Entropy and discrete random variables

Oliver Johnson
O.Johnson@bristol.ac.uk
https://people.maths.bris.ac.uk/∼maotj
Twitter: BristOliver

School of Mathematics, University of Bristol, UK

CIRM, 28th August 2017

General plan

General plan

- Gaussian random variables are lovely.

General plan

- Gaussian random variables are lovely.
    1. They are maximum entropy (under variance condition).

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).
  3. They are case of equality in EPI (and sharp log-Sobolev type results).

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).
  3. They are case of equality in EPI (and sharp log-Sobolev type results).
  4. Nice convexity properties (information geometry).

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).
  3. They are case of equality in EPI (and sharp log-Sobolev type results).
  4. Nice convexity properties (information geometry).
- Present some disjointed ideas about how to generalize to discrete random variables: Poisson, geometric, binomial etc.

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).
  3. They are case of equality in EPI (and sharp log-Sobolev type results).
  4. Nice convexity properties (information geometry).
- Present some disjointed ideas about how to generalize to discrete random variables: Poisson, geometric, binomial etc.
- Need 'log-concavity plus'?

General plan

- Gaussian random variables are lovely.
  1. They are maximum entropy (under variance condition).
  2. They are stable (class closed under addition and scaling).
  3. They are case of equality in EPI (and sharp log-Sobolev type results).
  4. Nice convexity properties (information geometry).
- Present some disjointed ideas about how to generalize to discrete random variables: Poisson, geometric, binomial etc.
- Need 'log-concavity plus'?
- Single summary in Johnson O. (2017) *Entropy and Thinning of Discrete Random Variables*. In: Carlen E., Madiman M., Werner E. (eds) Convexity and Concentration. IMA vol 161.

Outline of talk

- Poisson max entropy
- Entropy monotonicity
- Discrete log-Sobolev
- Entropy concavity

'Log-concavity plus' - idea 1 (ULC)

'Log-concavity plus' - idea 1 (ULC)

- Write Π_λ for the Poisson(λ) mass function.

'Log-concavity plus' - idea 1 (ULC)

- Write Π_λ for the Poisson(λ) mass function.

Definition (Pemantle, Liggett)

For any λ, define class of ultra-log-concave V with mass function P_V supported on $\mathbb{Z}_+$ satisfying

$$\mathbf{ULC}(\lambda) = \{V : \mathbb{E}V = \lambda \text{ and } P_V(v)/\Pi_\lambda(v) \text{ is log-concave}\}.$$

'Log-concavity plus' - idea 1 (ULC)

- Write Π_λ for the Poisson(λ) mass function.

Definition (Pemantle, Liggett)

For any λ, define class of ultra-log-concave V with mass function P_V supported on $\mathbb{Z}_+$ satisfying

$$\mathbf{ULC}(\lambda) = \{V : \mathbb{E}V = \lambda \text{ and } P_V(v)/\Pi_\lambda(v) \text{ is log-concave}\}.$$

- Class includes Bernoulli sums and Poisson.

'Log-concavity plus' - idea 1 (ULC)

- Write Π_λ for the Poisson(λ) mass function.

Definition (Pemantle, Liggett)

For any λ, define class of ultra-log-concave V with mass function P_V supported on $\mathbb{Z}_+$ satisfying

$$\mathbf{ULC}(\lambda) = \{V : \mathbb{E}V = \lambda \text{ and } P_V(v)/\Pi_\lambda(v) \text{ is log-concave}\}.$$

- Class includes Bernoulli sums and Poisson.
- Class preserved on summation.

Maximum entropy and $\mathbf{ULC}(\lambda)$

Theorem (OJ: *Stoch. Proc. Appl.* 2007, pp.791-802)

If $X \in \mathbf{ULC}(\lambda)$ *and* $Y \sim \Pi_\lambda$ *then the entropy* H *satisfies*

$$H(X) \leq H(Y),$$

with equality if and only if $X \sim \Pi_\lambda$.

Maximum entropy and **ULC**(λ)

Theorem (OJ: *Stoch. Proc. Appl.* 2007, pp.791-802)

*If $X \in$ **ULC**(λ) and $Y \sim \Pi_\lambda$ then the entropy H satisfies*

$$H(X) \leq H(Y),$$

with equality if and only if $X \sim \Pi_\lambda$.

- See also Harremoës, 2001.

