Autoreferat

Michał D. Lemańczyk

2020/12/29

1 Abstrakt

Niech $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$, gdzie $X_i \in \mathcal{X}$ a $\mathcal{X}$ jest (mierzalną) przestrzenią stanów, będzie procesem stochastycznym. Niniejsza rozprawa doktorska koncentruje się na ***procesach czasów powrotu*** $\mathbf{R} = (R_i)_{i\in\mathbb{Z}}$ kolejnych powrotów X_i do A oraz ich roli zarówno w teorii prawdopodobieństwa, jak i w teorii ergodycznej. Przypomnijmy, że dla danego podzbioru $A \subset \mathcal{X}$ odpowiadający mu proces czasów powrotu jest zdefiniowany jako

$$R_i = \begin{cases} \inf\{j \geq 0 : X_j \in A\}, & i = 0, \\ \inf\{j > R_{i-1} : X_j \in A\}, & i \geq 1, \\ \sup\{j < R_{i+1} : X_j \in A\}, & i \leq -1. \end{cases} \tag{1.1}$$

Głównym rezultatem rozprawy w teorii prawdopodobieństwa jest dowód nierówności Bernsteina dla funkcjonałów addytywnych ogólnych, **niekoniecznie** silnie aperiodycznych, łańcuchów Markowa, co daje odpowiedź na pytanie sformułowane w pracy [3] (patrz [A3]). Dowodzimy również pewnej nowej wersji nierówności Bernsteina dla 1-zależnych procesów (klasa ta jest silnie związana z łańcuchami Markowa dzięki tzw. technice regeneracji). Główne rezultaty w teorii ergodycznej dotyczą dokładnych wzorów, bądź nierówności, związanych z entropią (ang. entropy rate) punktowego iloczynu procesów (patrz [A1]). Stają się one narzędziem do rozwiązania kilku otwartych problemów. Podajemy nowy, jawny wzór na ciśnienie topologiczne układów $\mathscr{B}$-wolnych oraz, w pewnych przypadkach, dowodzimy jedyności stanów równowagi dla układu wyznaczonego przez $\mathscr{B}$ (co rozszerza rezultaty o wewnętrznej ergodyczności udowodnione w [46, 33]). Odpowiadamy na pytanie postawione w [46] o braku własności Gibbsa dla miary o maksymalnej entropii (patrz [A2]). W końcu, odpowiadamy na kilka pytań dotyczących entropii układów $\mathscr{B}$-wolnych z pracy [33] (patrz [A1]).

Część rezultatów rozprawy jest nowa, pozostałe rezultaty pochodzą z następujących trzech artykułów:

[A1] J. Kułaga-Przymus and M.D. Lemańczyk. Entropy rate of product of independent processes. *Preprint: arXiv:2004.07648*, 2020.

[A2] J. Kułaga-Przymus and M.D. Lemańczyk. Hereditary subshifts whose measure of maximal entropy has no Gibbs property. *Ukaże się w Colloquium Mathematicum, arXiv:2004.07643*, 2020.

[A3] M.D. Lemańczyk. General Bernstein-like inequality for additive functionals of Markov chains. *Journal of Theoretical Probability*, 2020.

2 Definicje i oznaczenia

Dla uproszczenia rozważań przyjmujemy, że wszystkie zmienne losowe są zdefiniowane na wspólnej przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$, przy czym przez ***zmienną losową*** X rozumiemy dowolną funkcję mierzalną, przyjmującą wartości w mierzalnej **przestrzeni stanów** $\mathcal{X}$ (w skrócie $X \in \mathcal{X}$). ***(Dyskretnym) procesem*** **(stochastycznym)** $\mathbf{X} = (X_i)_{i\in T}$ (T jest zawsze równe albo $\mathbb{N}$, albo $\mathbb{Z}$) nazywamy rodzinę zmiennych losowych X_i przyjmujących wartości we **wspólnej** przestrzeni stanów $\mathcal{X}$.

Czasami zastępujemy (domyślną) miarę probabilistyczną $\mathbb{P}$ przez jej wersję warunkową $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot \cap A) / \mathbb{P}(A)$, gdzie $A \in \mathcal{F}$, przy czym $\mathbb{P}(A) > 0$. W szczególności $\mathbb{E}_A$ oznacza średnią warunkową liczoną względem $\mathbb{P}_A$.

