon for a wide class of functionals and convexity conditions?

A general setup

Question

Is there a general phenomenon for a wide class of functionals and convexity conditions?

Setup

General functional of density (ϕ-functional):

$$I_\phi(f) = \int_{\mathbb{R}^n} \phi(f(x))dx.$$

A general setup

Question

Is there a general phenomenon for a wide class of functionals and convexity conditions?

Setup

General functional of density (ϕ-functional):

$$I_\phi(f) = \int_{\mathbb{R}^n} \phi(f(x))dx.$$

General convexity condition (ψ-convexity): given a non-increasing function ψ,

$$f(x) = \psi(g(x)), \quad g \text{ convex}.$$

Target

Target

Find tight upper and lower bounds for the ϕ-functional of ψ-convex densities f:

$$LB(n, \phi, \psi, f_{\max}) \leq I_\phi(f) \leq UB(n, \phi, \psi, f_{\max}).$$

Target

Target

Find tight upper and lower bounds for the ϕ-functional of ψ-convex densities f:

$$LB(n, \phi, \psi, f_{\max}) \leq I_\phi(f) \leq UB(n, \phi, \psi, f_{\max}).$$

Main parameters:

- n: dimensionality of the density;

Target

Target

Find tight upper and lower bounds for the ϕ-functional of ψ-convex densities f:

$$LB(n, \phi, \psi, f_{\max}) \leq I_\phi(f) \leq UB(n, \phi, \psi, f_{\max}).$$

Main parameters:

- n: dimensionality of the density;
- ϕ: given function used in the target functional;

Target

Target

Find tight upper and lower bounds for the ϕ-functional of ψ-convex densities f:

$$LB(n, \phi, \psi, f_{\max}) \leq I_\phi(f) \leq UB(n, \phi, \psi, f_{\max}).$$

Main parameters:

- n: dimensionality of the density;
- ϕ: given function used in the target functional;
- ψ: given function used in the convexity condition;

Target

Target

Find tight upper and lower bounds for the ϕ-functional of ψ-convex densities f:

$$LB(n, \phi, \psi, f_{\max}) \leq I_\phi(f) \leq UB(n, \phi, \psi, f_{\max}).$$

Main parameters:

- n: dimensionality of the density;
- ϕ: given function used in the target functional;
- ψ: given function used in the convexity condition;
- $f_{\max}$: the sup-norm of the density.

Main inequality

Theorem (Main Inequality)

Let $b \triangleq \psi^{-1}(f_{\max})$, and $F_k, G_k : [b, \infty) \to \mathbb{R}$ be real-valued functions vanishing at the infinity such that

$$(-1)^k \frac{d^k}{dx^k} F_k(x) = \phi(\psi(x)), \quad (-1)^k \frac{d^k}{dx^k} G_k(x) = \psi(x),$$

for $k = 0, 1, \cdots, n$.

Main inequality

Theorem (Main Inequality)

Let $b \triangleq \psi^{-1}(f_{\max})$, and $F_k, G_k : [b, \infty) \to \mathbb{R}$ be real-valued functions vanishing at the infinity such that

$$(-1)^k \frac{d^k}{dx^k} F_k(x) = \phi(\psi(x)), \quad (-1)^k \frac{d^k}{dx^k} G_k(x) = \psi(x),$$

for $k = 0, 1, \cdots, n$. If there exists a real number A such that:

(i) $F_n(b) - A \cdot G_n(b) \leq 0$;

(ii) $F_0(b) - A \cdot G_0(b) \leq 0$;

(iii) The function $x \mapsto F_0(x) - A \cdot G_0(x)$ has at most one zero on $[b, \infty)$;

then $I_\phi(f) \leq A$.

