Stability of the Shannon-Stam Inequality

Dan Mikulincer

Students Probability Day, 2019

Weizmann Institute of Science
Joint work with Ronen Eldan

Relative Entropy

The central quantity we will deal is relative entropy:

Definition (Relative Entropy)

Let $X \sim \mu$, $Y \sim \nu$ be random vectors in $\mathbb{R}^d$, define the entropy of X, relative to Y as

$$\mathrm{Ent}(X||Y) = \mathrm{Ent}(\mu||\nu) := \begin{cases} \int_{\mathbb{R}^d} \ln\left(\frac{d\mu}{d\nu}\right) d\mu & \text{if } \mu \ll \nu \\ \infty & \text{otherwise} \end{cases}.$$

The Shannon-Stam Inequality

In $48'$ Shannon noted the following inequality, which was later proved by Stam, in $56'$.

Theorem (Shannon-Stam Inequality)

Let X, Y be random vectors in $\mathbb{R}^d$ and let $G \sim \mathcal{N}(0, \mathrm{I})$ be a random vector with the law of the standard Gaussian. Then, for any $\lambda \in [0,1]$

$$\mathrm{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y||G) \leq \lambda \mathrm{Ent}(X||G) + (1-\lambda)\mathrm{Ent}(Y||G).$$

Moreover, equality holds if and only if X and Y are Gaussians with identical covariances.

Remark: Shannon and Stam actually proved an equivalent form of the inequality, called the entropy power inequality. The equivalence was observed by Lieb in 78'.

The Shannon-Stam Inequality

In $48'$ Shannon noted the following inequality, which was later proved by Stam, in $56'$.

Theorem (Shannon-Stam Inequality)

Let X, Y be random vectors in $\mathbb{R}^d$ and let $G \sim \mathcal{N}(0, \mathrm{I})$ be a random vector with the law of the standard Gaussian. Then, for any $\lambda \in [0, 1]$

$$\mathrm{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y || G) \leq \lambda \mathrm{Ent}(X||G) + (1-\lambda)\mathrm{Ent}(Y||G).$$

Moreover, equality holds if and only if X and Y are Gaussians with identical covariances.

Remark: Shannon and Stam actually proved an equivalent form of the inequality, called the entropy power inequality. The equivalence was observed by Lieb in 78'.

Stability

Define the deficit

$$\delta_\lambda(X, Y) = \lambda \mathrm{Ent}(X||G) + (1-\lambda)\mathrm{Ent}(Y||G) - \mathrm{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y||G).$$

The question of stability deals with approximate equality cases.

Question

Suppose that $\delta_\lambda(X, Y)$ is small, must X and Y be 'close' to Gaussian vectors, which are themselves 'close' to each other?

We will now show that the deficit can be bounded in terms of a stochastic process and that in certain cases this gives a positive answer to the above question.

Stability

Define the deficit

$$\delta_\lambda(X, Y) = \lambda \mathrm{Ent}(X||G) + (1-\lambda)\mathrm{Ent}(Y||G) - \mathrm{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y||G).$$

The question of stability deals with approximate equality cases.

Question

Suppose that $\delta_\lambda(X, Y)$ is small, must X and Y be 'close' to Gaussian vectors, which are themselves 'close' to each other?

We will now show that the deficit can be bounded in terms of a stochastic process and that in certain cases this gives a positive answer to the above question.

Stability

Define the deficit

$$\delta_\lambda(X, Y) = \lambda \mathrm{Ent}(X||G) + (1-\lambda)\mathrm{Ent}(Y||G) - \mathrm{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y||G).$$

The question of stability deals with approximate equality cases.

Question

Suppose that $\delta_\lambda(X, Y)$ is small, must X and Y be 'close' to Gaussian vectors, which are themselves 'close' to each other?

We will now show that the deficit can be bounded in terms of a stochastic process and that in certain cases this gives a positive answer to the above question.