Dla danego ciągu $\mathbf{x} \in \mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$ lub procesu $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ oraz skończonego zbioru indeksów $N \subset \mathbb{Z}$, gdzie $N = \{i_1, \ldots, i_k\}$, $i_1 < i_2 < \cdots < i_k$, określamy

$$x_N = (x_{i_1}, \ldots, x_{i_n}), \qquad X_N = (X_{i_1}, \ldots, X_{i_n})$$

(z oczywistymi modyfikacjami, gdy zbiór N jest nieskończony). Ponadto, dla dowolnych k, $l \in \mathbb{Z}$, wprowadzamy ***przedziały całkowitoliczbowe*** $[k,l] = \{k, k+1, \ldots, l\}$, $(-\infty, k] = \{\ldots, k-2, k-1, k\}$, $[l, \infty) = \{l, l+1, l+2, \ldots\}$, przy czym $[k,l] = \varnothing$, o ile $k > l$. Zatem, dla przykładu, $X_{[1,n]} = (X_1, \ldots, X_n)$.

Wszystkie standardowe operacje wykonywane w zbiorze liczb rzeczywistych $\mathbb{R}$ rozszerzamy „po współrzędnych" na ciągi i procesy. Dla przykładu $\mathbf{x} + \mathbf{y} = \mathbf{w}$, gdzie $w_i = x_i + y_i$, $\mathbf{X} \cdot \mathbf{Y} = (X_i Y_i)_{i\in T}$ lub $\mathbf{X} \leq \mathbf{Y}$, gdy $X_i \leq Y_i$ p.w. dla wszystkich $i \in T$.

Niech $\mathcal{X}$ będzie przestrzenią topologiczną (gdy $\mathcal{X}$ jest przestrzenią co najwyżej przeliczalną, to rozpatrujemy ją z topologią dyskretną). Przypomnijmy, że odwzorowanie $S\colon \mathcal{X}^{\mathbb{Z}} \to \mathcal{X}^{\mathbb{Z}}$ dane przez

$$S\,(x_i)_{i\in\mathbb{Z}} = (x_{i+1})_{i\in\mathbb{Z}}$$

nazywamy lewostronnym ***przesunięciem*** lub (częściej) **shiftem**. Parę $(\mathscr{X}, S)$ (lub krócej $\mathscr{X}$) nazywamy ***podshiftem***, gdy $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$ jest domknięty i S-niezmienniczy ($S\mathscr{X} \subset \mathscr{X}$). Zauważmy, że shift S jest ciągły. Ponadto, działa on na funkcjach poprzez składanie, $Sf = f \circ S$, oraz na miarach probabilistycznych poprzez branie obrazu miary, $S\mu(\mathbf{A}) = \mu(S^{-1}A)$. Ponadto, jeśli $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ jest procesem, to $S\mathbf{X} = (X_{i+1})_{i\in T}$ (czasami nazywamy $S\mathbf{X}$ ***procesem przesuniętym***). Przypomnijmy, że proces $\mathbf{X}$ nazywamy ***stacjonarnym***, gdy $\mathbf{X} \sim S\mathbf{X}$ (większość procesów, które będziemy rozpatrywać w kontekście dynamicznym, będzie stacjonarna). Proces $\mathbf{X}$ nazywamy ***ergodycznym***, gdy dla dowolnej funkcji $f \in L_1(\mathbb{P})$ zachodzi wersja **mocnego prawa wielkich liczb**, a mianowicie

$$\frac{1}{n}\sum_{i=0}^{n-1} f\left(S^i \mathbf{X}\right) \xrightarrow[L_1(\mathbb{P})]{\mathbb{P}\ a.s.} \mathbb{E} f\left(\mathbf{X}\right).$$

Proces $\mathbf{Y}$ nazywamy ***faktorem*** procesu $\mathbf{X}$, jeśli istnieje funkcja (mierzalna) $g\colon \mathcal{X}^T \to \mathcal{Y}$ taka, że

$$Y_n = S^n g(\mathbf{X}) = g(S^n \mathbf{X}), \qquad n \in \mathbb{Z}. \tag{2.1}$$