Main inequality

Theorem (Main Inequality)

Let $b \triangleq \psi^{-1}(f_{\max})$, and $F_k, G_k : [b, \infty) \to \mathbb{R}$ be real-valued functions vanishing at the infinity such that

$$(-1)^k \frac{d^k}{dx^k} F_k(x) = \phi(\psi(x)), \quad (-1)^k \frac{d^k}{dx^k} G_k(x) = \psi(x),$$

for $k = 0, 1, \cdots, n$. If there exists a real number A such that:

(i) $F_n(b) - A \cdot G_n(b) \leq 0$;

(ii) $F_0(b) - A \cdot G_0(b) \leq 0$;

(iii) The function $x \mapsto F_0(x) - A \cdot G_0(x)$ has at most one zero on $[b, \infty)$;

then $I_\phi(f) \leq A$. Similarly, $I_\phi(f) \geq A$ if both $\leq$ in conditions (i) and (ii) are replaced by $\geq$.

Tightness of the main inequality

Recall that

$$(-1)^k \frac{d^k}{dx^k} F_k(x) = \phi(\psi(x)), \quad (-1)^k \frac{d^k}{dx^k} G_k(x) = \psi(x).$$

Tightness of the main inequality

Recall that

$$(-1)^k \frac{d^k}{dx^k} F_k(x) = \phi(\psi(x)), \quad (-1)^k \frac{d^k}{dx^k} G_k(x) = \psi(x).$$

Theorem (Tightness)

If condition (iii) holds for the following choices of A and A', then

$$\sup\{I_\phi(f) : f \text{ is } \psi\text{-convex}\} = A \triangleq \max\left\{\frac{F_0(b)}{G_0(b)}, \frac{F_n(b)}{G_n(b)}\right\},$$

$$\inf\{I_\phi(f) : f \text{ is } \psi\text{-convex}\} = A' \triangleq \min\left\{\frac{F_0(b)}{G_0(b)}, \frac{F_n(b)}{G_n(b)}\right\}.$$

Example I: differential entropy

Log-concave density

Let $\phi(x) = -x \log x$ and $\psi(x) = e^{-x}$. Then

$$b = \log\left(\frac{1}{f_{\max}}\right), \quad F_k(x) = (x+k)e^{-x}, \quad G_k(x) = e^{-x}.$$

Clearly condition (iii) holds, and the main inequality gives

$$\log\left(\frac{1}{f_{\max}}\right) = \frac{F_0(b)}{G_0(b)} \leq h(f) \leq \frac{F_n(b)}{G_n(b)} = \log\left(\frac{1}{f_{\max}}\right) + n.$$

Example I: differential entropy

Log-concave density

Let $\phi(x) = -x \log x$ and $\psi(x) = e^{-x}$. Then

$$b = \log\left(\frac{1}{f_{\max}}\right), \quad F_k(x) = (x+k)e^{-x}, \quad G_k(x) = e^{-x}.$$

Clearly condition (iii) holds, and the main inequality gives

$$\log\left(\frac{1}{f_{\max}}\right) = \frac{F_0(b)}{G_0(b)} \le h(f) \le \frac{F_n(b)}{G_n(b)} = \log\left(\frac{1}{f_{\max}}\right) + n.$$

β-concave density

For $\beta > n$, similar algebra for $\psi(x) = x^{-\beta}$ gives

$$\log\left(\frac{1}{f_{\max}}\right) \le h(f) \le \log\left(\frac{1}{f_{\max}}\right) + \sum_{i=1}^{n} \frac{\beta}{\beta - i}.$$

Example II: Rényi entropy

Log-concave density

Let $\phi(x) = x^\alpha$ and $\psi(x) = e^{-x}$. Then

$$b = \log\left(\frac{1}{f_{\max}}\right), \quad F_k(x) = \alpha^{-k} e^{-\alpha x}, \quad G_k(x) = e^{-x}.$$

Clearly condition (iii) holds, and the main inequality gives

$$\log\left(\frac{1}{f_{\max}}\right) \le h_\alpha(f) \le \log\left(\frac{1}{f_{\max}}\right) + \frac{n \log \alpha}{\alpha - 1}.$$

Example II: Rényi entropy

Log-concave density

Let $\phi(x) = x^\alpha$ and $\psi(x) = e^{-x}$. Then

$$b = \log\left(\frac{1}{f_{\max}}\right), \quad F_k(x) = \alpha^{-k} e^{-\alpha x}, \quad G_k(x) = e^{-x}.$$

