On the entropy of rectifiable and stratified measures

Juan Pablo Vigneaux

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

GSI 2023 - Information Theory and Statistics

Saint-Malo, August 30th, 2023

Outline

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

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**: collection $\mathfrak{E}$ of subsets of E (called "events"), closed under complementation and countable unions, with $\emptyset \in \mathfrak{E}$.
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 pieces 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; $\rho \ll \mu$ with density f.

Proposition (AEP)

Suppose that the entropy $H_\mu(\rho) := -\int_E f \ln f \, \mathrm{d}\mu = \mathbb{E}_\rho \left(-\ln \frac{\mathrm{d}\rho}{\mathrm{d}\mu} \right) < \infty$.
Define: for every $\delta > 0$,

$$W_\delta^{(n)} = \left\{ (x_1, \ldots, x_n) \in E_X^n \; : \; \left| -\frac{1}{n} \ln \prod_{i=1}^{n} f(x_i) - H_\mu(\rho) \right| < \delta \right\}. \quad (1)$$

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

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

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. $\rho_1 \ll \mathcal{L}^d$, 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

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

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

A first asymptotic analysis

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

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

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

See $\mathbf{y}$ as realization of $(Y_1, ..., Y_n) \sim \mathrm{Bin}(n, q)$, which entails concentration of probability around its mean: for $\mathbf{y}$ "typical", $N(1;\mathbf{y}) \approx nq$.

A corresponding "typical stratum" $E_{y_1} \times \cdots E_{y_n}$ that supports $\rho_{y_1} \otimes \cdots \otimes \rho_{y_n}$ has roughly dimension nqd.

Meaning of the chain rule

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

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

Outline

m-Hausdorff measure

It gives a notion of m-dimensional volume:

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

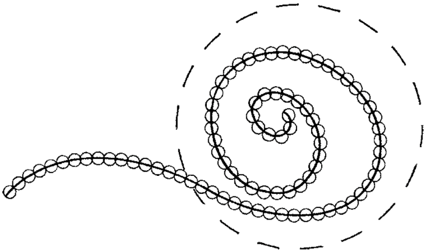

Figure: Case $m = 1$: the sum of diameters of smaller balls give a better approximation of the curve's length.

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.

Rectifiable measures

Let ρ be a locally finite and regular measure and s a nonnegative real number.

Theorem (Marstrand)

If the limiting density $\Theta_s(\rho, x) := \lim_{r \downarrow 0} \rho(B(x, r))/(w_s r^s)$ *exists and is strictly positive and finite for* ρ*-almost every* x*, then* s *is an integer not greater than* n*.*

Later Preiss proved that such a measure is also s-rectifiable in the following sense.

Definition

A measure ν on $\mathbb{R}^d$ is called m-rectifiable if $\nu \ll \mathcal{H}^m$ and there exists a countably $(\mathcal{H}^m, m)$-rectifiable Borel set E such that $\nu(\mathbb{R}^d \setminus E) = 0$.

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

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

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.
$E_Y = \{1, ..., k\}$.

Then:

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

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

Because of the weak law of large numbers: $N(i;\mathbf{y})/n \to q_i$ in probability. So the set $A_\eta^{(n)}$ of $\mathbf{y} \in E_Y^n$ such that $|N(i;\mathbf{y})/n - q_i| < \eta$, for all $i \in E_Y$, is "typical": concentrates almost all the probability when n is big.

$\# A_\eta^{(n)} \approx \exp(nH_\#(q_1, ..., q_k))$, cf. Cisiszár and Körner's strong typicality. Moreover,

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

For $\mathbf{y} \in A_\eta^{(n)}$, 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

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

Back to Renyi

For a stratified measure as above, $\rho = \sum_{i=1}^{k} q_i \rho_i$ with ρ_i being m_i-rectifiable, Renyi's information dimension is $D = \sum_{i=1}^{k} q_i m_i$ (provided compacity; otherwise still open).

However, the D-dimensional entropy differs from $H_\mu(\rho)$.