Stability of the Shannon-Stam Inequality

Dan Mikulincer
Students Probability Day, 2019
Weizmann Institute of Science
Joint work with Ronen Eldan
The central quantity we will deal with 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

$$
\text{Ent}(X \| Y) = \text{Ent}(\mu \| \nu) := \begin{cases} 
\int_{\mathbb{R}^d} \ln \left( \frac{d\mu}{d\nu} \right) d\mu & \text{if } \mu \ll \nu \\
\infty & \text{otherwise}
\end{cases}.
$$
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 $G \sim \mathcal{N}(0, I)$ be a random vector with the law of the standard Gaussian. Then, for any $\lambda \in [0, 1]$

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

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 $G \sim \mathcal{N}(0, I)$ be a random vector with the law of the standard Gaussian. Then, for any $\lambda \in [0, 1]$

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

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’.
Define the deficit

\[ \delta_\lambda(X, Y) = \lambda \text{Ent}(X \| G) + (1-\lambda) \text{Ent}(Y \| G) - \text{Ent}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y \| G). \]

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

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

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

\[ \delta_\lambda(X, Y) = \lambda \text{Ent}(X \| G) + (1 - \lambda) \text{Ent}(Y \| G) - \text{Ent}(\sqrt{\lambda}X + \sqrt{1 - \lambda}Y \| G). \]

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. Föllmer (1984) and then Lehec (2011) have shown that there exists a process $\Gamma^X_t$, such that

1. $\int_0^1 \Gamma^X_t dB_t$ has the law of $X$.

2. $\text{Ent}(X||G) = \frac{1}{2} \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma^X_t)^2 \right]}{1-t} dt$.

3. If $H^X_t$ is another process such that $\int_0^1 H^X_t dB_t$ has the law of $X$,

$$\int_0^1 \frac{\mathbb{E} \left[ (1 - H^X_t)^2 \right]}{1-t} dt \geq \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma^X_t)^2 \right]}{1-t} dt.$$
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. Föllmer (1984) and then Lehec (2011) have shown that there exists a process $\Gamma^X_t$, such that

- $\int_0^1 \Gamma^X_t dB_t$ has the law of $X$.
- $\text{Ent}(X\|G) = \frac{1}{2} \int_0^1 \mathbb{E}\left[\frac{(1 - \Gamma^X_t)^2}{1 - t}\right] dt$.
- If $H^X_t$ is another process such that $\int_0^1 H^X_t dB_t$ has the law of $X$, then
  \[
  \int_0^1 \mathbb{E}\left[\frac{(1 - H^X_t)^2}{1 - t}\right] dt \geq \int_0^1 \mathbb{E}\left[\frac{(1 - \Gamma^X_t)^2}{1 - t}\right] dt.
  \]
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. Föllmer (1984) and then Lehec (2011) have shown that there exists a process $\Gamma^X_t$, such that

- $\int_0^1 \Gamma^X_t dB_t$ has the law of $X$.

- $\text{Ent}(X||G) = \frac{1}{2} \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma^X_t)^2 \right]}{1-t} dt$.

- If $H^X_t$ is another process such that $\int_0^1 H^X_t dB_t$ has the law of $X$,

$$\int_0^1 \frac{\mathbb{E} \left[ (1 - H^X_t)^2 \right]}{1-t} dt \geq \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma^X_t)^2 \right]}{1-t} dt.$$
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. Föllmer (1984) and then Lehec (2011) have shown that there exists a process $\Gamma_t^X$, such that

- $\frac{1}{2} \int_0^1 \Gamma_t^X dB_t$ has the law of $X$.
- $\text{Ent}(X||G) = \frac{1}{2} \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma_t^X)^2 \right]}{1-t} dt$.

- If $H_t^X$ is another process such that $\int_0^1 H_t^X dB_t$ has the law of $X$, then
  $$\int_0^1 \frac{\mathbb{E} \left[ (1 - H_t^X)^2 \right]}{1-t} dt \geq \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma_t^X)^2 \right]}{1-t} dt.$$
Now, for $X, Y$ random variables, take two independent Brownian motions $B^X_t, B^Y_t$ and $\Gamma^X_t, \Gamma^Y_t$ 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

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

for some Brownian motion $B_t$. 


Now, for \( X, Y \) random variables, take two independent Brownian motions \( B^X_t, B^Y_t \) and \( \Gamma^X_t, \Gamma^Y_t \) 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

\[
\frac{X + Y}{\sqrt{2}} = \frac{1}{\sqrt{2}} \left( \int_0^1 \Gamma^X_t \, dB^X_t + \int_0^1 \Gamma^Y_t \, dB^Y_t \right) \overset{\text{law}}{=} \int_0^1 \sqrt{\frac{(\Gamma^X_t)^2 + (\Gamma^Y_t)^2}{2}} \, dB_t.
\]

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

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

where $G$ is another standard Gaussian.