Föllmer Martingales

We focus on the one dimensional case and $\lambda = \frac{1}{2}$.
Let X be centered random variable, and let B_t denote a standard Brownian motion. Fölmmer (1984) and then Lehec (2011) have shown that there exists a process Γ_t^X, such that

- $\int_0^1 \Gamma_t^X dB_t$ has the law of X.
- $\mathrm{Ent}(X||G) = \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t}dt$.
- If H_t^X is another process such that $\int_0^1 H_t^X dB_t$ has the law of X,

$$\int_0^1 \frac{\mathbb{E}\left[(1-H_t^X)^2\right]}{1-t}dt \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t}dt.$$

Föllmer Martingales

We focus on the one dimensional case and $\lambda = \frac{1}{2}$.
Let X be centered random variable, and let B_t denote a standard Brownian motion. Fölmmer (1984) and then Lehec (2011) have shown that there exists a process Γ_t^X, such that

- $\int_0^1 \Gamma_t^X dB_t$ has the law of X.
- $\mathrm{Ent}(X||G) = \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t}dt$.
- If H_t^X is another process such that $\int_0^1 H_t^X dB_t$ has the law of X,

$$\int_0^1 \frac{\mathbb{E}\left[(1-H_t^X)^2\right]}{1-t}dt \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t}dt.$$

Föllmer Martingales

We focus on the one dimensional case and $\lambda = \frac{1}{2}$.

Let X be centered random variable, and let B_t denote a standard Brownian motion. Fölmmer (1984) and then Lehec (2011) have shown that there exists a process Γ_t^X, such that

- $\int_0^1 \Gamma_t^X dB_t$ has the law of X.
- $\mathrm{Ent}(X||G) = \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t} dt$.
- If H_t^X is another process such that $\int_0^1 H_t^X dB_t$ has the law of X,

$$\int_0^1 \frac{\mathbb{E}\left[(1-H_t^X)^2\right]}{1-t} dt \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t} dt.$$

Föllmer Martingales

We focus on the one dimensional case and $\lambda = \frac{1}{2}$.

Let X be centered random variable, and let B_t denote a standard Brownian motion. Fölmmer (1984) and then Lehec (2011) have shown that there exists a process Γ_t^X, such that

- $\int_0^1 \Gamma_t^X dB_t$ has the law of X.
- $\mathrm{Ent}(X||G) = \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t} dt$.
- If H_t^X is another process such that $\int_0^1 H_t^X dB_t$ has the law of X,

$$\int_0^1 \frac{\mathbb{E}\left[(1-H_t^X)^2\right]}{1-t} dt \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{1-t} dt.$$

Bounding the Deficit

Now, for X, Y random variables, take two independent Brownian motions B_t^X, B_t^Y and Γ_t^X, Γ_t^Y as above. Note that if G_1 and G_2 are standard Gaussians, then for any $a, b \in \mathbb{R}$

$$aG_1 + bG_2 \stackrel{\text{law}}{=} \sqrt{a^2 + b^2}\,G,$$

where G is another standard Gaussian.

This implies

$$\frac{X+Y}{\sqrt{2}} = \frac{1}{\sqrt{2}}\left(\int\limits_0^1 \Gamma_t^X dB_t^X + \int\limits_0^1 \Gamma_t^Y dB_t^Y\right) \stackrel{\text{law}}{=} \int\limits_0^1 \sqrt{\frac{(\Gamma_t^X)^2 + (\Gamma_t^Y)^2}{2}}\,dB_t.$$

for some Brownian motion B_t.

Bounding the Deficit

Now, for X, Y random variables, take two independent Brownian motions B_t^X, B_t^Y and Γ_t^X, Γ_t^Y as above. Note that if G_1 and G_2 are standard Gaussians, then for any $a, b \in \mathbb{R}$

$$aG_1 + bG_2 \overset{\text{law}}{=} \sqrt{a^2 + b^2}\, G,$$

where G is another standard Gaussian.

