# Gaussian mixtures: convexity properties and CLT rates for the entropy and Fisher information

Alexandros Eskenazis (CNRS, Sorbonne & Cambridge),
Lampros Gavalakis [1] (Gustave Eiffel)

Athens, Greece | July 12, 2024

[1] L.G. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 101034255.

# Outline

# Entropy

$X_1, X_2$ independent RVs

- Entropy Power Inequality (EPI)

$$h(\sqrt{1-t}X_1 + \sqrt{t}X_2) \geq (1-t)h(X_1) + th(X_2)$$

# Entropy

$X_1, X_2$ independent RVs

- Entropy Power Inequality (EPI)

$$h(\sqrt{1-t}X_1 + \sqrt{t}X_2) \geq (1-t)h(X_1) + th(X_2)$$

- For i.i.d.

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

minimized at

$$t = 0, 1$$

# Entropy

$X_1, X_2$ independent RVs

- Entropy Power Inequality (EPI)

$$h(\sqrt{1-t}X_1 + \sqrt{t}X_2) \geq (1-t)h(X_1) + th(X_2)$$

- For i.i.d.

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

  minimized at

$$t = 0, 1$$

Q: Where is it maximized?

# Entropy

$t = \frac{1}{2}$?

# Entropy

$t = \frac{1}{2}$? NOT in general

# Entropy

$t = \frac{1}{2}$? NOT in general

**Conjecture (Ball, Nayar, Tkocz, 2016)**

*For $X_1, X_2$ i.i.d. log-concave*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave*

# Entropy

$t = \frac{1}{2}$? NOT in general

**Conjecture (Ball, Nayar, Tkocz, 2016)**

*For $X_1, X_2$ i.i.d. log-concave*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave*

- EPI special case
- No special case known (of $X_1, X_2$)!

# Entropy: Gaussian Mixtures

Gaussian mixture (GM)

$$X = Y \cdot Z,$$

$Z \sim \mathcal{N}(0, 1)$, $\mathbb{P}(Y > 0) = 1$ independent

# Entropy: Gaussian Mixtures

Gaussian mixture (GM)

$$X = Y \cdot Z,$$

$Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Y > 0) = 1$ independent

$$g_X(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

# Entropy: Gaussian Mixtures

Gaussian mixture (GM)

$$X = Y \cdot Z,$$

$Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Y > 0) = 1$ independent

$$g_X(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

E.g.

- $g_X \propto e^{-|x|^p}$ (p- stable, $p \in (0,2]$)

# Entropy: Gaussian Mixtures

Gaussian mixture (GM)

$$X = Y \cdot Z,$$

$Z \sim \mathcal{N}(0,1)$, $\mathbb{P}(Y > 0) = 1$ independent

$$g_X(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

E.g.

- $g_X \propto e^{-|x|^p}$ (p- stable, $p \in (0,2]$)
- Cauchy $\propto \frac{1}{\pi(1+x^2)} \Rightarrow \frac{Z_1}{|Z_2|}$

# Entropy: Gaussian Mixtures

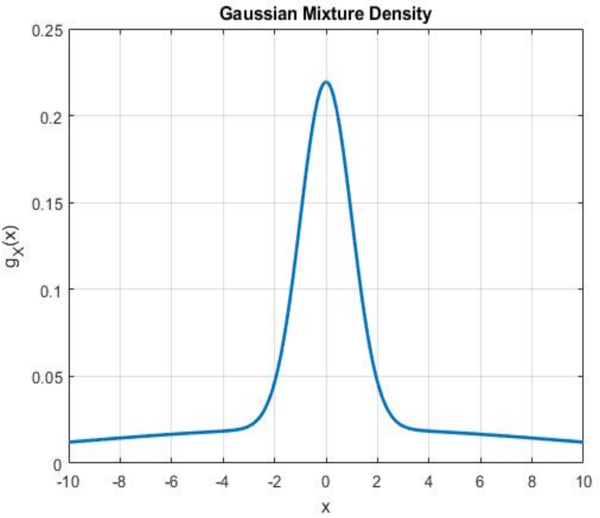



Figure: $\mathbb{P}(Y = 1) = \mathbb{P}(Y = 10) = 0.5$

# Entropy: Gaussian Mixtures

(Eskenazis, Nayar, Tkocz, 2018): A symmetric RV is GM iff $x \mapsto g_X(\sqrt{x})$ is completely monotonic on $(0, \infty)$