This implies

$$\frac{X + Y}{\sqrt{2}} = \frac{1}{\sqrt{2}} \left( \int_0^1 \Gamma^X_t dB^X_t + \int_0^1 \Gamma^Y_t dB^Y_t \right) \xrightarrow{\text{law}} \int_0^1 \sqrt{\frac{(\Gamma^X_t)^2 + (\Gamma^Y_t)^2}{2}} dB_t.$$

for some Brownian motion $B_t$. 
Bounding the Deficit

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

Consequently,

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

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

Using concavity of the square root then shows

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

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

Consequently,

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

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

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

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 \( H_t = \sqrt{\frac{(\Gamma^X_t)^2 + (\Gamma^Y_t)^2}{2}} \), \( \text{Ent} \left( \frac{X + Y}{\sqrt{2}} \| G \right) \leq \frac{1}{2} \int_0^1 \mathbb{E} \left[ (1 - H_t)^2 \right] dt \).

Consequently,

\[
2\delta_1^\frac{1}{2}(X, Y) \geq \int_0^1 \frac{\mathbb{E} \left[ (1 - \Gamma^Y_t)^2 \right]}{2(1 - t)} + \frac{\mathbb{E} \left[ (1 - \Gamma^X_t)^2 \right]}{2(1 - t)} - \frac{\mathbb{E} \left[ (1 - H_t)^2 \right]}{1 - t} dt
\]

\[
= \int_0^1 \frac{2\mathbb{E}[H_t] - \mathbb{E}[\Gamma^X_t] - \mathbb{E}[\Gamma^Y_t]}{1 - t} dt.
\]

Using concavity of the square root then shows

\[
\delta_1^\frac{1}{2}(X, Y) \gtrsim \int_0^1 \mathbb{E} \left[ \frac{(\Gamma^X_t - \Gamma^Y_t)^2}{(1 - t)(\Gamma^X_t + \Gamma^Y_t)} \right] dt.
\]
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 $\Gamma^X_t \leq 1$ almost surely.

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

$$\delta_{1/2} (X, Y) \geq \int_0^1 \mathbb{E} \left[ \frac{(\Gamma^X_t - \Gamma^Y_t)^2}{1-t} \right] dt$$

We use this to derive a quantitative stability bound.
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

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

We use this to derive a quantitative stability bound.
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 $\Gamma^X_t \leq 1$ almost surely.

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

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

We use this to derive a quantitative stability bound.
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 $\Gamma^X_t \leq 1$ almost surely.

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

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

We use this to derive a quantitative stability bound.
\[
\int_0^1 \mathbb{E} \left[ \frac{(\Gamma_t^X - \Gamma_t^Y)^2}{1-t} \right] dt \\
\geq \int_0^1 \text{Var}(\Gamma_t^X) dt + \int_0^1 \text{Var}(\Gamma_t^Y) dt + \int_0^1 \left( \mathbb{E} \left[ \Gamma_t^X \right] - \mathbb{E} \left[ \Gamma_t^Y \right] \right)^2 dt \\
\geq \mathcal{W}_2^2(X, G_1) + \mathcal{W}_2^2(Y, G_2) + \mathcal{W}_2^2(G_1, G_2).
\]

Here, \( \mathcal{W}_2 \) denotes the Wasserstein distance and 
\[
G_1 = \int_0^1 \mathbb{E}[\Gamma_t^X] dB_t^X, \quad G_2 = \int_0^1 \mathbb{E}[\Gamma_t^Y] dB_t^Y
\]
are Gaussians.
\[
\int_0^1 \mathbb{E} \left[ \frac{(\Gamma^X_t - \Gamma^Y_t)^2}{1 - t} \right] dt \\
\geq \int_0^1 \text{Var}(\Gamma^X_t) dt + \int_0^1 \text{Var}(\Gamma^Y_t) dt + \int_0^1 \left( \mathbb{E} \left[ \Gamma^X_t \right] - \mathbb{E} \left[ \Gamma^Y_t \right] \right)^2 dt \\
\geq \mathcal{W}^2_2(X, G_1) + \mathcal{W}^2_2(Y, G_2) + \mathcal{W}^2_2(G_1, G_2).
\]

Here, \( \mathcal{W}_2 \) denotes the Wasserstein distance and \( G_1 = \int_0^1 \mathbb{E}[\Gamma^X_t] dB_t^X \), \( G_2 = \int_0^1 \mathbb{E}[\Gamma^Y_t] dB_t^Y \) are Gaussians.
Thank You