Information Dimension

Mina Karzand

Massachusetts Institute of Technology

November 16, 2011

Information Dimension Definition

DOF of Interference Channel

Let X would be a real-valued random variable. For $m \in \mathbb{N}$, the $m-$ point uniform quantized version of X is shown by

$$\langle X \rangle_m = \frac{\lfloor mX \rfloor}{m}$$

Thus, $\langle X \rangle_m \in \mathbb{Z}/m$
Lower Information Dimension:

$$\underline{d}(X) = \liminf_{m\to\infty} \frac{H(\langle X \rangle_m)}{\log m}$$

Upper Information Dimension:

$$\overline{d}(X) = \limsup_{m\to\infty} \frac{H(\langle X \rangle_m)}{\log m}$$

If $\underline{d}(X) = \overline{d}(X)$, then

$$d(X) = \lim_{m \to \infty} \frac{H(\langle X \rangle_m)}{\log m}$$

Entropy of dimension $d(X)$:

$$\hat{H}(X) = \lim_{m \to \infty} [H(\langle X \rangle_m) - d(X) \log m]$$

- If $H(\lfloor X^n \rfloor) < \infty$, then $0 \le \underline{d}(X^n) \le \overline{d}(X^n) \le n$.

- If $H(\lfloor X^n \rfloor) < \infty$, then $0 \leq \underline{d}(X^n) \leq \overline{d}(X^n) \leq n$.
- If $\mathbb{E}[\log(1 + |X|)] < \infty$, then $\overline{d}(X) < \infty$.

- If $H(\lfloor X^n \rfloor) < \infty$, then $0 \le \underline{d}(X^n) \le \overline{d}(X^n) \le n$.
- If $\mathbb{E}[\log(1 + |X|)] < \infty$, then $\overline{d}(X) < \infty$.
- It is sufficient to restrict to the exponential subsequence $m = 2^l$. Define $[.]_l \triangleq \langle . \rangle_m$,

$$d(X) = \lim_{n \to \infty} \frac{H([X]_l)}{l}$$

- If $H(\lfloor X^n \rfloor) < \infty$, then $0 \leq \underline{d}(X^n) \leq \overline{d}(X^n) \leq n$.
- If $\mathbb{E}[\log(1 + |X|)] < \infty$, then $\overline{d}(X) < \infty$.
- It is sufficient to restrict to the exponential subsequence $m = 2^l$. Define $[.]_l \triangleq \langle . \rangle_m$,

$$d(X) = \lim_{n \to \infty} \frac{H([X]_l)}{l}$$

- *Translation Invariance*, $\forall x^n \in \mathbb{R}^n$,

$$d(x^n + X^n) = d(X^n)$$

- *Translation Invariance*, $\forall x^n \in \mathbb{R}^n$,

$$d(x^n + X^n) = d(X^n)$$

- *Scale Invariance*, $\forall \alpha \neq 0$,

$$d(\alpha X^n) = d(X^n)$$

- *Translation Invariance*, $\forall x^n \in \mathbb{R}^n$,

$$d(x^n + X^n) = d(X^n)$$

- *Scale Invariance*, $\forall \alpha \neq 0$,

$$d(\alpha X^n) = d(X^n)$$

- If X^n and Y^n are independent,

$$\begin{aligned} \max\{d(X^n), d(Y^n)\} &\leq d(X^n + Y^n) \\ &\leq d(X^n) + d(Y^n) \end{aligned}$$

- *Translation Invariance*, $\forall x^n \in \mathbb{R}^n$,

$$d(x^n + X^n) = d(X^n)$$

- *Scale Invariance*, $\forall \alpha \neq 0$,

$$d(\alpha X^n) = d(X^n)$$

- If X^n and Y^n are independent,

$$\begin{aligned} \max\{d(X^n), d(Y^n)\} &\leq d(X^n + Y^n) \\ &\leq d(X^n) + d(Y^n) \end{aligned}$$

- If $\{X_i\}$ are independent and $d(X_i)$ exists for all i,

$$d(X^n) = \sum_{i=1}^{n} d(X_i)$$

- *Translation Invariance*, $\forall x^n \in \mathbb{R}^n$,

$$d(x^n + X^n) = d(X^n)$$

- *Scale Invariance*, $\forall \alpha \neq 0$,

$$d(\alpha X^n) = d(X^n)$$

- If X^n and Y^n are independent,

$$\begin{aligned} \max\{d(X^n), d(Y^n)\} &\leq d(X^n + Y^n) \\ &\leq d(X^n) + d(Y^n) \end{aligned}$$

