Notes on the Kolmogorov-Sinai theorem

François Le Maître

May 22, 2017

Abstract

We present an elementary proof of the Kolmogov-Sinai theorem, following the books of Glasner [Gla03] and Downarowicz [Dow11] and the lecture notes of Rohlin [Rok67].

Let (X,μ) be a standard probability space. We will look at everything up to measure zero. In particular we will identity two countable partitions of X into Borel subsets if they coincide up to measure zero.

1 Entropy

Let α be a countable partition of X into Borel subsets. Define the **entropy** of α as

$$H(\alpha) = -\sum_{A\in\alpha} \mu(A)\ln\mu(A).$$

This is the average amount of information on x one gets from knowing in which atom of the partition a μ-random element $x \in X$ is.

The function $\phi : t \mapsto -t\ln t$ is strictly concave: its derivative is $\phi'(t) = -\ln t - 1$ which is strictly decreasing. From this two important properties can be derived.

Proposition 1. *The entropy over partitions into k subsets is at most $\ln k$, and this value is attained when all the atoms have the same measure ($1/k$).*

Proof. The function ϕ is strictly convex, so given $p_1,...,p_k \geqslant 0$, we have

$$\phi\left(\sum_{i=1}^{k}\frac{p_i}{k}\right) \geqslant \sum_{i=1}^{k}\frac{\Phi(p_i)}{k}$$

with equality if an only if the p_i's are all equal. The left hand term is equal to $\frac{\ln n}{n}$ and the right hand term is equal to $-\frac{\sum_{i=1}^n p_i\ln p_i}{n}$ so multiplying both by n we get the desired result. □

Proposition 2. *As a function on the convex set of probability measures over a countable set, entropy is strictly concave.*

Proof. Let $t \in [0,1]$, consider two probability measures μ and ν over a fixed countable set K. Then

$$\begin{aligned} H(t\mu + (1-t)\nu) &= \sum_{i \in K} \phi(t\mu(i) + (1-t)\nu(i)) \\ &\geqslant \sum_{i \in K} t\phi(\mu(i)) + (1-t)\phi(\nu(i)) \\ &\geqslant tH(\mu) + (1-t)H(\nu). \end{aligned}$$

We conclude that the entropy is concave. That it is strictly concave follows the fact that ϕ is strictly concave by the same arguement. $\square$

One of the fundamental properties of entropy is subadditivity: if α and β are two countable partitions, their **join** is the partition $\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}$, and we have

$$H(\alpha \vee \beta) \leqslant H(\alpha) + H(\beta).$$

To prove this property, we will introduce relative entropy $H(\alpha|\beta)$, which is also important in its own right. This measures the amount information one gets from knowing in which atom of the partition a μ-random element $x \in X$ lies when we already know it was lying in some atom B of β:

$$H(\alpha|\beta) = -\sum_{B \in \beta} \sum_{A \in \alpha} \mu(A \cap B) \ln \left(\frac{\mu(A \cap B)}{\mu(B)} \right).$$

Using the fact that $\ln \mu \left(\frac{\mu(A \cap B)}{\mu(B)} \right) = \ln(\mu(A \cap B)) - \ln(\mu(B))$ and $\sum_{A \in \alpha} \mu(A \cap B) = \mu(B)$, we get the formula

$$H(\alpha|\beta) = H(\alpha \vee \beta) - H(\beta)$$

which we rewrite as

$$H(\alpha \vee \beta) = H(\alpha|\beta) + H(\beta). \tag{1}$$

So entropy is an increasing function on partitions:

$$H(\alpha \vee \beta) \geqslant H(\beta) \tag{2}$$

Given a subset $B \subseteq X$ and a partition α, we define $H_B(\alpha)$ as the entropy of the partition induced by α on B, where B is equipped with the *probability* measure μ_B defined by $\mu_B(A) = \frac{\mu(A)}{\mu(B)}$. Now observe that by definition

$$H(\alpha|\beta) = \sum_{B \in \beta} \mu(B) H_B(\alpha). \tag{3}$$

Using the concavity of the entropy as a function of probability measures on finite sets (for each B the partition α defines a probability measure μ_B on α and the convex combination $\sum \mu(B)\mu_B$ is equal to the probability measure induced by μ on α), we conclude that

