Typicality for stratified measures

Juan Pablo Vigneaux

Department of Mathematics
California Institute of Technology
Pasadena, CA, USA
vigneaux@caltech.edu

Zürich, September 6th, 2023

Outline

1. Measure theory and limit laws
2. *Geometric* measure theory
3. Stratified measures
4. Open problems

Typicality

A random binary word $\mathbf{X} = (X_1, ..., X_n)$, with i.i.d. symbols, $X_i \sim \text{Ber}(q)$, is expected to have nq ones and $n(1-q)$ zeroes.

There are roughly $\binom{n}{nq} \approx e^{nH(q)}$ of such words, and each has probability $\approx e^{-nH(q)}$.

Entropy: $H(q) := -q \ln q - (1-q)\ln(1-q)$.

More formally, one might introduce a set of "typical realizations" of $\mathbf{X}$ in two ways:

- **Weak or entropic**: Sequences $\mathbf{x} = (x_1, ..., x_n) \in \{0,1\}^n$ such that $\left|-\frac{1}{n}\ln \mathbb{P}(X_1 = x_1)\cdots\mathbb{P}(X_n = x_n) - H(q)\right| < \delta$.
- **Strong** (when $0 < q < 1$): Sequences $\mathbf{x} \in \{0,1\}^n$ such that $\left|\frac{1}{n}N(1;\mathbf{x}) - q\right| < \delta$.

Law of large numbers: $-\frac{1}{n}\sum_{i=1}^{n}\mathbb{P}(X_i = x_i) \to H(q)$ and $\frac{1}{n}N(1;\mathbf{x}) \to q$ in probability.

Measure theory

Introduced by Lebesgue, Borel, etc. to understand integration.

Kolmogorov (1933) used it to axiomatize *probability* and formalize limit theorems.

Fundamental ingredients:

1. Set E of "elementary outcomes".

Measure theory

Introduced by Lebesgue, Borel, etc. to understand integration.

Kolmogorov (1933) used it to axiomatize *probability* and formalize limit theorems.

Fundamental ingredients:

1. Set E of "elementary outcomes".
2. σ-**algebra**: distinguished collection $\mathfrak{E}$ of subsets of E, called "events".

Measure theory

Introduced by Lebesgue, Borel, etc. to understand integration.

Kolmogorov (1933) used it to axiomatize *probability* and formalize limit theorems.

Fundamental ingredients:

1. Set E of "elementary outcomes".
2. σ-**algebra**: distinguished collection $\mathfrak{E}$ of subsets of E, called "events".
3. **Measure**: a function $\mu : \mathfrak{E} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\bigsqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$;
   the measure is σ-**finite** is E can be partitioned into countably many events of finite μ-measure.

Measure theory

Introduced by Lebesgue, Borel, etc. to understand integration.

Kolmogorov (1933) used it to axiomatize *probability* and formalize limit theorems.

Fundamental ingredients:

1. Set E of "elementary outcomes".
2. σ-**algebra**: distinguished collection $\mathfrak{E}$ of subsets of E, called "events".
3. **Measure**: a function $\mu : \mathfrak{E} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\bigsqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$;
   the measure is σ-**finite** is E can be partitioned into countably many events of finite μ-measure.
4. **Probability measure**: a measure ρ such that $\rho(E) = 1$.

Measure theory

Introduced by Lebesgue, Borel, etc. to understand integration.

Kolmogorov (1933) used it to axiomatize *probability* and formalize limit theorems.

Fundamental ingredients:

1. Set E of "elementary outcomes".
2. σ-**algebra**: distinguished collection $\mathfrak{E}$ of subsets of E, called "events".
3. **Measure**: a function $\mu : \mathfrak{E} \to [0, \infty]$ such that $\mu(\emptyset) = 0$ and $\mu(\bigsqcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i)$;
   the measure is σ-**finite** is E can be partitioned into countably many events of finite μ-measure.
4. **Probability measure**: a measure ρ such that $\rho(E) = 1$.
5. **Absolute continuity** $\rho \ll \mu$: there is a real-valued integrable function $f \equiv \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ such that $\rho(A) = \int_A f \, \mathrm{d}\mu$ for any $A \in \mathfrak{E}$.

