Entropia w układach dynamicznych

Środowiskowe studia doktoranckie
Uniwersytet Jagielloński
Kraków, marzec-kwiecień 2013

Tomasz Downarowicz

Część II
Entropia topologiczna i zasada wariacyjna

1 Wstęp

Zacznijmy od początku.

Zadajemy komuś pytanie, na które możliwa jest pewna skończona liczba l odpowiedzi. Na przykład. „Jak masz na imię?” Możliwe odpowiedzi, to l-elementowe rozbicie przestrzeni „wszystkiego, co się może potem wydarzyć”. Jak wiemy, entropia tego rozbicia (średnia ilość informacji z otrzymanej odpowiedzi) zależy od rozkładu prawdopodobieństwa na przestrzeni i nie przekracza $\log l$ w przypadku, gdy wszystkie odpowiedzi są jednakowo prawdopodobne.

Wiemy też jednak, że rozkład prawdopodobieństwa jest sprawą subiektywną, zależną od wiedzy obserwatora o sytuacji. Tak naprawdę w praktyce prawie nigdy nie jest ono *a priori* dokładnie zdefiniowane. Jeśli lekarz mówi, po badaniu USG ciężarnej kobiecie, że „na 80% będzie to dziewczynka”, to podany procent jest czystą spekulacją, oszacowaniem na oko, tym co się lekarzowi zdaje na podstwie niewyraźnego obrazka. Równie dobrze mógłby on powiedzieć 70% lub 90%. Jeśli później okaże się, że to jednak chłopiec, prawdopodobnie będzie starał się przekonać pacjentkę, że mówił 50%.

A zatem dobrze byłoby w praktyce nie opierać się na bliżej nieokreślonym prawdopodobieństwie lecz na czymś bardziej pewnym. Jedyną rzeczą, którą możemy ustalić na pewno jest liczba możliwych odpowiedzi l. Można więc przyjąc $\log l$ jako rodzaj *kombinatorycznej* entropii rozbicia. Zaletą takiego parametru jest to, że nie zależy ona od rozkładu prawdopodobieńtwa, a mimo to coś nam mówi o ich entropiach – jest on bowiem równy maksimum entropii przy różnych rozkładach.

$$\mathbf{H}_{\mathsf{comb}}(\mathcal{P}) = \max\{H_\mu(\mathcal{P}) : \mu - \text{miara probabilistyczna na } X\}.$$

Otóż, tę samą zależność będziemy próbowali przenieść na topologiczne układy dynamiczne w odniesieniu do entropii dynamicznej. Zależności tego typu noszą wspólne miano *zasad wariacyjnych.*

Pewien problem, na jaki się przy tym natkniemy, to występowanie w układach topologicznych efektu tzw. logiki rozmytej. Na przykład gdy ktoś robi sobie test ciążowy, to ma co prawda tylko dwie możliwości, które powinny być rozpoznawalne przy pomocy kolorów próbnika: biały – brak ciąży, czerwony – ciąża. Ale zdarza się (i to zapewne nader często), że kolor próbnika jest różowy. Jak obliczyć ilość uzyskanej wtedy informacji?

2 Entropia Kołmogorowa–Sinaia – uzupełnienie

Musimy najpierw uzupełnić naszą wiedzę na temat entropii dynamicznej w układach teorio-miarowych. Przypomnijmy, że teorio-miarowym układem dynamicznym nazywamy czwórkę $(X, \mathfrak{A}, \mu, T)$, gdzie $(X, \mathfrak{A}, \mu)$ jest standardową przestrzenią probabilistyczną, a $T :\to X$ – transformacją mierzalną zachowującą miarę. Jeśli w przestrzeni X wprowadzimy skończone rozbicie mierzalne $\mathcal{P} = \{A_1, \ldots, A_l\}$, to dostaniemy faktor symboliczny naszego ukladu zwany *procesem generowanym przez* $\mathcal{P}$. Dla takich procesów zdefiniowaliśmy entropię dynamiczną

$$h(\mu, T, \mathcal{P}) = \lim_{n\to\infty} \tfrac{1}{n} H(\mu, \mathcal{P}^n),$$

gdzie przez $\mathcal{P}^n$ oznaczaliśmy połączenie $\bigvee_{i=0}^{n} T^{-i}(\mathcal{P})$. Dla takiej entropii udowodniliśmy Twierdzenie Shannona–McMillana–Breimana.

Chcemy teraz uniezależnić się od rozbicia i określić entropię układu dynamicznego. Rolę tę spełnia tzw. *entropia Kołmogorowa–Sinaia* zdefiniowana następująco

Definicja 4.1.1 Entropią Kołmogorowa–Sinaja układu dynamicznego $(X, \mathfrak{A}, \mu, T)$ nazywamy liczbę

$$h(\mu, T) = \sup_{\mathcal{P}} h(\mu, T, \mathcal{P}),$$

gdzie $\mathcal{P}$ przebiega wszystkie $\mathfrak{A}$-mierzalne rozbicia skończone (lub, co daję równoważną definicję, rozbicia przeliczalne o skończonej entropii statycznej). Jeśli transformacja lub miara, lub obie te rzeczy są ustalone, to będziemy używać alternatywmych oznaczeń $h(\mu)$, $h(T)$, $h(\mathfrak{A})$.

Jako interpretację tego pojęcia możemy powiedzieć, że jeśli traktujemy rozbicie jako rodzaj *rozdzielczości teorio-miarowej* naszej obserwacji układu, i jeśli już rozumiemy entropię procesu przy ustalonym rozbiciu, to entropia Kołmogorowa–Sinaja określa maksymalną entropię procesu, jaką możemy uzyskać w układzie dynamicznym zmieniając dowolnie tę rozdzielczość. Okazuje się bowiem, że chociaż entropie statyczne rozbić skończonych nie są ograniczone, to jednak ich entropie dynamiczne mogą okazać się ograniczone. Tak więc $h(\mu, T)$ to *maksymalna średnia ilość informacji po przestrzeni i czasie, jaką dostacza nam układ niezależnie od tego z jaką rozdzielczością go obserwujemy.*

Jeśli teraz $\mathfrak{B}$ jest podniezmienniczym sigma-ciałem, to możemy zdefiniować entropię warunkową układu względem faktora:

Definicja 4.1.5 Entropią warunkową układu dynamicznego $(X, \mathfrak{A}, \mu, T)$ względem faktora (czyli sigma ciała podniezmienniczego) $\mathfrak{B}$ nazywamy liczbę

$$h(\mu, T|\mathfrak{B}) = \sup_{\mathcal{P}} h(\mu, T, \mathcal{P}|\mathfrak{B}).$$

Alternatywne oznaczenia, to $h(\mu|\nu)$, $h(T|S)$, $h(\mathfrak{A}|\mathfrak{B})$, gdzie ν i S oznaczają odpowiednio miarę i transformację na faktor-przestrzeni atomów sigma-ciała $\mathfrak{B}$.

Fakt 4.1.6 $h(\mathfrak{A}|\mathfrak{B}) + h(\mathfrak{B}) = h(\mathfrak{A})$.