$$H(\alpha|\beta) \leqslant H(\alpha) \tag{4}$$

From the formula (1), we can now conclude that we have subadditivity as announced:

$$H(\alpha \vee \beta) \leqslant H(\alpha) + H(\beta) \tag{5}$$

Finally, let us observe that we have an equality in (5) if and only if we have equality in (4), which by strict concavity happens only when for all $A \in \alpha$, the quantity $\frac{\mu(A \cap B)}{\mu(B)}$ does not depend on B, i.e. when α and β are *independent.*

2 Entropy of a measure-preserving transformation

Let us recall Fekete's lemma: given a sequence (u_n) of reals, if for every $n, m \in \mathbb{N}$ we have $u_{n+m} \leqslant u_n + u_m$ (i.e. the sequence is *subadditive*) then $\lim_n \frac{u_n}{n}$ exists and is equal to $\inf_n \frac{u_n}{n}$ (see e.g. [Wal82, Thm. 4.9]).

We now have all the tools to define the entropy of a measure preserving transformation T. First, given a countable partition α, its T-entropy is defined by

$$h(\alpha, T) = \inf_n \frac{H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)}{n} = \lim_{n \to +\infty} \frac{H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right)}{n} \tag{6}$$

where the limit exists and the last equality holds by virtue of Fekete's subadditive lemma.

Definition 3. Let T be a measure-preserving transformation. A partition α is **dynamically generating** (with respect to T) if the smallest T-invariant σ-algebra containing α is the Borel σ-algebra of X.

Example 4. Consider $X = K^{\mathbb{Z}}$ where K is a countable set (the base space) equipped with a probability measure ν and the invariant measure is $\mu = \nu^{\otimes \mathbb{Z}}$. The **Bernoulli shift** on X is the transformation T defined by $T((x_n)_{n \in \mathbb{N}}) = (x_{n-1})_{n \in \mathbb{Z}}$. Every $k \in K$ defines a measurable set $A_k := \{(x_n)_{n \in \mathbb{Z}} : x_0 = k\}$. The partition $\alpha = (A_k)_{k \in K}$ is then a dynamically generating partition. Note moreover that

$$h(\alpha, T) = H(\alpha) = -\sum_{k \in K} \nu(k) \ln \nu(k).$$

The k**-shift** is the Bernoulli shift over a set K with k elements where ν is the equidistributed probability measure (each element has measure $1/k$). Before Kolmogorov's work, it was a famous open problem wether the 3-shift could be conjugate to the 2-shift.

We can now state the Kolmogorov-Sinai theorem.

Theorem 5. *Let T be a measure-preserving transformation. Suppose α is a finite generate dynamically generating partition. Then*

$$h(T) = h(\alpha, T).$$

So for a Bernoulli shift, the entropy is equal to the entropy of the base space. In particular, the 2-shift is not conjugate to the 3-shift. The theorem will follow from the following more technical result which we will prove later.

Theorem 6. *Suppose* (α_n) *is an increasing family of partitions (meaning that* α_{n+1} *refines* α_n*)) such that the sigma-algebra generated by* $\bigcup_{n\in\mathbb{N}} \alpha_n$ *is equal to the Borel* σ*-algebra of* X*. Then*

$$h(T) = \lim_{n\to+\infty} h(\alpha_n, T).$$

Proof of Thm. 5. Since α is a generating partition, the sequence of partitions (α_n) defined by

$$\alpha_n := \bigvee_{i=-n}^{n} T^{-i}\alpha$$

satisfies the asumptions of Thm. 6. But for every $n \in \mathbb{N}$ we have

$$h(\alpha_n, T) = \lim_{N\to+\infty} \frac{H\left(\bigvee_{j=0}^{N-1} T^{-j}\alpha_n\right)}{N}$$

