Some results in maximum entropy

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

Statistics Group, University of Bristol, UK

EPFL/MLV meeting, Lausanne, 8th–9th September 2008

Maximum entropy distributions

Maximum entropy distributions

- Maximum entropy distributions often take simple form.

Maximum entropy distributions

- Maximum entropy distributions often take simple form.
- Fix mean and variance:

Maximum entropy distributions

- Maximum entropy distributions often take simple form.
- Fix mean and variance:
  continuous entropy h maximised by the normal.

Maximum entropy distributions

- Maximum entropy distributions often take simple form.
- Fix mean and variance:
  continuous entropy h maximised by the normal.
- Positive support and fixed mean:

Maximum entropy distributions

- Maximum entropy distributions often take simple form.
- Fix mean and variance:
  continuous entropy h maximised by the normal.
- Positive support and fixed mean:
  discrete entropy H maximised by geometric.

Standard proof by Gibbs inequality

Theorem

For any function f, on fixing $\sum p(x)f(x)$, the maximum entropy mass function is $\phi(x) = \alpha \exp(-\beta f(x))$.

Standard proof by Gibbs inequality

Theorem

For any function f, on fixing $\sum p(x) f(x)$, the maximum entropy mass function is $\phi(x) = \alpha \exp(-\beta f(x))$.

Proof.

$$\begin{aligned} -\sum_x p(x) \log \phi(x) &= \sum_x p(x) \left(-\log \alpha + \beta f(x)\right) \\ &= -\sum_x \phi(x) \log \phi(x) \end{aligned}$$

Standard proof by Gibbs inequality

Theorem

For any function f, on fixing $\sum p(x) f(x)$, the maximum entropy mass function is $\phi(x) = \alpha \exp(-\beta f(x))$.

Proof.

$$\begin{aligned} -\sum_x p(x) \log \phi(x) &= \sum_x p(x) \left(-\log \alpha + \beta f(x)\right) \\ &= -\sum_x \phi(x) \log \phi(x) \end{aligned}$$

Hence (in fact enough that $-\sum p \log \phi \leq -\sum \phi \log \phi$):

$$\begin{aligned} -H(p) + H(\phi) &= \sum_x p(x) \log p(x) - \sum_x p(x) \log \phi(x) \\ &= D(p \| \phi) \geq 0. \end{aligned}$$

More complicated classes

More complicated classes

- Gibbs formalism answers the wrong question?

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?
- What about Poisson distribution Π_λ with mean λ?

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?
- What about Poisson distribution Π_λ with mean λ?
- Natural well-behaved class over which Π_λ maximises entropy?

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?
- What about Poisson distribution Π_λ with mean λ?
- Natural well-behaved class over which Π_λ maximises entropy?
- Similarly, stable laws, in particular Cauchy, parameter c.

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?
- What about Poisson distribution Π_λ with mean λ?
- Natural well-behaved class over which Π_λ maximises entropy?
- Similarly, stable laws, in particular Cauchy, parameter c.
- Has anyone ever calculated $\mathbb{E} \log X!$ (Poisson) or $\mathbb{E} \log(c^2 + X^2)$ (Cauchy)

More complicated classes

- Gibbs formalism answers the wrong question?
- Answer very nice for Gaussians, but misleading?
- What about Poisson distribution Π_λ with mean λ?
- Natural well-behaved class over which Π_λ maximises entropy?
- Similarly, stable laws, in particular Cauchy, parameter c.
- Has anyone ever calculated $\mathbb{E} \log X!$ (Poisson) or $\mathbb{E} \log(c^2 + X^2)$ (Cauchy) ... other than to find entropy?

Manifesto

"Find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws . . ."

– Gnedenko and Korolev

Manifesto

"Find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws . . ."

– Gnedenko and Korolev

- Want 'natural' conditions for maximum entropy.

Manifesto

"Find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws ..."

– Gnedenko and Korolev

- Want 'natural' conditions for maximum entropy.
- Want them to depend on (pseudo)moments or other simple conditions.