Clearly condition (iii) holds, and the main inequality gives

$$\log\left(\frac{1}{f_{\max}}\right) \le h_\alpha(f) \le \log\left(\frac{1}{f_{\max}}\right) + \frac{n \log \alpha}{\alpha - 1}.$$

β-concave density

For $\min\{\beta, \alpha\beta\} > n$, similar algebra for $\psi(x) = x^{-\beta}$ gives

$$\log\left(\frac{1}{f_{\max}}\right) \le h_\alpha(f) \le \log\left(\frac{1}{f_{\max}}\right) + \frac{1}{\alpha - 1} \sum_{i=1}^{n} \log \frac{\alpha\beta - i}{\beta - i}.$$

Example III: truncated density

Let $\phi_t(x) = \min\{x, t\}$, with $0 < t < f_{\max}$. Then

$$I_{\phi_t}(f) = \int_{\mathbb{R}^n} \min\{f(x), t\} dx.$$

Example III: truncated density

Let $\phi_t(x) = \min\{x, t\}$, with $0 < t < f_{\max}$. Then

$$I_{\phi_t}(f) = \int_{\mathbb{R}^n} \min\{f(x), t\} dx.$$

Log-concave density

For $X \sim \text{Poisson}(\log(f_{\max}/t))$,

$$\mathbb{P}(X = 0) \leq I_{\phi_t}(f) \leq \mathbb{P}\left(X \leq n\right).$$

Example III: truncated density

Let $\phi_t(x) = \min\{x, t\}$, with $0 < t < f_{\max}$. Then

$$I_{\phi_t}(f) = \int_{\mathbb{R}^n} \min\{f(x), t\} dx.$$

Log-concave density

For $X \sim \text{Poisson}(\log(f_{\max}/t))$,

$$\mathbb{P}(X = 0) \leq I_{\phi_t}(f) \leq \mathbb{P}\left(X \leq n\right).$$

β-concave density

For integer $\beta > n$ and $Y \sim \text{Binomial}(\beta, 1 - (t/f_{\max})^{1/\beta})$,

$$\mathbb{P}(Y = 0) \leq I_{\phi_t}(f) \leq \mathbb{P}\left(Y \leq n\right).$$

Side results on probability theory

Corollary 1: variational representation of Binomial CDF

Fix any integers $n \geq k$ and $\lambda > 0$. Let $\mathcal{F}_{n,k}$ be the set of all n-concave densities on $\mathbb{R}^k$ with unit sup-norm, then

$$\mathbb{P}(\mathrm{Binomial}(n, 1 - e^{-\lambda/n}) \leq k) = \sup_{f \in \mathcal{F}_{n,k}} \int_{\mathbb{R}^k} \min\{f(x), e^{-\lambda}\} dx.$$

Side results on probability theory

Corollary 1: variational representation of Binomial CDF

Fix any integers $n \geq k$ and $\lambda > 0$. Let $\mathcal{F}_{n,k}$ be the set of all n-concave densities on $\mathbb{R}^k$ with unit sup-norm, then

$$\mathbb{P}(\text{Binomial}(n, 1 - e^{-\lambda/n}) \leq k) = \sup_{f \in \mathcal{F}_{n,k}} \int_{\mathbb{R}^k} \min\{f(x), e^{-\lambda}\} dx.$$

Corollary 2: an increasing Poisson limit

For any $\lambda > 0$, the random variables

$$X_n \sim \text{Binomial}(n, 1 - e^{-\lambda/n})$$

is the series of Binomial random variables Binomial(n, p_n) with largest success probability p_n such that X_n weakly converges to $X \sim \text{Poisson}(\lambda)$ and each X_n is stochastically dominated by X.

Proof sketch of the main inequality

An elementary proof starting from the first principle:

Proof sketch of the main inequality

An elementary proof starting from the first principle:

1. Let $h(u) = \mathrm{Vol}_n(\{x : g(x) \le u\})^{1/n}$, then the convexity of g implies that h is non-negative, increasing, and concave.