Note that $\alpha_n = \bigvee_{i=-n}^{n} T^{-i}\alpha$ so

$$\bigvee_{j=0}^{N-1} T^{-j}\alpha_n = \bigvee_{i=-n}^{N-1+n} T^{-i}\alpha.$$

Now since T^n is measure-preserving transformation we have

$$H\left(\bigvee_{i=-n}^{N-1+n} T^{-i}\alpha\right) = H\left(\bigvee_{i=0}^{N-1+2n} T^{-i}\alpha\right)$$

We conclude that

$$\begin{aligned} h(\alpha_n, T) &= \lim_{N\to+\infty} \frac{H\left(\bigvee_{i=0}^{N-1+2n} T^{-i}\alpha\right)}{N} \\ &= \lim_{N\to+\infty} \frac{N+2n}{N} \frac{H\left(\bigvee_{i=0}^{N-1+2n} T^{-i}\alpha\right)}{N+2n} \\ &= h(\alpha, T) \end{aligned}$$

The conclusion now follows from Thm. 6. □

Note that the proof used the fact that in $\mathbb{Z}$, big invervals like $[0, N]$ are almost invariant under translation, which is one of the characterizations of *amenability.* We now need to prove Thm. 6. An important tool will be the following inequality for arbitrary countable partitions α and β and measure preserving transformation T:

$$h(T, \beta) \leqslant H(\beta|\alpha) + h(T, \alpha) \tag{7}$$

Note that this inequality is true if we replace $h(T, \cdot)$ by $H(\cdot)$: indeed by equation (1) we have

$$H(\beta|\alpha) + H(\alpha) = H(\alpha \vee \beta) = H(\alpha|\beta) + H(\beta) \geqslant H(\beta).$$

Assuming this inequality, we now sketch the proof of Thm. 6.

Sketch of proof of Thm. 6. Let (α_n) be an increasing family of partitions such that the sigma-algebra generated by $\bigcup_{n\in\mathbb{N}} \alpha_n$ is equal to the Borel σ-algebra of X. It is not hard to check that the sequence $(h(\alpha_n, T))_{n\in\mathbb{N}}$ is increasing (use that if β refines α then $H(\alpha) \leqslant H(\beta)$ as per equation (5)). So the limit $\lim_{n\to+\infty} h(\alpha_n, T)$ exists, and by the definition of $h(T)$ it satisfies

$$h(T) \geqslant \lim_{n\to+\infty} h(\alpha_n, T).$$

To prove the reverse inequality, let β be a finite partition and let $n \in \mathbb{N}$. By inequality (7) we have

$$h(\beta) \leqslant H(\beta|\alpha_n) + h(\alpha_n).$$

Now since the σ-algebra generated by α_n is equal to the whole Borel σ-algebra, we should have $H(\beta|\alpha_n) \to 0$ and hence

$$h(\beta) \leqslant \lim_{n\to+\infty} h(\alpha_n).$$

So $h(T) \leqslant \lim_{n\to+\infty} h(\alpha_n)$ as wanted. $\square$

There are two things we need to justify in order to make the above argument valid:

- the inequality (7) and
- the fact that $H(\beta|\alpha_n) \to 0$ (see Prop. 10.

In order to do this, we will first generalize (in)equalities (1), (4) and (5) to the relative setting in the next section so as to obtain inequality (7), and in section 4 we will prove that $H(\beta|\alpha_n) \to 0$.

3 More on relative entropy

Let us first do the relative version of the computations which led us to formula (1). We have

$$\begin{aligned}
H(\alpha \vee \beta|\gamma) &= - \sum_{(A,B,C)\in\alpha\times\beta\times\gamma} - \ln \mu(A\cap B\cap C) \ln\left(\frac{\mu(A\cap B\cap C)}{\mu(C)}\right) \\
&= - \sum_{(A,B,C)\in\alpha\times\beta\times\gamma} \mu(A\cap B\cap C) \ln\left(\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}\frac{\mu(B\cap C)}{\mu(C)}\right) \\
&= - \sum_{(A,B,C)\in\alpha\times\beta\times\gamma} \mu(A\cap B\cap C) \ln\left(\frac{\mu(A\cap B\cap C)}{\mu(B\cap C)}\right) \\
&\quad - \sum_{(A,B,C)\in\beta\times\gamma} \mu(A\cap B\cap C) \ln\left(\frac{\mu(B\cap C)}{\mu(C)}\right).
\end{aligned}$$

