

THE SHANNON-MCMILLAN-BREIMAN THEOREM

BY THOMAS HUGHES FOR MATH 607D, APRIL 2015

(X, \mathcal{A}, μ) a probability space, $T: X \rightarrow X$ a MP transformation, α a finite measurable partition of X , $\alpha = \{\alpha_1, \dots, \alpha_m\}$.

You will recall the information function of α : $I_\alpha: X \rightarrow \mathbb{R}_+$,

$$I_\alpha(x) := - \sum_{A \in \alpha} \mathbf{1}_A(x) \log \mu(A) = \sum_{i=1}^m \mathbf{1}_{\alpha_i}(x) \log \mu(\alpha_i)$$

i.e. $I_\alpha(x) = -\log \mu(\alpha_i)$ if $x \in \alpha_i$.

The information function clearly satisfies:

$$E(I_\alpha) = \sum_{i=1}^m \mu(\alpha_i) \log \mu(\alpha_i) = H(\alpha), \text{ the entropy of } \alpha.$$

If entropy is the "measure of how much a given partition "surprises you" under a measure, the information function tells you how surprising a given outcome in X is under the partition and measure.

If $\mathcal{B} \subset \mathcal{A}$ a sub- σ -algebra, define conditional information

$$I_{\alpha|\mathcal{B}} = - \sum_{i=1}^m \mathbf{1}_{\alpha_i}(x) \log \mu(\alpha_i | \mathcal{B}).$$

Recall: if $\beta \subset \alpha$ a partition and $\sigma(\beta) = \mathcal{B}$, then

$$H(\alpha | \mathcal{B}) = H(\alpha | \beta) = E(I_{\alpha|\mathcal{B}}).$$

Properties for $I_{\alpha} \sim \text{Properties for } H(\alpha)$

Most of our usual properties of entropy can be derived from similar properties for information simply by taking expectation.

$$\text{e.g. (1)} \quad I_{\alpha \vee \beta} = I_{\alpha} + I_{\beta | \alpha} \quad \sim = I_{\beta | \alpha(\alpha)}$$

$$\sim \quad H(\alpha \vee \beta) = H(\alpha) + H(\beta | \alpha)$$

(Proof only in presence of STRONG EXTERNAL PRESSURE $\neq 0$)

An important result we briefly saw regarding continuity of conditional information:

Theorem: α a finite partition. $F_1 \subset F_2 \subset F_3 \subset \dots$ an increasing sequence of sub- σ -algebras and $\sigma(\bigcup_{n=1}^{\infty} F_n) =: F$.

$$\text{Then } I_{\alpha | F_n} \rightarrow I_{\alpha | F} \text{ a.s. and in } L'$$

The a.s. convergence is from the martingale convergence theorem.

For L' convergence we require:

$$(\forall A_i, \alpha(A_i | F_n) \rightarrow \alpha(A | F) \text{ a.s.})$$

Lemma: $\sup_{n \geq 1} I_{\alpha | F_n} \in L'$

Let $f^* = \sup_{n \geq 1} I_{\alpha | F_n}$. Let $\epsilon > 0$ and $A \in \alpha$.

$$\text{Defining } C_{\alpha, A}(\epsilon) := \int_{F_n} P(A | F_n) dP$$

Let $f_{n,A} := -\log m(A/F_n)$ and define

$$C_{n,A}(\epsilon) := \{x \in X : f_{1,A}(x) \leq \epsilon, \dots, f_{n-1,A}(x) \leq \epsilon, f_{n,A}(x) > \epsilon\}$$

Then $C_{n,A}(\epsilon) \cap C_{m,A}(\epsilon) = \emptyset$, and since $C_{n,A}(\epsilon) \in F_n$, by the definition of conditional expectation

$$m(A \cap C_{n,A}(\epsilon)) = \int_{C_{n,A}(\epsilon)} \chi_A dm = \int_{C_{n,A}(\epsilon)} \chi_{A/F_n} dm$$

$$= \int_{C_{n,A}(\epsilon)} \mu(A/F_n) dm$$

$$= \int_{C_{n,A}(\epsilon)} e^{-f_{n,A}} dm \leq e^{-\epsilon} m(C_{n,A}(\epsilon))$$

by def. of $C_{n,A}(\epsilon)$.

