Fast Evaluation of Modified Renyi Entropy for Fractal Analysis

Jaromir Kukal & Martin Dlask

FNSPE, CTU in Prague

May 22, 2022

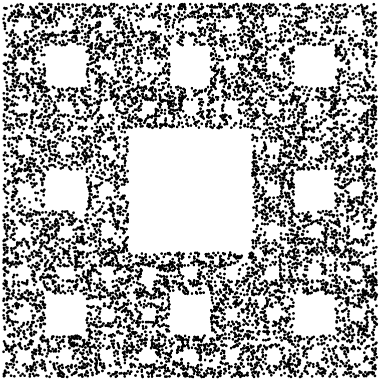
Main Ideas

- Boxcounting offers biased estimate of capacity dimension D_0
- Particle counting for biased estimation of D_α
- Particle counting is sensitive to grid translation and rotation
- Investigation of finite sample from given uncountable set
- Basic tools: Parzen density estimate & Renyi entropy
- New term: Modified Renyi entropy H^*_α
- H^*_α is translation, rotation and mirroring invariant
- Monte Carlo estimation of H^*_α and D^*_α
- D^*_α estimate can be also scaling invariant
- Effective implementation of ϵ-querry using k-d tree
- Proven relationships: $D^*_0 = D_0, \quad D^*_2 = D_2$

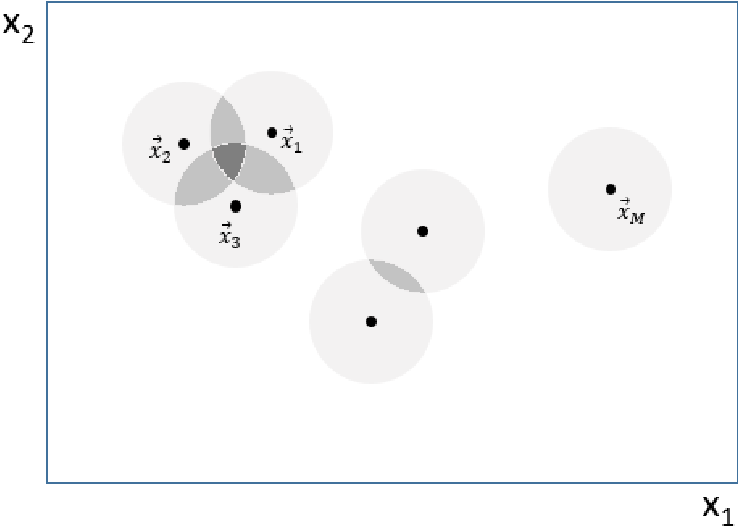
Dlask, M., Kukal, J., Translation and Rotation Invariant Method of Renyi Dimension Estimation, Chaos, Solitons & Fractals, Elsevier, 114(C), 536-541, 2018.

Uniform Sampling from Fractal Set

- Investigated structure $\mathcal{F} \subset \mathbb{R}^n$
- Uniform sampling $\mathbf{x} \sim \mathrm{U}(\mathcal{F})$
- Sample $\Phi = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_M\} \subset \mathcal{F}$



TRM Invariant Parzen Estimate



Trivial Formulas First

- n-dimensional ball $\mathcal{B}(\mathbf{x},\epsilon) = \{\mathbf{y} \in \mathbb{R}^n : \|\mathbf{y}-\mathbf{x}\|_2 \le \epsilon\}$
- Ball volume $V_* = V_n \cdot \epsilon^n$
- Kernel density $\mathrm{f}_0(\mathbf{x},\epsilon) = \mathrm{I}(\|\mathbf{x}\|_2 \le \epsilon)/V_*$
- Parzen formula $\mathrm{f}(\mathbf{x},\Phi,\epsilon) = M^{-1}\sum_{k=1}^{M} \mathrm{f}_0(\mathbf{x}-\mathbf{x}_k,\epsilon)$
- Point degeneracy $\mathrm{G}(\mathbf{x},\Phi,\epsilon) = \sum_{k=1}^{M} \mathrm{I}(\|\mathbf{x}-\mathbf{x}_k\|_2 \le \epsilon)$
- Simplified PDF

$$\mathrm{f}(\mathbf{x},\Phi,\epsilon) = \frac{1}{M \cdot V_*}\sum_{k=1}^{M} \mathrm{I}(\|\mathbf{x}-\mathbf{x}_k\|_2 \le \epsilon) = \frac{\mathrm{G}(\mathbf{x},\Phi,\epsilon)}{M \cdot V_*}$$

- Minkowski sausage $\mathcal{S} \approx \mathcal{S}_M = \bigcup_{k=1}^{M} \mathcal{B}(\mathbf{x}_k,\epsilon)$
- Degeneracy range $\mathrm{G}(\mathbf{x},\Phi,\epsilon) \in \{1,\dots,M\}$ for $\mathbf{x} \in \mathcal{S}_M$