# Entropy: Gaussian Mixtures

(Eskenazis, Nayar, Tkocz, 2018): A symmetric RV is GM iff $x \mapsto g_X(\sqrt{x})$ is completely monotonic on $(0, \infty)$

**Theorem**

*Let $X_1, X_2$ be independent GMs. Then*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave in $t \in [0, 1]$*

# Entropy: Gaussian Mixtures

(Eskenazis, Nayar, Tkocz, 2018): A symmetric RV is GM iff $x \mapsto g_X(\sqrt{x})$ is completely monotonic on $(0, \infty)$

**Theorem**

*Let $X_1, X_2$ be independent GMs. Then*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave in $t \in [0, 1]$*

- Proof: Bernstein and Hölder $\Rightarrow -\log f_t(\sqrt{\cdot})$ is concave + variational formula for entropy

# Entropy: Gaussian Mixtures

(Eskenazis, Nayar, Tkocz, 2018): A symmetric RV is GM iff $x \mapsto g_X(\sqrt{x})$ is completely monotonic on $(0, \infty)$

**Theorem**

*Let $X_1, X_2$ be independent GMs. Then*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave in $t \in [0, 1]$*

- Proof: Bernstein and Hölder $\Rightarrow -\log f_t(\sqrt{\cdot})$ is concave + variational formula for entropy
- In fact $(a_1^2, \ldots, a_n^2) \longmapsto h_\alpha\Big(\sum_{i=1}^n a_i X_i\Big)$ is concave for $\sum_{i=1}^n a_i^2 = 1$ and every $\alpha \geq 1$

# Entropy: Gaussian Mixtures

(Eskenazis, Nayar, Tkocz, 2018): A symmetric RV is GM iff $x \mapsto g_X(\sqrt{x})$ is completely monotonic on $(0, \infty)$

**Theorem**

*Let $X_1, X_2$ be independent GMs. Then*

$$t \mapsto h(\sqrt{1-t}X_1 + \sqrt{t}X_2)$$

*is concave in $t \in [0, 1]$*

- Proof: Bernstein and Hölder $\Rightarrow -\log f_t(\sqrt{\cdot})$ is concave + variational formula for entropy
- In fact $(a_1^2, \ldots, a_n^2) \longmapsto h_\alpha\left(\sum_{i=1}^n a_i X_i\right)$ is concave for $\sum_{i=1}^n a_i^2 = 1$ and every $\alpha \geq 1$
- In [Costa '85]: strengthened EPI for stable laws

# Fisher Information

$X \sim f$ "smooth" on $\mathbb{R}^d$

Fisher informtaion

$$I(X) = \int_{\mathbb{R}^d} \frac{|\nabla f(X)|^2}{f(X)} dx$$

# Fisher Information

$X \sim f$ "smooth" on $\mathbb{R}^d$

Fisher informtaion

$$I(X) = \int_{\mathbb{R}^d} \frac{|\nabla f(X)|^2}{f(X)} dx$$

Fisher information matrix

$$\mathcal{I}(X)_{i,j} = \int_{\mathbb{R}^d} \frac{\partial_i f(X) \partial_j f(X)}{f(X)} dx$$