- If $\{X_i\}$ are independent and $d(X_i)$ exists for all i,

$$d(X^n) = \sum_{i=1}^{n} d(X_i)$$

- If X^n, Y^n and Z^n are independent, then

$$d(X^n + Y^n + Z^n) + d(Z^n) \leq d(X^n + Z^n) + d(Y^n) + d(Z^n)$$

A probability distribution can be uniquely represented as the mixture

$$v = pv_d + qv_c + rv_s$$

v_d: purely atomic prob. measure (discrete part)

v_c: absolutely continuous probability measure

v_s: probability measure singular with respect to Lebesgue measure

$p + q + r = 1$

Theorem: Let X be a random variable s.t. $H(\lfloor X \rfloor) < \infty$. Its distribution can be represented as

$$v = (1-\rho)v_d + \rho v_c$$

Then $d(X) = \rho$ and

$$\hat{H}(X) = (1-\rho)H(v_d) + \rho h(v_c) + h_b(\rho)$$

Renyi entropy of order α of a discrete random variable:

$$H_\alpha(Y) = \begin{cases} \sum_y p_y \log \frac{1}{p_y}, & \alpha = 1 \\ \log \frac{1}{\max_y p_y}, & \alpha = \infty \\ \frac{1}{1-\alpha} \log \left(\sum_y p_y^\alpha \right), & \alpha \neq 1, \infty. \end{cases}$$

Renyi entropy of order α of a discrete random variable:

$$H_\alpha(Y) = \begin{cases} \sum_y p_y \log \frac{1}{p_y}, & \alpha = 1 \\ \log \frac{1}{\max_y p_y}, & \alpha = \infty \\ \frac{1}{1-\alpha} \log \left(\sum_y p_y^\alpha \right), & \alpha \neq 1, \infty. \end{cases}$$

$$\underline{d}_\alpha(X) = \liminf_{m\to\infty} \frac{H_\alpha(\langle X \rangle_m)}{\log m}$$

$$\overline{d}_\alpha(X) = \limsup_{m\to\infty} \frac{H_\alpha(\langle X \rangle_m)}{\log m}$$

Renyi entropy of order α of a discrete random variable:

$$H_\alpha(Y) = \begin{cases} \sum_y p_y \log \frac{1}{p_y}, & \alpha = 1 \\ \log \frac{1}{\max_y p_y}, & \alpha = \infty \\ \frac{1}{1-\alpha} \log \left(\sum_y p_y^\alpha\right), & \alpha \neq 1, \infty. \end{cases}$$

$$\underline{d}_\alpha(X) = \liminf_{m\to\infty} \frac{H_\alpha(\langle X\rangle_m)}{\log m}$$

$$\overline{d}_\alpha(X) = \limsup_{m\to\infty} \frac{H_\alpha(\langle X\rangle_m)}{\log m}$$

$$\hat{H}_\alpha(X) = \lim_{m\to\infty} \left[H_\alpha(\langle X\rangle_m) - d_\alpha(X)\log m\right]$$

Theorem: Let X be a real random variable, satisfying the property $H_{\alpha(\lfloor X \rfloor) < \infty}$ with the distribution represented as:

$$v = pv_d + qv_c + rv_s$$

Then,

- For $\underline{\alpha > 1}$: If $p > 0$ (X has a discrete component), then $d_\alpha(X) = 0$ and $\hat{H}_\alpha(X) = H_\alpha(v_d) + \frac{\alpha}{1-\alpha} \log p$.

Theorem: Let X be a real random variable, satisfying the property $H_{\alpha(\lfloor X \rfloor) < \infty}$ with the distribution represented as:

$$v = pv_d + qv_c + rv_s$$

Then,

- For $\underline{\alpha > 1}$: If $p > 0$ (X has a discrete component), then $d_\alpha(X) = 0$ and $\hat{H}_\alpha(X) = H_\alpha(v_d) + \frac{\alpha}{1-\alpha} \log p$.
- For $\underline{\alpha < 1}$: If $q > 0$(X has an absolutely continuous part), then $d_\alpha(X) = 1$ and $\hat{H}_\alpha(X) = h_\alpha(v_c) + \frac{\alpha}{a-\alpha} \log q$

Dyadic expansion of X can be written as

$$X = \sum_{j=1}^{\infty} (X)_j 2^{-j}$$

There is a one to one correspondence between X and the binary random process $\{(X)_j, j \in \mathbb{N}\}$

$$\underline{d}(X) = \liminf_{i \to \infty} \frac{H((X)_1, (X)_2, ..., (X)_i)}{i}$$

$$\overline{d}(X) = \limsup_{i \to \infty} \frac{H((X)_1, (X)_2, ..., (X)_i)}{i}$$

Random variables whose lower and upper information dimension differ can be constructed from processes with different lower and upper entropy rate.