Law of Large numbers and AEP

Setting: $(E_X, \mathfrak{B}, \mu)$ measure space; μ σ-finite; ρ proba; $\rho \ll \mu$; $f = \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$.

Entropy: $H_\mu(\rho) := \mathbb{E}_\rho\left(-\ln \frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right) = -\int_E f \ln f \,\mathrm{d}\mu$.

A sequence $(X_1, ..., X_n)$ of i.i.d. realizations of ρ has law $\rho^{\otimes n}$ (product) and density $f^{\otimes n}(x_1, ..., x_n) = f(x_1) \cdots f(x_n)$.

Law of Large numbers and AEP

Setting: $(E_X, \mathfrak{B}, \mu)$ measure space; μ σ-finite; ρ proba; $\rho \ll \mu$; $f = \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$.

Entropy: $H_\mu(\rho) := \mathbb{E}_\rho\left(-\ln \frac{\mathrm{d}\rho}{\mathrm{d}\mu}\right) = -\int_E f \ln f \,\mathrm{d}\mu$.

A sequence $(X_1, \dots, X_n)$ of i.i.d. realizations of ρ has law $\rho^{\otimes n}$ (product) and density $f^{\otimes n}(x_1, \dots, x_n) = f(x_1) \cdots f(x_n)$.

Proposition (AEP)

Suppose that the entropy $H_\mu(\rho) < \infty$. For every $\delta > 0$,

$$W_\delta^{(n)} := \left\{ \mathbf{x} \in E_X^n \ : \ \left| -\frac{1}{n} \ln f^{\otimes n}(\mathbf{x}) - H_\mu(\rho) \right| < \delta \right\}. \tag{1}$$

Then, for every $\varepsilon > 0$, provided n big enough,

1. $\rho^{\otimes n}(W_\delta^{(n)}) > 1 - \varepsilon$ *and*
2. $(1-\varepsilon)e^{n(H_\mu(\rho)-\delta)} \leq \mu^{\otimes n}(W_\delta^{(n)}) \leq e^{n(H_\mu(\rho)+\delta)}$.

Examples AEP

1. E finite,
   μ counting measure,
   $f = \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ probability mass function,
   $H_\mu(\rho) = -\sum_{i\in E} f(i) \ln f(i)$ discrete entropy,
   $\mu^{\otimes n}$ counting measure too.
   Therefore: $\#W_\delta^{(n)} \approx \exp(-n \sum_{i\in E} f(i) \ln f(i))$.

Examples AEP

1. E finite,
   μ counting measure,
   $f = \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ probability mass function,
   $H_\mu(\rho) = -\sum_{i \in E} f(i) \ln f(i)$ discrete entropy,
   $\mu^{\otimes n}$ counting measure too.
   Therefore: $\# W_\delta^{(n)} \approx \exp(-n \sum_{i \in E} f(i) \ln f(i))$.
2. $E = \mathbb{R}^d$,
   $\mu = \mathcal{L}^d$ Lebesgue measure (d-volume),
   $f = \frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ probability density function,
   $H_\mu(\rho) = -\int_E f \ln f$ differential entropy,
   $\nu^{\otimes n} = \mathcal{L}^{nd}$, the nd-dimensional volume.
   Therefore: $\mathrm{vol}(W_\delta^{(n)}) \approx \exp(-n \int_E f \log f)$.

Discrete-continuous mixture

Consider a probability measure $\rho = q\rho_1 + (1-q)\rho_0$ on $E_X = \mathbb{R}^d$, where:

1. $q \in [0, 1]$,
2. $\rho_1 \ll \mu_1 = L^d$, the Lebesgue measure.
3. $\rho_0 \ll \mu_0$, where μ_0 is the counting measure on a countable set $S \subset \mathbb{R}^d$.

Remark: $\rho \ll \mu_1 + \mu_0$.