This implies

$$\frac{X+Y}{\sqrt{2}} = \frac{1}{\sqrt{2}} \left(\int_0^1 \Gamma_t^X dB_t^X + \int_0^1 \Gamma_t^Y dB_t^Y \right) \overset{\text{law}}{=} \int_0^1 \sqrt{\frac{(\Gamma_t^X)^2 + (\Gamma_t^Y)^2}{2}}\, dB_t.$$

for some Brownian motion B_t.

Bounding the Deficit

Now, for X, Y random variables, take two independent Brownian motions B_t^X, B_t^Y and Γ_t^X, Γ_t^Y as above. Note that if G_1 and G_2 are standard Gaussians, then for any $a, b \in \mathbb{R}$

$$aG_1 + bG_2 \overset{\text{law}}{=} \sqrt{a^2 + b^2}\, G,$$

where G is another standard Gaussian.

This implies

$$\frac{X + Y}{\sqrt{2}} = \frac{1}{\sqrt{2}} \left(\int_0^1 \Gamma_t^X dB_t^X + \int_0^1 \Gamma_t^Y dB_t^Y \right) \overset{\text{law}}{=} \int_0^1 \sqrt{\frac{(\Gamma_t^X)^2 + (\Gamma_t^Y)^2}{2}} dB_t.$$

for some Brownian motion B_t.

Bounding the Deficit

If $H_t = \sqrt{\frac{(\Gamma_t^X)^2+(\Gamma_t^Y)^2}{2}}$, $\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}||G\right) \leq \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt.$

Consequently,

$$2\delta_{\frac{1}{2}}(X,Y) \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^Y)^2\right]}{2(1-t)} + \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{2(1-t)} - \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$$
$$= \int_0^1 \frac{2\mathbb{E}[H_t] - \mathbb{E}[\Gamma_t^X] - \mathbb{E}[\Gamma_t^Y]}{1-t}.$$

Using concavity of the square root then shows

$$\delta_{\frac{1}{2}}(X,Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{(1-t)(\Gamma_t^X+\Gamma_t^Y)}\right]dt.$$

Bounding the Deficit

If $H_t = \sqrt{\frac{(\Gamma_t^X)^2+(\Gamma_t^Y)^2}{2}}$, $\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}||G\right) \leq \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$.

Consequently,

$$2\delta_{\frac{1}{2}}(X,Y) \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^Y)^2\right]}{2(1-t)} + \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{2(1-t)} - \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$$

$$= \int_0^1 \frac{2\mathbb{E}[H_t] - \mathbb{E}[\Gamma_t^X] - \mathbb{E}[\Gamma_t^Y]}{1-t}.$$

Using concavity of the square root then shows

$$\delta_{\frac{1}{2}}(X,Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{(1-t)(\Gamma_t^X + \Gamma_t^Y)}\right] dt.$$

Bounding the Deficit

If $H_t = \sqrt{\frac{(\Gamma_t^X)^2+(\Gamma_t^Y)^2}{2}}$, $\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}||G\right) \leq \frac{1}{2}\int_0^1 \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$.

Consequently,

$$2\delta_{\frac{1}{2}}(X,Y) \geq \int_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^Y)^2\right]}{2(1-t)} + \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{2(1-t)} - \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$$

$$= \int_0^1 \frac{2\mathbb{E}[H_t] - \mathbb{E}[\Gamma_t^X] - \mathbb{E}[\Gamma_t^Y]}{1-t}.$$

Using concavity of the square root then shows

$$\delta_{\frac{1}{2}}(X,Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{(1-t)(\Gamma_t^X+\Gamma_t^Y)}\right]dt.$$

Bounding the Deficit

If $H_t = \sqrt{\frac{(\Gamma_t^X)^2+(\Gamma_t^Y)^2}{2}}$, $\mathrm{Ent}\left(\frac{X+Y}{\sqrt{2}}||G\right) \leq \frac{1}{2}\int\limits_0^1 \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$.