Thinning discrete random variables

Definition (Rényi)

Given Y, define the α-thinned version of Y by

$$T_\alpha Y = \sum_{i=1}^{Y} B_i,$$

where $B_1, B_2 \ldots$ i.i.d. Bernoulli(α), independent of Y.

Thinning discrete random variables

Definition (Rényi)

Given Y, define the α-thinned version of Y by

$$T_\alpha Y = \sum_{i=1}^{Y} B_i,$$

where $B_1, B_2 \ldots$ i.i.d. Bernoulli(α), independent of Y.

- Thinning operation preserves several parametric families

Thinning discrete random variables

Definition (Rényi)

Given Y, define the α-thinned version of Y by

$$T_\alpha Y = \sum_{i=1}^{Y} B_i,$$

where $B_1, B_2 \ldots$ i.i.d. Bernoulli(α), independent of Y.

- Thinning operation preserves several parametric families
- T_α is discrete equivalent of scaling by $\sqrt{\alpha}$?

Monotonicity of entropy

Theorem (OJ–Yu, *IEEE Trans. Inform Thy.* 2010)

Given positive α_i *such that* $\sum_{i=1}^{n+1} \alpha_i = 1$*, and writing* $\alpha^{(j)} = 1 - \alpha_j$*, then for any independent ULC* X_i*,*

$$nH\left(\sum_{i=1}^{n+1} T_{\alpha_i} X_i\right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} H\left(\sum_{i \neq j} T_{\alpha_i/\alpha^{(j)}} X_i\right).$$

Monotonicity of entropy

Theorem (OJ–Yu, *IEEE Trans. Inform Thy.* 2010)

Given positive α_i *such that* $\sum_{i=1}^{n+1} \alpha_i = 1$*, and writing* $\alpha^{(j)} = 1 - \alpha_j$*, then for any independent ULC* X_i*,*

$$nH\left(\sum_{i=1}^{n+1} T_{\alpha_i} X_i\right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} H\left(\sum_{i \neq j} T_{\alpha_i/\alpha^{(j)}} X_i\right).$$

- Exact analogue of Artstein/Ball/Barthe/Naor result,

$$nh\left(\sum_{i=1}^{n+1} \sqrt{\alpha_i} X_i\right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} h\left(\sum_{i \neq j} \sqrt{\alpha_i/\alpha^{(j)}} X_i\right),$$

replacing scalings by thinnings.

Monotonicity of entropy

Theorem (OJ–Yu, *IEEE Trans. Inform Thy.* 2010)

Given positive α_i such that $\sum_{i=1}^{n+1} \alpha_i = 1$, and writing $\alpha^{(j)} = 1 - \alpha_j$, then for any independent ULC X_i,

$$nH\left(\sum_{i=1}^{n+1} T_{\alpha_i} X_i\right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} H\left(\sum_{i \neq j} T_{\alpha_i/\alpha^{(j)}} X_i\right).$$

- Exact analogue of Artstein/Ball/Barthe/Naor result,

$$nh\left(\sum_{i=1}^{n+1} \sqrt{\alpha_i} X_i\right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} h\left(\sum_{i \neq j} \sqrt{\alpha_i/\alpha^{(j)}} X_i\right),$$

  replacing scalings by thinnings.
- Yu implicitly proved corresponding result for relative entropy, no restriction on X_i.

'Log-concavity plus' - idea 2 (c-log-concavity)

'Log-concavity plus' - idea 2 (c-log-concavity)

Definition (Caputo et al.?)

Given probability mass function P_V supported on $\mathbb{Z}_+$, write

$$\mathcal{E}^{(V)}(x) := \frac{P_V(x)^2 - P_V(x-1)P_V(x+1)}{P_V(x)P_V(x+1)} = \frac{P_V(x)}{P_V(x+1)} - \frac{P_V(x-1)}{P_V(x)}.$$

'Log-concavity plus' - idea 2 (c-log-concavity)

Definition (Caputo et al.?)

Given probability mass function P_V supported on $\mathbb{Z}_+$, write

$$\mathcal{E}^{(V)}(x) := \frac{P_V(x)^2 - P_V(x-1)P_V(x+1)}{P_V(x)P_V(x+1)} = \frac{P_V(x)}{P_V(x+1)} - \frac{P_V(x-1)}{P_V(x)}.$$

Condition (c-log-concavity)

If $\mathcal{E}^{(V)}(x) \geq c$ for all $x \in Z_+$, we say V is c-log-concave.