Lemma

$$H_\mu(\rho) = H_\#(q) + qH_{\mu_1}(\rho_1) + (1-q)H_{\mu_0}(\rho_0).$$

Renyi's dimension and entropy

Given an arbitrary probability measure ρ on $\mathbb{R}^d$, Renyi (1959) first discretized it through a measurable partition of $\mathbb{R}^d$ into cubes with vertices in $\mathbb{Z}^d/n$, getting laws ρ_n with countable support.

If

$$H_{\#}(\rho_n) = D \ln n + h + o(1)$$

for some $D, h \in \mathbb{R}$, Renyi calls D the **information dimension** and h the D-**dimensional entropy** of the measure ρ.

When $\rho = q\rho_1 + (1-q)\rho_0$ is a discrete continuous mixture:

$$D = qd \quad \text{and} \quad h = H_{\#}(q) + qH_{\mu_1}(\rho_1) + (1-q)H_{\mu_0}(\rho_0).$$

Topological meaning of D?

A first asymptotic analysis

Set $E_0 = S$ and $E_1 = \mathbb{R}^d \setminus S$. Remark $\mu_1|_{E_1} = \mu_1$ and $\rho_1|_{E_1} = \rho_1$.

Partition $E_X^n = (\mathbb{R}^d)^n$ into strata $E_{y_1} \times \cdots E_{y_n}$, for any $\mathbf{y} = (y_1, ..., y_n) \in E_Y^n$, where $E_Y = \{0, 1\}$.

Then

$$\rho^{\otimes n} = \sum_{\mathbf{y} \in E_Y^n} q^{N(1;\mathbf{y})}(1-q)^{n-N(1;\mathbf{y})} \rho_{y_1} \otimes \cdots \otimes \rho_{y_n}.$$

A first asymptotic analysis

Set $E_0 = S$ and $E_1 = \mathbb{R}^d \setminus S$. Remark $\mu_1|_{E_1} = \mu_1$ and $\rho_1|_{E_1} = \rho_1$.

Partition $E_X^n = (\mathbb{R}^d)^n$ into strata $E_{y_1} \times \cdots E_{y_n}$, for any $\mathbf{y} = (y_1, \ldots, y_n) \in E_Y^n$, where $E_Y = \{0, 1\}$.

Then

$$\rho^{\otimes n} = \sum_{\mathbf{y} \in E_Y^n} q^{N(1;\mathbf{y})}(1-q)^{n-N(1;\mathbf{y})} \rho_{y_1} \otimes \cdots \otimes \rho_{y_n}.$$

See $\mathbf{y}$ as realization of $(Y_1, \ldots, Y_n) \sim \text{Ber}(q)^{\otimes n}$. Then $N(1; \mathbf{y}) \sim \text{Bin}(n, q)$, which entails concentration of probability around its mean: for $\mathbf{y}$ strongly typical, $N(1; \mathbf{y}) \approx nq$.

Consequently, a corresponding "typical stratum" $E_{y_1} \times \cdots E_{y_n}$ that "carries" $\rho_{y_1} \otimes \cdots \otimes \rho_{y_n}$ has roughly dimension nqd.

Meaning of the chain rule

Recall

$$H_\mu(\rho) = H_\#(q) + qH_{\mu_1}(\rho_1) + (1-q)H_{\mu_0}(\rho_0).$$

What is the meaning of this expression in terms of the AEP?

Recall

$$H_\mu(\rho) = H_\#(q) + qH_{\mu_1}(\rho_1) + (1-q)H_{\mu_0}(\rho_0).$$

What is the meaning of this expression in terms of the AEP?

It holds that

$$\underbrace{e^{nH_\mu(\rho)}}_{\approx \mu^{\otimes n}\text{-volume of } W^{(n)}(\rho)} = \underbrace{e^{nH_\#(q,1-q)}}_{\substack{\approx \#\text{ of typical } \mathbf{y} \\ =\#\text{typical strata}}} \underbrace{e^{n(qH_{\mu_1}(\rho_1)+(1-q)H_{\mu_0}(\rho_0))}}_{\substack{\approx\text{volume of typical realizations} \\ \text{in each stratum}}},$$

Outline

A problem

Let ρ be a probability measure on $\mathbb{R}^d$.