Consequently,

$$2\delta_{\frac{1}{2}}(X,Y) \geq \int\limits_0^1 \frac{\mathbb{E}\left[(1-\Gamma_t^Y)^2\right]}{2(1-t)} + \frac{\mathbb{E}\left[(1-\Gamma_t^X)^2\right]}{2(1-t)} - \frac{\mathbb{E}\left[(1-H_t)^2\right]}{1-t}dt$$

$$= \int\limits_0^1 \frac{2\mathbb{E}[H_t] - \mathbb{E}[\Gamma_t^X] - \mathbb{E}[\Gamma_t^Y]}{1-t}.$$

Using concavity of the square root then shows

$$\delta_{\frac{1}{2}}(X,Y) \gtrsim \int\limits_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{(1-t)(\Gamma_t^X + \Gamma_t^Y)}\right] dt.$$

Log-Concave Measures

We say that X is strongly log-concave if it has a density f such that $-\ln(f)'' \geq 1$.

Fact: if X is strongly log-concave then $\Gamma_t^X \leq 1$ almost surely.

So, if both X and Y are strongly log-concave

$$\delta_{\frac{1}{2}}(X, Y) \gtrsim \int\limits_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$

We use this to derive a quantitative stability bound.

Log-Concave Measures

We say that X is strongly log-concave if it has a density f such that $-\ln(f)'' \geq 1$.

Fact: if X is strongly log-concave then $\Gamma_t^X \leq 1$ almost surely.

So, if both X and Y are strongly log-concave

$$\delta_{\frac{1}{2}}(X, Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$

We use this to derive a quantitative stability bound.

Log-Concave Measures

We say that X is strongly log-concave if it has a density f such that $-\ln(f)'' \geq 1$.
Fact: if X is strongly log-concave then $\Gamma_t^X \leq 1$ almost surely.
So, if both X and Y are strongly log-concave

$$\delta_{\frac{1}{2}}(X, Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$

We use this to derive a quantitative stability bound.

Log-Concave Measures

We say that X is strongly log-concave if it has a density f such that $-\ln(f)'' \geq 1$.
Fact: if X is strongly log-concave then $\Gamma_t^X \leq 1$ almost surely.
So, if both X and Y are strongly log-concave

$$\delta_{\frac{1}{2}}(X, Y) \gtrsim \int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$

We use this to derive a quantitative stability bound.

Log-Concave Measures

$$\int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$

$$\geq \int_0^1 \mathrm{Var}(\Gamma_t^X) dt + \int_0^1 \mathrm{Var}(\Gamma_t^Y) dt + \int_0^1 \left(\mathbb{E}\left[\Gamma_t^X\right] - \mathbb{E}\left[\Gamma_t^Y\right]\right)^2 dt$$

$$\geq \mathcal{W}_2^2(X, G_1) + \mathcal{W}_2^2(Y, G_2) + \mathcal{W}_2^2(G_1, G_2).$$

Here, $\mathcal{W}_2$ denotes the Wasserstein distance and

$G_1 = \int_0^1 \mathbb{E}[\Gamma_t^X] dB_t^X$, $G_2 = \int_0^1 \mathbb{E}[\Gamma_t^Y] dB_t^Y$ are Gaussians.

$$\int_0^1 \mathbb{E}\left[\frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t}\right] dt$$
$$\geq \int_0^1 \mathrm{Var}(\Gamma_t^X) dt + \int_0^1 \mathrm{Var}(\Gamma_t^Y) dt + \int_0^1 \left(\mathbb{E}\left[\Gamma_t^X\right] - \mathbb{E}\left[\Gamma_t^Y\right]\right)^2 dt$$
$$\geq \mathcal{W}_2^2(X, G_1) + \mathcal{W}_2^2(Y, G_2) + \mathcal{W}_2^2(G_1, G_2).$$

Here, $\mathcal{W}_2$ denotes the Wasserstein distance and $G_1 = \int_0^1 \mathbb{E}[\Gamma_t^X] dB_t^X, G_2 = \int_0^1 \mathbb{E}[\Gamma_t^Y] dB_t^Y$ are Gaussians.

Thank You