Since $\sum_{A\in\alpha} \mu(A\cap B\cap C) = \mu(B\cap C)$ we obtain the relative version of (1)

$$H(\alpha \vee \beta|\gamma) = H(\alpha|\beta\vee\gamma) + H(\beta|\gamma) \tag{8}$$

In particular, relative entropy is an increasing function over partitions:

$$H(\alpha \vee \beta|\gamma) \geqslant H(\beta|\gamma) \tag{9}$$

We now define a conditional version of H_C for $C \subseteq X$ of positive measure: we let $H_C(\alpha|\beta)$ be the relative entropy of the partitions induced by α and β on C equipped with the probability measure μ_C.

Observe that we have the following immediate consequence of (3):

$$H(\alpha \vee \beta|\gamma) = \sum_{C\in\gamma} \mu(C) H_C(\alpha \vee \beta). \tag{10}$$

Then by formula (1) we have

$$H(\alpha \vee \beta|\gamma) = \sum_{C\in\gamma} \mu(C) H_C(\alpha|\beta) + \sum_{C\in\gamma} \mu(C) H_C(\beta)$$
$$H(\alpha \vee \beta|\gamma) = \sum_{C\in\gamma} \mu(C) H_C(\alpha|\beta) + H(\beta|\gamma)$$

So using 8 we can identify

$$H(\alpha|\beta \vee \gamma) = \sum_{C\in\gamma} \mu(C) H_C(\alpha|\beta) \tag{11}$$

By concavity of entropy (Prop. 2) we obtain the relative analogue of (4):

$$H(\alpha|\beta \vee \gamma) \leqslant H(\alpha|\beta) \tag{12}$$

Finally, using equation (10) and subadditivity of entropy (inequation (5)), we obtain the subadditivity of relative entropy :

$$H(\alpha \vee \beta|\gamma) \leqslant H(\alpha|\gamma) + H(\beta|\gamma). \tag{13}$$

Proposition 7. *Inequality (7) holds: for every measure-preserving transformation T and every countable partitions α and β we have*

$$h(T,\beta) \leqslant H(\beta|\alpha) + h(T,\alpha)$$

Proof. Let $n \in \mathbb{N}$, then by inequality (2) we have

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) \leqslant H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta \vee \bigvee_{i=0}^{n-1} T^{-i}\alpha\right)$$

so by equation (1)

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) \leqslant H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta \,\middle|\, \bigvee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right). \tag{14}$$

Now by subadditivity of relative entropy (inequality (13)) we have

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta \,\middle|\, \bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leqslant \sum_{i=0}^{n-1} H\left(T^{-i}\beta \,\middle|\, \bigvee_{j=0}^{n-1} T^{-j}\alpha\right).$$

For every $i \in \{0, ..., n-1\}$, inequality (12) yields

$$H\left(T^{-i}\beta \middle| \bigvee_{i=0}^{n-1} T^{-i}\alpha\right) \leqslant H(T^{-i}\beta|T^{-i}\alpha) = H(\beta|\alpha).$$

We can thus deduce from inequality (14) that

$$H\left(\bigvee_{i=0}^{n-1} T^{-i}\beta\right) \leqslant nH(\beta|\alpha) + H\left(\bigvee_{i=0}^{n-1} T^{-i}\alpha\right).$$

Dividing all by n and letting n tend to $+\infty$ we obtain the desired inequality (7). $\square$

4 Using the topology on the space of partitions

We will now justify that if α_n is an increasing sequence of finite partitions whose union generates the whole σ-algebra of X, then for every finite partition β we have $H(\beta|\alpha_n) \to 0$.