Cantor Distribution

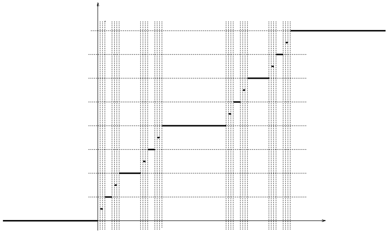
$$
\begin{aligned}
C_0 &= [0,1] \\
C_1 &= [0,1/3] \cup [2/3,1] \\
C_2 &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\
C_3 &= \cdots
\end{aligned}
$$

The support of the Cantor distribution is the Cantor set $\cap_{i=1}^{\infty} C_i$.

Cantor Distribution

$$
\begin{aligned}
C_0 &= [0, 1] \\
C_1 &= [0, 1/3] \cup [2/3, 1] \\
C_2 &= [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1] \\
C_3 &= \cdots
\end{aligned}
$$

The support of the Cantor distribution is the Cantor set $\cap_{i=1}^{\infty} C_i$.



Degrees of freedom of the interference channel

Channel Model:
K-user real-valued memoryless Gaussian interference channel with a fixed deterministic channel matrix $\mathbf{H} = [h_{ij}]$ (known at encoder and decoder), where at each symbol epoch the $i-$th user transmits X_i and the $i - th$ decoder receives

$$Y_i = \sum_{j=1}^{k} \sqrt{snr} h_{ij} X_j + N_i$$

where $\{X_i, N_i\}_{i=1}^{K}$ are independent with $\mathbb{E}[X_i^2] \leq 1$ and $N_i \sim \mathcal{N}(0,1)$.

Sum-rate capacity:

$$\overline{C}(\mathbf{H}, snr) \triangleq \max\left\{\sum_{i=1}^{K} R_i : R^K \in \mathcal{C}(\mathbf{H}, snr)\right\}$$

Degrees of freedom or the multiplexing gain

$$DOF(\mathbf{H}) = \lim_{snr\to\infty} \frac{\overline{C}(\mathbf{H}, snr)}{\frac{1}{2}\log snr}$$

Theorem: Let X be independent of N which is standard normal random variable. Denote

$$I(X, snr) = \triangleq I(X; \sqrt{snr}X + N)$$

Then,

$$\lim_{snr \to \infty} \frac{I(X, snr)}{\frac{1}{2} \log snr} = d(X)$$

Mutual information is maximized **asymptotically** *by any absolutely continuous input distribution, where* $d(X) = 1$*.*

Information dimension under projection

- Almost every projection preserves the dimension.
- But, computing the dimension for individual projections is in general difficult.

Theorem: Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ with $m \leq n$. Then for any X^n,

$$d(\mathbf{A}X^n) \leq \min\{d(X^n), rank(\mathbf{A})\}$$

Theorem: Let $\alpha \in (1, 2]$ and $m \leq n$. Then for almost every $\mathbf{A} \in \mathcal{R}^{m \times n}$,

$$d_\alpha(\mathbf{A}X^n) = \min\{d_\alpha(X^n), m\}$$

Theorem: Let,

$$dof(X^K, \mathbf{H}) \triangleq \sum_{i=1}^{K} \left[d\left(\sum_{j=1}^{K} h_{ij} X_j \right) - d\left(\sum_{j \neq i} h_{ij} X_j \right) \right]$$

Then,

$$DOF(\mathbf{H}) = \sup_{X^K} dof(X^K, \mathbf{H})$$

where the supremum is over independent $X_1, X_2, ..., X_K$ such that

$$H(\lfloor X_i \rfloor) \leq C$$

for some fixed $C > 0$.

Theorem: Let,

$$dof(X^K, \mathbf{H}) \triangleq \sum_{i=1}^{K} \left[d\left(\sum_{j=1}^{K} h_{ij} X_j \right) - d\left(\sum_{j \neq i} h_{ij} X_j \right) \right]$$

Then,

$$DOF(\mathbf{H}) = \sup_{X^K} dof(X^K, \mathbf{H})$$

where the supremum is over independent $X_1, X_2, ..., X_K$ such that

$$H(\lfloor X_i \rfloor) \leq C$$

for some fixed $C > 0$.
Applies to non-Gaussian noise as long as finite non-Gaussianness, $D(N||N_G) < \infty$.

$$dof(X^K, \mathbf{H}) \triangleq \sum_{i=1}^{K} \left[\underbrace{d\left(\sum_{j=1}^{K} h_{ij} X_j\right)}_{\text{info. dim. of the i-th user}} - \underbrace{d\left(\sum_{j \neq i} h_{ij} X_j\right)}_{\text{info. dim. of the interference}} \right]$$

$$\overline{C}(\mathbf{H}, snr) = \lim_{n\to\infty} \frac{1}{n} \sup_{X_1^n,\dots,X_K^n} \sum_{i=1}^{K} I(X_i^n; Y_i^n)$$