Dowód: Równość jest trywialna, gdy $h(\mathfrak{B}) = \infty$, gdyż jest oczywiste z definicji, że $h(\mathfrak{A}) \geqslant h(\mathfrak{B})$. W pozostałych przypadkach mamy udowodnić wzór „substraktywny" $h(\mathfrak{A}|\mathfrak{B}) = h(\mathfrak{A}) - h(\mathfrak{B})$. Niech $\mathcal{P}$ i $\mathcal{Q}$ przebiegają wszystkie rozbicia skończone odpowiednio $\mathfrak{A}$- i $\mathfrak{B}$-mierzalne. Wtedy

$$h(\mathfrak{A}) - h(\mathfrak{B}) = \sup_{\mathcal{P}} h(\mathcal{P}) - \sup_{\mathcal{Q}} h(\mathcal{Q}) = \inf_{\mathcal{Q}} \sup_{\mathcal{P}} (h(\mathcal{P}) - h(\mathcal{Q})) =$$
$$\inf_{\mathcal{Q}} \sup_{\mathcal{P}} (h(\mathcal{P} \vee Q) - h(\mathcal{Q})) \geqslant \sup_{\mathcal{P}} \inf_{\mathcal{Q}} (h(\mathcal{P} \vee Q) - h(\mathcal{Q})) =$$
$$\sup_{\mathcal{P}} \inf_{\mathcal{Q}} (h(\mathcal{P}|Q)) = \sup_{\mathcal{P}} h(\mathcal{P}|\mathfrak{B}) = h(\mathfrak{A}|\mathfrak{B}).$$

Z drugiej strony, mamy też

$$h(\mathfrak{A}|\mathfrak{B}) = \sup_{\mathcal{P}} \inf_{\mathcal{Q}} (h(\mathcal{P} \vee Q) - h(\mathcal{Q})) \geqslant \sup_{\mathcal{P}} \inf_{\mathcal{Q}} (h(\mathcal{P}) - h(\mathcal{Q})) = h(\mathfrak{A}) - h(\mathfrak{B}).$$

□

Podamy teraz listę własności entropii Kołmogorowa–Sinaja (bez dowodów, które są natychmiastowymi konsekwencjami analogicznych własności dla entropii procesów). Poniżej, $\mathfrak{B}, \mathfrak{C}$ i $\mathfrak{D}$ są poniezmienniczymi pod-sigma-ciałami $\mathfrak{A}$.

Fakt 4.1.7

$$h(\mathfrak{B} \vee \mathfrak{C}|\mathfrak{D}) = h(\mathfrak{B}|\mathfrak{C} \vee \mathfrak{D}) + h(\mathfrak{C}|\mathfrak{D}),$$
$$\mathfrak{B} \succcurlyeq \mathfrak{C} \implies h(\mathfrak{B}|\mathfrak{D}) \geqslant h(\mathfrak{C}|\mathfrak{D})$$
$$\mathfrak{C} \succcurlyeq \mathfrak{D} \implies h(\mathfrak{B}|\mathfrak{C}) \leqslant h(\mathfrak{B}|\mathfrak{D})$$
$$h(\mathfrak{B} \vee \mathfrak{C}|\mathfrak{D}) \leqslant h(\mathfrak{B}|\mathfrak{D}) + h(\mathfrak{C}|\mathfrak{D}),$$
$$h(\mathfrak{B}|\mathfrak{D}) \leqslant h(\mathfrak{B}|\mathfrak{C}) + h(\mathfrak{C}|\mathfrak{D}).$$

□

Mamy też „zasadę potęgową" (tu również dowód wynika natychmiast z analogicznej zasady dla procesów).

Fakt 4.1.14 Dla każdego $n \geqslant 0$ (a dla działań odwracalnych również dla n ujemnych) mamy

$$h(T^n) = |n| h(T).$$

Dowód wynika natychmiast z Faktu 2.4.19 przez nałożenie supremum po rozbiciach. □

Twierdzenie (Kołmogorowa–Sinaja) Jeśli $\mathcal{P}$ jest generatorem (tzn. $\mathcal{P}^{\mathbb{N}_0} = \mathfrak{A}$), to

$$h(\mu, T) = h(\mu, T, \mathcal{P}).$$

Dowód wynika natychmiast z Faktu 2.4.1. □

3 Rozdzielczość topologiczna

Przechodzimy do części topologicznej naszych rozważań. Przez *topologiczny układ dynamiczny* będziemy rozumieć parę (X, T), gdzie X jest przestrzenią metryczną zwartą, a $T : X \to X$ jest transformacją ciągłą. Przede wszystkim, trzeba wiedzieć, że na mocy twierdzenia o punkcie stałym Bogolubowa–Kryłowa w takim układzie zawsze istnieje przynajmniej jedna miara (borelowska, probabilistyczna) T-niezmiennicza. Wtedy układ $(X, \mathfrak{A}_\mu, \mu, T)$ jest teorio-miarowym układem dynamicznym ($\mathfrak{A}_\mu$ oznacza sigma-ciało zbiorów borelowskich uzupełnione dla miary μ). Miar niezmienniczych może być wiele (nawet nieprzeliczalnie wiele), tak więc jeden układ topologiczny najczęściej integruje w sobie wiele układów teorio-miarowych.

W zasadzie chcielibyśmy wprowadzić topologiczną funkcję informacji i entropię topologiczną w oparciu o wcześniej zasygnalizowaną ideę entropii kobinatorycznej, opartej na liczeniu (niepustych) elementów rozbicia odpowiadającego odpowiedziom na jakieś pytanie (wyniki jakiegoś pomiaru). W pewnych przypadkach rzeczywiście można tak zrobić i my wrócimy do tego prostego pomysłu przy omawianiu entropii ukladów zero-wymiarowych. Jednak w ogólnym przypadku przestrzeni metrycznej zwartej patrzenie na rozbicia mierzalne kłóci się ze strukturą topologiczną przestrzeni – po prostu funkcje charakterystyczne elementów rozbicia na ogół nie są ciągłe, tak więc rozbicie „rozrywa" przestrzeń zmieniając niejako jej topologię. Mówiąc ściślej, odwzorowanie faktorujące z naszego układu w układ symboliczny uzyskany przy pomocy rozbicia nie jest wtedy ciągłe. Aby lepiej zrozumieć ideę funkcji informacji i entropii topologicznej musimy omówić dokładniej interpretację pojęcia „rozdzielczości topologicznej".

Przypuśćmy, że dokonujemy pomiaru jakiejś wielkości, która może przyjmować wartości z odcinka $[0, 1]$. Nasze zdolności rozdzielcze są ograniczone, dlatego nie będziemy odróżniać wyników mało od siebie odległych, powiedzmy bliższych sobie niż pewien ϵ. Wtedy „klasą nierozróżnialności" wyniku x będzie odcinek otwarty $(x - \epsilon, x + \epsilon)$. Zauważmy, że klasy różnych wyników nie są rozłączne — relacja nierozróżnialności, choć jest zwrotna i symetryczna, nie jest przechodnia, a więc nie jest relacją równoważności.