'Log-concavity plus' - idea 2 (c-log-concavity)

Definition (Caputo et al.?)

Given probability mass function P_V supported on $\mathbb{Z}_+$, write

$$\mathcal{E}^{(V)}(x) := \frac{P_V(x)^2 - P_V(x-1)P_V(x+1)}{P_V(x)P_V(x+1)} = \frac{P_V(x)}{P_V(x+1)} - \frac{P_V(x-1)}{P_V(x)}.$$

Condition (c-log-concavity)

If $\mathcal{E}^{(V)}(x) \geq c$ for all $x \in Z_+$, we say V is c-log-concave.

- If $V = \Pi_\lambda$ then $\mathcal{E}^{(V)}(x) \equiv 1/\lambda$.

'Log-concavity plus' - idea 2 (c-log-concavity)

Definition (Caputo et al.?)

Given probability mass function P_V supported on $\mathbb{Z}_+$, write

$$\mathcal{E}^{(V)}(x) := \frac{P_V(x)^2 - P_V(x-1)P_V(x+1)}{P_V(x)P_V(x+1)} = \frac{P_V(x)}{P_V(x+1)} - \frac{P_V(x-1)}{P_V(x)}.$$

Condition (c-log-concavity)

If $\mathcal{E}^{(V)}(x) \geq c$ for all $x \in Z_+$, we say V is c-log-concave.

- If $V = \Pi_\lambda$ then $\mathcal{E}^{(V)}(x) \equiv 1/\lambda$.
- V ULC $\Longrightarrow$ c-log-concave with $c = P_V(1)/P_V(0)$.

Discrete LSI (OJ `arxiv:1507.06268`, *pending Ann. IHP*)

Theorem

Fix c-log-concave V supported on $\mathbb{Z}_+$. For any positive f:

$$\mathrm{Ent}_V(f) \leq \frac{1}{c} \sum_{x=0}^{\infty} V(x) f(x+1) \left(\log \left(\frac{f(x+1)}{f(x)} \right) - 1 + \frac{f(x)}{f(x+1)} \right)$$

Discrete LSI (OJ `arxiv:1507.06268`, *pending Ann. IHP*)

Theorem

Fix c-log-concave V supported on $\mathbb{Z}_+$. For any positive f:

$$\mathrm{Ent}_V(f) \leq \frac{1}{c} \sum_{x=0}^{\infty} V(x) f(x+1) \left(\log \left(\frac{f(x+1)}{f(x)} \right) - 1 + \frac{f(x)}{f(x+1)} \right)$$

- Is sharp: equality holds for $V = \Pi_\lambda$, $f(x) = \exp(ax+b)$.

Discrete LSI (OJ arxiv:1507.06268, *pending Ann. IHP*)

Theorem

Fix c-log-concave V supported on $\mathbb{Z}_+$. For any positive f:

$$\mathrm{Ent}_V(f) \leq \frac{1}{c} \sum_{x=0}^{\infty} V(x) f(x+1) \left(\log \left(\frac{f(x+1)}{f(x)} \right) - 1 + \frac{f(x)}{f(x+1)} \right)$$

- Is sharp: equality holds for $V = \Pi_\lambda$, $f(x) = \exp(ax+b)$.
- Proved using Caputo, Dai Pra, Posta's discrete Bakry-Émery theory.

- Strengthens and generalizes previous log-Sobolev inequalities of 1. Wu 2. Caputo et al 3. Bobkov and Ledoux.

- Strengthens and generalizes previous log-Sobolev inequalities of 1. Wu 2. Caputo et al 3. Bobkov and Ledoux.
- e.g. linearizing log and taking $V = \Pi_\lambda$ recover

$$\mathrm{Ent}_V(f) \leq \lambda \sum_{x=0}^{\infty} V(x) \frac{\Delta f(x)^2}{f(x)}.$$

- Strengthens and generalizes previous log-Sobolev inequalities of 1. Wu 2. Caputo et al 3. Bobkov and Ledoux.
- e.g. linearizing log and taking $V = \Pi_\lambda$ recover

$$\mathrm{Ent}_V(f) \leq \lambda \sum_{x=0}^{\infty} V(x) \frac{\Delta f(x)^2}{f(x)}.$$

- e.g. Recover the fact that if p has mean λ

$$D(p \| \Pi_\lambda) \leq \lambda D(p^{\#} \| p),$$

where $p^{\#}$ is the size-biased version of p.