$$\text{Hence } m(A \cap \{f^* > \epsilon\}) = \sum_{n=1}^{\infty} m(A \cap C_{n,A}(\epsilon)) \leq e^{-\epsilon} \cdot \underbrace{\sum_{n=1}^{\infty} m(C_{n,A}(\epsilon))}_{< 1 \text{ since they're disjoint}} \leq e^{-\epsilon}.$$

$$\text{Hence } \int f^* dm = \int_0^\infty m(A \cap \{f^* > \epsilon\}) d\epsilon < 1 \text{ since they're disjoint}$$

$$= \sum_{\alpha \in A} \int_0^\infty m(A \cap \{f^* > \epsilon\}) d\epsilon$$

$$\leq |\alpha| \int_0^\infty e^{-\epsilon} d\epsilon < \infty$$

(If you are more careful, show $\int f^* dm \leq H(\alpha) + 1$) \square

Notation:

- $D_n = [0, n-1]^d$
- $\alpha^{\bar{d}} : \bigvee_{g \in G} T^{-g} \alpha$

- For $d=1$, $\alpha_m^n = \alpha^{\{m, m+1, \dots, n-1\}} = \bigvee_{k=m}^{n-1} T^{-k} \alpha$

- When $\mathbb{Z}^d \rightarrow$ ordered lexicographically,
 $\bar{\alpha} = \{\bar{i} \in \mathbb{Z}^d : i < \bar{o}\}$

- $\mathcal{I}_n(\tau) =$ invert. σ -algebra of (α, τ)

Theorem (Shannon-McMillan-Breiman): (X, \mathcal{A}, μ) probability space
 T a MP \mathbb{Z}^d -action, α a finite partition of X (measurable).
 Then

$$\lim_{n \rightarrow \infty} \frac{1}{n^d} I_{\alpha^D_n} = E[I_{\alpha | \alpha^{P^-}} | \mathcal{I}_n(\tau)]$$

a.s. and in L'

"Corollary 1": (Part Thm)

$$h(\alpha, T) = H(\alpha | \alpha^{P^-})$$

"Corollary 2": If T is ergodic, then

$$\frac{1}{n^d} I_{\alpha^D_n} \xrightarrow{\text{a.s. and in } L'} h(\alpha, T)$$

(Please have for the how cool is that effect.)

Remark: Proofs for $d=1$ and $d > 1$ are very different.

The arguments for $d > 1$ are much more technical.

(...)

PROOF ($d=1$): first note, in our notation

$$\alpha^{D_n} = \alpha_0^n \quad \text{and} \quad \alpha^P = \alpha_1^\infty \quad (\text{Show } \frac{1}{n} I_{\alpha_0^n} \rightarrow I_{\alpha_1^\infty})$$

By (1) (the first property of I_{α_1} we have),

$$I_{\alpha_0^n} = I_{\alpha_1} \circ \alpha_0^n = I_{\alpha_1} \circ T^{-1} \alpha_0^{n-1}$$

$$= I_{T^{-1} \alpha_0^{n-1}} + I_{\alpha_1 | \alpha_0^n}$$

$$= I_{\alpha_0^{n-1}} \circ T + I_{\alpha_1 | \alpha_0^n}$$

$$\begin{aligned} \text{since } 1_A \circ T(x) &= 1 \iff T^S(x) \in A \\ &\iff x \in T^{-S}(A) \end{aligned}$$

Applying this recursively,

$$I_{\alpha_0^n} = I_{\alpha_1 | \alpha_0^n} + I_{\alpha_1 | \alpha_0^{n-1}} \circ T + \dots + I_{\alpha_1 | \alpha_0^1} \circ T^{n-1} + I_{\alpha_1 | \alpha_0^0} \circ T^{n-1}$$

Now add and subtract what you want:

$$\frac{1}{n} I_{\alpha_0^n} = \sum_{k=0}^{n-1} I_{\alpha_1 | \alpha_0^{n-k}} \circ T^k$$

$$= \frac{1}{n} \sum_{k=0}^{n-1} I_{\alpha_1 | \alpha_0^\infty} \circ T^k + \frac{1}{n} \sum_{k=0}^{n-1} (I_{\alpha_1 | \alpha_0^{n-k}} - I_{\alpha_1 | \alpha_0^\infty}) \circ T^k$$

$$= B_n + S_n$$

Claim: (i) $R_n \xrightarrow[C']{\text{erg.}} E[I_{\alpha/\alpha_i^\infty} | \mathcal{I}_n(\tau)]$