If there is an m-dimensional manifold E, with $m < d$, such that $\rho(\mathbb{R}^d \setminus E) = 0$, then $\rho \not\ll \mathcal{L}^d$. Hence the entropy $H_{\mathcal{L}^m}(\rho)$ is not defined. But ρ my have a density with respect to the natural m-dimensional volume.

A problem

Let ρ be a probability measure on $\mathbb{R}^d$.

If there is an m-dimensional manifold E, with $m < d$, such that $\rho(\mathbb{R}^d \setminus E) = 0$, then $\rho \not\ll \mathcal{L}^d$. Hence the entropy $H_{\mathcal{L}^m}(\rho)$ is not defined. But ρ my have a density with respect to the natural m-dimensional volume.

Moreover, if $\rho_\varepsilon \to \rho$ weakly as $\varepsilon \to 0$, where ρ_ε is a noisy version of ρ, it holds that $H_{\mathcal{L}^m}(\rho_\varepsilon) \to -\infty$ as $\varepsilon \to 0$.

This is consistent with the volume estimates presented above: the typical realizations of $\rho^{\otimes n}$ live in a space with vanishing $\mathcal{L}^{nd}$-volume.

m-Hausdorff measure

The m-Hausdorff measure gives a notion of m-dimensional volume:

$$\mathcal{H}^m(A) = \lim_{\delta \to 0} \inf_{\substack{\{S_i\}_{i \in \mathbb{N}} \\ A \subset \bigcup_i S_i, \, \text{diam } S_i < \delta}} \sum_{i \in I} \underbrace{w_m \left(\frac{\text{diam } S_i}{2} \right)^m}_{= \mathcal{L}^m(B(0, \frac{1}{2} \text{diam } S_i))}, \tag{2}$$

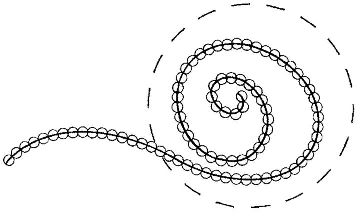

Figure: Case $m = 1$: the sum of diameters of smaller balls give a better approximation of the curve's length.

Rectifiable sets

In GMT: Manifolds $\rightsquigarrow$ Rectifiable sets (not necessarily smooth)
Smooth maps $\rightsquigarrow$ Lipschitz maps

Definition

A subset S of $\mathbb{R}^d$ is:

- *m-rectifiable*, for $m \leq d$, if it is the image of a bounded subset of $\mathbb{R}^m$ under a Lipschitz map;
- *countably m-rectifiable* if it is a countable union of m-rectifiable sets.
- *countably* $(\mathcal{H}^m, m)$*-rectifiable* if there exist countable m-rectifiable set containing $\mathcal{H}^m$-almost all of S.

Examples:

1. countably 0-rectifiable: countable set.
2. $\mathbb{R}^d$ is countably d-rectifiable.
3. An m-dimensional C^1 submanifold of $\mathbb{R}^d$ is countably m-rectifiable.

Another example: Manifolds

[Insert drawing]

Rectifiable measures

Let ρ be a locally finite and regular measure, and $s \in (0, \infty)$.

Limiting density: $\Theta_s(\rho, x) := \lim_{r \downarrow 0} \rho(B(x, r))/(w_s r^s)$

Theorem (Marstrand)

If $\Theta_s(\rho, x)$ exists and is strictly positive and finite for ρ-almost every x, then s is an integer not greater than n.

Theorem (Preiss)

... Moreover, there exists a countably $(\mathcal{H}^m, m)$-rectifiable Borel set E such that $\rho \ll \mathcal{H}^m|_E$.

In particular, $\rho(\mathbb{R}^d \setminus E) = 0$. Such measures are called **rectifiable**.

Koliander, Pichler, Riegler, and Hlawatsch (2016) studied them from an information-theoretic viewpoint

Proposition

If S is a countably $(\mathcal{H}^m, m)$-rectifiable subset of $\mathbb{R}^d$, then

$$S \subset S_0 \cup \bigcup_{i=0}^{\infty} f_i(K_i)$$

where:

- S_0 *is* $\mathcal{H}^m$*-null,*
- $(K_i)_i$ *compact subsets of* $\mathbb{R}^m$*,*
- $(f_i)_i$ *of Lipschitz functions from* $\mathbb{R}^m$ *to* $\mathbb{R}^d$*.*

In fact, it also holds that S is contained in the union of an $\mathcal{H}^m$-null set and countably many C^1 manifolds.