Jeśli g jest bi-mierzalną bijekcją, to mówimy, że $\mathbf{X}$ i $\mathbf{Y}$ są ***izomorficzne.*** Podkreślmy, że powyższe (trochę niestandardowe) definicje (ergodyczność, faktor, izomorfizm) są zgodne z odpowiednimi definicjami w układach dynamicznych. W końcu, dla danego podshiftu $\mathscr{X}$, przez $\mathcal{M}_{\mathscr{X}}$ oznaczamy ***zbiór (rozkładów) wszystkich procesów stacjonarnych*** $\mathbf{X}$ takich, że $\mathbb{P}\left(\mathbf{X} \in \mathscr{X}\right) = 1$. Ponadto $\mathcal{M}^e_{\mathscr{X}} \subset \mathcal{M}_{\mathscr{X}}$ oznacza ***podzbiór procesów ergodycznych***.

3 Nierówności koncentracyjne

Nierówności koncentracyjne odgrywają ważną rolę w wielu dziedzinach. Dla przykładu wymieńmy tylko statystykę (estymacja kowariancji, grupowanie (ang. clustering), grafy losowe, detekcję społecznościową, uzupełnianie macierzy i wiele innych) oraz dział algorytmów losowych (np. lemat Johnsona-Lindenstraussa, cięcia maksymalne dla grafów itd.). Jako świetne źródło zarówno ścisłej teorii, ale też intuicji i zastosowań polecamy książkę Vershynina (patrz [55]). Wyjaśnimy teraz podstawowe idee.

Ustalmy zmienną losową $X \in \mathbb{R}^d$ dla pewnego $d \in \mathbb{N}$ oraz zbiór $A \subset \mathbb{R}^d$ taki, że $\mathbb{P}\left(X \in A\right) \geq 1/2$. Dla danego $t \in \mathbb{R}_+$, niech $A_t = \{x \in \mathbb{R}^d \mid |A - x| \leq t\}$ bedzie t-otoczką A, a $A^c_t = \mathbb{R}^d \setminus A_t$ jej dopełnieniem. Pytamy teraz jak, szybko X koncentruje się wokół zbiorów o mierze $1/2$ lub, innymi słowy, jaki jest optymalny wybór funkcji $\alpha\colon \mathbb{R} \to \mathbb{R}_+$ takiej, że $\alpha(t) \xrightarrow{t\to\infty} 0$ oraz

$$\mathbb{P}\left(X \in A^c_t\right) \leq \alpha(t) \;\text{ zachodzi } \textbf{jednostajnie} \text{ względem wyboru } A. \tag{3.1}$$

Rozważmy dla przykładu $X \in \mathbb{R}^d$ jednostajnie rozłożoną na sferze jednostkowej $S = \{x \in \mathbb{R}^d \mid |x| = 1\}$. Okazuje się, że w tym przypadku możemy wziąć $\alpha(t) = \exp(-\frac{d-1}{2}t^2)$. Oznacza to, że jeśli wybierzemy dowolny zbiór $A \subset S$ taki, że $\mathbb{P}(X \in A) \geq 1/2$ (np. „czapeczkę"), to prawdopodobieństwo tego, że zaobserwujemy X w odległości $\sqrt{-\frac{2}{d-1}\ln c}$ od A jest co najmniej równe c. Innymi słowy możemy przewidzieć, gdzie pojawi się X!

Jednowymiarowym odpowiednikiem (3.1) jest znalezienie optymalnej funkcji $\alpha(t)$ spełniającej

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \alpha(t). \tag{3.2}$$

Ta nierówność oddaje naszą intuicję, że rzeczywiste zmienne losowe koncentrują się wokół ich średniej. Ponadto im szybciej $\alpha(t)$ maleje, tym lepiej $\mathbb{E}X$ aproksymuje X. Jeśli dla przykładu $\alpha(t) \sim \frac{1}{t^2}$, to z (relatywnie) dużym prawdopodobieństwem X może się mocno odchylać od swojej średniej. Jednakże jeśli $\alpha(t)$ maleje ekstremalnie szybko (np. wykładniczo, tzn. $\alpha(t) \sim c^{-t^a}$ dla pewnych $a \in \mathbb{R}_+$ oraz $c > 1$), to możemy oczekiwać, że dla danej niezależnej próbki z X, $X_1, X_2, \ldots, X_n$, z dużym prawdopodobieństwem, większość z X_i będzie się pojawiać blisko $\mathbb{E}X$.

