Entropy and thinning of discrete random variables

Joint work with Daly, Harremoës, Hillion, Kontoyiannis, Yu

Oliver Johnson
O.Johnson@bristol.ac.uk
http://www.stats.bris.ac.uk/∼maotj
Twitter: @BristOliver

Statistics Group, University of Bristol

13th April 2015

Outline of talk

1. Maximum entropy
2. Poincaré inequality
3. Entropy monotonicity
4. Entropy concavity

Key class of discrete random variables: ULC

- Write Π_λ for Poisson(λ) mass function.

Definition (Pemantle, Liggett)

For any λ, define class of ultra-log-concave V with mass function P_V satisfying

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

- Equivalently $vP_V(v)^2 \geq (v+1)P_V(v+1)P_V(v-1)$, for all v.
- Class includes Bernoulli sums and Poisson.
- Class preserved on summation.
- Discrete version of Bakry-Émery condition?

Equivalent characterization of **ULC**(λ)

Definition (Kontoyiannis–Harremoës–OJ:
IEEE Trans. Inform. Theory 2005, pp.466–472)

For random variable V with mean λ, define scaled score function

$$\rho_V(v) = \frac{(v+1)P_V(v+1)}{\lambda P_V(v)} - 1,$$

and scaled Fisher information $K(V) = \lambda \mathbb{E} \rho_V(V)^2$.

- **ULC**(λ) equivalent to decreasing score ρ_V.

Properties of K (from KHJ paper)

- For $V_1, \ldots, V_n$ independent, the score of the sum

$$\rho_{V_1+\ldots+V_n}(x) = \mathbb{E}\left[\sum_i \alpha_i \rho_{V_i}(V_i) \middle| V_1 + \ldots + V_n = x \right],$$

where $\alpha_i = \lambda_{V_i} / (\lambda_{V_1} + \ldots + \lambda_{V_n})$.
- Hence

$$K(V_1 + \ldots + V_n) \leq \sum_i \alpha_i K(V_i).$$

- Leads to bounds of right order in Poisson approximation.
- Log-Sobolev (Bobkov/Ledoux) – for V with mean λ:

$$D(V \| \Pi_\lambda) \leq K(V).$$

Maximum entropy and **ULC**(λ)

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

If $X \in \textbf{ULC}(\lambda)$ *and* $Y \sim \Pi_\lambda$ *then*

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

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 operation preserves several parametric families e.g: Poisson, binomial, negative binomial.
- Believe T_α is the discrete equivalent of scaling by $\sqrt{\alpha}$.
- 'Mean-preserving transform' $T_\alpha X + T_{1-\alpha} Y$ equivalent to 'variance-preserving transform' $\sqrt{\alpha} X + \sqrt{1-\alpha} Y$ in continuous case? (Matches maximum entropy condition).

Key identity: gradient form

- Write $\Delta f(x) = f(x+1) - f(x)$ and $\Delta^* g(x) = g(x-1) - g(x)$.

Lemma

Write $P_\alpha(z) = \mathbb{P}(T_\alpha X + T_{1-\alpha} Y = z)$, where $X \sim P$ with mean λ and $Y \sim \Pi_\lambda$ then

$$\frac{\partial}{\partial \alpha} P_\alpha(z) \quad = \quad \frac{\lambda}{\alpha} \Delta^* (P_\alpha(z) \rho_\alpha(z)),$$

where ρ_α is the score of $T_\alpha X + T_{1-\alpha} Y$ in the KHJ sense above.

Proof of Maximum Entropy Property

- Then

$$-\frac{\partial}{\partial\alpha}\sum_z P_\alpha(z)\log\Pi_\lambda(z) = \frac{\lambda}{\alpha}\sum_z \Delta^*(P_\alpha(z)\rho_\alpha(z))\log\Pi_\lambda(z)$$
$$= \frac{\lambda}{\alpha}\sum_z P_\alpha(z)\rho_\alpha(z)\log\left(\frac{z+1}{\lambda}\right)$$

- $X \in \mathbf{ULC}(\lambda)$ makes this $= \mathrm{Cov}\,(\mathrm{decreasing}, \mathrm{increasing}) \leq 0$.
- $X \in \mathbf{ULC}(\lambda)$ makes $-\sum_z P_\alpha(z)\log\Pi_\lambda(z)$ decreasing in α.
- Since $P_0 = \Pi_\lambda$, and $P_1 = P$, deduce that

$$H(X) \leq -\sum_z P(z)\log\Pi_\lambda(z) \leq -\sum_z \Pi_\lambda(z)\log\Pi_\lambda(z) = H(Y).$$

Discrete Poincaré constant

Definition

Writing $\Delta f(x) = f(x+1) - f(x)$, then for random variable V

$$R_V = \sup_{g \in \mathcal{G}(V)} \left\{ \frac{\mathbb{E}[g(V)^2]}{\mathbb{E}[\Delta g(V)^2]} \right\},$$

where the supremum is taken over the set

$$\mathcal{G}(V) = \{ g : \mathbb{Z}^+ \mapsto \mathbb{R} \text{ with } \mathbb{E}[g(V)^2] < \infty \text{ and } \mathbb{E}[g(V)] = 0 \}.$$

- That is $\mathbb{E}[g(V)^2] \leq R_V \mathbb{E}\left[\Delta g(V)^2\right]$.