- $X_i^n = [X_{i,1}, X_{i,2}, \dots, X_{i,n}]$: $i-$th input user.
- sup is over independent $X_1^n, \dots, X_K^n$.
-

$$\begin{aligned} I(X_i^n; Y_i^n) &= I(X_1^n, \dots, X_K^n; Y_i^n) - I(X_1^n, \dots, X_K^n; Y_i^n | X_i^n) \\ &= I\left(\sum_{j=1}^{K} h_{ij} X_j^n, snr\right) - I\left(\sum_{j\neq i} h_{ij} X_j^n, snr\right) \end{aligned}$$

$$
\begin{aligned}
DOF(\mathbf{H}) = & \lim_{snr\to\infty} \lim_{n\to\infty} \sup_{X_1^n,\dots,X_K^n} \frac{1}{\frac{n}{2}\log snr} \\
& \sum_{i=1}^{K} \left\{ I\left(\sum_{j=1}^{K} h_{ij} X_j^n, snr \right) - I\left(\sum_{j\neq i} h_{ij} X_j^n, snr \right) \right\} \\
= & \lim_{n\to\infty} \sup_{X_1^n,\dots,X_K^n} \lim_{snr\to\infty} \frac{1}{\frac{n}{2}\log snr} \\
& \sum_{i=1}^{K} \left\{ I\left(\sum_{j=1}^{K} h_{ij} X_j^n, snr \right) - I\left(\sum_{j\neq i} h_{ij} X_j^n, snr \right) \right\}
\end{aligned}
$$

$$DOF(\mathbf{H}) = \lim_{snr\to\infty} \lim_{n\to\infty} \sup_{X_1^n,\dots,X_K^n} \frac{1}{\frac{n}{2}\log snr} \sum_{i=1}^{K} \left\{ I\left(\sum_{j=1}^{K} h_{ij}X_j^n, snr\right) - I\left(\sum_{j\neq i} h_{ij}X_j^n, snr\right) \right\}$$

$$= \lim_{n\to\infty} \sup_{X_1^n,\dots,X_K^n} \lim_{snr\to\infty} \frac{1}{\frac{n}{2}\log snr} \sum_{i=1}^{K} \left\{ I\left(\sum_{j=1}^{K} h_{ij}X_j^n, snr\right) - I\left(\sum_{j\neq i} h_{ij}X_j^n, snr\right) \right\}$$

$$I(., snr) = \frac{d(.)}{2}\log snr + o(\log snr)$$

$$DOF(\mathbf{H}) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{K} \left\{ d\left(\sum_{j=1}^{K} h_{ij} X_j^n, snr \right) - d\left(\sum_{j\neq i} h_{ij} X_j^n, snr \right) \right\}$$

SINGLE LETTERIZATION AND EXAMPLES

Two user IC

$$DOF\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right)$$

$$= \sup_{X_1 \perp X_2} d(aX_1 + bX_2) + d(cX_1 + dX_2) - d(bX_2) - d(cX_1)$$

$$= \begin{cases} 0, & a = d = 0 \\ 2, & a \neq 0, d \neq 0, b = c = 0 \\ 1, & \text{otherwise} \end{cases}$$

Many-to-one IC:

$$DOF\left(\begin{bmatrix} h_{11} & h_{12} & h_{13} & \cdots & h_{1K} \\ 0 & h_{22} & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & h_{KK} \end{bmatrix}\right) = K-1$$

Achieved by choosing X_1 discrete and the rest absolutely continuous.

One-to-Many IC:

$$DOF\left(\begin{bmatrix} h_{11} & 0 & \cdots & 0 & 0 \\ h_{21} & h_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & & \ddots & 0 \\ h_{K1} & 0 & \cdots & 0 & h_{KK} \end{bmatrix}\right) = K - 1$$

Achieved by choosing X_1 discrete and the rest absolutely continuous.

MAC:

$$DOF\left(\begin{bmatrix}1 & \cdots & 1\\ \vdots & \ldots & \vdots\\ 1 & \cdots & 1\end{bmatrix}\right) = 1$$

Information Dimension and Rate Distortion Theory

For scalar source and MSE distortion, whenever $d(X)$ exists and is finite, as $D \to 0$

$$R_X(D) = \frac{d(X)}{2} \log \frac{1}{D} + o(\log D)$$

Information Dimension and Rate Distortion Theory

For scalar source and MSE distortion, whenever $d(X)$ exists and is finite, as $D \to 0$

$$R_X(D) = \frac{d(X)}{2} \log \frac{1}{D} + o(\log D)$$

- X is discrete and $H(X) < \infty$:
$$R_X(D) = H(X) + o(1)$$
- X is continuous and $h(X) > -\infty$:
$$R_X(D) = \frac{1}{2} \log \frac{1}{2\pi e D} + h(X) + o(1)$$
- X is discrete-continuous mixed:
$$R_X(D) = \frac{\rho}{2} \log \frac{1}{D} + \hat{H}(X) + o(1)$$