The Shannon-McMillan-Breiman Theorem

The Shannon-McMillan-Breiman Theorem uses entropy to measure how large sets in the n-th joint $\mathcal{P}_n$ are. Typically, they decrease exponentially and the exponential rate is exactly the measure-theoretical entropy.

Shannon-McMillan-Breiman Theorem: Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving transformation and $\mathcal{P}$ a (countable or finite) partition with $H(\mathcal{P}) < \infty$ Let $\mathcal{P}_n = \bigvee_{k=0}^{n-1} T^{-k}(\mathcal{P})$ and $\mathcal{P}_n(x)$ the element of $\mathcal{P}_n$ containing x. Then

$$-\lim_{n\to\infty} \frac{1}{n} \log \mu(\mathcal{P}_n(x)) = h(\mathcal{P}, T) \quad \mu\text{-a.e.}$$

Apart from proving this theorem, we will discuss an application called Lochs' Theorem, on the number of known digits of expansions of reals.

The Shannon-McMillan-Breiman Theorem

Define the information function

$$I_{\mathcal{P}}(x) := -\log \mu(\mathcal{P}(x)) = -\sum_{P \in \mathcal{P}} 1_P(x) \log \mu(P),$$

with respect to which we have $H(\mathcal{P}) = \mathbb{E}(I_{\mathcal{P}})$. Inserting this in the definition of the entropy, we obtain

$$h(\mathcal{P}, T) = \lim_{n \to \infty} \frac{1}{n} H(\mathcal{P}_n) = \lim_{n \to \infty} \int_X \frac{1}{n} I_{\mathcal{P}_n}(x) \, d\mu.$$

The Shannon-McMillan-Breiman Theorem says that in fact the integrand converges to $h(\mathcal{P}, T)$ μ-a.e.

The Shannon-McMillan-Breiman Theorem

The proof requires some more technical tools: conditional expectation, conditional entropy and the Martingale theorem.

For a measure preserving system $(X, \mathcal{B}, \mu, T)$, some measurable function $f : X \to \mathbb{R}$ and σ-algebra $\mathcal{C}$ (possibly $\mathcal{C} = \mathcal{B}$, possibly $\mathcal{C}$ coarser than $\mathcal{B}$), we can define the conditional expectation $\mathbb{E}_\mu(f|\mathcal{C})$ as the unique $\mathcal{C}$-measurable function $\bar{f}$ such that

$$\int_C \bar{f}\, d\mu = \int_C f\, d\mu \text{ for all } C \in \mathcal{C}.$$

- Recall that $\mathcal{C}$-measurable means that $\bar{f}^{-1}([t, \infty)) \in \mathcal{C}$ for all $t \in \mathbb{R}$, and therefore $\bar{f}$ must be constant on all atoms of $\mathcal{C}$.
- Note that conditional expectation is a function, and (unlike expectation or conditional probability) not a number. It is the function $\bar{f}$ such that for each atom C,

$$\bar{f}(x) = \frac{1}{\mu(C)} \int_C f\, d\mu \quad \text{for } \mu\text{-a.e. } x \in C.$$

The Shannon-McMillan-Breiman Theorem

The finer the σ-algebra $\mathcal{C}$, the more $\bar{f}$ looks like f. This is expressed in the following version of the

Theorem (Martingale Convergence Theorem)

If $(\mathcal{C}_n)_n$ is a sequence of σ-algebras such that $\mathcal{C}_{n+1}$ refines $\mathcal{C}_n$ and $\mathcal{C} = \lim_{n\to\infty} \mathcal{C}_n := \bigvee_{n=1}^{\infty} \mathcal{C}_n$, then for every $f \in L^1(\mu)$

$$\mathbb{E}_\mu(f|\mathcal{C}_n) \to \mathbb{E}_\mu(f|\mathcal{C}) \quad \mu\text{-a.e. as } n \to \infty.$$

We skip the proof.