Discrete Poincaré inequality

Theorem (Daly–OJ:
Stat. Probab. Letters 2013, pp.511-518)

For $V \in \mathbf{ULC}(\lambda)$*:*

$$\mathrm{Var}\,(V) \leq R_V \leq \mathbb{E}V.$$

- Lower bound true for all variables (take $g(t) = t - \mathbb{E}V$).
- Upper bound special case of more general result.
- Define size-biased V^* by
  $\mathbb{P}(V^* = v) = (v+1)\mathbb{P}(V = v+1)/(\mathbb{E}V)$.
- We prove $R_V \leq \mathbb{E}V$ if V^* stochastically dominated by V.
- ULC property means V^* dominated by V in likelihood ratio ordering (stronger).
- Can this be extended to log-Sobolev inequality?

Monotonicity for continuous variables

Theorem (Artstein–Ball–Barthe–Naor)

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

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

- See also Tulino–Verdú and Madiman–Barron.
- Choosing $\alpha_i = 1/(n+1)$ for IID X_i shows $h\left(\sum_{i=1}^n X_i/\sqrt{n}\right)$ is monotone increasing in n.
- Equivalently relative entropy $D\left(\sum_{i=1}^n X_i/\sqrt{n} \middle\| Z\right)$ is monotone decreasing in n.
- Means CLT is equivalent of 2nd Law of Thermodynamics?

Entropy power and scaling

- Define $\mathcal{E}(t) = h(N(0,t)) = \frac{1}{2}\log_2(2\pi e t)$.
- Define entropy power $v(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)}/(2\pi e)$.
- Key fact about scaling:

$$v(\sqrt{\alpha}X) = \alpha v(X).$$

- This holds since $h(\sqrt{\alpha}X) = h(X) + \frac{1}{2}\log\alpha$, and $v(\sqrt{\alpha}X) = 2^{2h(\sqrt{\alpha}X)}/(2\pi e)$.

ECI and EPI

Theorem (Strengthened EPI)

Given independent continuous Y_i with finite variance, the entropy powers satisfy

$$nv\left(\sum_{i=1}^{n+1} Y_i\right) \geq \sum_{j=1}^{n+1} v\left(\sum_{i\neq j} Y_i\right),$$

with equality if and only if all the Y_i are Gaussian.

Lemma

Strengthened EPI equivalent to monotonicity.

- Proof uses key scaling fact above.

Discrete Monotonicity

- Write $D(X)$ for $D(X \| \Pi_{\mathbb{E}X})$.
- By convex ordering arguments, Yu showed that for i.i.d. X_i:
  1. relative entropy $D\left(\sum_{i=1}^n T_{1/n} X_i\right)$ is monotone decreasing in n,
  2. for ULC X_i the entropy $H\left(\sum_{i=1}^n T_{1/n} X_i\right)$ is monotone increasing in n.
- In fact, implicit in work of Yu is following stronger theorem:

Theorem

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 X_i,

$$nD\left(\sum_{i=1}^{n+1} T_{\alpha_i} X_i\right) \leq \sum_{j=1}^{n+1} \alpha^{(j)} D\left(\sum_{i \neq j} T_{\alpha_i/\alpha^{(j)}} X_i\right).$$

Monotonicity of entropy

Theorem (OJ–Yu: *IEEE Trans Inform Theory* 2010, pp.5387-5395)

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.

Discrete EPI

- Define $\mathcal{E}(t) = H(\Pi_t)$, an increasing, concave function.
- Define $V(X) = \mathcal{E}^{-1}(H(X))$.

Conjecture

Consider independent discrete X and Y. Then

$$V(X+Y) \geq V(X) + V(Y),$$

with equality if and only if X and Y are Poisson.

- Turns out not to be true!
- Even natural restrictions e.g. ULC, Bernoulli sums don't help
- Counterexample (not mine!): $X \sim Y$, $P_X(0) = 1/6$, $P_X(1) = 2/3$, $P_X(2) = 1/6$.

Scaling result

- Scaling result does partially hold.

Theorem (OJ–Yu)

Consider ULC X. For any α,

$$V(T_\alpha X) \geq \alpha V(X).$$

- Nonetheless (IMHO) no convincing single form of discrete EPI.

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

- 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

- 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

- Suggests we introduce mixtures of mass functions:

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

- Here $\alpha_t(k) = k/n$ for all k and t and $v = n(q-p)$.
- Again see a gradient form for derivative.
- Analogue of continuous transport, deduce discrete Benamou–Brenier formula.

Entropy concavity theorem

- Introduce conditions on α (monotonicity and generalized log-concavity).
- If these conditions hold, can prove entropy is concave.
- Able to deduce as special case:

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 `arXiv:1303.3381` – *Ann. Probab.* to appear (2015?)
- General case in recent preprint `arXiv:1503.01570`

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 $2 - 4q + 2^q = 0$.