W związku z tym proponowane są dwa sposoby obliczania funkcji informacji z takiego pomiaru. Pierwszy to policzyć ile maksymalnie wyników możemy rozróżnić i uznać logarytm z tej liczby jako (stałą na całej przestrzeni) funkcję informacji. Inna możliwość, to policzyć ile minimalnie klas nierozróżnialności wystarczy, aby pokryć całą przestrzeń (i przyjąć logarytm tej liczby jaką naszą stałą funkcję informacji). Oba sposoby różnią się nieznacznie (na przykład jeśli ϵ jest nieco większy od $\frac{1}{2}$ to maksymalnie możemy rozróżnić dwa elementy – np. 0 i 1, ale już jedna klasa – np. punktu $\frac{1}{2}$ – pokrywa cały odcinek). Inny sposób liczenia, ktory jast jakby ekstraktem z powyższych dwóch sposobów, poznamy za chwilę.

4 Pokrycia otwarte

Przez *pokrycie* będziemy rozmumieć rodzinę zbiorów otwartych, których suma jest całą przestrzenią X. Na przykład $\mathcal{U}_{(1,\epsilon)}$ oznaczać będzie pokrycie wszystkimi kulami o promieniu ϵ. Formalnie nie ma przeszkód, aby elementem pokrycia był zbiór pusty. *Podpokryciem* pokrycia $\mathcal{U}$ nazwiemy każdą podrodzinę $\mathcal{V} \subset \mathcal{U}$, która jest pokryciem. Ze zwartości wynika, że każde pokrycie posiada podpokrycie skończone, dlatego można zdefiniować parametr skończony $N(\mathcal{U})$ jako minimalną liczność podpokrycia skończonego. Podpokrycie $\mathcal{V}$ o tej liczności spełnia $N(\mathcal{U}) = N(\mathcal{V}) = \#\mathcal{V}$. Pokrycie spełniające ostatnią równość $N(\mathcal{V}) = \#\mathcal{V}$ nazwiemy *optymalnym.* Pokrycie $\mathcal{V}$ jest optymalne wtedy i tylko wtedy, gdy dla każdego $V \in \mathcal{V}$ istnieje punkt pokryty wyłącznie przez V (oczywiście zbiór pusty nie może być elementem pokrycia optymalnego).

Powiemy, że pokrycie $\mathcal{V}$ jest *wpisane* w pokrycie $\mathcal{U}$ (co zapiszemy przez $\mathcal{V} \succcurlyeq \mathcal{U}$, jeśli każdy element rozbicia $\mathcal{V}$ jest zawarty w pewnym elemencie rozbicia $\mathcal{U}$. Natomiast *połączenie* rozbić $\mathcal{U}, \mathcal{V}$ zdefiniowane jest tak samo jak dla rozbić:

$$\mathcal{U} \vee \mathcal{V} = \{U \cap V : U \in \mathcal{U},\ V \in \mathcal{V}\}.$$

Zwróćmy uwagę na istotne różnice pomiędzy relacją $\succcurlyeq$ i operacją $\vee$ dla rozbić dla pokryć. Po pierwsze jeśli $\mathcal{V} \succcurlyeq \mathcal{U}$, to wcale nie musi być tak, że każdy element $\mathcal{U}$ jest sumą elementów $\mathcal{V}$. Co prawda zawsze jest $\mathcal{U} \vee \mathcal{V} \succcurlyeq \mathcal{U}$, ale aby zachodziła tu równość nie wystarczy, żeby $\mathcal{V} \succcurlyeq \mathcal{U}$ (potrzebny jest o wiele mocniejszy warunek); na przykład $\mathcal{U} \vee \mathcal{U}$ na ogół nie równa się $\mathcal{U}$. Po drugie liczność $\mathcal{V}$ może być mniejsza nić liczność $\mathcal{U}$; na przykład każde popokrycie pokrycia $\mathcal{U}$ jest weń wpisane. Mamy jednak następujące zależności dotyczące paramertu $N(\mathcal{U})$:

Fakt 6.1.2

$$\begin{gathered}
\mathcal{U} \succcurlyeq \mathcal{V} \implies N(\mathcal{U}) \geqslant N(\mathcal{V}), \\
N(\mathcal{U} \vee \mathcal{V}) \leqslant N(\mathcal{U}) N(\mathcal{V}), \\
N(\mathcal{U} \vee \mathcal{U}) = N(\mathcal{U}), \\
N(T^{-1}(\mathcal{U})) \leqslant N(\mathcal{U}).
\end{gathered}$$

W ostatnim punkcie mamy na myśli transformację ciągłą $T : X \to X$. Zauważmy, że przeciwobraz pokrycia jest pokryciem. Ponadto, jeśli T jest surjekcją, to $N(T^{-1}(\mathcal{U})) = N(\mathcal{U})$.

Dowody są elementarne. □

W kontekście układu dynamicznego zadanego przez transformację T jak wyżej będziemy pisać, podobnie jak dla rozbić

$$\mathcal{U}^n = \bigvee_{i=0}^{n} T^{-i}(\mathcal{U}).$$

Z pokryciem $\mathcal{U}$ zwiążemy jeszcze dwa parametry: średnicę i liczbę Lebesgue'a:

Definicja *Średnicą* pokrycia $\mathcal{U}$ oznaczoną $\mathsf{diam}(\mathcal{U})$ nazwiemy supremum średnic jego elementów. Jego *liczba Lebesgue'a*, oznaczana $\mathsf{Leb}(\mathcal{U})$, to maksymalna $\delta > 0$, taki że każda kula o promieniu δ mieści się w całości w którymś elemencie pokrycia $\mathcal{U}$ (nietrudno wykazać, że w przestrzeni zwartej taka liczba dodatnia istnieje).

Oczywiście $\mathsf{Leb}(\mathcal{U}) \leqslant \mathsf{diam}(\mathcal{U})$ oraz

$$\mathcal{U}_{(1,\mathsf{Leb}(\mathcal{U}))} \succcurlyeq \mathcal{U} \succcurlyeq \mathcal{U}_{(1,\mathsf{diam}(\mathcal{U}))}.$$

5 Informacja i entropia topologiczna – definicje

W układzie dynamicznym (X, T) wprowadzamy ciąg metryk d^n wzorem

$$d^n(x,y) = \max_{i=0,\ldots,n-1} d(T^i x, T^i y)$$

Oczywiście $d^1 = d$ i metryki te rosną wraz n, jednak wszystkie są sobie równoważne. Kule w tej metryce oznaczać będziemy przez $B^n(x,\epsilon)$ i nazywać (n,ϵ)-*kulami (Bowena)*. Pokrycie wszystkimi (n,ϵ)-kulami oznaczymy przez $\mathcal{U}_{(n,\epsilon)}$.

Podamy teraz dwie definicje informacji i entropii topologicznej (Dinaburg 1970, Bowen 1971):

Zbiór E nazwiemy (n,ϵ)-*rozdzielonym*, jeśli $d^n(x,y) \geqslant \epsilon$ dla dowolnych $x, y \in F$, $x \neq y$. Ze zwartości wynika łatwo, że przy ustalonych n i ϵ, liczności zbiorów (n,ϵ)-rozdzielonych są wspólnie ograniczone przez pewną liczbę skończoną, którą oznaczymy przez $s(n,\epsilon)$. Zbiór (n,ϵ)-rozdzielony o tej liczności nazwiemy *maksymalnym*.

