KNOW Center

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

On the Information Dimension of Random Variables and Stochastic Processes

Bernhard C. Geiger
Joint work with Tobias Koch

Authors and Funders

uc3m | Universidad Carlos III de Madrid

FWF Der Wissenschaftsfonds.

Unterstützt von / Supported by Alexander von Humboldt Stiftung/Foundation

iDev40

ECSEL Joint Undertaking

Part of this work was funded by the iDev40 project. The iDev40 project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 783163. The JU receives support from the European Unions Horizon 2020 research and innovation programme. It is co-funded by the consortium members, grants from Austria, Germany, Belgium, Italy, Spain and Romania.

Rényi Information Dimension[1]

X is L-dimensional and real-valued

$$d(X) \triangleq \lim_{m\to\infty} \frac{H([X]_m)}{\log m}$$

where

$$[X]_m \triangleq \frac{\lfloor mX \rfloor}{m}$$

and

$$H(Z) \triangleq -\sum_z \mathbb{P}(Z=z) \log \mathbb{P}(Z=z).$$

[1] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

Rényi Information Dimension[1]

X is L-dimensional and real-valued

$$d(X) \triangleq \lim_{m\to\infty} \frac{H([X]_m)}{\log m}$$

where

$$[X]_m \triangleq \frac{\lfloor mX \rfloor}{m}$$

and

$$H(Z) \triangleq -\sum_{z} \mathbb{P}(Z=z) \log \mathbb{P}(Z=z).$$

(we assume throughout that the limit exists and is finite)

[1] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

Properties of Information Dimension[2,3,4]

- Bounded:

$$0 \leq d(X) \leq L$$

[2] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

[3] Wu and Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression", 2010

[4] Wu, "Shannon Theory for Compressed Sensing", 2011

Properties of Information Dimension[2,3,4]

- Bounded:

$$0 \leq d(X) \leq L$$

- Lipschitz Maps: ($\Rightarrow$ Scale & Translation Invariance)

$$d(f(X)) \leq d(X)$$

[2] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

[3] Wu and Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression", 2010

[4] Wu, "Shannon Theory for Compressed Sensing", 2011

Properties of Information Dimension[2,3,4]

- Bounded:

$$0 \leq d(X) \leq L$$

- Lipschitz Maps: ($\Rightarrow$ Scale & Translation Invariance)

$$d(f(X)) \leq d(X)$$

- Subadditive:

$$d(X, Y) \leq d(X) + d(Y)$$

with equality if $X \perp Y$

[2] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

[3] Wu and Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression", 2010

[4] Wu, "Shannon Theory for Compressed Sensing", 2011

The Discrete, the Continuous, and the Singular[5]

- If X has a discrete distribution, then $d(X) = 0$.

[5] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

The Discrete, the Continuous, and the Singular[5]

- If X has a discrete distribution, then $d(X) = 0$.
- If X has an absolutely continuous distribution, then $d(X) = L$.

[5] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

The Discrete, the Continuous, and the Singular[5]

- If X has a discrete distribution, then $d(X) = 0$.
- If X has an absolutely continuous distribution, then $d(X) = L$.
- "It can be shown that $[d(X) = K < L]$ for absolutely continuous probability distributions on sufficiently smooth K-dimensional manifolds lying in $\mathbb{R}^L$."

[5] Rényi, "On the Dimension and Entropy of Probability Distributions", 1959

Gaussian Case

Theorem

If X is Gaussian and has covariance matrix C_X, then

$$d(X) = \mathrm{rank}(C_X).$$

Gaussian Case

Theorem

If X has covariance matrix C_X, then

$$d(X) \leq \mathrm{rank}(C_X)$$

with equality if X is Gaussian.

Information Dimension is Relevant:

Communications & Information Theory:

- Rate-distortion theory[6,7]
- Almost lossless analog compressed sensing[8]
- DoF of Gaussian interference channels[9,10]

Dynamical Systems Theory:

- Characterization of Chaotic Attractors[11]

---

[6] Kawabata and Dembo, "The rate-distortion dimension of sets and measures", 1994

[7] Koch, "The Shannon Lower Bound Is Asymptotically Tight", 2016

[8] Wu and Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression", 2010

[9] Wu, Shamai (Shitz), and Verdú, "Information Dimension and the Degrees of Freedom of the Interference Channel", 2015

[10] Stotz and Bölcskei, "Degrees of Freedom in Vector Interference Channels", 2016

[11] Farmer, Ott, and Yorke, "The dimension of chaotic attractors", 1983

Generalization to Stochastic Processes

$\{\mathbf{X}_t,\ t \in \mathbb{Z}\}$ is an L-variate, real-valued, stationary process

$$d(\{\mathbf{X}_t\}) \triangleq \lim_{m\to\infty} \frac{\overline{H}(\{[\mathbf{X}_t]_m\})}{\log m}$$

where

$$\overline{H}(\{[\mathbf{X}_t]_m\}) \triangleq \lim_{n\to\infty} \frac{H([\mathbf{X}_1]_m, \dots, [\mathbf{X}_n]_m)}{n}.$$