Manifesto

"Find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws ..."

– Gnedenko and Korolev

- Want 'natural' conditions for maximum entropy.
- Want them to depend on (pseudo)moments or other simple conditions.
- Want stability – that is, classes preserved on summation.

Manifesto

> *"Find conditions under which certain limit laws appearing in limit theorems of probability theory possess extremal entropy properties. Immediate candidates to be subjected to such analysis are, of course, stable laws ..."*
>
> – Gnedenko and Korolev

- Want 'natural' conditions for maximum entropy.
- Want them to depend on (pseudo)moments or other simple conditions.
- Want stability – that is, classes preserved on summation.
- Work in progress – with Harremoës, Kontoyiannis and Madiman.

Poisson distribution

Poisson distribution

- Harremoës (2001) defines

$$B_n(\lambda) = \left\{ S : \mathbb{E}S = \lambda, S = \sum_{i=1}^{n} X_i, \quad X_i \text{ independent Bernoulli} \right\}.$$

Poisson distribution

- Harremoës (2001) defines

$$B_n(\lambda) = \left\{ S : \mathbb{E}S = \lambda, S = \sum_{i=1}^{n} X_i, \quad X_i \text{ independent Bernoulli} \right\}.$$

- Harremoës (2001) proved Π_λ has maximum entropy property:

$$\sup_{S \in \bigcup_n B_n(\lambda)} H(S) = H(\Pi_\lambda) \text{ for any } \lambda.$$

Poisson distribution

- Harremoës (2001) defines

$$B_n(\lambda) = \left\{ S : \mathbb{E}S = \lambda, S = \sum_{i=1}^n X_i, \quad X_i \text{ independent Bernoulli} \right\}.$$

- Harremoës (2001) proved Π_λ has maximum entropy property:

$$\sup_{S \in \bigcup_n B_n(\lambda)} H(S) = H(\Pi_\lambda) \text{ for any } \lambda.$$

- We give new proof, and larger closed class $\mathbf{ULC}(\lambda)$.

Ultra-log-concavity

Definition

For any λ, define the class of random variables V with mass function p_V satisfying

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

Ultra-log-concavity

Definition

For any λ, define the class of random variables V with mass function p_V satisfying

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

That is

$$i p_V(i)^2 \geq (i+1) p_V(i+1) p_V(i-1), \text{ for all } i.$$

Equivalent characterization of **ULC**(λ)

'Entropy and the Law of Small Numbers' (I. Kontoyiannis, P. Harremoës, O. Johnson)
IEEE Trans. Inform. Theory, Vol 51/2, 2005, pages 466–472

Equivalent characterization of **ULC**(λ)

'Entropy and the Law of Small Numbers' (I. Kontoyiannis, P. Harremoës, O. Johnson)
IEEE Trans. Inform. Theory, Vol 51/2, 2005, pages 466–472

Definition

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

$$\rho_V(i) = \frac{(i+1)p_V(i+1)}{\lambda p_V(i)} - 1,$$

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

Equivalent characterization of **ULC**(λ)

'Entropy and the Law of Small Numbers' (I. Kontoyiannis, P. Harremoës, O. Johnson)
IEEE Trans. Inform. Theory, Vol 51/2, 2005, pages 466–472

Definition

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

$$\rho_V(i) = \frac{(i+1)p_V(i+1)}{\lambda p_V(i)} - 1,$$

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

- Equivalently **ULC**(λ) is class of random variables V with mean λ and decreasing score ρ_V.

Properties of **ULC**

Lemma

(None of these are new results)

Properties of **ULC**

Lemma

(None of these are new results)

1. *Poisson* $\Pi_\lambda \in \mathbf{ULC}(\lambda)$.

Properties of **ULC**

Lemma

(None of these are new results)

1. *Poisson* $\Pi_\lambda \in \mathbf{ULC}(\lambda)$.
2. *For independent* $U \in \mathbf{ULC}(\lambda)$ *and* $V \in \mathbf{ULC}(\mu)$, $U + V \in \mathbf{ULC}(\lambda + \mu)$.