Proof sketch of the main inequality

An elementary proof starting from the first principle:

1. Let $h(u) = \mathrm{Vol}_n(\{x : g(x) \leq u\})^{1/n}$, then the convexity of g implies that h is non-negative, increasing, and concave.
2. Express both $I_\phi(f)$ and $\int_{\mathbb{R}^n} f(x)dx$ in terms of h:

$$I_\phi(f) = \int_b^\infty h(u)^n \phi'(\psi(u))(-\psi'(u))du,$$
$$1 = \int_b^\infty h(u)^n(-\psi'(u))du.$$

Proof sketch of the main inequality

An elementary proof starting from the first principle:

1. Let $h(u) = \mathrm{Vol}_n(\{x : g(x) \le u\})^{1/n}$, then the convexity of g implies that h is non-negative, increasing, and concave.
2. Express both $I_\phi(f)$ and $\int_{\mathbb{R}^n} f(x)dx$ in terms of h:

$$I_\phi(f) = \int_b^\infty h(u)^n \phi'(\psi(u))(-\psi'(u))du,$$
$$1 = \int_b^\infty h(u)^n(-\psi'(u))du.$$

3. Define $H_k(x) = F_k(x) - A \cdot G_k(x)$, then $I_\phi(f) \le A$ is equivalent to

$$\int_b^\infty h(u)^n H_0'(u)du \ge 0.$$

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u)du \geq 0.$$

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u)du \geq 0.$$

1. Base case $k = 0$: $S_0 = -H_n(b) \geq 0$ by condition (i);

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u)du \geq 0.$$

1. Base case $k = 0$: $S_0 = -H_n(b) \geq 0$ by condition (i);
2. Inductive step: integration by parts gives

$$S_{k+1} = \underbrace{-h(b)^{k+1} H_{n-k-1}(b)}_{\geq 0} + (k+1) \int_b^\infty h(u)^k h'(u) H'_{n-k}(u)du.$$

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u)du \geq 0.$$

1. Base case $k = 0$: $S_0 = -H_n(b) \geq 0$ by condition (i);
2. Inductive step: integration by parts gives

$$S_{k+1} = \underbrace{-h(b)^{k+1} H_{n-k-1}(b)}_{\geq 0} + (k+1) \int_b^\infty h(u)^k h'(u) H'_{n-k}(u) du.$$

- If $H'_{n-k}(u) \geq 0$ for all $u \geq b$, then the second term ≥ 0;

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u) du \geq 0.$$

1. Base case $k = 0$: $S_0 = -H_n(b) \geq 0$ by condition (i);
2. Inductive step: integration by parts gives

$$S_{k+1} = \underbrace{-h(b)^{k+1} H_{n-k-1}(b)}_{\geq 0} + (k+1) \int_b^\infty h(u)^k h'(u) H'_{n-k}(u) du.$$

   - If $H'_{n-k}(u) \geq 0$ for all $u \geq b$, then the second term ≥ 0;
   - Otherwise, condition (i)-(iii) implies $H'_{n-k}(u)$ has a unique zero $u = z \in [b, \infty)$, then $(h'(u) - h'(z)) H'_{n-k}(u) \geq 0$ and

$$S_{k+1} \geq (k+1) h'(z) \int_b^\infty h(u)^k H'_{n-k}(u) du = (k+1) h'(z) S_k \geq 0.$$

Proof sketch of the main inequality (cont'd)

Will prove by induction on $k = 0, 1, \cdots, n$ that

$$S_k \triangleq \int_b^\infty h(u)^k H'_{n-k}(u) du \geq 0.$$

1. Base case $k = 0$: $S_0 = -H_n(b) \geq 0$ by condition (i);
2. Inductive step: integration by parts gives

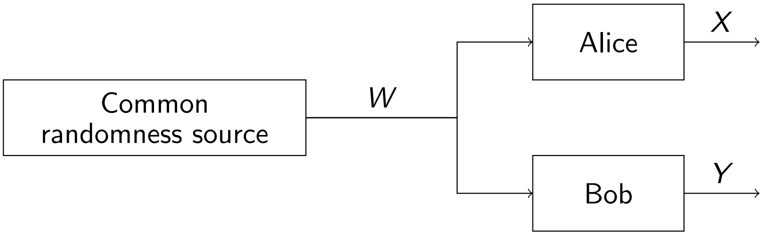
$$S_{k+1} = \underbrace{-h(b)^{k+1} H_{n-k-1}(b)}_{\geq 0} + (k+1) \int_b^\infty h(u)^k h'(u) H'_{n-k}(u) du.$$