Uwaga 3.1. Zauważmy, że dzięki nierówności Markowa można zawsze wziąć $\alpha(t) = \mathrm{Var}\,(X)/t^2$ w (3.2). W szczególności, jeśli t ma rząd ***standardowego odchylenia*** dla przykładu $t = \sqrt{2\mathrm{Var}\,(X)}$, to $\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq 1/2$, lub równoważnie, $\mathbb{P}(|X - \mathbb{E}X| < t) \geq 1/2$. Po pierwsze, ta obserwacja daje natychmiast naturalny, konstruktywny przykład zbioru A z (3.1). Po drugie, nieco nieformalnie możemy powiedzieć, że standardowe odchylenie $\sqrt{\mathrm{Var}\,(X)}$ jest dobrą jednostką mierzenia koncentracji miary. Dla przykładu praktyczna zasada w statystyce (ang. rule of thumb) mówi, że niemal cała masa zmiennej losowej o rozkładzie normalnym jest w odległości 2-3 takich „jednostek" od jej średniej.

Nas interesuje wersja (3.2) (zwana nierównością Bernsteina), w której zarówno $\alpha(t)$ jak i X mają specjalną postać. Przypomnijmy najpierw strukturę takiej nierówności w najprostszym przypadku, w którym zmienne są ograniczone oraz i.i.d.

Twierdzenie 3.2 (Klasyczna nierówność Bernsteina)**.** *Jeśli $(\xi_i)_i$ jest ciągiem i.i.d. rzeczywistych, scentrowanych zmiennych losowych takim, że $\|\xi_i\|_\infty \leq M$, to dla $\sigma^2 = \mathbb{E}\xi_i^2$ oraz dowolnego $t > 0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2n\sigma^2 + \frac{2}{3}Mt}\right). \tag{3.3}$$

Przeanalizujmy teraz (nieco nieformalnie) prawą stronę w (3.3). Zauważmy, że jako funkcja argumentu t prezentuje ona dwa typy zachowań: **dla "małych" t** dominuje ***zachowanie gaussowskie*** (rzędu $\exp(-ct^2)$ dla pewnego $c > 0$), a mianowicie,

$$2\exp\left(-\frac{t^2}{2n\sigma^2}\right). \tag{3.4}$$

Z kolei **dla "dużych" t *zachowanie wykładnicze*** (rzędu $\exp(-ct)$ dla pewnego $c > 0$) zaczyna odgrywać główną rolę, a konkretnie część $2\exp\left(-\frac{t}{\frac{2}{3}M}\right)$. Przez „duże" i „małe" t rozumiemy zakres tych t, dla których jeden z wyrazów, $2n\sigma^2$ lub $\frac{2}{3}Mt$, „mocno" dominuje nad drugim (w takim przypadku będziemy pisać trochę nieformalnie, że odpowiednio $2n\sigma^2 \ll \frac{2}{3}Mt$ lub $2n\sigma^2 \gg \frac{2}{3}Mt$).

„Część (3.4)" nierówności (3.3) jest zwykle nazywana ***częścią gaussowską nierówności Bernsteina***. Wyjaśnijmy teraz pochodzenie tej nazwy. Załóżmy, że η_i są niezależnymi zmiennymi losowymi gaussowskimi o zerowej średniej i wariancji σ^2, tzn. $\eta_i \sim \mathcal{N}(0, \sigma^2)$. Wiadomym jest, że zachodzą zależności

$$\sum_{i=1}^n \eta_i \sim \mathcal{N}(0, n\sigma^2) \qquad \text{oraz} \qquad \mathbb{P}\left(\left|\sum_{i=1}^n \eta_i\right| \geq t\right) \approx 2\exp\left(-\frac{t^2}{2n\sigma^2}\right).$$

Zatem jeśli $2n\sigma^2 \gg \frac{2}{3}Mt$, to twierdzenie 3.2 po prostu mówi, że najgorszy przypadek (jeśli chodzi o część gaussowską) powstaje, gdy $\xi_i = \eta_i$, innymi słowy,

$$\mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \geq t\right) \lesssim \mathbb{P}\left(\left|\sum_{i=1}^n \eta_i\right| \geq t\right). \tag{3.5}$$