Definicja 6.1.1 Definiujemy kolejno

$$\mathbf{H}_1(n,\epsilon) = \log s(n,\epsilon)$$

$$\mathbf{h}_1(T,\epsilon) = \limsup_{n\to\infty} \frac{1}{n}\mathbf{H}_1(n,\epsilon),$$

$$\mathbf{h}_1(T) = \lim_{\epsilon\to 0} \uparrow \mathbf{h}_1(T,\epsilon).$$

Interpretujemy to następująco: liczba $s(n,\epsilon)$ to maksymalna ilość orbit, jaką jesteśmy w stanie rozróżnić posługując się urządzeniem obserwacyjnym o rozdzielczości ϵ. Zatem możemy przyjąć logarytm tej liczby jako (stałą na całej przestrzeni) funkcję informacji uzyskanej w n krokach. Dalej już postępujemy dokładnie tak samo jak w przypadku miarowym: $\mathbf{H}_1(n,\epsilon)$ jest średnią informacją w n krokach (teraz nie mamy ustalonej miary, ale do uśrednienia stałej nie trzeba jej ustalać – po prostu jest to ta sam stała), $\mathbf{h}_1(T,\epsilon)$ to średni przyrost entropii w jednym kroku (nie mamy jednak zagwarantowanego istnienia granicy) — jest to odległy analog entropii dynamicznej procesu przy ustalonym rozbiciu, wreszcie $\mathbf{h}_1(T)$ to supremum tego, co można uzyskać dowolnie poprawiając rozdzielczość, a więc analog entropii Kołmogorowa–Sinaja.

Kolejna defincja (również tych samych autorów) korzysta z pojęcia rozpinania. Zbiór E nazwiemy (n,ϵ)-rozpinającym, jeśli stanowi on ϵ-sieć w metryce d^n, czyli jeśli dla dowolnego $x \in X$ istnieje $y \in E$, taki że $d(x,y) < \epsilon$. Innymi słowy (n,ϵ)-kule wokól elementów E pokrywają X (są podpokryciem $\mathcal{U}_{(n,\epsilon)}$). Rzecz jasna, istnieją skończone zbiory (n,ϵ)-rozpinające i ich minimalną liczność oznaczymy przez $r(n,\epsilon)$. Zbiór rozpinający o tej liczności nazwiemy *minimalnym*.

Definicja 6.1.2 Definiujemy kolejno

$$\begin{aligned} \mathbf{H}_2(n,\epsilon) &= \log r(n,\epsilon) \\ \mathbf{h}_2(T,\epsilon) &= \limsup_{n\to\infty} \frac{1}{n}\mathbf{H}_2(n,\epsilon), \\ \mathbf{h}_2(T) &= \lim_{\epsilon\to 0} \uparrow \mathbf{h}_2(T,\epsilon). \end{aligned}$$

Interpretacja jest niemal identyczna, jak poprzednio, z tą tylko różnicą, że liczba $r(n,\epsilon)$ to minimalna ilość n-orbit, jaka wystarcza, aby zakwalifikować dowolną inną n-orbitę jako nierozróżnialną z jedną z nich. To też jest w pewnym sensie liczba rozróżnialnych n-orbit w układzie.

Podamy teraz trzecią definicję, która jest najogólniejsza, nie korzysta bowiem z pojęcia metryki (a zatem można ją stosować nawet w przestrzeniach niemetryzowalnych). Historycznie rzecz ujmując pojawia się ona najwcześniej (Adler–Konheim–McAndrew 1965), jednak dobrze jest widzieć ją jako uogólnienie defincji poprzedniej.

Definicja 6.1.3 Definiujemy kolejno

$$\begin{aligned} \mathbf{H}_3(n,\mathcal{U}) &= \log N(\mathcal{U}^n) \\ \mathbf{h}_3(T,\mathcal{U}) &= \lim_{n\to\infty} \frac{1}{n}\mathbf{H}_3(n,\mathcal{U}), \\ \mathbf{h}_3(T) &= \sup_{\mathcal{U}} \mathbf{h}_3(T,\mathcal{U}). \end{aligned}$$

Tym razem możemy napisać granicę, gdyż ciąg $\mathbf{H}_3(\mathcal{U}^n)$ jest podaddytywny, co wynika łatwo z wcześniejszych zależności: $N(\mathcal{U} \vee \mathcal{V}) \leqslant N(\mathcal{U})N(\mathcal{V})$ oraz $N(T^{-1}(\mathcal{U})) \leqslant N(\mathcal{U})$.

Interpretacja jest następująca: elementy pokrycia $\mathcal{U}$ to „klasy nierozróżnialności". Liczba $N(\mathcal{U}^n)$ to, jak poprzednio, minimalna liczba orbit, jaka wystarcza, aby zakwalifikować dowolną inną n-orbitę jako nierozróżnialną z jedną z nich. Poprzednia defnicja jest prawie szczególnym przypadkiem tej pokryciowej, jeśli $\mathcal{U}^n$ zastąpimy przez $\mathcal{U}_{(n,\epsilon)}$. Nie jest to ściśle szczególny przypadek, gdyż pokrycie $\mathcal{U}_{(n,\epsilon)}$ jest jedynie podpokryciem, a nie tym samym co $\mathcal{U}^n_{(1,\epsilon)}$.

Uwaga. W przestrzeni metrycznej zawsze można znaleźć ciąg pokryć $\mathcal{U}_k$, taki że dla każdego innego pokrycia $\mathcal{U}$ dostatecznie dalekie $\mathcal{U}_k$ jest weń wpisane. Dodatkowo możemy żądać, aby $\mathcal{U}_{k+1} \succcurlyeq \mathcal{U}_k$. Mówimy wtedy o *rozdrabniającym* *ciągu pokryć*. W takim przypadku supremum po $\mathcal{U}$ w ostatniej definicji można zastąpić granicą wstępującą po $\mathcal{U}_k$.

Entropię topologiczną definiuje się w oparciu o następujące twierdzenie, jako wspólną wartość liczb $\mathbf{h}_1(T)$, $\mathbf{h}_2(T)$ i $\mathbf{h}_3(T)$ i oznacza przez $\mathbf{h}(T)$.