Conditional Entropy

Definition: Motivated by conditional measure $\mu(P|Q) = \frac{\mu(P \cap Q)}{\mu(Q)}$, we define conditional entropy of a measure μ as

$$H_\mu(\mathcal{P}|\mathcal{Q}) = -\sum_{Q_j \in \mathcal{Q}} \mu(Q_j) \sum_{P_i \in \mathcal{P}} \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)} \log \frac{\mu(P_i \cap Q_j)}{\mu(Q_j)}. \qquad (1)$$

Before trying to interpret this notion, let us first list some properties that follow directly from the definition and Jensen's inequality:

Proposition: Given measures μ, μ_i and two partitions $\mathcal{P}$ and $\mathcal{Q}$,

1. $H_\mu(\mathcal{P} \vee \mathcal{Q}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$;
2. $H_\mu(\mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q} \mid \mathcal{P})$, and hence $h_\mu(T, \mathcal{Q}) = h_\mu(T, \mathcal{P}) + H_\mu(\mathcal{Q} \mid \mathcal{P})$.
3. $\sum_{i=1}^n p_i H_{\mu_i}(\mathcal{P}) \leq H_{\sum_{i=1}^n p_i \mu_i}(\mathcal{P})$ for each probability vector $(p_1, \dots, p_n)$.

Conditional Information Function

Similarly to conditional entropy, we define the conditional information function

$$I_{\mathcal{P}|\mathcal{Q}}(x) := -\sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} 1_{P\cap Q}(x)\log\frac{\mu(P\cap Q)}{\mu(Q)}.$$

Comparing this to the definition of conditional entropy, we get

$$\int_X I_{\mathcal{P}|\mathcal{Q}}\,d\mu = -\sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}}\mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu(Q)} = H_\mu(\mathcal{P}|\mathcal{Q}). \tag{2}$$

One can check (using the previous proposition and the definition) that

$$I_{\mathcal{P}\vee\mathcal{Q}} = I_{\mathcal{P}} + I_{\mathcal{Q}|\mathcal{P}}. \tag{3}$$

Conditional Information Function

By the definition of conditional expectation and because $1_P 1_Q = 1_{P\cap Q}$ we have

$$\begin{aligned}
-\log \mathbb{E}_\mu(1_{\mathcal{P}(x)}|\mathcal{Q}) &= -\log \mathbb{E}_\mu\Big(\sum_{P\in\mathcal{P}} 1_P|\mathcal{Q}\Big) \\
&= -\log \sum_{Q\in\mathcal{Q}} \frac{1}{\mu(Q)} \int_Q \sum_{P\in\mathcal{P}} 1_P \, d\mu \\
&= -\log \sum_{P\in\mathcal{P}} \sum_{Q\in\mathcal{Q}} 1_{P\cap Q} \frac{\mu(P\cap Q)}{\mu(Q)} \int_Q 1_P \, d\mu \\
&= I_{\mathcal{P}|\mathcal{Q}}(x).
\end{aligned}$$

Proof of the Shannon-McMillan-Breiman Theorem

We are now ready to do the proof of the Shannon-Breiman-McMillan Theorem.

Proof: Write $g_k(x) = I_{\mathcal{P}|\bigvee_{j=1}^{k-1} T^{-j}\mathcal{P}}(x)$ for $k \geq 2$ and $g_1(x) = I_{\mathcal{P}}$.
Then by (3)

$$
\begin{aligned}
I_{\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) &= I_{\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}}(x) + I_{\mathcal{P}|\bigvee_{j=1}^{n-1} T^{-j}\mathcal{P}}(x) \\
&= I_{\bigvee_{j=0}^{n-2} T^{-j}\mathcal{P}}(Tx) + g_n(x) \\
&= I_{\bigvee_{j=1}^{n-2} T^{-j}\mathcal{P}}(Tx) + I_{\mathcal{P}|\bigvee_{j=1}^{n-2} T^{-j}\mathcal{P}}(Tx) + g_n(x) \\
&= I_{\bigvee_{j=0}^{n-3} T^{-j}\mathcal{P}}(T^2x) + g_{n-1}(Tx) + g_n(x) \\
&\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
&= g_1(T^{n-1}(x)) + \cdots + g_{n-1}(T(x)) + g_n(x) \\
&= \sum_{j=0}^{n-1} g_{n-j}(T^j x).
\end{aligned}
$$