Teraz omówimy problem optymalności w (3.3). Z centralnego twierdzenia granicznego (CTG),

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i \quad \Rightarrow \quad \mathcal{N}(0,\sigma^2),$$

co z grubsza można przepisać jako

$$\sum_{i=1}^{n}\xi_i \sim \mathcal{N}(0,n\sigma^2).$$

Ta obserwacja połączona z (3.5) pokazuje, że nierówność Bernsteina (3.3) jest w istocie rzeczy (asymptotycznie) optymalną nierównością koncentracyjną (przynajmniej, gdy mowa o części gaussowskiej). Do tego faktu będziemy odnosić się, mówiąc (nieco nieprecyzyjnie), że ***nierówność Bernsteina jest optymalna***.

Opuśćmy teraz świat niezależnych zmiennych losowych. Załóżmy, że $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ jest dowolnym rzeczywistym procesem takim, że $\sup_i \|X_i\|_\infty < \infty$, $\mathbb{E}X_i = 0$, dla którego chcemy udowodnić nierówność typu Bernsteina. Załóżmy dodatkowo, że CTG zachodzi dla $\mathbf{X}$, tzn.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i \quad \Rightarrow \quad \mathcal{N}(0,\sigma_\infty^2) \qquad \text{dla pewnej liczby } \sigma_\infty^2 \geq 0 \tag{3.6}$$

(σ_∞^2 jest nazywana ***wariancją asymptotyczną***). Naturalnym jest oczekiwanie, że (np. dla procesów silnie mieszających) σ_∞^2 można policzyć jako

$$\sigma_\infty^2 = \lim_{n\to\infty}\operatorname{Var}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\right) = \lim_{n\to\infty}\frac{1}{n}\operatorname{Var}\left(\sum_{i=1}^{n}X_i\right) \tag{3.7}$$

i następujący analog nierówności (3.3) powinien zachodzić dla odpowiedniego wyboru stałej $C \geq 0$:

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}X_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2n\sigma_\infty^2 + CMt}\right), \tag{3.8}$$

gdzie $M = \sup_i \|X_i\|_\infty$. Zauważmy, że aby zapewnic sobie, że (3.8) oddaje zachowanie wynikające z CTG (3.6) musimy położyć nacisk na to, aby C była rzędu o(n), gdy $n \to \infty$. Z drugiej strony, ogólnie, możemy sobie pozwolić na to, aby stała C zależała od pewnych własności procesu $\mathbf{X}$ (o ile $C = $ o(n)). Z takim przypadkiem się spotykamy dla łańcuchów Markowa, dla których C zależy od punktu startowego, funkcji przejścia i jest rzędu $\log n$ (ze szczegółami można się zapoznać w twierdzeniu 3.4). Podobnie jak w przypadku niezależnym będziemy mówili, że ***nierówność Bernsteina jest optymalna***, o ile nierówność (3.8) oddaje zachowanie CTG (3.6).

3.1 Nierówność Bernsteina dla m-zależnych zmiennych losowych

Przypomnijmy, że proces $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ jest m-zależny dla $m \in \mathbb{N}$, gdy

$$X_{(-\infty,k]} \amalg X_{[k+m+1,\infty)}, \qquad \forall_{k\in\mathbb{Z}}. \tag{3.9}$$

Klasa m-zależnych zmiennych losowych była badana w wielu pracach, np. [1, 38, 10, 27, 32, 13], jednak wydaje się, że problem optymalnej nierówności Bernsteina nie był do tej pory rozpatrywany (przynajmniej nie ma do tej pory wyników, które by się zbliżały do optymalnych). Wspomnijmy, że ciągi 1-zależne są mocno związane z łańcuchami Markowa, dzięki metodzie rozdzielania (ang. the splitting method) wymyślonej przez Athreyę i Neya, która rozbija łańcuch Markowa na bloki 1-zależne. W szczególności, każda wersja nierówności Bernsteina dla procesów 1-zależnych niemal natychmiastowo daje pewną nierówność dla łańcuchów Markowa (ale nie odwrotnie). Z drugiej strony istnieje przypuszczenie, że każdy 1-zależny proces stacjonarny jest w rzeczywistości 1-faktorem pewnego 1-zależnego łańcucha Markowa, co (gdyby było prawdziwe) ustalałoby ładną odpowiedniość pomiędzy tymi klasami procesów.