- Strengthens and generalizes previous log-Sobolev inequalities of 1. Wu 2. Caputo et al 3. Bobkov and Ledoux.
- e.g. linearizing log and taking $V = \Pi_\lambda$ recover

$$\mathrm{Ent}_V(f) \leq \lambda \sum_{x=0}^{\infty} V(x) \frac{\Delta f(x)^2}{f(x)}.$$

- e.g. Recover the fact that if p has mean λ

$$D(p \| \Pi_\lambda) \leq \lambda D(p^\# \| p),$$

  where $p^\#$ is the size-biased version of p.
- (This is a log-Sobolev inequality of Wu, reproved more directly by Yu).

Concavity of entropy: Shepp–Olkin conjecture

Concavity of entropy: Shepp–Olkin conjecture

- Consider n independent Bernoulli random variables, with parameters $\mathbf{p} = (p_1, \ldots, p_n)$.

Concavity of entropy: Shepp–Olkin conjecture

- Consider n independent Bernoulli random variables, with parameters $\mathbf{p} = (p_1, \ldots, p_n)$.
- Their sum has mass function $f_{\mathbf{p}}(k)$ for $k = 0, 1, \ldots, n$.

Concavity of entropy: Shepp–Olkin conjecture

- Consider n independent Bernoulli random variables, with parameters $\mathbf{p} = (p_1, \ldots, p_n)$.
- Their sum has mass function $f_{\mathbf{p}}(k)$ for $k = 0, 1, \ldots, n$.
- Consider entropy of $f_{\mathbf{p}}$, defined by

$$H(\mathbf{p}) := -\sum_{k=0}^{n} f_{\mathbf{p}}(k) \log f_{\mathbf{p}}(k).$$

Concavity of entropy: Shepp–Olkin conjecture

- Consider n independent Bernoulli random variables, with parameters $\mathbf{p} = (p_1, \ldots, p_n)$.
- Their sum has mass function $f_{\mathbf{p}}(k)$ for $k = 0, 1, \ldots, n$.
- Consider entropy of $f_{\mathbf{p}}$, defined by

$$H(\mathbf{p}) := -\sum_{k=0}^{n} f_{\mathbf{p}}(k) \log f_{\mathbf{p}}(k).$$

Conjecture (Shepp–Olkin (1981))

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ is concave.

Concavity of entropy: Shepp–Olkin conjecture

- Consider n independent Bernoulli random variables, with parameters $\mathbf{p} = (p_1, \ldots, p_n)$.
- Their sum has mass function $f_{\mathbf{p}}(k)$ for $k = 0, 1, \ldots, n$.
- Consider entropy of $f_{\mathbf{p}}$, defined by

$$H(\mathbf{p}) := -\sum_{k=0}^{n} f_{\mathbf{p}}(k) \log f_{\mathbf{p}}(k).$$

Conjecture (Shepp–Olkin (1981))

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ is concave.

- Sufficient to consider concavity for affine t, i.e. take

$$p_i(t) = p_i(0)(1-t) + p_i(1)t.$$

Known cases

Known cases

- Folklore: $n = 1$.

Known cases

- Folklore: $n = 1$.
- Shepp–Olkin (1981): $n = 2, 3$ (claim with no proof).

Known cases

- Folklore: $n = 1$.
- Shepp–Olkin (1981): $n = 2, 3$ (claim with no proof).
- Shepp–Olkin (1981): for all i, $p_i(t) = t$ (binomial case).

Known cases

- Folklore: $n = 1$.
- Shepp–Olkin (1981): $n = 2, 3$ (claim with no proof).
- Shepp–Olkin (1981): for all i, $p_i(t) = t$ (binomial case).
- Yu–Johnson (2009): for all i, either $p_i(0) = 0$ or $p_i(1) = 0$.

Known cases

- Folklore: $n = 1$.
- Shepp–Olkin (1981): $n = 2, 3$ (claim with no proof).
- Shepp–Olkin (1981): for all i, $p_i(t) = t$ (binomial case).
- Yu–Johnson (2009): for all i, either $p_i(0) = 0$ or $p_i(1) = 0$.
- Hillion (2012): for all i, either $p_i(t) = t$ or $p_i(t)$ constant (binomial translation case).