   - If $H'_{n-k}(u) \geq 0$ for all $u \geq b$, then the second term ≥ 0;
   - Otherwise, condition (i)-(iii) implies $H'_{n-k}(u)$ has a unique zero $u = z \in [b, \infty)$, then $(h'(u) - h'(z)) H'_{n-k}(u) \geq 0$ and

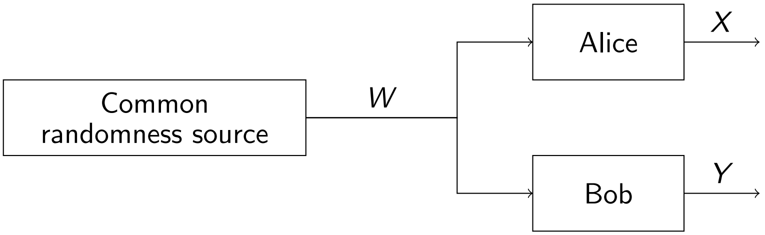
$$S_{k+1} \geq (k+1) h'(z) \int_b^\infty h(u)^k H'_{n-k}(u) du = (k+1) h'(z) S_k \geq 0.$$

3. Finally, $S_n \geq 0$ gives the desired result. $\square$

Application to distributed simulation

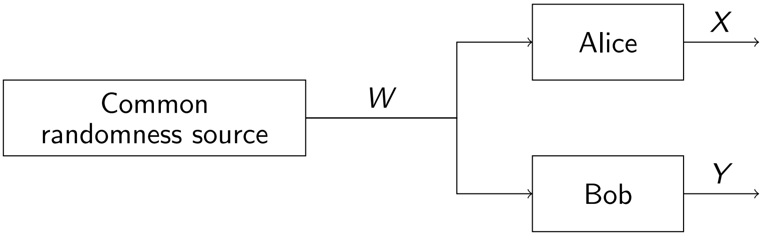


Application to distributed simulation



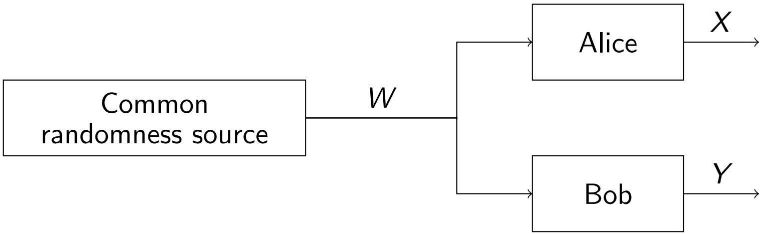
- $W = (W_1, W_2, \cdots)$ is i.i.d. Bern$(1/2)$ sequence known to both Alice and Bob

Application to distributed simulation



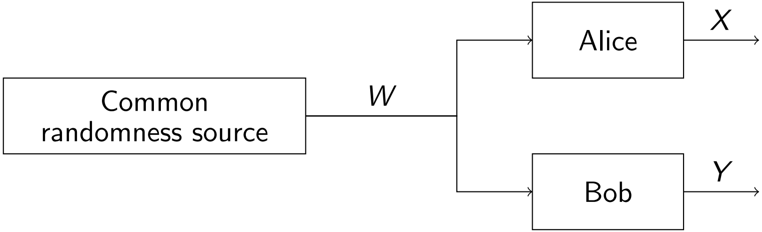
- $W = (W_1, W_2, \cdots)$ is i.i.d. Bern(1/2) sequence known to both Alice and Bob
- Alice and Bob each has local randomness, independent of each other and W

Application to distributed simulation



- $W = (W_1, W_2, \cdots)$ is i.i.d. Bern(1/2) sequence known to both Alice and Bob
- Alice and Bob each has local randomness, independent of each other and W
- Given a stopping time L and $(W_1, \cdots, W_L) = w$, Alice and Bob generates $X \sim q(x|w)$, $Y \sim q(y|w)$