Niech $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ będzie stacjonarnym m-zależnym ciągiem scentrowanych i ograniczonych zmiennych losowych. W tym przypadku ***wariancja asymptotyczna*** (porównaj z (3.7)) jest dana przez

$$\sigma_\infty^2 = \lim_{n\to\infty} \frac{1}{n}\mathrm{Var}\,(X_1 + X_2 + \cdots + X_n) = \mathbb{E}X_0^2 + 2\sum_{i=1}^{m} \mathbb{E}X_0X_i. \tag{3.10}$$

Nasz główny wynik jest dosyć techniczny i dlatego zaprezentujemy tutaj jedynie wniosek dla 1-faktorów m-zależnych l-łańcuchów Markowa (które odgrywają ważną rolę w teorii ciągów m-zależnych). Przypominamy, że process $\mathbf{X} \in \mathcal{X}^{\mathbb{Z}}$ jest nazywany ***k-faktorem*** procesu $\mathbf{Y} \in \mathcal{Y}^{\mathbb{Z}}$ (porównaj z (2.1)), gdy istnieje funkcja $f\colon \mathcal{X}^k \to \mathcal{Y}$ taka, że

$$Y_i = f(X_i, X_{i+1}, \ldots, X_{i+k-1}).$$

Proces $\mathbf{Y}$ nazywa się ***l-łańcuchem Markowa*** gdy, dla dowolnego $k \in \mathbb{Z}$, znając teraźniejszość $X_{[k,k+l-1]}$, przyszłość $X_{(k+l,\infty]}$ jest niezależna od przeszłości $X_{(-\infty,k-1]}$.

Twierdzenie 3.3. *Niech $X_i = f(Y_i)$, gdzie f jest ograniczoną funkcja mierzalną oraz $\mathbf{Y} = (Y_i)$ jest stacjonarnym, m-zależnym l-łańcuchem Markowa. Wówczas*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| \geq t\right) \leq 2(m+1)\exp\left(-\frac{t^2}{c_{m,l}(n+m+l)\sigma_\infty^2 + d_m tM}\right), \tag{3.11}$$

gdzie $c_{m,l} = 2(1+\frac{3}{2\log(2m+2l)})^2(m+l)$, $d_{m,l} = \frac{4}{3}(1+\frac{3}{2\log(2m+2l)})(m+l)(m+1)$, $M = \|X_i\|_\infty$ oraz σ_∞^2 jest dana w (3.10).

Nowością w twierdzeniu 3.3 jest użycie wariancji asymptotycznej σ_∞^2 w gaussowskiej części nierówności Bernsteina zamiast używania $\sigma^2 = \mathbb{E}X_0^2$, tzn. wariancji pojedynczej zmiennej losowej. W istocie rzeczy jest łatwo otrzymać wersję nierówności (3.11), w której σ_∞^2 jest zastąpiona przez σ^2. Zauważmy jeszcze, że z nierówności Schwarza wynika, że zawsze mamy $\sigma_\infty^2 \leq (m+1)\sigma^2$. Z drugiej strony może się zdarzyć, że $\sigma_\infty^2 \ll \sigma^2$, tzn. σ_∞^2 może być dowolnie mała w porównaniu z σ^2 (w przypadku ekstremalnym, $\sigma_\infty^2 = 0 < \sigma^2$). Podsumowując, nasza nierówność Bernsteina (3.11) jest optymalna (z dokładnością do stałych zależących od l i m) i warto ją stosować jedynie, gdy $\sigma_\infty^2 \ll \sigma^2$.