Twierdzenie 6.1.8 W topologicznym układzie dynamicznym zachodzą równości

$$\mathbf{h}_1(T) = \mathbf{h}_2(T) = \mathbf{h}_3(T).$$

Dowód: Jak wiemy, zbiór E jest (n,ϵ)-rozpinający, wtedy i tylko wtedy, gdy (n,ϵ)-kule wokół jego elementów stanowią podpokrycie $\mathcal{U}_{(n,\epsilon)}$, które z kolei jest wpisane (jako podpokrycie) w $\mathcal{U}^n_{(1,\epsilon)}$. Zatem

$$r(n,\epsilon) = N(\mathcal{U}_{(n,\epsilon)}) \geqslant N^n(\mathcal{U}_{(1,\epsilon)}).$$

Dalej, jeśli dla jakiegoś pokrycia $\epsilon \leqslant \mathsf{Leb}(\mathcal{U})$, to $\mathcal{U}_{(1,\epsilon)} \succcurlyeq \mathcal{U}$, więc

$$N^n(\mathcal{U}_{(1,\epsilon)}) \geqslant N(\mathcal{U}^n).$$

Następnie zauważmy, że maksymalny zbiór (n,ϵ)-rozdzielony musi być (n,ϵ)-rozpinający, zatem

$$s(n,\epsilon) \leqslant r(n,\epsilon).$$

Wreszcie, jeśli $\mathsf{diam}(\mathcal{V}) < \epsilon$, to elementy $\mathcal{V}^n$ zawierają po co najwyżej jednym elemencie ze zbioru (n,ϵ)-rozdzielonego, co implikuje, że

$$N(\mathcal{V}^n) \geqslant s(n,\epsilon).$$

Z powyższych nierówności wynika, że

$$\mathbf{h}_3(T,\mathcal{V}) \geqslant \mathbf{h}_1(T,\epsilon) \geqslant \mathbf{h}_2(T,\epsilon) \geqslant \mathbf{h}_3(T,\mathcal{U}).$$

Teraz wystarczy nałożyć w suprema: najpierw po $\mathcal{V}$, potem po ϵ, na końcu po $\mathcal{U}$, i dostajemy tezę. □

Uwaga: Teraz widać, że jeśli w definicjach $\mathbf{h}_1(T,\epsilon)$ i $\mathbf{h}_2(T,\epsilon)$ w miejsce $\limsup_n$ zastosujemy $\liminf_n$, to wartości entropii $\mathbf{h}_1(T,)$ i $\mathbf{h}_2(T)$ nie ulegną zmianie.

6 Własności entropii topologicznej

Podukładem układu (X,T) nazywamy ddowlny podzbiór domknięty $Y \subset X$ taki, że $T(Y) \subset Y$ (czyli podniezmienniczy). Wtedy (Y,T) (formalnie powinno się pisać $T|_Y$) jest układem dynamicznym. Z kolei faktorem (topologicznym) układu (X,T) nazywamy dowolny inny układ (Y,S) jeśli istnieje odwzorowanie faktorujące **ciągłe** z X na Y (defnicja odwzorowania faktorującego jest taka sama jak w przypadku teorio-miarowym).

Fakt 6.2.1, 6.2.2 Entropia podukładu i entropia faktora są nie większe od entropii danego układu.

Dowód: Maksymalny zbiór (n,ϵ)-rodzielony w podukładzie jest (n,ϵ)-rodzielony w całym układzie (być może tu już nie jest maksymalny). Przeciwobraz pokrycia optymalnego w faktorze jest pokryciem optymalnym w rozszerzeniu i ma tę samą liczność (bo odwzorowanie faktorujące jest surjekcją). Supremum w definicji entropii w rozszerzeniu uwzględnia między innymi pokrycia podniesione z faktora (ale nie tylko te). To już implikuje żądaną nierówność. $\square$

Fakt 6.2.3

$$\mathbf{h}(T^n) = |n|\mathbf{h}(T).$$

Dowód przebiega identycznie jak dla entropii miarowej. $\square$

Fakt 6.2.4

$$\mathbf{h}(T,\mathcal{U}^n) = \mathbf{h}(T,\mathcal{U}).$$

Dowód. Co prawda nie zachodzi równość pokryć $(\mathcal{U}^n)^m$ i $\mathcal{U}^{n+m}$, ale pokrycia te mają tą samą liczbę $N(\cdot)$ (na tego samego powodu, co $N(\mathcal{U}\vee\mathcal{U}) = N(\mathcal{U})$). Tak więc dzieląc ich logarytm przez m i przechodząc z m do nieskończoności otrzymamy tą samą granicę co dla $\mathcal{U}^m$. $\square$

Wniosek (Analog tw. Kołmogorowa–Sinaja): Jeśli $\mathcal{U}$ jest *generatorem topologicznym* (tzn. ciąg $\mathcal{U}^n$ jest rozdrabniający), to $\mathbf{h}(T) = \mathbf{h}(T,\mathcal{U})$.

7 Miary niezmiennicze

Twierdzenie (Bogolubov–Kryłow 1937) W każdym topologicznym układzie dynamicznym (X,T) istnieje przynajmniej jedna (borelowska probabilistyczna) miara T-niezmiennicza (tzn., taka że $\mu(T^{-1}(A)) = \mu(A)$ dla każdego zbioru borelowskiego A).

Dowód: Z twierdzenia Riesza możemy traktować miary probabilistyczne jako funkcjonały nieujemne unormowane na $C(X)$. Z twierdzenia Banacha-Alaoglu, zbiór miar probabilistycznych $\mathcal{P}(X)$ jest zwarty w *-słabej topologii. Jest on również wypukły, a T działający na miarach przeprowadza $\mathcal{P}(X)$ w siebie i jest w tej topologii ciągły. Wybieramy dowolną miarę $\nu \in \mathcal{P}(X)$ i patrzymy na ciąg średnich

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} T^i\nu.$$

Są to elementy $\mathcal{P}(X)$. Zauważmy, że

$$\|\mu_n - T\mu_n)\| = \|\tfrac{\nu}{n} + \tfrac{T^n\nu}{n}\| \leqslant \tfrac{2}{n},$$

Niech μ będzie dowolnym punktem skupienia (w *-słabej topologii) ciągu μ_n. Ze zwartości $\mathcal{P}(X)$, taka miara probabilistyczna istnieje. Wtedy, dla dowolnej unormowanej funkcji $f \in C(X)$ i dowolnego $\epsilon > 0$ istnieje $n > \frac{1}{\epsilon}$, takie że $|\int f\,d\mu - \int f\,d\mu_n| < \epsilon$ oraz $|\int f \circ T\,d\mu - \int f \circ T\,d\mu_n| < \epsilon$. Przypomnijmy też, jak działa operator T na miarach:

$$\int f\,dT\mu = \int f \circ T\,d\mu.$$

Zatem, mamy

$$\begin{aligned} &\left|\int f\,d\mu - \int f\,dT\mu\right| \leqslant \\ &\left|\int f\,d\mu - \int f\,d\mu_n\right| + \left|\int f\,d\mu_n - \int f\,dT\mu_n\right| + \left|\int f\,dT\mu_n - \int f\,dT\mu\right| \leqslant \\ &\qquad\qquad \epsilon + \|f\| \cdot \|\mu_n - T\mu_n\| + \epsilon \leqslant 2\epsilon + \tfrac{2}{n} < 4\epsilon. \end{aligned}$$

Ponieważ ϵ jest dowolny, wykazaliśmy, że $\int f\,d\mu - \int f\,dT\mu = 0$. To oznacza, że $\mu - T\mu$ jest funkcjonałem zerowym na $C(X)$, co implikuje, że jest to po prostu miara zerowa, zatem $\mu = T\mu$, czyli wskazaliśmy miarę T-niezmienniczą. $\square$

Zbiór miar niezmienniczych będziemy oznaczać przez $\mathcal{P}_T(X)$. Jest on również wypukły i *-słabo zwarty (i oczywiście niepusty). Metrykę w tym zbiorze (równoważną z *-słabą topologią) można zadać w następujący sposób. Trzeba wybrać i ustalić ciąg funkcji unormowanych $(f_n)_{n\geqslant 1}$ o tej własności, że zbiór $\{f_n \circ T^k : n \geqslant 1, k \geqslant 0\}$ jest liniowo gęsty w $C(X)$. Następnie ustalić ciąg sumowalny liczb dodatnich $(c_n)_n \geqslant 1$. I wtedy możemy położyć

$$d^*(\mu, \nu) = \sum_{n=1}^{\infty} c_n \left|\int f_n\,d\mu - \int f_n\,d\nu\right|.$$

De facto warunek sumowalności można osłabić. Wystaczy, żeby szereg funkcyjny $\sum_n c_n|f_n|$ był zbieżny punktowo i wspólnie ograniczony. Szczegóły uzasadnienia tego stwierdzenia pominiemy.