Proof of the Shannon-McMillan-Breiman Theorem

Let $g = \lim_{n\to\infty} g_n$, which exists μ-a.e. and belongs to $L^1(\mu)$ because of the Martingale Convergence Theorem. We write the previous equality as

$$\frac{1}{n} I_{\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) = \frac{1}{n}\sum_{j=0}^{n-1} g(T^j x) + \frac{1}{n}\sum_{j=0}^{n-1}(g_{n-j} - g)(T^j x).$$

Since μ is ergodic, the first sum converges μ-a.e. to $\int_X g\, d\mu$, which is equal to $H_\mu(\mathcal{P}|\bigvee_{j=1}^{\infty} T^{-j}\mathcal{P})$ by (2), which in turn is equal to $h(\mathcal{P}, T)$.

Proof of the Shannon-McMillan-Breiman Theorem

For the second sum, we define

$$G_N = \sup_{k \geq N} |g_k - g| \quad \text{and} \quad g^* = \sup_{n \geq 1} g_n.$$

Then $0 \leq G_N \leq g + g^*$ and $g + g^* \in L^1(\mu)$; this is because $\int g_n \, d\mu = H_\mu(\mathcal{P} | \bigvee_{j=1}^{n-1} \mathcal{P})$ is decreasing in n. Moreover, $G_N \to 0$ μ-a.e., so by the Dominated Convergence Theorem,

$$\lim_{N \to \infty} \int_X G_N \, d\mu = \int_X \lim_{N \to \infty} G_N \, d\mu = 0$$

Proof of the Shannon-McMillan-Breiman Theorem

Now for any $N \geq 1$ and $n \geq N$ we split the second sum:

$$
\begin{aligned}
&\frac{1}{n}\sum_{j=0}^{n-1}(g_{n-j} - g)(T^j x) \\
&\quad = \frac{1}{n}\sum_{j=0}^{n-N-1}(g_{n-j} - g)(T^j x) + \frac{1}{n}\sum_{j=n-N}^{n-1}(g_{n-j} - g)(T^j x) \\
&\quad \leq \frac{1}{n}\sum_{j=0}^{n-N-1} G_N(T^j x) + \frac{1}{n}\sum_{j=n-N}^{n-1}(g_{n-j} - g)(T^j x).
\end{aligned}
$$

First take the limit $n \to \infty$. The the second sum tends to zero, and by the Ergodic Theorem, the first sum tends to $\int_X G_N \, d\mu$. Finally, taking $N \to \infty$, also $\int_X G_N \, d\mu \to 0$. Hence $I_{\bigvee_{j=0}^{n-1} T^{-j}\mathcal{P}}(x) \to h(\mathcal{P}, T)$ μ-a.e., as required. This finishes the proof.

Lochs' Theorem

Lochs' Theorem: For Lebesgue-a.e. $x \in (0,1)$, the number $c(d)$ of terms of the continued fraction expansion of x that are required to determine the first d decimal places satisfies

$$\lim_{d \to \infty} \frac{c(d)}{d} = \frac{6 \log 2 \log 10}{\pi^2} \approx 0.97027014.$$

The proof relies on the fact that the terms a_n, are obtained as symbolic itineraries of a particular dynamical system, namely the Gauß map $G(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ for continued fractions and the map $T : x \mapsto 10x \bmod 1$ for decimal expansion.

We can do this for other expansions too. For example, if $b(d)$ is the number of binary digits necessary to determine the d-th decimal, then

$$\lim_{d \to \infty} \frac{b(d)}{d} = \frac{\log 10}{\log 2} \approx 3.32189.$$

Proof of Lochs' Theorem

Proof: That $c = c(d)$ digits of the continued fraction determine d decimal digital means that the c-cylinder $\tilde{Z}_c(x)$ of the Gauss map G is contained in the d-cylinder $Z_d(x)$ of the (Lebesgue measure preserving) map $T : x \mapsto 10x \bmod 1$, but not in the $d+1$-cylinder. Since the invariant measure μ of the Gauß map has density $\frac{d\mu(x)}{dx} = \frac{1}{\log 2}\frac{1}{1+x}$, we find

$$\frac{\log 2}{10}\,\mathrm{Leb}(Z_d) \le \mu(\tilde{Z}_c(x)) \le 2\log 2\,\mathrm{Leb}(Z_d(x)). \tag{4}$$