3.2 Nierówność Bernsteina dla ogólnych łańcuchów Markowa

Załóżmy, że $\mathbf{X} = (X_n)_{n\in\mathbb{N}}$ jest łańcuchem Markowa określonym na przestrzeni probabilistycznej $(\Omega, \mathcal{F}, \mathbb{P})$, przyjmującym wartości w przestrzeni mierzalnej (przeliczalnie generowanej) $(\mathcal{X}, \mathcal{B})$, z funkcją przejścia $P\colon \mathcal{X}\times\mathcal{B} \to [0,1]$. Ponadto załóżmy, że $\mathbf{X}$ jest **ψ-nieredukowalny**, **aperiodyczny** i **dopuszcza jedyną niezmienniczą miarą probabilistyczną** π. Jak zwykle, dla rozkładu początkowego μ na $\mathcal{X}$, piszemy $\mathbb{P}_\mu(\mathbf{X} \in \cdot)$ dla rozkładu łańcucha, dla którego X_0 ma rozkład μ. Będziemy pisać $\mathbb{P}_x$ zamiast $\mathbb{P}_{\delta_x}$, gdzie δ_x oznacza miarę Diraca w punkcie x.

Mówimy, że $\mathbf{X}$ jest ***geometrycznie ergodyczny***, gdy istnieją liczba dodatnia $\rho < 1$ oraz funkcja rzeczywista $G\colon \mathcal{X} \to \mathbb{R}$ należąca do $L_1(\pi)$ takie, że dla dowolnego punktu startowego $x \in \mathcal{X}$ oraz $n \in \mathbb{N}$,

$$\|P^n(x,\cdot) - \pi(\cdot)\|_{TV} \leq G(x)\rho^n, \tag{3.12}$$

gdzie $\|\cdot\|_{TV}$ oznacza normę totalnego wahania miary, $P^n(\cdot,\cdot)$ zaś jest n-krokową funkcją przejścia łańcucha.

Nasz główny rezultat jest następujący:[1]

Twierdzenie 3.4. *Niech $\mathbf{X}$ będzie geometrycznie ergodycznym łańcuchem Markowa z przestrzenią stanów $\mathcal{X}$ oraz niech π będzie jego jedyną stacjonarną miarą probabilistyczną. Ponadto, niech $f\colon \mathcal{X} \to \mathbb{R}$ będzie ograniczoną funkcją mierzalną taką, że $\mathbb{E}_\pi f = 0$. Niech $x \in \mathcal{X}$. Wówczas możemy znaleźć stałe $K, \tau > 0$ zależące jedynie od x oraz funkcji przejścia $P(\cdot,\cdot)$ takie, że dla wszystkich $t > 0$,*

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \leq K\exp\left(-\frac{t^2}{32n\sigma_{Mrv}^2 + \tau t\|f\|_\infty \log n}\right),$$

[1] Poniżej, dla pewnej wygody, kładziemy $\log(\cdot) = \ln(\cdot \vee e)$, gdzie $\ln(\cdot)$ oznacza logarytm naturalny.

gdzie

$$\sigma^2_{Mrv} = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i=1}^{\infty} \mathrm{Cov}_\pi(f(X_0), f(X_i)) \tag{3.13}$$

oznacza wariancję asymptotyczną procesu $(f(X_i))_i$.

Uwaga 3.5. Stałe K i τ są jawne. Ponadto w rozprawie podajemy ogólniejsze wersje twierdzenia 3.4 (dla nieograniczonych zmiennych losowych).

Wyjaśnijmy teraz, w jaki sposób twierdzenie 3.4 można zobaczyć w szerszym kontekście. Przypomnijmy sobie klasyczną nierówność Bernsteina w przypadku ograniczonym z twierdzenia 3.2. Prawo wielkich liczb dla łańcuchów Markowa (patrz [11, 44, 43]) gwarantuje nam, że przy założeniach i oznaczeniach twierdzenia 3.4, sumy $\frac{1}{\sqrt{n}}\sum_{i=0}^{n-1} f(X_i)$ zbiegają wg rozkładu do rozkładu normalnego $\mathcal{N}(0, \sigma^2_{Mrv})$. Stąd nierówność otrzymana w twierdzeniu 3.4 jest odbiciem (z dokładnością do stałych) asymptotycznie normalnego zachowania sum $\frac{1}{\sqrt{n}}\sum f(X_i)$, podobnie do przypadku klasycznej nierówności Bernsteina w kontekście i.i.d. Dalej, wyraz $\log n$, który pojawia się w naszej nierówności, jest konieczny – jeśli dla wszystkich $t > 0$ mamy