Uwaga: Jeśli $\pi : (X,T) \to (Y,S)$ jest „odwzorowaniem faktorującym" między topologicznymi układami dynamicznymi (czyli ciągłą surjekcją z X na Y spełniającą $\pi \circ T = S \circ \pi$, to odwzorowanie indukowane na miarach (również oznaczmy je przez π i przypomnijmy, że $(\pi\mu)(A) = \mu(\pi^{-1}(A))$ dla zbioru A borelowskiego w Y) jest ciągłą afiniczną *surjekcją* z $\mathcal{P}_T(X)$ na $\mathcal{P}_S(Y)$.

Dowód: Niech $\mu \in \mathcal{P}_T(X)$. Weźmy zbiór A borelowski w Y. Mamy

$$S(\pi\mu)(A) = \mu(\pi^{-1}S^{-1}(A)) = \mu(T^{-1}\pi^{-1}(A)) = \mu(\pi^{-1}(A)) = (\pi\mu)(A),$$

czyli $\pi\mu$ jest S-niezmiennicza. Ciągłość w *-słabych topologiach: Niech μ_n zbiegają słabo do μ. Wtedy dla dowolnej $f \in C(Y)$ mamy

$$\int f\,d(\pi\mu_n) = \int f \circ \pi\,d\mu_n \to \int f \circ \pi\,d\mu = \int f\,d(\pi\mu).$$

Afiniczność jest oczywista. Nietrywialna jest tylko surjektywność. Niech $\nu \in \mathcal{P}_S(Y)$. Istnieje miara (niekoniecznie niezmiennicza) $\mu \in \mathcal{P}(X)$, taka że $\pi\mu = \nu$. Wynika to wprost z twierdzenia Hahna–Banacha o przedłużaniu funkcjonału: ν zadaje na podprzestrzeni $\{f \circ \pi : f \in C(Y)\} \subset C(X)$ funkcjonał $F_\nu(f \circ \pi) = \int f\,d\nu$. Ten funkcjonał po przedłużeniu do miary nieujemnej unormowanej na $C(X)$ będzie szukaną miarą μ. Teraz postępujemy tak, jak w dowodzie twierdzenia Bogolubova–Kryłowa; średnie $\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} T^i\mu$ mają punkt skupienia μ_0 będący miarą niezmienniczą. Ponieważ $\pi\mu$ jest miarą S-niezmienniczą ν, więc

$$\pi\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} \pi T^i\mu = \frac{1}{n}\sum_{i=0}^{n-1} S^i\pi\mu = \frac{1}{n}\sum_{i=0}^{n-1} S^i\nu = \frac{1}{n}\sum_{i=0}^{n-1} \nu = \nu$$

(dla każdego n), a z ciągłości π na miarach, również $\pi\mu_0 = \nu$, co kończy dowód. $\square$

Jeśli μ jest miarą niezmienniczą w topologicznym układzie dynamicznym (X,T), to otrzymujemy teorio-miarowy układ dynamiczny $(X, \mathfrak{A}_\mu, \mu, T)$, gdzie $\mathfrak{A}_\mu$ jest sigma-ciałem zbiorów borelowskich uzupełnionym względem miary μ (uzupełnienie stosujemy tylko po to, żeby otrzymać *standardową* przestrzeń probabilistyczną). Układ ten posiada swoją entropię Kołmogorowa–Sinaja $h(\mu,T)$. Pamiętajmy, że na ogół układ topologiczny może posiadać wiele miar niezmienniczych. Związek pomiędzy entropią topologiczną układu, a entropiami Kołmogorowa–Sinaja jego miar niezmienniczych ustala poniższe twierdzenie, uważane za jedno z kluczowych (obok twierdzenia Shannona–McMillana–Breimana) w teorii entropii układów dynamicznych.

Twierdzenie (Zasada wariacyjna)

$$\mathbf{h}(T) = \sup\{h(\mu,T) : \mu \in \mathcal{P}_T(X)\}.$$

Dowód podamy w kolejnych rozdziałach, ale ograniczymy się do przypadku, gdy X jest przestrzenią zero-wymiarową.

8 Dynamika i miary niezmiennicze w wymiarze zero

Aby zrozumieć dynamikę w przestrzenich zero-wymiarowych trzeba przede wszystkim zrozumieć dynamikę symboliczną.

Definicja *Układem symbolicznym* nazwiemy dowolny układ (X,T), gdzie $X \subset \Lambda^{\mathbb{N}_0}$ jest domkniętym zbiorem podniezmienniczym na transformację „przesunięcie” (ang. *shift*) $\sigma((x_n)_{n\in\mathbb{N}_0}) = (x_{n+1})_{n\in\mathbb{N}_0}$, Λ jest zbiorem skończonym (zwanym *alfabetem*), a T jest właśnie tą transformacją (obciętą do X).

Powyżej, w zbiór Λ traktujemy jako przestrzeń dyskretną (jest ona zwarta), a w $\Lambda^{\mathbb{N}_0}$ stosujemy topolgię produktową (która, na mocy tw Tichonowa, też jest zwarta).

Zauważmy, że w tzw. pełnym układzie symbolicznym (w którym $X_\Lambda = \Lambda^{\mathbb{N}_0}$) rozbicie na cylindry nad współrzędną zerową $\{[a] : a \in \Lambda\}$ (które również oznaczymy przez Λ) jest rozbiciem na zbiory otwarto-domknięte, w szczególności jest to więc pokrycie otwarte. Pokrycie to jest generatorem topologicznym. Jeśli teraz ograniczymy się do podukładu X (a więc dowolnego układu symbolicznego), to rozbicie $\Lambda|_X$ (Λ zrelatywizowane do X – dalszej części będziemy pomijać pisanie „$|_X$") jest nadal jego rozbiciem. To samo dotyczy (zrelatywizowanych) pokryć Λ^n. Zatem każde z takich pokryć ma jedyne podpokrycie optymalne otrzymane poprzez odrzucenie zbiorów pustych. Czyli parametr $N(\Lambda^n)$ (na X) liczy ile cylindrów z Λ^n kroi się niepusto z X. Dlatego wprowadzimy oznaczenie

$$\Lambda^n(X) = \{B \in \Lambda^n : [B] \cap X \neq \emptyset\}.$$

Interpretacja tego zbioru jest taka, że są to bloki długości n nad alfabetem Λ które *występują* w X (wystarczy aby wystąpił on w co najmniej jednym elemencie $(x_n)_{n \in \mathbb{N}_0} \in X$ na jakiejkowliek pozycji – wtedy stosując wielokrotnie „shift" zobaczymy go w jakimś elemencie X na pozycjach od zera do $n-1$, czyli właśnie $[B] \cap X \neq \emptyset$). Jak już powiedzieliśmy, zbiór $\Lambda^n(X)$ stanowi optymalne podpokrycie X pokrycia Λ^n, a ponieważ Λ jest generatorem topologicznym (również w X), więc mamy poniższy niezwykle prosty wzór na entropię topologiczną:

$$\mathbf{h}(X,T) = \mathbf{h}(X,T,\Lambda) = \lim_n \tfrac{1}{n} \log \#\Lambda^n(X).$$