Properties of **ULC**

Lemma

(None of these are new results)

1. *Poisson* $\Pi_\lambda \in \mathbf{ULC}(\lambda)$.
2. *For independent* $U \in \mathbf{ULC}(\lambda)$ *and* $V \in \mathbf{ULC}(\mu)$, $U + V \in \mathbf{ULC}(\lambda + \mu)$.
3. $B_\infty(\lambda) \subseteq \mathbf{ULC}(\lambda)$.

Maximum entropy and **ULC**(λ)

O.T. Johnson 'Log-concavity and the maximum entropy property of the Poisson distribution'
Stoch. Proc. Appl. Vol 117/6, 2007, pages 791-802.

Maximum entropy and **ULC**(λ)

O.T. Johnson 'Log-concavity and the maximum entropy property of the Poisson distribution'
Stoch. Proc. Appl. Vol 117/6, 2007, pages 791-802.

Theorem

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

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

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

Adding and thinning

Definition

Adding and thinning

Definition

1. Given random variable X, define $S_\beta X \sim X + \Pi_\beta$

Adding and thinning

Definition

1. Given random variable X, define $S_\beta X \sim X + \Pi_\beta$
2. Given random variable Y, define the α-thinned rv

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

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

Adding and thinning

Definition

1. Given random variable X, define $S_\beta X \sim X + \Pi_\beta$
2. Given random variable Y, define the α-thinned rv

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

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

3. Given λ, define the combined map

$$U_\alpha = S_{\lambda(1-\alpha)} \circ T_\alpha.$$

Note: if X has mean λ then $U_\alpha X$ has mean λ.

Key properties in the proof

Lemma

If $Y \in \mathbf{ULC}(\mu)$ then $U_\alpha Y \in \mathbf{ULC}(\mu)$.

Key properties in the proof

Lemma

If $Y \in \textbf{ULC}(\mu)$ then $U_\alpha Y \in \textbf{ULC}(\mu)$.

Lemma

U has semigroup structure: $U_{\alpha_1\alpha_2} = U_{\alpha_1} \circ U_{\alpha_2}$.

Key properties in the proof

Lemma

If $Y \in \mathbf{ULC}(\mu)$ then $U_\alpha Y \in \mathbf{ULC}(\mu)$.

Lemma

U has semigroup structure: $U_{\alpha_1 \alpha_2} = U_{\alpha_1} \circ U_{\alpha_2}$.

Lemma

Take X with mean λ. Writing $P_\alpha(z) = \mathbb{P}(U_\alpha X = z)$, then

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

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

Proof of Maximum Entropy Property

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

Proof of Maximum Entropy Property

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

- This $= \mathrm{Cov}\,(\mathrm{decreasing}, \mathrm{increasing}) \leq 0$.

Proof of Maximum Entropy Property

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

- This $= \mathrm{Cov}\,(\mathrm{decreasing}, \mathrm{increasing}) \leq 0$.
- $X \in \mathbf{ULC}(\lambda)$ makes $-\sum_x P_\alpha(x) \log \Pi_\lambda(x)$ a decreasing function of α.

Proof of Maximum Entropy Property (cont.)

Proof of Maximum Entropy Property (cont.)

- Since $U_0 X \sim \Pi_\lambda$, and $U_1 X = X$, deduce that

$$-\sum_x P(x) \log \Pi_\lambda(x) \leq -\sum_x \Pi_\lambda(x) \log \Pi_\lambda(x).$$

Proof of Maximum Entropy Property (cont.)