# Fisher Information

$X \sim f$ "smooth" on $\mathbb{R}^d$

Fisher informtaion

$$I(X) = \int_{\mathbb{R}^d} \frac{|\nabla f(X)|^2}{f(X)} dx$$

Fisher information matrix

$$\mathcal{I}(X)_{i,j} = \int_{\mathbb{R}^d} \frac{\partial_i f(X) \partial_j f(X)}{f(X)} dx$$

Cramér-Rao (C-R):

$$I(X) \geq \frac{1}{\mathrm{Var}(X)} \quad (d = 1)$$

# Fisher Information

## Proposition

*Let* $X = Y \cdot Z$ *a GM*, $\mathrm{Var}(X) = \mathbb{E}[Y^2]$

$$\frac{1}{\mathbb{E}[Y^2]} \overset{C-R}{\leq} I(X) \leq \mathbb{E}\left[\frac{1}{Y^2}\right]$$

# Fisher Information

## Proposition

*Let* $X = Y \cdot Z$ *a GM*, $\mathrm{Var}(X) = \mathbb{E}[Y^2]$

$$\frac{1}{\mathbb{E}[Y^2]} \overset{C-R}{\leq} I(X) \leq \mathbb{E}\left[\frac{1}{Y^2}\right]$$

Proof.

$$f(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right], f'(x) = -x\mathbb{E}\left[\frac{1}{Y^2\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

# Fisher Information

## Proposition

*Let* $X = Y \cdot Z$ *a GM*, $\mathrm{Var}(X) = \mathbb{E}[Y^2]$

$$\frac{1}{\mathbb{E}[Y^2]} \overset{C-R}{\leq} I(X) \leq \mathbb{E}\left[\frac{1}{Y^2}\right]$$

Proof.

$$f(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right], f'(x) = -x\mathbb{E}\left[\frac{1}{Y^2\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

$$C - S \quad \Rightarrow \left(f'(x)\right)^2 \leq x^2 f(x) \mathbb{E}\left[\frac{1}{Y^4\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

# Fisher Information

## Proposition

*Let* $X = Y \cdot Z$ *a* *GM*, $\mathrm{Var}(X) = \mathbb{E}[Y^2]$

$$\frac{1}{\mathbb{E}[Y^2]} \overset{C-R}{\leq} I(X) \leq \mathbb{E}\left[\frac{1}{Y^2}\right]$$

Proof.

$$f(x) = \mathbb{E}\left[\frac{1}{\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right], f'(x) = -x\mathbb{E}\left[\frac{1}{Y^2\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

$$C - S \quad \Rightarrow \left(f'(x)\right)^2 \leq x^2 f(x)\mathbb{E}\left[\frac{1}{Y^4\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right]$$

$$I(X) = \int \frac{(f'(x))^2}{f(x)} dx \leq \int x^2 \mathbb{E}\left[\frac{1}{Y^4\sqrt{2\pi Y^2}} e^{-\frac{x^2}{2Y^2}}\right] dx = \mathbb{E}\left[\frac{1}{Y^2}\right] \qquad \square$$

# III. Fisher Information

(GM) in $\mathbb{R}^d$: $\quad X = \mathbf{Y} \cdot Z$,
$Z \sim \mathcal{N}(0, \mathrm{I}_{d\times d})$, $\mathbf{Y} \in \mathbb{R}^{d\times d}$, symmetric, $\mathbb{P}(\mathbf{Y} \succeq 0) = 1$

(GM) in $\mathbb{R}^d$: $X = \mathbf{Y} \cdot Z$,
$Z \sim \mathcal{N}(0, \mathrm{I}_{d\times d})$, $\mathbf{Y} \in \mathbb{R}^{d\times d}$, symmetric, $\mathbb{P}(\mathbf{Y} \succeq 0) = 1$
Cramér-Rao (C-R):

$$\mathcal{I}(X) \succeq \left(\mathbb{E}\mathbf{Y}\mathbf{Y}^T\right)^{-1}$$

# III. Fisher Information

(GM) in $\mathbb{R}^d$: $\quad X = \mathbf{Y} \cdot Z$,

$Z \sim \mathcal{N}(0, \mathrm{I}_{d\times d})$, $\mathbf{Y} \in \mathbb{R}^{d\times d}$, symmetric, $\mathbb{P}(\mathbf{Y} \succeq 0) = 1$