Carriers

Fact: an m-rectifiable measure ν is absolutely continuous with respect to the restricted measure $\mathcal{H}^m|_{E^*}$, where E^* is countably m-rectifiable. Call any such E^* a *carrier*.

In the case of discrete continuous mixtures, the sets E_0 and E_1 were carriers.

Lemma

Let S_i be a carrier of an m_i-rectifiable measure ν_i (for $i = 1, 2$). Then

- *S_i has Hausdorff dimension m_i;*
- *$S_1 \times S_2$ is a carrier of $\nu_1 \otimes \nu_2$, of Hausdorff dimension $m_1 + m_2$. Additionally,*

$$\mathcal{H}^{m_1+m_2}|_{S_1 \times S_2} = \mathcal{H}^{m_1}|_{S_1} \otimes \mathcal{H}^{m_2}|_{S^2}.$$

Outline

Stratified measures

Definition (k-stratified measure)

A measure ν on $\mathbb{R}^d$ is *k-stratified*, for $k \in \mathbb{N}^*$, if there are integers $(m_i)_{i=1}^k$ such that $0 \leq m_1 < m_2 < ... < m_k \leq d$ and ν can be expressed as a sum $\sum_{i=1}^k \nu_i$, where each ν_i is a nonzero m_i-rectifiable measure.

Examples:

- 1-stratified = rectifiable: discrete measure, continuous measure, measure carried by a manifold.
- 2-stratified: discrete-continuous mixtures

Standard form: one can find a sequence $(E_i)_{i=1}^k$ of *disjoint* rectifiable subsets of $\mathbb{R}^d$ such that E_i is *countably m_i-rectifiable* for each i, and $\nu = \sum_{i=1}^k q_i \nu_i$ with ν_i probability measures, $\nu_i \ll \mathcal{H}^{m_i}|_{E_i}$.

When ν probability measure, then $(q_1, ..., q_n)$ probability vector.

Products

$\rho = \sum_{i=1}^{k} q_i \rho_i$ probability measure in standard form.
$E_Y = \{1, \ldots, k\}$.

Then:

$$\rho^{\otimes n} = \sum_{\mathbf{y}=(y_1,\ldots,y_n)\in E_Y^n} q_1^{N(1;\mathbf{y})} \cdots q_k^{N(k;\mathbf{y})} \rho_{y_1} \otimes \cdots \otimes \rho_{y_n}.$$

Each stratum $\Sigma_{\mathbf{y}} = E_{y_1} \times \cdots \times E_{y_n}$ has Hausdorff dimension $m(\mathbf{y}) = \sum_{i=1}^{n} m_{y_i}$.

The measure $\rho_{y_1} \otimes \cdots \otimes \rho_{y_n}$ is absolutely continuous w.r.t. $\mu_{y_1} \otimes \cdots \otimes \mu_{y_n}$, which by the result above is the $m(\mathbf{y})$-Hausdorff measure on $\Sigma_{\mathbf{y}}$.

Asymptotic concentration

Recall: $N(i;\mathbf{y})/n \to q_i$ in probability.
The set $A^{(n)}_{\delta'}$ of $\mathbf{y} \in E^n_Y$ such that $|N(i;\mathbf{y})/n - q_i| < \delta'$, for all $i \in E_Y$, concentrates almost all the probability when n is big (strong typicality).

Number of typical strata $= \#A^{(n)}_{\delta'} \approx \exp(nH_{\#}(q_1, ..., q_k))$.