(ii) $S_n \xrightarrow[C']{\text{erg.}} \underline{\underline{0}}$

$$(i) R_n = \frac{1}{n} \sum_{k=0}^{n-1} I_{\alpha/\alpha_i^\infty} \circ T^k$$

$$\xrightarrow[C']{\text{erg.}} E[I_{\alpha/\alpha_i^\infty} | \mathcal{I}_n(\tau)]$$

by the ergodic theorem (since $I_{\alpha/\alpha_i^\infty} \in C'$).
 (that was easy)

$$(ii) S_n \xrightarrow[C']{\text{erg.}} \underline{\underline{0}}.$$

$$\underline{C'}: \int |S_n| d_\alpha = \frac{1}{n} \int \left| \sum_{k=0}^{n-1} (I_{\alpha/\alpha_i^{n-k}} - I_{\alpha/\alpha_i^\infty}) \right| d_\alpha$$

$$\leq \frac{1}{n} \sum_{k=0}^{n-1} \int |I_{\alpha/\alpha_i^{n-k}} + I_{\alpha/\alpha_i^\infty}| d_\alpha$$

by previous theorem, $I_{\alpha/\alpha_i^n} \rightarrow I_{\alpha/\alpha_i^\infty}$ in C' (since $\sigma(\alpha_i^n) \cap \sigma(\alpha_i^\infty)$)

$$\therefore \int |I_{\alpha/\alpha_i^{n-k}} + I_{\alpha/\alpha_i^\infty}| d_\alpha \rightarrow 0,$$

$$\text{hence } \sum_{k=0}^{n-1} \int |I_{\alpha/\alpha_i^{n-k}} + I_{\alpha/\alpha_i^\infty}| d_\alpha = o(n),$$

$$\therefore \int |S_n| d_\alpha \leq \frac{o(n)}{n} \rightarrow 0 \text{ as } n \rightarrow \infty.$$

c.s.: let $h_N := \sup_{n \geq N} |I_{\alpha|d_1^n} - I_{\alpha|d_1^\infty}|$

$$\forall N \geq 1, h_N \leq \sup_{n \geq 1} |I_{\alpha|d_1^n}|$$

so $\forall \alpha, h_N \in C'$ by previous Lemma (in continuity theorem).

Moreover, the same Theorem gives $h_N \xrightarrow{n \rightarrow \infty} 0$ c.s.

Theorems of properties and the conditional DCT give

$$\lim_{N \rightarrow \infty} E[h_N | \mathcal{I}_n(\tau)] = E[\lim_{N \rightarrow \infty} h_N | \mathcal{I}_n(\tau)] \\ = 0 \quad (\star)$$

Consider $S_n := \limsup_{n \rightarrow \infty} |S_n|$

$$\begin{aligned} & \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} |I_{\alpha|d_1^{n-k}} - I_{\alpha|d_1^\infty}| \circ T^k \\ (\text{fix } N \text{ some } n) & = \limsup_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{k=0}^{n-N} |I_{\alpha|d_1^{n-k}} - I_{\alpha|d_1^\infty}| \circ T^k \right. \\ & \quad \left. + \frac{1}{n} \sum_{k=n-N+1}^{n-1} |I_{\alpha|d_1^{n-k}} - I_{\alpha|d_1^\infty}| \circ T^k \right) \\ & \leq |I_{\alpha|d_1^n}, \dots, I_{\alpha|d_1^1}| \end{aligned}$$

$$\leq \limsup_{n \rightarrow \infty} \left(\underbrace{\left(\frac{1}{n} \sum_{k=0}^{n-N} h_N \circ T^k \right)}_{\longrightarrow E[h_N | \mathcal{I}_n(\tau)]} + \underbrace{\left(\frac{1}{n} \sum_{k=n-N+1}^{n-1} h_i \circ T^k \right)}_{h_i \text{ integrable + only f.m. terms (dividing by larger } n)} \right)$$

by ergodic theorem.



See last page for more justification

Hence for all $N \geq 1$,

$$\limsup_{n \rightarrow \infty} |S_n| \leq E[\chi_N / I_n(\tau)]$$

$\rightarrow 0 \text{ a.s. as } N \rightarrow \infty$.