Cramér-Rao (C-R):

$$\mathcal{I}(X) \succeq \left(\mathbb{E}\mathbf{Y}\mathbf{Y}^T\right)^{-1}$$

Q:

$$\mathcal{I}(X) \preceq \mathbb{E}\big((\mathbf{Y}\mathbf{Y}^T)^{-1}\big)?$$

# III. Fisher Information

(GM) in $\mathbb{R}^d$: $\quad X = \mathbf{Y} \cdot Z$,
$Z \sim \mathcal{N}(0, \mathrm{I}_{d\times d})$, $\mathbf{Y} \in \mathbb{R}^{d\times d}$, symmetric, $\mathbb{P}(\mathbf{Y} \succeq 0) = 1$
Cramér-Rao (C-R):

$$\mathcal{I}(X) \succeq \left(\mathbb{E}\mathbf{Y}\mathbf{Y}^T\right)^{-1}$$

Q:

$$\mathcal{I}(X) \preceq \mathbb{E}\big((\mathbf{Y}\mathbf{Y}^T)^{-1}\big)?$$

$R(x, y) = \frac{x^2}{y}$ is jointly convex on $\mathbb{R}^2_+ \Rightarrow I(X) = \int_{\mathbb{R}^d} \frac{|\nabla f(X)|^2}{f(X)} dx$ is a convex functional of $f$

# III. Fisher Information

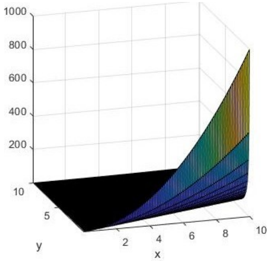
(GM) in $\mathbb{R}^d$: $\quad X = \mathbf{Y} \cdot Z$,
$Z \sim \mathcal{N}(0, \mathrm{I}_{d\times d})$, $\mathbf{Y} \in \mathbb{R}^{d\times d}$, symmetric, $\mathbb{P}(\mathbf{Y} \succeq 0) = 1$
Cramér-Rao (C-R):

$$\mathcal{I}(X) \succeq \left(\mathbb{E}\mathbf{Y}\mathbf{Y}^T\right)^{-1}$$

Q:

$$\mathcal{I}(X) \preceq \mathbb{E}\big((\mathbf{Y}\mathbf{Y}^T)^{-1}\big)?$$

$R(x,y) = \frac{x^2}{y}$ is jointly convex on $\mathbb{R}^2_+ \Rightarrow I(X) = \int_{\mathbb{R}^d} \frac{|\nabla f(X)|^2}{f(X)} dx$ is a convex functional of $f$



# Fisher Information

## Proposition (Bobkov, '22)

*The Fisher information satisfies "Jensen's" inequality: let $\pi$ be a probability measure on the space of probability densities in $\mathbb{R}^n$. Then, if $f$ is any mixture of densities $f = \int g\pi(dg)$,*

$$I(f) \leq \int I(g)\pi(dg)$$

# Fisher Information

## Proposition (Bobkov, '22)

*The Fisher information satisfies "Jensen's" inequality: let $\pi$ be a probability measure on the space of probability densities in $\mathbb{R}^n$. Then, if $f$ is any mixture of densities $f = \int g\pi(dg)$,*

$$I(f) \leq \int I(g)\pi(dg)$$

## Proposition

*The Fisher information matrix is convex as a matrix-valued functional and satisfies "Jensen's" inequality: if $f$ is any mixture of densities $f = \int g\pi(dg)$,*

$$\mathcal{I}(f) \preceq \int \mathcal{I}(g)\pi(dg)$$

*provided that $\int \|\mathcal{I}(g)\|_{\mathrm{OP}} < \infty$*

# Fisher Information

*Proof.* $R(x, \lambda) = \frac{xx^T}{\lambda}$ is jointly operator convex + Bobkov's proof with $\langle \mathcal{I}(X)x, x \rangle$ $\square$