Motivating example: binomial case

Example

Motivating example: binomial case

Example

- Write spatial derivative $\Delta^* f(k) = f(k) - f(k-1)$.

Motivating example: binomial case

Example

- Write spatial derivative $\Delta^* f(k) = f(k) - f(k-1)$.
- For $0 \leq p < q \leq 1$, define $p(t) = p(1-t) + qt$.

Motivating example: binomial case

Example

- Write spatial derivative $\Delta^* f(k) = f(k) - f(k-1)$.
- For $0 \leq p < q \leq 1$, define $p(t) = p(1-t) + qt$.
- Write $f_t(k) = \binom{n}{k} p(t)^k (1-p(t))^{n-k}$.

Motivating example: binomial case

Example

- Write spatial derivative $\Delta^* f(k) = f(k) - f(k-1)$.
- For $0 \leq p < q \leq 1$, define $p(t) = p(1-t) + qt$.
- Write $f_t(k) = \binom{n}{k} p(t)^k (1-p(t))^{n-k}$.
- Simple calculation (e.g. Mateev, Shepp–Olkin) shows:

$$\frac{\partial f_t(k)}{\partial t} = \Delta^* \left(n(q-p)\mathrm{Bin}_{n-1,p(t)}(k) \right).$$

Motivating example: binomial case

Example

- Write spatial derivative $\Delta^* f(k) = f(k) - f(k-1)$.
- For $0 \leq p < q \leq 1$, define $p(t) = p(1-t) + qt$.
- Write $f_t(k) = \binom{n}{k} p(t)^k (1-p(t))^{n-k}$.
- Simple calculation (e.g. Mateev, Shepp–Olkin) shows:

$$\frac{\partial f_t(k)}{\partial t} = \Delta^* \left(n(q-p) \mathrm{Bin}_{n-1,p(t)}(k) \right).$$

- Rewrite using an idea of Yu ('hypergeometric thinning'):

$$\mathrm{Bin}_{n-1,p}(k) = \frac{(k+1)}{n} \mathrm{Bin}_{n,p}(k+1) + \left(1 - \frac{k}{n}\right) \mathrm{Bin}_{n,p}(k).$$

Motivating example: binomial case (cont.)

Example

Motivating example: binomial case (cont.)

Example

- Suggests we introduce mixtures of mass functions:

$$\begin{aligned} \frac{\partial f_t(k)}{\partial t} &= \Delta^* \left(v G_t^{(\boldsymbol{\alpha})}(k) \right), \\ \text{for} \quad G_t^{(\boldsymbol{\alpha})}(k) &= \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k). \end{aligned}$$

Motivating example: binomial case (cont.)

Example

- Suggests we introduce mixtures of mass functions:

$$\begin{aligned} \frac{\partial f_t(k)}{\partial t} &= \Delta^* \left(v G_t^{(\boldsymbol{\alpha})}(k) \right), \\ \text{for} \quad G_t^{(\boldsymbol{\alpha})}(k) &= \alpha_t(k+1) f_t(k+1) + (1-\alpha_t(k)) f_t(k). \end{aligned}$$

- i.e. for binomial example take $\alpha_t(k) = k/n$ for all k and t and $v = n(q-p)$.

Motivating example: binomial case (cont.)

Example

- Suggests we introduce mixtures of mass functions:

$$\begin{aligned} \frac{\partial f_t(k)}{\partial t} &= \Delta^* \left(v G_t^{(\boldsymbol{\alpha})}(k) \right), \\ \text{for} \quad G_t^{(\boldsymbol{\alpha})}(k) &= \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k). \end{aligned}$$

- i.e. for binomial example take $\alpha_t(k) = k/n$ for all k and t and $v = n(q-p)$.
- Analogue of continuous transport, deduce discrete Benamou–Brenier formula.

Discrete Benamou–Brenier formula

Definition

Discrete Benamou–Brenier formula

Definition

- Write $\mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ for the set of probability mass functions $f_t(k)$, given end constraints $f_t(k)|_{t=0} = f_0(k)$ and $f_t(k)|_{t=1} = f_1(k)$.