We fix once and for all $k \in \mathbb{N}$ and study the set $\mathcal{P}_k$ of partitions of X into k (measurable) subsets. We view such partitions as *sets* of subsets rather than tuples. Given such a partition α let $\mathcal{E}(\alpha)$ be the set of enumerations of α, i.e. bijective maps

$$f : \{0, ..., k-1\} \to \alpha.$$

We then have a natural metric d_μ on the set $\mathcal{P}_k$ of such partitions:

$$d_\mu(\alpha, \beta) = \min_{(f,g)\in\mathcal{E}(\alpha)\times\mathcal{E}(\beta)} \sum_{i=0}^{k-1} \mu(f(i) \triangle g(i)).$$

Given an arbitrary partition α, we let $\langle\alpha\rangle$ be the set of all subsets obtained as reunions of elements of α. We let $\mathcal{P}_k(\alpha)$ be the set of all partitions of X into k subsets belonging to $\langle\alpha\rangle$.

Lemma 8. *Suppose (α_n) is an increasing sequence of finite partitions whose union generates the whole σ-algebra of X, then $\bigcup_{n\in\mathbb{N}} \mathcal{P}_k(\alpha_n)$ is dense in $\mathcal{P}_k$ for the metric d_μ.*

Proof. We leave it to the reader to check that it suffices to show that for every $\epsilon > 0$ and every measurable $B \subseteq X$ there is $A \in \bigcup_{n\in\mathbb{N}}\langle\alpha_n\rangle$ such that $\mu(A \triangle B) < \epsilon$. One then just needs to check that the set $\mathcal{F}$ of measurable $B \subseteq X$ such that there is $A \in \bigcup_{n\in\mathbb{N}} \langle\alpha_n\rangle$ such that $\mu(A \triangle B) < \epsilon$ is a σ-algebra (since it obviously contains $\bigcup_{n\in\mathbb{N}} \alpha_n$ the conclusion will follow).

Observe that $\mathcal{F}$ is stable under complements for $\mu(X \setminus A \triangle X \setminus B) = \mu(A \triangle B)$. Now let (B_n) be a sequence of elements of $\mathcal{F}$, we want to approximate $B := \bigcup_n B_n$ by some $A \in \bigcup_{n\in\mathbb{N}} \langle\alpha_n\rangle$. We first approximate each B_n by some $A_n \in \bigcup_{n\in\mathbb{N}} \langle\alpha_n\rangle$ up to an $\epsilon/2^n$ error, and observe that if $A = \bigcup_n A_n$ then we have $\mu(A \triangle B) \leqslant \epsilon$ and

$$\lim_{n\to+\infty} \mu(A \setminus \bigcup_{i=0}^{n} A_n) = 0$$

so that B is 2ϵ-approximated by some $A' \in \bigcup_n \langle\alpha_n\rangle$ as wanted. $\square$

We now define another metric d_H on the space of partitions into k subsets by

$$d_H(\alpha, \beta) = \max(H(\beta|\alpha), H(\alpha|\beta)).$$

Observe that $H(\beta|\alpha) = 0$ if and only if for all $A \in \alpha$ and all $B \in \beta$, we have $\frac{\mu(A\cap B)}{\mu(A)} \in \{0, 1\}$. This means that α refines β, in particular $d_H(\alpha, \beta) = 0$ if and only if $\alpha = \beta$.

Symmetry of d_H is clear, and finally the triangle inequality follows from

$$H(\gamma|\alpha) \leqslant H(\gamma \vee \beta|\alpha) = H(\gamma|\beta \vee \alpha) + H(\beta|\alpha) \leqslant H(\gamma|\beta) + H(\beta|\alpha)$$

so d_H is indeed a metric.

The metric d_H is actually equivalent to d_μ, but we will only show that it is coarser than d_μ since that is all we need. For more properties of this metric including its equivalence to d_μ, see [Dow11, Sec. 1.7].