The Shannon-McMillan-Breiman Theorem gives

$$\begin{aligned}
\frac{h_{\mathrm{Leb}}(T)}{h_\mu(G)} &= \lim_{d\to\infty} \frac{-\log \mathrm{Leb}(Z_d(x))}{d} \frac{c(d)}{-\log \mu(\tilde{Z}_c(x))} \\
&= \lim_{d\to\infty} \frac{c(d)}{d} \lim_{d\to\infty} \frac{\log \mathrm{Leb}(Z_d(x))}{\log \mu(\tilde{Z}_c(x))} \qquad \mathrm{Leb}\text{-a.e.}
\end{aligned}$$

Proof of Lochs' Theorem

Combining this with (4), we obtain

$$\begin{aligned}\frac{h_{\mathrm{Leb}}(T)}{h_\mu(G)} &\le \lim_{d\to\infty}\frac{c(d)}{d}\,\frac{d\log 10}{d\log 10-\log(2\log 2)}\\ &\le \lim_{d\to\infty}\frac{c(d)}{d}\left(1+\frac{\log(2\log 2)}{d\log 10-\log(2\log 2)}\right).\end{aligned}$$

By the same token

$$\frac{h_{\mathrm{Leb}}(T)}{h_\mu(G)} \ge \lim_{d\to\infty}\frac{c(d)}{d}\left(1-\frac{\log\log 2}{d\log 10-\log(\frac{\log 2}{10})}\right).$$

Proof of Lochs' Theorem

Hence the limit

$$\lim_{d\to\infty} \frac{c(d)}{d} = \frac{h_{\mathrm{Leb}}(T)}{h_\mu(G)} \quad \mathrm{Leb}-\text{a.e.}$$

The entropy $h_{\mathrm{Leb}}(T) = \log 10$, because the map $([0,1], \mathrm{Leb}, T)$ is isomorphic to the $(\frac{1}{10}, \dots, \frac{1}{10})$-Bernoulli shift.

The entropy $h_\mu(G) = \frac{\pi^2}{6 \log 2}$ is trickier to prove, but it can be done as follows. The Rokhlin formula says that for absolutely continuous measures

$$h_\mu(T) = \int_X \log |T'| \, d\mu.$$

Recalling that $\frac{d\mu}{dx} = \frac{1}{\log 2} \frac{1}{1+x}$, we get

$$h_\mu(G) = \frac{2}{\log 2} \int_0^1 \frac{\log 1/x}{1+x} \, dx.$$

Proof of Lochs' Theorem

Use $\frac{1}{1+x} = \sum_{k=0}^{\infty}(-x)^k$ and integration by parts:

$$
\begin{aligned}
\int_0^1 \frac{\log x}{1+x}\,dx &= \sum_{k=0}^{\infty}\int_0^1 (-x)^k \log x\,dx \\
&= \sum_{k=0}^{\infty}\left[-\frac{(-x)^{k+1}}{k+1}\log x\right]_0^1 + \int_0^1 \frac{(-x)^k}{k+1}\,dx \\
&= \sum_{k=0}^{\infty}\frac{(-1)^{k+1}}{(k+1)^2} \\
&= \sum_{k=1}^{\infty}\frac{1}{(2k)^2} - \sum_{k=1}^{\infty}\frac{1}{(2k-1)^2} \\
&= 2\sum_{k=1}^{\infty}\frac{1}{(2k)^2} - \sum_{k=1}^{\infty}\frac{1}{k^2} = -\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k^2} = -\frac{\pi^2}{12}.
\end{aligned}
$$

Lochs' Theorem

Inserting $\int_0^1 \frac{\log x}{1+x}\, dx = -\frac{\pi^2}{12}$ in

$$h_\mu(G) = -\frac{2}{\log 2}\int_0^1 \frac{\log x}{1+x}\, dx.$$

we arrive at $h_\mu(G) = \frac{\pi^2}{6\log 2}$. This concludes the proof.

This number $\frac{\pi^2}{6\log 2}$ is sometimes called Khinchin-Lévy's constant. The original proof by Paul Lévy from 1936 which doesn't use Rokhlin's formula, was adjusted by Khinchin.