Podobnie, jeśli ustalimy dowolną miarę niezmienniczą $\mu \in \mathcal{P}_T(X)$, to miara ta jest również miarą niezmienniczą pełnego układu symbolicznego (X_Λ, σ), i Λ traktowana teraz jako rozbicie mierzalne jest generatorem (teorio-miarowym). Zatem mamy wzór

$$h(\mu,T) = h(\mu,\sigma) = h(\mu,\sigma,\Lambda).$$

Tak więc licząc entropię czy to topologiczną, czy to Kołmogorowa–Sinaja jakiejś miary niezmienniczej, wystarczy patrzeć na „pokrycio-rozbicie" Λ.

Jeśli teraz mamy dowolny układ zero-wymiarowy, to co prawda nie musi on być układem symbolicznym (do tegu musiałby jeszcze być on *ekspanywny*), ale zawsze istnieje w nim ciąg rozbić (pokryć) otwarto-domkniętych, które łącznie generują zarówno topologię jak i sigma-ciało zbiorów borelowskich. Aby odróżnić te rozbicia od zwykłych rozbić (które są tylko mierzalne) oznaczymy je przez Λ_k. Każde takie rozbicie generuje proces, a zarazem układ symboliczny, w którym Λ_k staje się alfabetem. Przez $\Lambda_k(X)$ oznaczymy, jak poprzednio, zbiór bloków, które wstępują w tym układzie symbolicznym. Mamy wtedy podobne wzory na entropie, jak dla układów symbolicznych, z tym tylko, że we wzorach pojawi się rozsnąca granica po k:

$$\mathbf{h}(X,T) = \lim_k \uparrow \mathbf{h}(X,T,\Lambda_k) = \lim_k \uparrow \lim_n \tfrac{1}{n} \log \#\Lambda_k^n(X),$$

oraz, dla każdej miary niezmienniczej μ na X,

$$h(\mu,T) = \lim_k \uparrow h(\mu,T,\Lambda_k).$$

Przyjrzyjmy się jeszcze miarom niezmienniczym w układzie symbolicznym. Ponieważ są to miary niezmiennicze również w pełnym układzie symbolicznym nad danym alfabetem, można od razu założyć, że patrzymy na układ $(\Lambda^{\mathbb{N}_0}, \sigma)$. Każda miara niezmiennicza przypisuje wartości cylindrom nad blokami skończonymi i wartości te nie zależą od miejsca zaczepienia cylindra. Zatem miara jest zdeterminowana poprzez swoje wartości na cylindrach zaczeopinych na współrzędenej zerowej. Zgodnie z konwencją, zbiór takich cylindrów długości n będziemy po prostu oznaczać przez Λ^n. Topologię *-słabą w zbiorze miar niezmienniczych układu symbolicznego można zmetryzować przy pomocy takiej oto metryki:

$$d^*(\mu, \nu) = \lim_n \sum_{B \in \Lambda^n} |\mu(B) - \nu(B)|$$

(granica istnieje, gdyż powyższy ciąg jest ograniczony przez 2 i nietrudno przekonać się, że jest niemalejący). Wynika z tego następująca interpretacja bliskości miar: dwie miary μ i ν są „blisko” jeśli wszystkim dostatecznie długim cylindrom nadają podobne wartości:

$$\forall_{\epsilon>0}\, \exists_{n,\delta}\ \ (\forall_{B\in\Lambda^n}\ |\mu(B) - \nu(B)| < \delta) \implies d^*(\mu, \nu) < \epsilon$$

oraz

$$\forall_{\epsilon>0}\, \exists_{n,\delta}\ \ d^*(\mu, \nu) < \epsilon \implies (\forall_{B\in\Lambda^n}\ |\mu(B) - \nu(B)| < \delta)$$

(jednak dobór n i δ do ϵ może być w obu przypadkach inny, dlatego nie piszemy jednego zdania logicznego z równoważnością).

Udowodnimy teraz następujący prosty fakt

Fakt 7.2.4 Funkcja entropii Kołmogorowa–Sinaja $\mu \mapsto h(\mu, \sigma)$ jest górnie półciągła w topolgii *-słabej na zbiorze wszystkich miar niezmienniczych pełnego układu symbolicznego $(\Lambda^{\mathbb{N}_0}, \sigma)$.

Dowód: Mamy

$$h(\mu, \sigma) = h(\mu, \sigma, \Lambda) = \lim_n \downarrow \frac{1}{n} \sum_{B\in\Lambda^n} \eta(\mu(B)).$$

Dla każdego cylindra B funkcja $\mu \mapsto \mu(B)$ jest ciągła (gdyż $\mathbf{1}_B$ jest ciągła jako funkcja charakterystyczna zbioru otwarto-domkniętego, a $\mu(B)$ to całka z tej funkcji). Funkcja η jest ciągła. Dalej mamy sumę skończoną i dzielenie przez stałą. Zatem mamy tu granicę malejącą ciągu funkcji ciągłych, a to jest funckja górnie półciągła. $\square$

9 Zasada wariacyjna w wymiarze zero

Podamy teraz dowód zasady wariacyjnej w układach zero-wymiarowych. Dowód w przypadku ogólnym jest o wiele bardziej skomplikowany. Istnieje też sposób aby twierdzenie to uogólnić z przypadku zero-wymiarowego na dowolny, jednak on również wymaga skomplikowanej konstrukcji tzw. zero-wymiarowych rozszerzeń pryncypialnych. Dlatego w ramach tego kursu ograniczymy się do przypadku zero-wymiarowego.

Główna część dowodu dotyczy układów symbolicznych.