Application to distributed simulation



- $W = (W_1, W_2, \cdots)$ is i.i.d. Bern(1/2) sequence known to both Alice and Bob
- Alice and Bob each has local randomness, independent of each other and W
- Given a stopping time L and $(W_1, \cdots, W_L) = w$, Alice and Bob generates $X \sim q(x|w)$, $Y \sim q(y|w)$

Target of exact simulation

Minimize $\mathbb{E}[L]$ over all possible generators such that $q(x,y) = p(x,y)$.

Exact common information

Definition (Exact Common Information)

$$G(X;Y) = \min_{X-W-Y} H(W).$$

Theorem (Kumar–Li–El Gamal 2014)

$$G(X;Y) \le \min_{\text{Generators}} \mathbb{E}[L] < G(X;Y) + 2.$$

Exact common information

Definition (Exact Common Information)

$$G(X;Y) = \min_{X-W-Y} H(W).$$

Theorem (Kumar–Li–El Gamal 2014)

$$G(X;Y) \leq \min_{\text{Generators}} \mathbb{E}[L] < G(X;Y) + 2.$$

Theorem (Li–El Gamal 2017)

If the probability density function of (X, Y) on $\mathbb{R}^2$ (with respect to the Lebesgue measure) is log-concave, then

$$I(X;Y) \leq G(X;Y) \leq I(X;Y) + 24.$$

The role of convex geometry

Question

Why do we need log-concave densities?

The role of convex geometry

Question

Why do we need log-concave densities?

- Prékopa–Leindler: log-concavity is preserved under marginalization

The role of convex geometry

Question

Why do we need log-concave densities?

- Prékopa–Leindler: log-concavity is preserved under marginalization
- Differential entropy is highly constrained under log-concavity

The role of convex geometry

Question

Why do we need log-concave densities?

- Prékopa–Leindler: log-concavity is preserved under marginalization
- Differential entropy is highly constrained under log-concavity
- Similar properties also hold for other convexity notions

The role of convex geometry

Question

Why do we need log-concave densities?

- Prékopa–Leindler: log-concavity is preserved under marginalization
- Differential entropy is highly constrained under log-concavity
- Similar properties also hold for other convexity notions

Theorem

If the probability density function of (X, Y) on $\mathbb{R}^2$ (with respect to the Lebesgue measure) is β-concave with $\beta \geq 2 + \varepsilon$, then

$$I(X; Y) \leq G(X; Y) \leq I(X; Y) + C(\varepsilon).$$

Generalization to n agents

Definition (General Exact Common Information)

$$G(X_1; X_2; \cdots; X_n) = \min_{X_1 \perp X_2 \perp \cdots \perp X_n | W} H(W).$$

Generalization to n agents

Definition (General Exact Common Information)

$$G(X_1; X_2; \cdots; X_n) = \min_{X_1 \perp X_2 \perp \cdots \perp X_n | W} H(W).$$

Definition (Dual Total Correlation)

$$I_D(X_1; X_2; \cdots; X_n) = H(X) - \sum_{i=1}^{n} H(X_i | X_{\setminus i}).$$

Generalization to n agents

Definition (General Exact Common Information)

$$G(X_1; X_2; \cdots; X_n) = \min_{X_1 \perp X_2 \perp \cdots \perp X_n | W} H(W).$$

Definition (Dual Total Correlation)

$$I_D(X_1; X_2; \cdots; X_n) = H(X) - \sum_{i=1}^{n} H(X_i | X_{\setminus i}).$$

Theorem

If the probability density function of $X = (X_1, X_2, \cdots, X_n)$ on $\mathbb{R}^n$ (with respect to the Lebesgue measure) is β-concave with $\beta \geq n + \varepsilon$,

$$I_D(X_1; X_2; \cdots; X_n) \leq G(X_1; X_2; \cdots; X_n) \leq I_D(X_1; X_2; \cdots; X_n) + C(\varepsilon) n^2.$$

Thank you!

Contact: yjhan@stanford.edu