- Since $U_0 X \sim \Pi_\lambda$, and $U_1 X = X$, deduce that

$$-\sum_x P(x) \log \Pi_\lambda(x) \leq -\sum_x \Pi_\lambda(x) \log \Pi_\lambda(x).$$

- Deduce that

$$-H(P) + H(\Pi_\lambda) \geq D(P \| \Pi_\lambda) \geq 0.$$

Similar ideas work for compound Poisson

Definition

Fix cluster distribution Q and write Q^{*y} for the yth convolution power of Q. Given distribution P of number of clusters, the corresponding Q-compound mass function

$$C_Q P(x) = \sum_y P(y) Q^{*y}(x).$$

Similar ideas work for compound Poisson

Definition

Fix cluster distribution Q and write Q^{*y} for the yth convolution power of Q. Given distribution P of number of clusters, the corresponding Q-compound mass function

$$C_Q P(x) = \sum_y P(y) Q^{*y}(x).$$

Example

Compound Poisson mass function

$$C_Q \Pi_\lambda(x) = \sum_y \Pi_\lambda(y) Q^{*y}(x).$$

Compound score and Fisher information

Definition

Given mass function P with mean λ, define score function

$$\rho_{C_Q P}(x) = \frac{\sum_{y=0}^{\infty}(y+1)P(y+1)Q^{*y}(x)}{\lambda \sum_{y=0}^{\infty} P(y)Q^{*y}(x)} - 1,$$

Compound score and Fisher information

Definition

Given mass function P with mean λ, define score function

$$\rho_{C_Q P}(x) = \frac{\sum_{y=0}^{\infty} (y+1) P(y+1) Q^{*y}(x)}{\lambda \sum_{y=0}^{\infty} P(y) Q^{*y}(x)} - 1,$$

and define corresponding Fisher information

$$K_Q(X) = \lambda (\mathbb{E} Q) \mathbb{E} \rho_{C_Q P}(X)^2.$$

Compound score and Fisher information

Definition

Given mass function P with mean λ, define score function

$$\rho_{C_Q P}(x) = \frac{\sum_{y=0}^{\infty} (y+1) P(y+1) Q^{*y}(x)}{\lambda \sum_{y=0}^{\infty} P(y) Q^{*y}(x)} - 1,$$

and define corresponding Fisher information

$$K_Q(X) = \lambda (\mathbb{E} Q) \mathbb{E} \rho_{C_Q P}(X)^2.$$

- K_Q has similar subadditivity/monotonicity properties to those of 'simple Fisher information' K – hence compound Poisson approximation bounds.

Maximum entropy property

Theorem (JKM)

If Q and $C_Q\Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

Maximum entropy property

Theorem (JKM)

If Q and $C_Q\Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

- Proof v similar - semigroup acts on cluster distribution P.

Maximum entropy property

Theorem (JKM)

If Q and $C_Q\Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

- Proof v similar - semigroup acts on cluster distribution P.
- Again, key property is decreasing score $\rho_{C_Q P}$.

Maximum entropy property

Theorem (JKM)

If Q and $C_Q \Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

- Proof v similar - semigroup acts on cluster distribution P.
- Again, key property is decreasing score $\rho_{C_Q P}$.
- Hard part is proving conditions for $C_Q \Pi_\lambda$ to be LC.

Maximum entropy property

Theorem (JKM)

If Q and $C_Q\Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

- Proof v similar - semigroup acts on cluster distribution P.
- Again, key property is decreasing score $\rho_{C_Q P}$.
- Hard part is proving conditions for $C_Q\Pi_\lambda$ to be LC.
- Works if Q is Bernoulli or geometric.

Maximum entropy property

Theorem (JKM)

If Q and $C_Q\Pi_\lambda$ are both log-concave, then for any $P \in \mathbf{ULC}(\lambda)$

$$H(C_Q P) \leq H(C_Q \Pi_\lambda).$$

- Proof v similar - semigroup acts on cluster distribution P.
- Again, key property is decreasing score $\rho_{C_Q P}$.
- Hard part is proving conditions for $C_Q\Pi_\lambda$ to be LC.
- Works if Q is Bernoulli or geometric.
- Similar theorem holds for compound Binomial distribution.

Real Entropy Power Inequality

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.
- Define $V(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)/(2\pi e)}$.