$\therefore S_n \rightarrow 0 \text{ a.s.}$

□

(d=1: 1957 ; d>1: 1983)

Comments on d>1 part:

① $\sup_{n \geq 1} (n^{-\frac{1}{d}} I_{\omega}^{(0n)}) \in \ell'$

Proof by contradiction: Assume $\limsup n^{\frac{1}{d}} I_{\omega}^{(0n)} \neq \liminf n^{\frac{1}{d}} I_{\omega}^{(0n)}$

$m(A), A \in \mathcal{B} \text{ a.s.} := \text{on a set } F \text{ with } m(F) > 0.$

$$\begin{aligned} \text{on } x \in F \Rightarrow m(I_{\omega}^{(0n)(x)}) &> e^{-n^{\frac{1}{d}a}} \quad \text{i.o.} \\ m(I_{\omega}^{(0n)(x)}) &< e^{-n^{\frac{1}{d}b}} \quad \text{i.o.} \end{aligned}$$

for constants $a, b > 0, b > a$, st. on F $\limsup n^{\frac{1}{d}} I_{\omega}^{(0n)} < a < b < \liminf n^{\frac{1}{d}} I_{\omega}^{(0n)}$

② $\alpha = \{A_1, \dots, A_m\}; Q = \{1, \dots, m\}$

$\gamma x \in Q^{D_n}$ is called an n -name, and an element x has

the n -name $\gamma \in Q^{D_n}$ st. $T^i x \in A_{\gamma_i}$ for $i \in D_n$.

However, by Lemma 1, the N-norm of almost all $x \in F$ is in Γ_N for sufficiently large N , i.e. it is in $\Gamma_N \setminus B_N$.

\therefore eventually (for each N) for almost all $x \in F$,

$$\mu(I_{d^N}(x)) \geq e^{-N^d(D - \frac{L-\epsilon}{\eta})}$$

This completes our initial properties of F . \square

\square

(probably best to check the case yourself.)

Krengel, Ergodic Theory: Theorem 2.5

A note on the notation:

$\mu(I_d(x)) = \mu(A)$ where $A \in \mathcal{L}$ is the partition cell containing x .

- $n(\gamma) = n\{\alpha \in X : \gamma \text{ is } \alpha \text{; } n\text{-name}\}$.

- γ is fat if $n(\gamma) > e^{-an^d}$

- Call a name γ ϵ -covered by $\{\gamma_j\}$ if the domain of γ_j covers up to a missing fraction of ϵ of the domain W of γ , and γ_j agrees with γ on $W_j \cap W$. If $W_j \subset W$ and the W_j are disjoint, $\{\gamma_j\}$ ϵ -tiles γ .

Lemma 3: $\epsilon > 0$, $m \geq 1$, c.e. $\alpha \in F$, $\exists N(\omega)$ s.t. $N \geq N(\omega)$
 \rightarrow the N -name of α can be ϵ -filled by fat names of sidelength $\geq m$.

Lemma 4: m large enough, ϵ small enough, then $|\tilde{P}_N| \leq e^{N^d(b+c)/2}$
 holds for large N ,

where $\tilde{P}_N = \left\{ \gamma \text{ is } N\text{-name} : \gamma \text{ can be } \epsilon\text{-filled by fat names with sidelength } \geq m \right\}$

Now: let $B_N = \left\{ \gamma \in \tilde{P}_N : n(\gamma) < e^{-N^d(b - (\frac{b-c}{4}))} \right\}$

Then $n\left(\{\alpha \in X : \alpha \text{; } N\text{-name is in } B_N\}\right) \leq |\tilde{P}_N| e^{-N^d(b - (\frac{b-c}{4}))}$
 $\leq e^{-N^d(\frac{b-c}{4})}$

Borel-Cantelli \Rightarrow the N -name of almost all $\alpha \in X$
 is eventually not in B_N .

More Trichotomy (Wishart):

Let (X, \mathcal{B}, μ, T) be a measure preserving transformation and f be a positive function on X . Notice that

$$\int_X f d\mu = \int_0^\infty \mu(f > t) dt.$$

That implies that $\sum_{i=1}^\infty \mu(f > ir) < \infty$ for all $r > 0$. Since T is measure preserving this implies

$$\sum_{i=0}^\infty \mu\left(\frac{1}{i}f \circ T^i > r\right) < \infty.$$

By Borel-Cantelli lemma

$$\mu\left(\frac{1}{i}f \circ T^i > r \text{ infinitely often}\right) = 0.$$

Therefore

$$\limsup_{i \rightarrow \infty} \frac{1}{i} f \circ T^i \leq r.$$

But r was arbitrary. Thus

$$\lim_{i \rightarrow \infty} \frac{1}{i} f \circ T^i = 0.$$

The result immediately follows for L^1 functions.