Moreover,

$$\rho^{\otimes n} \approx \rho^{(n)} := \sum_{\mathbf{y}\in A^{(n)}_{\delta'}} q_1^{N(1;\mathbf{y})} \cdots q_k^{N(k;\mathbf{y})} \underbrace{\rho_{y_1} \otimes \cdots \otimes \rho_{y_n}}_{\ll \mathcal{H}^{m(\mathbf{y})}|_{\Sigma_\mathbf{y}}} .$$

For $\mathbf{y} \in A^{(n)}_{\delta'}$, it holds that $m(\mathbf{y}) \approx n \sum_{i=1}^k q_i m_i$,
and that conditional entropy $H(X|Y) := \sum_{i=1}^k q_i H_{\mu_i}(\rho_i)$ is such that

$$\underbrace{\mu^{\otimes n}}_{\mathcal{H}^{m(\mathbf{y})}}(W^{(n)}_\delta(\rho) \cap \Sigma_\mathbf{y}) \approx \exp(nH(X|Y)).$$

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A_{\delta'_n}^{(n)}$ strongly typical sequences with $\delta'_n = C_{(q_1,\ldots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A^{(n)}_{\delta'_n}$ strongly typical sequences with $\delta'_n = C_{(q_1,\ldots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.
- $T^{(n)}_{\delta,\delta'_n}(\mathbf{y}) = W^{(n)}_{\delta} \cap \Sigma_{\mathbf{y}}$ for $\mathbf{y} \in A^{(n)}_{\delta'_n}$, doubly typical sequences.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A^{(n)}_{\delta'_n}$ strongly typical sequences with $\delta'_n = C_{(q_1,\ldots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.
- $T^{(n)}_{\delta,\delta'_n}(\mathbf{y}) = W^{(n)}_{\delta} \cap \Sigma_{\mathbf{y}}$ for $\mathbf{y} \in A^{(n)}_{\delta'_n}$, doubly typical sequences.
- $m(\mathbf{y}) = \sum_{i=1}^{n} m_{y_i}$ dimension of $\Sigma_{\mathbf{y}}$ and $T^{(n)}_{\delta,\delta'_n}(\mathbf{y})$.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A_{\delta'_n}^{(n)}$ strongly typical sequences with $\delta'_n = C_{(q_1,\ldots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.
- $T_{\delta,\delta'_n}^{(n)}(\mathbf{y}) = W_{\delta}^{(n)} \cap \Sigma_{\mathbf{y}}$ for $\mathbf{y} \in A_{\delta'_n}^{(n)}$, doubly typical sequences.
- $m(\mathbf{y}) = \sum_{i=1}^{n} m_{y_i}$ dimension of $\Sigma_{\mathbf{y}}$ and $T_{\delta,\delta'_n}^{(n)}(\mathbf{y})$.

Theorem

For any $\varepsilon > 0$ and $\delta > 0$, if n big enough, then $d_{TV}(\rho^{\otimes n}, \rho^{(n)}) < \varepsilon$ where $\rho_{\delta,\delta'_n}^{(n)} = \sum_{\mathbf{y} \in A_{\delta'_n}^{(n)}} \rho^{\otimes n}|_{T_{\delta,\delta'_n}^{(n)}(\mathbf{y})}$.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A^{(n)}_{\delta'_n}$ strongly typical sequences with $\delta'_n = C_{(q_1,\ldots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.
- $T^{(n)}_{\delta,\delta'_n}(\mathbf{y}) = W^{(n)}_{\delta} \cap \Sigma_{\mathbf{y}}$ for $\mathbf{y} \in A^{(n)}_{\delta'_n}$, doubly typical sequences.
- $m(\mathbf{y}) = \sum_{i=1}^{n} m_{y_i}$ dimension of $\Sigma_{\mathbf{y}}$ and $T^{(n)}_{\delta,\delta'_n}(\mathbf{y})$.

Theorem

For any $\varepsilon > 0$ and $\delta > 0$, if n big enough, then $d_{TV}(\rho^{\otimes n}, \rho^{(n)}) < \varepsilon$ where $\rho^{(n)}_{\delta,\delta'_n} = \sum_{\mathbf{y} \in A^{(n)}_{\delta'_n}} \rho^{\otimes n}|_{T^{(n)}_{\delta,\delta'_n}(\mathbf{y})}$. Moreover, $\rho^{(n)}$ is stratified: a sum of m-rectifiable measures for m in $[n\mathbb{E}(D) - n^{1/2+\xi}, n\mathbb{E}(D) + n^{1/2+\xi}]$. where $\mathbb{E}(D) = \sum_{i=1}^{k} q_i m_i$.