Discrete Benamou–Brenier formula

Definition

- Write $\mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ for the set of probability mass functions $f_t(k)$, given end constraints $f_t(k)|_{t=0} = f_0(k)$ and $f_t(k)|_{t=1} = f_1(k)$.
- Write $\mathcal{A}$ for the set of $\alpha(k)$ with $\alpha_t(0) \equiv 0$, $\alpha_t(n) \equiv 1$ and with $0 \leq \alpha_t(k) \leq 1$ for all k.

Discrete Benamou–Brenier formula

Definition

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n by

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n by
  - $G_t^{(\boldsymbol{\alpha})}(k) = \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k)$,

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n by
  - $G_t^{(\boldsymbol{\alpha})}(k) = \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k)$,
  - $$\frac{\partial f_t}{\partial t}(k) = -\nabla_1 \left(v_{\alpha,t}(k) G_t^{(\boldsymbol{\alpha})}(k) \right),$$

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n by
  - $G_t^{(\boldsymbol{\alpha})}(k) = \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k)$,
  - $$\frac{\partial f_t}{\partial t}(k) = -\nabla_1 \left(v_{\alpha,t}(k) G_t^{(\boldsymbol{\alpha})}(k) \right),$$
  - $$V_n(f_0, f_1) = \left(\inf_{\substack{f_t \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1), \\ \alpha_t(k) \in \mathcal{A}}} \int_0^1 \left(\sum_{k=0}^{n-1} G_t^{(\boldsymbol{\alpha})}(k) v_{\alpha,t}(k)^2 \right) dt \right)^{1/2}.$$

Discrete Benamou–Brenier formula

Definition

- For $f_t(k) \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1)$ and $\boldsymbol{\alpha} \in \mathcal{A}$, define probability mass function $G_t^{(\boldsymbol{\alpha})}(k)$, velocity field $v_{\alpha,t}(k)$ and distance V_n by
  - $G_t^{(\boldsymbol{\alpha})}(k) = \alpha_t(k+1) f_t(k+1) + (1 - \alpha_t(k)) f_t(k)$,
  - $$\frac{\partial f_t}{\partial t}(k) = -\nabla_1 \left(v_{\alpha,t}(k) G_t^{(\boldsymbol{\alpha})}(k) \right),$$
  - $$V_n(f_0, f_1) = \left(\inf_{\substack{f_t \in \mathcal{P}_{\mathbb{Z}}(f_0, f_1), \\ \alpha_t(k) \in \mathcal{A}}} \int_0^1 \left(\sum_{k=0}^{n-1} G_t^{(\boldsymbol{\alpha})}(k) v_{\alpha,t}(k)^2 \right) dt \right)^{1/2}.$$
- Refer to any path achieving the infimum as a geodesic.

Discrete Benamou–Brenier formula

Definition

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

- *V_n is a metric for probability measures on $\{0, \ldots n\}$.*

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

- *V_n is a metric for probability measures on $\{0, \ldots n\}$.*
- *V_n defines a length space: for any geodesic f, distance $V_n(f_s, f_t) = |t-s| V_n(f_0, f_1)$.*

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

- *V_n is a metric for probability measures on $\{0, \dots n\}$.*
- *V_n defines a length space: for any geodesic f, distance $V_n(f_s, f_t) = |t-s| V_n(f_0, f_1)$.*
- *If there exists a constant speed path then*

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

- *V_n is a metric for probability measures on $\{0,\ldots n\}$.*
- *V_n defines a length space: for any geodesic f, distance $V_n(f_s,f_t) = |t-s| V_n(f_0,f_1)$.*
- *If there exists a constant speed path then*
  - *f_0 and f_1 are stochastically ordered.*

Discrete Benamou–Brenier formula

Definition

- Example: binomial path is geodesic with $v_{\alpha,t}(k) \equiv n(q-p)$.
- Call path with $v_{\alpha,t}(k)$ fixed in k and t a constant speed path.

Proposition

- *V_n is a metric for probability measures on $\{0, \ldots n\}$.*
- *V_n defines a length space: for any geodesic f, distance $V_n(f_s, f_t) = |t-s| V_n(f_0, f_1)$.*
- *If there exists a constant speed path then*
  - *f_0 and f_1 are stochastically ordered.*
  - *Wasserstein distance W_1 and V_n coincide.*

Framework for concavity of entropy

Framework for concavity of entropy

- Want conditions under which entropy is concave.

Framework for concavity of entropy

- Want conditions under which entropy is concave.
- Give conditions in terms of $\alpha_t(k)$ to generalize binomial case.