Generalization to Stochastic Processes

$\{\mathbf{X}_t,\ t \in \mathbb{Z}\}$ is an L-variate, real-valued, stationary process

$$d(\{\mathbf{X}_t\}) \triangleq \lim_{m\to\infty} \frac{\overline{H}(\{[\mathbf{X}_t]_m\})}{\log m}$$

where

$$\overline{H}(\{[\mathbf{X}_t]_m\}) \triangleq \lim_{n\to\infty} \frac{H([\mathbf{X}_1]_m, \dots, [\mathbf{X}_n]_m)}{n}.$$

(we assume throughout that the limits exist and are finite)

Properties of Information Dimension Rate

- Bounded:

$$0 \leq d(\{\mathbf{X}_t\}) \leq \lim_{n\to\infty} \frac{d(\mathbf{X}_1, \dots, \mathbf{X}_n)}{n} \leq d(\mathbf{X}_1) \leq L$$

Properties of Information Dimension Rate

- Bounded:

$$0 \leq d(\{\mathbf{X}_t\}) \leq \lim_{n\to\infty} \frac{d(\mathbf{X}_1, \dots, \mathbf{X}_n)}{n} \leq d(\mathbf{X}_1) \leq L$$

- Lipschitz Maps: ($\Rightarrow$ Scale & Translation Invariance)

$$d(\{f_t(\mathbf{X}_t)\}) \leq d(\{\mathbf{X}_t\})$$

Properties of Information Dimension Rate

- Bounded:

$$0 \leq d(\{\mathbf{X}_t\}) \leq \lim_{n\to\infty} \frac{d(\mathbf{X}_1, \dots, \mathbf{X}_n)}{n} \leq d(\mathbf{X}_1) \leq L$$

- Lipschitz Maps: ($\Rightarrow$ Scale & Translation Invariance)

$$d(\{f_t(\mathbf{X}_t)\}) \leq d(\{\mathbf{X}_t\})$$

- Subadditive:

$$d(\{\mathbf{X}_t, \mathbf{Y}_t\}) \leq d(\{\mathbf{X}_t\}) + d(\{\mathbf{Y}_t\})$$

with equality if $\{\mathbf{X}_t\} \perp \{\mathbf{Y}_t\}$

The Discrete, the Continuous, and the Bandlimited

Consider a scalar ($L = 1$) process $\{X_t\}$:

- If $\{X_t\}$ is discrete-valued, then $d(\{X_t\}) = 0$.

The Discrete, the Continuous, and the Bandlimited

Consider a scalar ($L = 1$) process $\{X_t\}$:

- If $\{X_t\}$ is discrete-valued, then $d(\{X_t\}) = 0$.
- If $\{X_t\}$ is continuous-valued and i.i.d., hen $d(\{X_t\}) = 1$.

The Discrete, the Continuous, and the Bandlimited

Consider a scalar ($L = 1$) process $\{X_t\}$:

- If $\{X_t\}$ is discrete-valued, then $d(\{X_t\}) = 0$.
- If $\{X_t\}$ is continuous-valued and i.i.d., hen $d(\{X_t\}) = 1$.
- If $\{X_t\}$ is Gaussian with bandlimited power spectral density S_X, is there a connection between $d(\{X_t\})$ and the bandwidth?

Gaussian Process

Corollary

If $\{X_t\}$ *is a scalar, Gaussian process with power spectral density* S_X*, then*

$$d(\{X_t\}) = \lambda\left(\{\theta\colon S_X(\theta) > 0\}\right).$$

Gaussian Process

Corollary

If $\{X_t\}$ *is a scalar, Gaussian process with power spectral density* S_X*, then*

$$d(\{X_t\}) = \lambda\left(\{\theta\colon S_X(\theta) > 0\}\right).$$

Example

Let $\{X_t\}$ *be Gaussian and have power spectral density* $S_X\colon [-\frac{1}{2}, \frac{1}{2}] \to \mathbb{R}^+$ *positive on* $[-\frac{1}{4}, \frac{1}{4}]$ *and zero elsewhere (low-pass process). Then,*

$$d(\{X_t\}) = \frac{1}{2}.$$

Gaussian Process (cont'd)

Theorem

If $\{\mathbf{X}_t\}$ *is Gaussian and has power spectral density* $S_{\mathbf{X}}$*, then*

$$d(\{\mathbf{X}_t\}) = \int_{-1/2}^{1/2} \operatorname{rank}\left(S_{\mathbf{X}}(\theta)\right) \mathrm{d}\theta.$$

$$\left(\mathbb{E}\left(\mathbf{X}_{t+\tau}\mathbf{X}_t^{\mathsf{T}}\right) - \mathbb{E}\left(\mathbf{X}_{t+\tau}\right)\mathbb{E}\left(\mathbf{X}_t^{\mathsf{T}}\right) = \int_{-1/2}^{1/2} S_{\mathbf{X}}(\theta)\mathrm{e}^{-\imath 2\pi\tau\theta}\mathrm{d}\theta\right)$$