- $\rho = \sum_{i=1}^{k} q_i \rho_i$ stratified measure in standard form.
- $A^{(n)}_{\delta'_n}$ strongly typical sequences with $\delta'_n = C_{(q_1,\dots,q_n)} n^{-1/2+\xi}$, with $\xi \in (0, 1/2)$.
- $T^{(n)}_{\delta,\delta'_n}(\mathbf{y}) = W^{(n)}_{\delta} \cap \Sigma_{\mathbf{y}}$ for $\mathbf{y} \in A^{(n)}_{\delta'_n}$, doubly typical sequences.
- $m(\mathbf{y}) = \sum_{i=1}^{n} m_{y_i}$ dimension of $\Sigma_{\mathbf{y}}$ and $T^{(n)}_{\delta,\delta'_n}(\mathbf{y})$.

Theorem

For any $\varepsilon > 0$ and $\delta > 0$, if n big enough, then $d_{TV}(\rho^{\otimes n}, \rho^{(n)}) < \varepsilon$ where $\rho^{(n)}_{\delta,\delta'_n} = \sum_{\mathbf{y}\in A^{(n)}_{\delta'_n}} \rho^{\otimes n}|_{T^{(n)}_{\delta,\delta'_n}(\mathbf{y})}$. Moreover, $\rho^{(n)}$ is stratified: a sum of m-rectifiable measures for m in $[n\mathbb{E}(D) - n^{1/2+\xi}, n\mathbb{E}(D) + n^{1/2+\xi}]$. where $\mathbb{E}(D) = \sum_{i=1}^{k} q_i m_i$. Finally,

1. *For any $\mathbf{y} \in A^{(n)}_{\delta'_n}$, $\frac{1}{n} \ln \mathcal{H}^{m(\mathbf{y})}(T^{(n)}_{\delta,\delta'_n}(\mathbf{y})) \leq H(X|Y) + (\delta + \delta'_n)$.*
2. *For any $\varepsilon > 0$, the set $B^{(n)}_{\varepsilon}$ of $\mathbf{y} \in\subset A^{(n)}_{\delta'_n}$ such that $\frac{1}{n} \ln \mathcal{H}^{m(\mathbf{y})}(T^{(n)}_{\delta,\delta'_n}(\mathbf{y})) > H(X|Y) - \varepsilon + (\delta + \delta'_n)$, satisfies $\limsup_{\delta\to 0^+} \limsup_{n\to\infty} \frac{1}{n} \ln |B^{(n)}_{\varepsilon}| = H(Y)$.*

Verdú and Wu proved:

$$\dim_I \rho = \lim_{\varepsilon \to 0} \mathbb{E}_\rho \left(\frac{\ln \rho(B(X, \varepsilon))}{\ln \varepsilon} \right), \quad (3)$$

where B denotes an Euclidean ball.

Preiss' theorem: a measure ρ is m-rectifiable if and only if the density $\Theta_m(\rho, x)$ exists and is finite and nonzero for ρ-almost every x. In particular,

$$\lim_{\varepsilon \to 0} \frac{\ln \rho(B(x, \varepsilon))}{\ln \varepsilon} = m \quad \rho\text{-a.e.}$$

Therefore $\dim_I \rho = m$, *provided one can exchange the limit and the expectation in* (3). We say ρ is **dimensional regular**.

For a stratified measure $\rho = \sum_{i=1}^k q_i \rho_i$, with ρ_i being m_i-rectifiable and dimensional regular, Renyi's information dimension is $D = \sum_{i=1}^k q_i m_i$.

Rectifiable entropy differs from Renyi's dimensional entropy

Theorem (Perez, Csiszar, ...)