Renyi Entropy

- Exponent range $\alpha \in [0,1) \cup (1,\infty)$
- Renyi entropy
  $$H_\alpha = \frac{1}{1-\alpha} \ln \mathrm{E}\ p^{\alpha-1}$$
- Hartley entropy $H_0 = \ln \mathrm{E}\ p^{-1}$
- Shannon entropy $H_1 = -\mathrm{E}\ \ln p$
- Collision entropy $H_2 = -\ln \mathrm{E}\ p$
- Min-entropy $H_\infty = -\ln \sup p$
- Event probabilities from grid
- Biased estimation

Novelty: Modified Renyi Entropy

- Main integral

$$\mathrm{J}(\Phi, \alpha, \epsilon) = \int_{\mathbf{x} \in \mathbb{R}^n} \mathrm{f}^{\alpha}(\mathbf{x}, \Phi, \epsilon) \mathrm{d}\mathbf{x}$$

- Referential integral

$$\mathrm{J}_0(\alpha, \epsilon) = \int_{\mathbf{x} \in \mathbb{R}^n} \mathrm{f}_0^{\alpha}(\mathbf{x}, \epsilon) \mathrm{d}\mathbf{x} = V_*^{1-\alpha}$$

- Modified Renyi entropy

$$\mathrm{H}_{\alpha}^*(\Phi, \epsilon) = \frac{\ln \mathrm{J}(\Phi, \alpha, \epsilon) - \ln \mathrm{J}_0(\alpha, \epsilon)}{1 - \alpha}$$

- Relationship to degeneracy

$$H_{\alpha}^* = \ln M + \frac{\ln \mathrm{E}\ G^{\alpha-1}}{1 - \alpha}$$

- Modified Hartley entropy
  $H_0^* = \ln M + \ln \mathrm{E}\ G^{-1}$
- Modified Shannon entropy
  $H_1^* = \lim_{\alpha \to 1} H_\alpha^* = \ln M - \mathrm{E}\ \ln G$
- Modified collision entropy
  $H_2^* = \ln M - \ln \mathrm{E}\ G$
- Modified min-entropy
  $H_\infty^* = \lim_{\alpha \to \infty} H_\alpha^* = \ln M - \ln \max G$
- Modified Renyi dimension
  $D_\alpha^* = \lim_{\epsilon \to 0^+} \frac{\mathrm{H}_\alpha^*(\epsilon)}{-\ln \epsilon}$

Resulting Procedure

Perform *NMC*-times for given α and fixed $\epsilon > 0$:

- Generate sample index $j \sim \mathrm{U}(\{1, \ldots, M\})$
- Generate point $\mathbf{x} \sim \mathrm{U}(\mathcal{B}(\mathbf{x}_j, \epsilon))$
- Calculate degeneracy $G = \sum_{k=1}^{M} \mathrm{I}(\|\mathbf{x} - \mathbf{x}_k\|_2 \leq \epsilon)$

Estimate adequate mean value E

Estimate modified Renyi entropy D_α^*

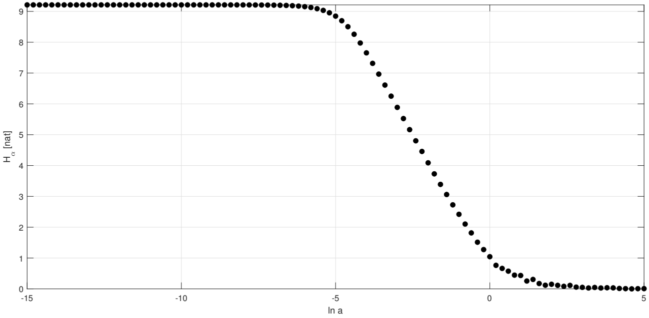
Using results for various ϵ values, we employ linear regression and estimate D_α^* as

$\mathrm{H}_\alpha^*(\epsilon) = C_\alpha - D_\alpha^* \ln \epsilon$

TSRM invariant estimation of D_α^*

How to Suppress Bias of D^*_α and Save Its MSE?

- Range of $H^*_\alpha \in [0, \ln M]$
- Fitting near inflex point
- Suggested setting $N \geq 10^7$, $NMC \geq 10^6$



How to Suppress Time Complexity?

- Effective implementation of ϵ-querry
- Batch construction of k-d tree with $\mathrm{T}(N) = \mathrm{O}(N \log^2 N)$
- Single querry complexity $\mathrm{T}(N, G) = \mathrm{O}(\log N) + G$
- Use only small values of ϵ
- Degeneracy constrain $G < N^\beta$ with $\beta \in (1/2, 3/4)$

Applications

- Sampling from IFS
- Sampling from function graph (fBm, Weierstrass, Wijs)
- Sampling from image structure (edges, watershed, binary morphology)
- Sampling from attractor (chaotic map, chaotic flow)
- Never apply to small sample
- Freely available tiny MATLAB library