Gaussian Process (cont'd)

Theorem

If $\{\mathbf{X}_t\}$ has power spectral density $S_{\mathbf{X}}$, then

$$d(\{\mathbf{X}_t\}) \leq \int_{-1/2}^{1/2} \mathrm{rank}\left(S_{\mathbf{X}}(\theta)\right) \mathrm{d}\theta$$

with equality if $\{\mathbf{X}_t\}$ is Gaussian.

$$\left(\mathbb{E}\left(\mathbf{X}_{t+\tau}\mathbf{X}_t^{\mathsf{T}}\right) - \mathbb{E}\left(\mathbf{X}_{t+\tau}\right)\mathbb{E}\left(\mathbf{X}_t^{\mathsf{T}}\right) = \int_{-1/2}^{1/2} S_{\mathbf{X}}(\theta)\mathrm{e}^{-\imath 2\pi\tau\theta}\mathrm{d}\theta\right)$$

Lebesgue Decomposition

Corollary

If $\{\mathbf{X}_t\}$ *has spectral distribution function*

$$F_{\mathbf{X}}(\theta) = F_{\mathbf{X}}^{ac}(\theta) + F_{\mathbf{X}}^{d}(\theta) + F_{\mathbf{X}}^{s}(\theta)$$

then

$$d(\{\mathbf{X}_t\}) = d(\{\mathbf{X}_t^{ac}\})$$

where $\{\mathbf{X}_t^{ac}\}$ *has spectral distribution function* $F_{\mathbf{X}}^{ac}$.

$$\left(\mathbb{E}\left(\mathbf{X}_{t+\tau}\mathbf{X}_t^{\mathsf{T}}\right) - \mathbb{E}\left(\mathbf{X}_{t+\tau}\right)\mathbb{E}\left(\mathbf{X}_t^{\mathsf{T}}\right) = \int_{-1/2}^{1/2} \mathrm{e}^{-\imath 2\pi\tau\theta}\mathrm{d}F_{\mathbf{X}}(\theta)\right)$$

Information Dimension Rate is Relevant, too:

Communications & Information Theory:

- Rate-distortion theory
- $\lim_{n\to\infty} \frac{d(\mathbf{X}_1,\dots,\mathbf{X}_n)}{n}$ is a necessary rate for almost error-free compressed sensing[12]
- $d(\{\mathbf{X}_t\})$ is a sufficient rate for asymptotically distortion-free compressed sensing[13,14]
- Fact $d(\{\mathbf{X}_t\}) < \lim_{n\to\infty} \frac{d(\mathbf{X}_1,\dots,\mathbf{X}_n)}{n}$ for, e.g., bandlimited Gaussian processes reveals fundamental difference between error-free and distortion-free compressed sensing

Dynamical Systems Theory:

- Causality? (back-up slides)

[12] Wu and Verdú, "Optimal Phase Transitions in Compressed Sensing", 2012

[13] Jalali and Poor, "Universal Compressed Sensing for Almost Lossless Recovery", 2017

[14] Rezagah et al., "Compression-Based Compressed Sensing", 2017

Conclusions

- Information dimension for stochastic processes
- Intricately connected with bandwidth
- Relevant quantity in asymptotically distortion-free compressed sensing
- Generalization to causality measure currently unclear

Proofs, results for non-existing limits:

1702.00645

Conclusions

- Information dimension for stochastic processes
- Intricately connected with bandwidth
- Relevant quantity in asymptotically distortion-free compressed sensing
- Generalization to causality measure currently unclear

Proofs, results for non-existing limits:

1702.00645

Thanks for your attention!

Potential Connection to Causality

$$d(\{\mathbf{X}_t\}|\{\mathbf{Y}_t\}) \triangleq \lim_{m\to\infty} \lim_{n\to\infty} \frac{H([\mathbf{X}_1]_m, \dots, [\mathbf{X}_n]_m | \{\mathbf{Y}_t\})}{n \log m}$$

$$d(\{\mathbf{X}_t\}||\{\mathbf{Y}_t\}) \triangleq \lim_{m\to\infty} \lim_{n\to\infty} \frac{H([\mathbf{X}_1]_m, \dots, [\mathbf{X}_n]_m | \{\mathbf{Y}_t,\ t \leq n\})}{n \log m}$$

(we are not sure what proper definitions should look like!)

Conjecture

$$d(\{\mathbf{X}_t\}|\{\mathbf{Y}_t\}) \leq d(\{\mathbf{X}_t\}||\{\mathbf{Y}_t\})$$

with equality if $\mathbf{X}_t = f(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \dots) + \mathbf{E}_t$.

Potential Connection to Causality (cont'd)

Open Questions:

- Proper definitions of $d(\{\mathbf{Y}_t\}|\{\mathbf{X}_t\})$ and $d(\{\mathbf{Y}_t\}||\{\mathbf{X}_t\})$
- Investigating the Gaussian case
- Connections with causal/non-causal Wiener filters in the Gaussian case?
- Connections with directed information/transfer entropy?