Proposition 9. *The topology defined by d_H is coarser than the one defined by d_μ.*

Proof. We have to show that the map id : $(\mathcal{P}_k, d_\mu) \to (\mathcal{P}_k, d_H)$ is continuous. Let $\alpha \in \mathcal{P}_k$ and $\delta > 0$. Let $\beta \in \mathcal{P}_k$, suppose $d_\mu(\alpha, \beta) < \epsilon$ for some $\epsilon > 0$ to be defined later. Enumerate $\alpha = \{A_i : i = 1, ..., k\}$ and $\beta = \{B_i : i = 1, ..., k\}$ so that

$$d_\mu(\alpha, \beta) = \sum_{i=1}^{k} \mu(A_i \bigtriangleup B_i) < \epsilon$$

Let $C = \bigcup_{i=1}^{k} A_i \bigtriangleup B_i$. Consider the partition $\gamma = \{C, X \setminus C\}$. We then have by inequality (9)

$$H(\alpha|\beta) \leqslant H(\alpha \vee \gamma|\beta)$$

So by equality (8) we deduce

$$H(\alpha|\beta) \leqslant H(\alpha|\gamma \vee \beta) + H(\gamma|\beta).$$

Now $H(\gamma|\beta) \leqslant H(\gamma) = H(\epsilon, 1 - \epsilon)$. On the other hand by formula (11) we have

$$H(\alpha|\gamma \vee \beta) = \mu(C)H_C(\alpha|\beta) + (1 - \mu(C))H_{X\setminus C}(\alpha|\beta).$$

Note that $H_{X\setminus C}(\alpha|\beta) = 0$ by the definition of C. Moreover $H_C(\alpha|\beta) \leqslant H_C(\alpha) \leqslant \ln k$ by Prop. 1. We thus have the inequality

$$H(\alpha|\beta) \leqslant \epsilon \ln(k) + H(\epsilon, 1 - \epsilon).$$

By symetry we also have $H(\beta|\alpha) \leqslant \epsilon \ln(k) + H(\epsilon, 1 - \epsilon)$, and it follows that if ϵ was chosen small enough, we have $d_H(\alpha, \beta) < \delta$ as wanted. □

We can finally prove the final piece of the proof of Thm. 6, and hence of the Kolmogorov-Sinai theorem.

Proposition 10. *Suppose (α_n) is an increasing sequence of finite partitions whose union generates the whole σ-algebra of X. Then for every finite partition β we have $H(\beta|\alpha_n) \to 0$.*

Proof. Let k be the cardinality of β. By Lemma 8 for every $n \in \mathbb{N}$ we find $\beta_n \in \mathcal{P}_k(\alpha_n)$ such that $d_\mu(\beta_n, \beta) \to 0$. By Proposition 9, we obtain that $d_H(\beta_n, \beta) \to 0$. Now for all $n \in \mathbb{N}$ we now have

$$\begin{aligned} H(\beta|\alpha_n) &\leqslant H(\beta \vee \beta_n|\alpha_n) \\ &= H(\beta|\beta_n \vee \alpha_n) + H(\beta_n|\alpha_n) \\ H(\beta|\alpha_n) &\leqslant H(\beta|\beta_n) + H(\beta_n|\alpha_n) \end{aligned}$$

But by the definition of d_H we have $H(\beta|\beta_n) \to 0$, and since α_n refines β_n we have $H(\beta_n|\alpha_n) = 0$. So $H(\beta|\alpha_n) \to 0$ as wanted. $\square$

Remark 11. One could of course give a direct proof without mentioning the metric d_H, but we felt it would be more transparent this way. Moreover, the metric d_H is important in its own right since it is a complete separable metric on the space of all countable partitions of finite entropy (see [Dow11, Fact 1.7.15]). We should also mention that there is a notion of entropy relative to a σ-algebra and that Prop. 10 can then be derived from a much more general statement (see [Gla03, Thm. 14.28]).

References

[Dow11] Tomasz Downarowicz. *Entropy in Dynamical Systems*, volume 18 of *New Mathematical Monographs*. Cambridge University Press, Cambridge, 2011.

[Gla03] Eli Glasner. *Ergodic Theory via Joinings*, volume 101 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, Rhode Island, February 2003.

[Rok67] V. A. Rokhlin. Lectures on the entropy theory of measure-preserving transformations. *Russian Mathematical Surveys*, 22(5):1, 1967.

[Wal82] Peter Walters. *An Introduction to Ergodic Theory*. Number 79 in Graduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg-Berlin, 1982.