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.
- Define $V(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)/(2\pi e)}$.

Theorem

Consider independent continuous random variables X and Y. Then

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

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

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.
- Define $V(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)/(2\pi e)}$.

Theorem

Consider independent continuous random variables X and Y. Then

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

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

- First stated by Shannon

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.
- Define $V(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)/(2\pi e)}$.

Theorem

Consider independent continuous random variables X and Y. Then

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

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

- First stated by Shannon
- Lots of proofs (Stam/Blachman, Dembo/Cover/Thomas, Tulino/Verdu/Guo)

Real Entropy Power Inequality

- Define $\mathcal{E}(t) = h(N(0,t)) = 1/2 \log_2(2\pi e t)$.
- Define $V(X) = \mathcal{E}^{-1}(h(X)) = 2^{2h(X)/(2\pi e)}$.

Theorem

Consider independent continuous random variables X and Y. Then

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

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

- First stated by Shannon
- Lots of proofs (Stam/Blachman, Dembo/Cover/Thomas, Tulino/Verdu/Guo)
- Restricted versions easier to prove? (Costa)

Natural conjecture

Natural conjecture

- Define $\mathcal{E}(t) = h(\mathrm{Po(t)})$, an increasing, concave function.

Natural conjecture

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

Natural conjecture

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

Conjecture

Consider independent discrete random variables X and Y. Then

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

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

Natural conjecture

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

Conjecture

Consider independent discrete random variables 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!

Natural conjecture

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

Conjecture

Consider independent discrete random variables 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

Natural conjecture

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

Conjecture

Consider independent discrete random variables 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!): $X \sim Y$,
  $P_X(0) = 1/6$, $P_X(1) = 2/3$, $P_X(2) = 1/6$.

Natural conjecture

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

Conjecture

Consider independent discrete random variables 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!): $X \sim Y$, $P_X(0) = 1/6$, $P_X(1) = 2/3$, $P_X(2) = 1/6$.
- A lot easier to make conjectures than prove things!

Thinned Entropy Power Inequality

Conjecture (TEPI)

Consider independent discrete ULC random variables X and Y. For any α, conjecture that

$$V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1-\alpha) V(Y).$$

Thinned Entropy Power Inequality

Conjecture (TEPI)

Consider independent discrete ULC random variables X and Y. For any α, conjecture that

$$V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1-\alpha) V(Y).$$

- Sharp for Poisson ULC

Thinned Entropy Power Inequality

Conjecture (TEPI)

Consider independent discrete ULC random variables X and Y. For any α, conjecture that

$$V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1-\alpha) V(Y).$$

- Sharp for Poisson ULC
- Or maybe not all α?

Thinned Entropy Power Inequality

Conjecture (TEPI)

Consider independent discrete ULC random variables X and Y. For any α, conjecture that

$$V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1-\alpha) V(Y).$$

- Sharp for Poisson ULC
- Or maybe not all α?
- Natural conjecture for more variables, implies monotonicity of entropy in thinning Poisson convergence regime.

Thinned Entropy Power Inequality

Conjecture (TEPI)

Consider independent discrete ULC random variables X and Y. For any α, conjecture that

$$V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1-\alpha) V(Y).$$

- Sharp for Poisson ULC
- Or maybe not all α?
- Natural conjecture for more variables, implies monotonicity of entropy in thinning Poisson convergence regime.
- Taking $Y \sim \Pi_0$, TEPI $\Rightarrow$ RTEPI below.

Restricted, Thinned Entropy Power Inequality

Conjecture (RTEPI)

Consider ULC random variable X. For any α, conjecture that

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

Restricted, Thinned Entropy Power Inequality

Conjecture (RTEPI)

Consider ULC random variable X. For any α, conjecture that

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

- Theorem: True for X Bernoulli(p).