Dowód zasady wariacyjnej dla układów symbolicznych: Dowód nierówności w jedną stronę jest natychmiastowy. Mamy pokazać, że entropia Kołmogorowa–Sinaja dowolnej miary niezmienniczej μ w układzie symbolicznym (X,σ) nad alfabetem Λ jest nie większa od jego entropii topologicznej. Mamy

$$h(\mu,T) = h(\mu,\sigma,\Lambda) = \lim_n \tfrac{1}{n} H(\mu,\Lambda^n).$$

Ponieważ miara μ jest niesiona przez X (dopełnienie X ma miarę zero), więc $H(\mu,\Lambda^n) = H(\mu,\Lambda^n(X)) \leqslant \log \#\Lambda^n(X)$, a co za tym idzie,

$$h(\mu,T) \leqslant \lim_n \tfrac{1}{n} \log \#\Lambda^n(X) = \mathbf{h}(X,T).$$

Dowód w drugą stronę jest nieco trudniejszy. Skonstruujemy miarę μ, taką że $h(\mu,T) = \mathbf{h}(X,T)$ (czyli miarę o maksymalnej entropii). Ustalmy n i niech X_n oznacza zbiór wszystkich ciągów uzyskanych jako nieskończone konkatenacje bloków z $\Lambda^n(X)$, z których pierwszy jest zaczepiony na współrzędnej zerowej. Oczywiście $X \subset X_n$. Jak łatwo widać, zbiór X_n jest niezmienniczy pod działaniem σ^n w sensie równości $\sigma^n(X_n) = X_n$ natomiast pod działaniem σ mamy taki oto cykl ciąg surjekcji

$$X_n \to \sigma(X_n) \to \sigma^2(X_n) \to \cdots \to \sigma^{n-1}(X_n) \to X_n.$$

Zatem każda z przestrzeni $\sigma^i(X_n)$ jest σ^n-niezmiennicza.

Na X_n mamy specjalną miarę σ^n-niezmienniczą, mianowicie miarę Bernoulliego $\mu_n^{(0)}$, która wszystkim blokom $B \in \Lambda^n(X)$ przypisuje równe wartości $(\#\Lambda^n(X))^{-1}$ (a konkatenacjom m takich bloków wartość $(\#\Lambda^n(X))^{-m}$). Zauważmy, że takie konkatenacje to właśnie połączenie m kolejnych przeciwobrazów przez σ_n rozbicia Λ^n, co możemy zapisać jako $(\Lambda^n)^m$ (tutaj pierwszy wykładnik – n – odnosi się do działania σ a drugi – m – do działania σ^n). Entropia tego rozbicia dla tej miary wynosi zatem

$$H(\mu_n^{(0)},(\Lambda^n)^m) = \log(\#\Lambda^n(X))^m = m \log \#\Lambda^n(X),$$

a entropia dynamiczna

$$h(\mu_n^{(0)},\sigma^n,\Lambda^n) = \lim_m \tfrac{1}{m} m \log \#\Lambda^n(X) = \log \#\Lambda^n(X).$$

Miara $\mu^{(0)}$ jest przenoszona przez kolejne iteracje σ^i $(i = 1,2,\dots,n-1)$ na miary σ^n-niezmiennicze $\mu^{(i)}$ niesione przez zbiory $\sigma^i(X_n)$, a następnie po n iteracjach wraca jako $\mu^{(0)}$ na X_n. Zatem każda z miar $\mu^{(i)}$ $(i = 0,\dots,n-1)$ jest faktorem teorio-miarowym każdej innej z tych miar (miary te są *słabo izomorficzne*). To wystarcza, aby stwierdzić, że mają one jednakowe entropie (pod działaniem σ^n). Miara μ_n zdefiniowana jako ich średnia

$$\mu_n = \frac{1}{n}\sum_{i=0}^{n-1}\mu_n^{(i)}$$

jest zarówno σ^n-, jak i σ-niezmiennicza. Jej entropię pod działaniem σ^n względem rozbicia Λ^n obliczymy z afiniczności entropii dynamicznej

$$h(\mu_n,\sigma^n,\Lambda^n) = \frac{1}{n}\sum_{i=0}^{n-1} h(\mu_n^{(i)},\sigma^n,\Lambda^n) = h(\mu_n^{(0)},\sigma^n,\Lambda^n) = \log\#\Lambda^n(X).$$

Z kolei do obliczenia entropii miary μ_n pod działaniem σ względem Λ zastosujemy zasadę potęgową:

$$h(\mu_n,\sigma,\Lambda) = \tfrac{1}{n}h(\mu_n,\sigma^n,\Lambda^n) = \tfrac{1}{n}\log\#\Lambda^n(X).$$

Ponieważ Λ jest generatorem, to jest to samo, co entropia Kołmogorowa–Sinaja $h(\mu_n,\sigma)$. Niech μ oznacza dowlny punkt skupienia ciągu μ_n (w zwartym zbiorze miar σ-niezmienniczych pełnego ukladu symbolicznego nad alfabetem Λ. Z górnej półciągłości funkcji entropii w układzie symbolicznym wynika, że

$$h(\mu,\sigma) \geqslant \liminf_n h(\mu_n,\sigma) = \lim_n \tfrac{1}{n}\log\#\Lambda^n(X) = \mathbf{h}(X,T).$$

Jeśli wykażemy, że mira μ jest niesiona przez X, to będzie to koniec dowodu. Trzeba więc pokazać, że dopełnienie X ma miarę μ zero. Dopełnienie to, jako zbiór otwarty w przestrzeni ośrodkowej, jest przeliczalną sumą zbiorów bazowych, czyli cylindrów otwarto-domkniętych. Wystarczy więc pokazać, że miara dowolnego cylindra C (dowolnej długości m) rozłącznego z X, czyli (jako blok) *nie występującego w* X (a więc nie należącego do $\Lambda^m(X)$) jest zero. Ponieważ miara zbioru otwarto-domkniętego jest ciągłą funkcją miary, wystarczy pokazać, że $\mu_n(C) \to 0$ po n, czyli, że $\mu_n(C)$ jest małe dla dużego n. Pokażemy o wiele więcej, że $\nu(C)$ jest małe dla dowolnej miary ν niesionej przez zbiór σ-niezmienniczy $X_n' = X_n \cup \sigma(X_n) \cup \cdots \cup \sigma^{n-1}(X_n)$. Wystarczy to pokazać dla miar ergodycznych, gdyż w każdym układzie topologicznym każda miara niezmiennicza jest granicą kombinacji liniowych miar ergodycznych (to wynika z twierdzenia Kreina–Milmana i tego, że miary ergodyczne, to to samo co punkty ekstremalne zbioru miar niezmienniczych). Zatem niech ν oznacza miarę ergodyczną na X_n'. Do oszacowania liczby $\nu(C)$ skorzystamy z twierdzenia ergodycznego. Ponieważ X_n jest podzbiorem X_n' miary dodatniej, istnieje $x \in X_n$ spełniający tezę twierdzenia ergodycznego dla miary μ_n i zbioru C, tzn. taki, że $\nu(C)$ jest równe średniej częstości odwiedzin orbity punktu x w cylindrze C. Ale ta częstość, to to samo co częstość, z jaką blok C występuje w ciągu symbolicznym, jakim jest x. Ponieważ C nie należy do $\Lambda^n(X)$, nie może on występować w żadnym bloku z rodziny $\Lambda^m(X)$, a skoro x jest konkatenacją takich bloków, to C może występować tylko na „złączeniach" bloków z rodziny $\Lambda^m(X)$, czyli pozycjach zaczepionych w miejscach poprzedzających takie złączenia (które występują okresowo co n) o co najwyżej m. To oznacza, że częstość jego występowania nie przekracza $\frac{m}{n}$, co jest, (przy ustalonym m i dowolnie dużym n) liczbą małą. To kończy cały dowód. $\square$

Instytut Matematyki i Informatyki, Politechnika Wrocławska
Wybrzeże Wyspiańskiego 27, 50-370 Wrocław
e-mail: downar@pwr.wroc.pl