Framework for concavity of entropy

- Want conditions under which entropy is concave.
- Give conditions in terms of $\alpha_t(k)$ to generalize binomial case.
- Recall that in that case, $\alpha_t(k) \equiv k/n$.

k-monotonicity and t-monotonicity conditions

Condition (k-MON)

Given t, we say that the $\alpha_t(k)$ are k-monotone at t if

$$\alpha_t(k) \leq \alpha_t(k+1) \quad \textit{for all } k = 0, \ldots, n-1.$$

Condition (t-MON)

Given t, we say that the $\alpha_t(k)$ are t-monotone at t if

$$\frac{\partial \alpha_t(k)}{\partial t} \geq 0 \quad \textit{for all } k = 0, \ldots, n.$$

GLC condition

Condition (GLC)

We say $f_t(k)$ is α-generalized log-concave at t, if for all $k = 0, \ldots, n-2$,

$$\begin{aligned} & \alpha_t(k+1)(1-\alpha_t(k+1)) f_t(k+1)^2 \\ \geq \quad & \alpha_t(k+2)(1-\alpha_t(k)) f_t(k) f_t(k+2). \end{aligned}$$

Theorem (Hillion–OJ 2016)

Consider constant speed path $f_t(k)$ and associated optimal $\alpha(t)$. If Conditions k-MON, t-MON and GLC hold at given $t = t^*$, the entropy $H(f_t)$ is concave in t at $t = t^*$.

Entropy concavity theorem

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker 'Condition 4'.

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker 'Condition 4'.
- Condition 4 holds for Shepp–Olkin paths.

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker 'Condition 4'.
- Condition 4 holds for Shepp–Olkin paths.

Theorem (Shepp–Olkin Theorem)

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ is concave.

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker ‘Condition 4’.
- Condition 4 holds for Shepp–Olkin paths.

Theorem (Shepp–Olkin Theorem)

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ *is concave.*

- Proved in recent works by Hillion–OJ

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker 'Condition 4'.
- Condition 4 holds for Shepp–Olkin paths.

Theorem (Shepp–Olkin Theorem)

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ is concave.

- Proved in recent works by Hillion–OJ
- Case where all p_i' have same sign in Ann. Probab. vol 44/1, 2016, pages 276-306

Entropy concavity theorem

- However, t-MON condition fails for some Shepp–Olkin paths.
- Entropy remains concave if replace by t-MON by weaker 'Condition 4'.
- Condition 4 holds for Shepp–Olkin paths.

Theorem (Shepp–Olkin Theorem)

For any n, the function $\mathbf{p} \mapsto H(\mathbf{p})$ is concave.

- Proved in recent works by Hillion–OJ
- Case where all p_i' have same sign in Ann. Probab. vol 44/1, 2016, pages 276-306
- General case in *Bernoulli* vol 23/4B, 2017, pages 3638-3649

Open problem

Conjecture (Generalized Shepp–Olkin conjecture)

Open problem

Conjecture (Generalized Shepp–Olkin conjecture)

1. *There is a critical q_R^* such that the q-Rényi entropy of all Bernoulli sums is concave for $q \leq q_R^*$, and the entropy of some interpolation is convex for $q > q_R^*$.*

Open problem

Conjecture (Generalized Shepp–Olkin conjecture)

1. *There is a critical q_R^* such that the q-Rényi entropy of all Bernoulli sums is concave for $q \leq q_R^*$, and the entropy of some interpolation is convex for $q > q_R^*$.*
2. *There is a critical q_T^* such that the q-Tsallis entropy of all Bernoulli sums is concave for $q \leq q_T^*$, and the entropy of some interpolation is convex for $q > q_T^*$.*

Open problem

Conjecture (Generalized Shepp–Olkin conjecture)

1. *There is a critical q_R^* such that the q-Rényi entropy of all Bernoulli sums is concave for $q \leq q_R^*$, and the entropy of some interpolation is convex for $q > q_R^*$.*
2. *There is a critical q_T^* such that the q-Tsallis entropy of all Bernoulli sums is concave for $q \leq q_T^*$, and the entropy of some interpolation is convex for $q > q_T^*$.*

Indeed we conjecture that $q_R^* = 2$ and $q_T^* = 3.65986\ldots$, the root of $2 - 4q + 2^q = 0$.