Weaker: Thinned entropy concavity inequality

Conjecture (TECI)

Consider independent ULC random variables X and Y. For any α, conjecture that

$$H(T_\alpha X + T_{1-\alpha} Y) \geq \alpha H(X) + (1-\alpha) H(Y).$$

Weaker: Thinned entropy concavity inequality

Conjecture (TECI)

Consider independent ULC random variables X and Y. For any α, conjecture that

$$H(T_\alpha X + T_{1-\alpha} Y) \geq \alpha H(X) + (1-\alpha) H(Y).$$

- TECI relates to Shepp-Olkin conjecture.

Weaker: Thinned entropy concavity inequality

Conjecture (TECI)

Consider independent ULC random variables X and Y. For any α, conjecture that

$$H(T_\alpha X + T_{1-\alpha} Y) \geq \alpha H(X) + (1-\alpha) H(Y).$$

- TECI relates to Shepp-Olkin conjecture.
- Concavity of $\mathcal{E}$ means TEPI $\Rightarrow$ TECI.

Weaker: Thinned entropy concavity inequality

Conjecture (TECI)

Consider independent ULC random variables X and Y. For any α, conjecture that

$$H(T_\alpha X + T_{1-\alpha} Y) \geq \alpha H(X) + (1-\alpha) H(Y).$$

- TECI relates to Shepp-Olkin conjecture.
- Concavity of $\mathcal{E}$ means TEPI $\Rightarrow$ TECI.
- Continuous versions EPI $\Leftrightarrow$ ECI (Dembo/Cover/Thomas).

Weaker: Thinned entropy concavity inequality

Conjecture (TECI)

Consider independent ULC random variables X and Y. For any α, conjecture that

$$H(T_\alpha X + T_{1-\alpha} Y) \geq \alpha H(X) + (1-\alpha) H(Y).$$

- TECI relates to Shepp-Olkin conjecture.
- Concavity of $\mathcal{E}$ means TEPI $\Rightarrow$ TECI.
- Continuous versions EPI $\Leftrightarrow$ ECI (Dembo/Cover/Thomas).
- Theorem: TECI holds when Y is Poisson.

Close to the TEPI?

Theorem

Consider independent ULC random variables X and Y. For any β, γ such that

$$\frac{\beta}{1-\gamma} \leq \frac{V(Y)}{V(X)} \leq \frac{1-\beta}{\gamma},$$

if RTEPI and TECI hold then

$$V(T_\beta X + T_\gamma Y) \geq \beta V(X) + \gamma V(Y).$$

Close to the TEPI?

Theorem

Consider independent ULC random variables X and Y. For any β, γ such that

$$\frac{\beta}{1-\gamma} \leq \frac{V(Y)}{V(X)} \leq \frac{1-\beta}{\gamma},$$

if RTEPI and TECI hold then

$$V(T_\beta X + T_\gamma Y) \geq \beta V(X) + \gamma V(Y).$$

- Note restrictions appearing on coefficients.

Close to the TEPI?

Theorem

Consider independent ULC random variables X and Y. For any β, γ such that

$$\frac{\beta}{1-\gamma} \leq \frac{V(Y)}{V(X)} \leq \frac{1-\beta}{\gamma},$$

if RTEPI and TECI hold then

$$V(T_\beta X + T_\gamma Y) \geq \beta V(X) + \gamma V(Y).$$

- Note restrictions appearing on coefficients.
- In general $\beta + \gamma < 1$.

Close to the TEPI?

Theorem

Consider independent ULC random variables X and Y. For any β, γ such that

$$\frac{\beta}{1-\gamma} \leq \frac{V(Y)}{V(X)} \leq \frac{1-\beta}{\gamma},$$

if RTEPI and TECI hold then

$$V(T_\beta X + T_\gamma Y) \geq \beta V(X) + \gamma V(Y).$$

- Note restrictions appearing on coefficients.
- In general $\beta + \gamma < 1$.
- Rephrase as

$$V(T_\beta X + T_\gamma Y + T_{1-\beta-\gamma}\Pi_0) \geq \beta V(X) + \gamma V(Y) + (1-\beta-\gamma)H(\Pi_0).$$