# Fisher Information

Proof. $R(x,\lambda) = \frac{xx^T}{\lambda}$ is jointly operator convex + Bobkov's proof with $\langle \mathcal{I}(X)x, x \rangle$ $\square$

$$\Rightarrow$$

# Fisher Information

Proof. $R(x,\lambda) = \frac{xx^T}{\lambda}$ is jointly operator convex + Bobkov's proof with $\langle \mathcal{I}(X)x, x\rangle$ $\square$

$$\Rightarrow$$

$$\mathcal{I}(\mathbf{Y}Z) \preceq \mathbb{E}_{\mathbf{Y}}\mathcal{I}(Z_{\mathbf{Y}}) = \mathbb{E}(\mathbf{Y}\mathbf{Y}^T)^{-1},$$

where $Z_\Sigma \sim \mathcal{N}(0, \Sigma\Sigma^T)$

# CLT rates

$(a_1, \ldots, a_n) : \sum_{i=1}^n a_i^2 = 1$, $X_i = Y_i Z_i$ GMs

$$S_n = \sum_{i=1}^n a_i X_i = \sum_{i=1}^n a_i Y_i Z_i \stackrel{d}{=} \left(\sum a_i^2 Y_i^2\right)^{1/2} Z$$

# CLT rates

$(a_1, \ldots, a_n) : \sum_{i=1}^n a_i^2 = 1$, $X_i = Y_i Z_i$ GMs

$$S_n = \sum_{i=1}^n a_i X_i = \sum_{i=1}^n a_i Y_i Z_i \stackrel{d}{=} \left(\sum a_i^2 Y_i^2\right)^{1/2} Z$$

- $$I(S_n) \leq \mathbb{E}\left[\frac{1}{\sum a_i^2 Y_i^2}\right]$$

# CLT rates

$(a_1, \ldots, a_n) : \sum_{i=1}^n a_i^2 = 1$, $X_i = Y_i Z_i$ GMs

$$S_n = \sum_{i=1}^n a_i X_i = \sum_{i=1}^n a_i Y_i Z_i \stackrel{d}{=} \left(\sum a_i^2 Y_i^2\right)^{1/2} Z$$

- $$I(S_n) \leq \mathbb{E}\left[\frac{1}{\sum a_i^2 Y_i^2}\right]$$

Standardised Fisher information:

$$J_{st}(S_n) := \mathrm{Var}(S_n) I(S_n) - 1 \leq \left(\mathbb{E} Y_1^2\right) \mathbb{E}\left[\frac{1}{\sum a_i^2 Y_i^2}\right] - 1$$

# CLT rates

## Corollary

*Suppose* $\mathbb{E}Y_1^{2+2\delta}, \mathbb{E}Y_1^{-2-2\delta} < \infty$. *Then*

$$J_{st}(S_n) \leq C_{Y_1} \|a\|_{2+2\delta}^{\frac{2\delta}{(1+\delta)}}$$

# CLT rates

**Corollary**

*Suppose* $\mathbb{E}Y_1^{2+2\delta}, \mathbb{E}Y_1^{-2-2\delta} < \infty$. *Then*

$$J_{st}(S_n) \le C_{Y_1} \|a\|_{2+2\delta}^{\frac{2\delta}{(1+\delta)}}$$

- E.g. $a_i = \frac{1}{\sqrt{n}}$, then

$$\|a\|_{2+2\delta}^{\frac{2\delta}{1+\delta}} = \left( \sum_{i=1}^{n} \frac{1}{n^{1+\delta}} \right)^{\frac{2\delta}{(1+\delta)(2+2\delta)}} = \left( \frac{1}{n} \right)^{\frac{\delta^2}{(1+\delta)^2}} \to 0$$