1. *μ be a σ-finite measure, ρ proba., $\rho \ll \mu$. Then*

$$\frac{\mathrm{d}\rho}{\mathrm{d}\mu} = \lim_n \sum_{\substack{\mathbf{j}\in\mathbb{Z}^d \\ \mu(C_{2^n,\mathbf{j}})>0}} \frac{\rho(C_{2^n,\mathbf{j}})}{\mu(C_{2^n,\mathbf{j}})} \chi_{C_{2^n,\mathbf{j}}} \tag{4}$$

almost surely and in L^1.

2. *$\rho \ll \mathcal{H}^m|_E$ for some $(\mathcal{H}^m, m)$-rectifiable Borel set E. Suppose $\mu := \mathcal{H}^m|_E$ is a finite measure and that $\frac{\mathrm{d}\rho}{\mathrm{d}\mu}$ is bounded. Then*

$$H_\mu(\rho) = \lim_n \left(H_\#(\rho_{2^n}) + \sum_{\substack{\mathbf{j}\in\mathbb{Z}^d \\ p_{2^n,\mathbf{j}}>0}} p_{2^n,\mathbf{j}} \ln \mu(C_{2^n,\mathbf{j}}) \right). \tag{5}$$

Outline

Rate-distortion function (RDF)

Given a probability measure ρ on $\mathbb{R}^d$, the (Euclidean, quadratic) RDF maps $D \in (0, \infty)$ to

$$R_\rho(D) = \inf_{\{K(y|x) : \int \|x-y\|_2^2 K(y|x), d\rho(x) < D\}} I(\mathbf{X}, \mathbf{Y}), \tag{6}$$

where $\mathbf{X} \sim \rho$, $K(\mathbf{y}|\mathbf{x})$ is a stochastic transition kernel that gives the distribution of $\mathbf{Y} \in \mathbb{R}^d$, and $I(\mathbf{X}, \mathbf{Y})$ is the mutual information introduced by Kolmogorov (via finite partitions).

Classical result: when $\mathbf{X} \sim \rho$ on $\mathbb{R}^d$ has a sufficiently smooth density f,

$$R_\rho(D) \sim -\frac{k}{2} \log(2\pi e D) - \int f \ln f \, \mathrm{d}x$$

as $D \to 0$.

Dembo and Kawabata: $\lim_{D\to 0} \frac{R_\rho(D)}{-\frac{1}{2} \log D} = \dim_I(\rho)$.

More on RDF

Charusaie *et al.*: define the **dimensional-rate bias** $b(\rho)$ as

$$\lim_{D \to 0} \left(R_\rho(D) - \underbrace{\left(-\frac{\dim_I(\rho)}{2} \log D + f(\dim_I \rho) \right)}_{= H_{\mathcal{L}^k}(\mathbf{X}) \text{ where } \mathbf{X} \sim \mathcal{N}_r(0, \alpha I) \text{ and } \mathbb{E}(\|\mathbf{X}\|^2) = D} \right) \qquad (7)$$

when it exists.

Analogous to Renyi's definition of the dimensional entropy.

They proved that when ρ is affinely singular (stratified, with each E_i a union of affine m_i-dimensional spaces), then $b(\rho) = H_\mu(\rho)$. In this case μ is a sum of Lebesgue measures.

How about general stratified measures?

Statistical estimation

How to device a statistical test to establish that a certain measure from which we can draw samples is rectifiable? And stratified?

Could we use that if ρ is m-rectifiable then

$$\lim_{\varepsilon \to 0} \frac{\ln \rho(B(x, \varepsilon))}{\ln \varepsilon} = m \quad \rho\text{-a.e.?}$$

Conclusion

- A measure-theoretic statement of the AEP for memoryless sources with general alphabets gives a unifying picture of the asymptotic behavior of discrete, continuous, discrete-continuous, rectifiable and stratified sources (measures).
- Geometric measure theory provides a very robust and general language to talk about measures that "live" on a geometric space.
- Typical realizations of a stratified measure concentrate on strata of a few typical dimensions, around a mean value that "coincides" with its information dimension.
- The entropy of a stratified measure satisfies a chain rule whose conditional term quantifies the typical realizations in each typical stratum.
- The standard entropy, defined via density, in general differs from Renyi's d-dimensional entropy.
- Open problems concerning: information dimension of rectifiable measures, asymptotic behavior of RDF, statistical test for rectifiability.