# CLT rates

## Corollary

*Suppose* $\mathbb{E} Y_1^{2+2\delta}, \mathbb{E} Y_1^{-2-2\delta} < \infty$. *Then*

$$J_{st}(S_n) \le C_{Y_1} \|a\|_{2+2\delta}^{\frac{2\delta}{(1+\delta)}}$$

- E.g. $a_i = \frac{1}{\sqrt{n}}$, then

$$\|a\|_{2+2\delta}^{\frac{2\delta}{1+\delta}} = \left( \sum_{i=1}^{n} \frac{1}{n^{1+\delta}} \right)^{\frac{2\delta}{(1+\delta)(2+2\delta)}} = \left( \frac{1}{n} \right)^{\frac{\delta^2}{(1+\delta)^2}} \to 0$$

For this example better rate in [Bobkov, Chistyakov, Gotze, '14]

# CLT rates

## Corollary

*Suppose* $\mathbb{E}Y_1^{2+2\delta}, \mathbb{E}Y_1^{-2-2\delta} < \infty$. *Then*

$$J_{st}(S_n) \leq C_{Y_1} \|a\|_{2+2\delta}^{\frac{2\delta}{(1+\delta)}}$$

- E.g. $a_i = \frac{1}{\sqrt{n}}$, then

$$\|a\|_{2+2\delta}^{\frac{2\delta}{1+\delta}} = \left(\sum_{i=1}^{n} \frac{1}{n^{1+\delta}}\right)^{\frac{2\delta}{(1+\delta)(2+2\delta)}} = \left(\frac{1}{n}\right)^{\frac{\delta^2}{(1+\delta)^2}} \to 0$$

  For this example better rate in [Bobkov, Chistyakov, Gotze, '14]
- $O(\|a\|_4^4)$ in [Artstein, Ball, Barthe, Naor, '04] for weighted sums under much stronger moment assumptions

# CLT rates

**Corollary (General dimension)**

*If* $\mathbb{E}\|\mathbf{Y}\mathbf{Y}^T\|_{\mathrm{OP}}^{1+\delta} < \infty$ *and* $\mathbb{E}\|\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}\|_{\mathrm{OP}}^{1+\delta} < \infty$,

$$\|\mathrm{Cov}(S_n)^{\frac{1}{2}}\mathcal{I}(S_n)\mathrm{Cov}(S_n)^{\frac{1}{2}} - \mathrm{I}_d\|_{\mathrm{OP}} \leq C(\mathbf{Y})\log^{\delta}(d+1)\|a\|_{2+2\delta}^{\frac{2\delta}{1+\delta}}$$

# CLT rates

**Corollary (General dimension)**

*If* $\mathbb{E}\|\mathbf{Y}\mathbf{Y}^T\|_{\mathrm{OP}}^{1+\delta} < \infty$ *and* $\mathbb{E}\|\left(\mathbf{Y}\mathbf{Y}^T\right)^{-1}\|_{\mathrm{OP}}^{1+\delta} < \infty$,

$$\|\mathrm{Cov}(S_n)^{\frac{1}{2}}\mathcal{I}(S_n)\mathrm{Cov}(S_n)^{\frac{1}{2}} - \mathrm{I}_d\|_{\mathrm{OP}} \leq C(\mathbf{Y})\log^{\delta}(d+1)\|a\|_{2+2\delta}^{\frac{2\delta}{1+\delta}}$$

- Convergence of the Fisher information matrix
- First such result!

# An Open Question

- Is

$$I\left(\sqrt{t}X_1 + \sqrt{1-t}X_2\right)$$

minimized at $t = \frac{1}{2}$ ? More generally, is

$$I\left(\sum_{i=1}^{n} a_i X_i\right)$$

minimized at $a_i = \frac{1}{\sqrt{n}}$?

Thank you!