$$\mathbb{P}_x\left(\left|\sum_{i=0}^{n-1} f(X_i)\right| > t\right) \le const \cdot \exp\left(-\frac{t^2}{const \cdot n\sigma^2 + const(x)\cdot a_n t\|f\|_\infty}\right) \tag{3.14}$$

dla pewnych $a_n = \mathrm{o(n)}$ oraz $\sigma \in \mathbb{R}$ ($const$ oznacza tu za każdym razem pewną stałą absolutną, podczas gdy $const(x)$ zależy jedynie od x i samego łańcucha Markowa), to musimy mieć $\sigma^2 \ge const \cdot \sigma^2_{Mrv}$. Co więcej, dla pewnych geometrycznie ergodycznych łańcuchów Markowa, a_n muszą wzrastać co najmniej logarytmicznie z n (patrz [2], Section 3.3).

Nierówności koncentracyjne dla łańcuchów Markowa i procesów zostały gruntownie zbadane w literaturze (patrz np. [2, 3, 5, 6, 12, 14, 18, 23, 28, 36, 35, 41, 40, 45, 50, 59]). Pewne wyniki są poświęcone koncentracji dla ogólnych funkcji łańcucha (zwykle takie rezultaty otrzymuje się przy różnych założeniach lipschitzowości, czy warunków ograniczenia różnic), inne idą w kierunku funkcjonałów addytywnych, które są również obiektem badań w rozprawie. Ogonowe nierówności dla funkcjonałów addytywnych są zwykle odpowiednikami nierówności Hoeffdinga czy Bernsteina. Te pierwsze nierówności nie biorą pod uwagę wariancji funkcjonału addytywnego i są wyrażane jedynie w terminach normy $\|f\|_\infty$. Często otrzymuje się je jako przypadki specjalne nierówności koncentracyjnych dla ogólnych funkcji (patrz np. [14, 45, 50]). Oszacowania typu Bernsteina postaci (3.14) są rozpatrywane np. w [2, 3, 5, 6, 12, 18, 23, 36, 35, 41, 40, 45, 59] i używają one zamienników wariancji σ^2, które niekoniecznie zgadzają się z graniczną wariancją σ^2_{Mrv}. W przypadku czasu ciągłego nierówności typu Bernsteina dla naturalnych odpowiedników funkcjonałów addytywnych (używające wariancji asymptotycznej) zostały otrzymane przy założeniu luki spektralnej lub warunków typu Lapunowa w [23, 36]. Dla dyskretnych łańcuchów Markowa nierówności otrzymane w [2, 3, 6, 12, 18] metodą regeneracji dają (3.14) (przy różnego rodzaju założeniach ergodyczności oraz z różnymi parametrami a_n) z σ^2, co zgadza się z σ^2_{Mrv} jedynie przy dodatkowym założeniu silnej aperiodyczności łańcucha. Z drugiej strony artykuły [41, 40, 50, 59] przynoszą ogólniejsze rezultaty, stosowalne dla ciągów zmiennych losowych (niekoniecznie Markowa), spełniających różne warunki mieszania. Zamienniki wariancji σ^2, które są używane w tych pracach, są bliskie wariancji asymptotycznej, ale na ogół nie są jej równe. Dla przykładu nierówność otrzymana w [41], która zachodzi w szczególności dla geometrycznie ergodycznych łańcuchów, używa (w naszych oznaczeniach) $\sigma^2 = \mathrm{Var}_\pi(f(X_0)) + 2\sum_{i=1}^{\infty}|\mathrm{Cov}_\pi(f(X_0), f(X_i))|$. Porównując to z (3.13), możemy zobaczyć, że $\sigma^2_{Mrv} \le \sigma^2$. W rzeczywistości możemy skonstruować przykłady, dla których iloraz tych dwóch wielkości jest dowolnie duży lub nawet $\sigma^2_{Mrv} = 0$ oraz $\sigma^2 > 0$. Pozycja [59] dostarcza nierówność dla jednostajnie geometrycznie ergodycznych procesów, uwzględniającą pewien ukryty zamiennik wariancji σ_n^2, który może być ograniczony z góry przez σ^2 na mocy [41] lub przez $\mathrm{Var}_\pi(f(X_0)) + C\|f\|_\infty \mathbb{E}_\pi|f(X_0)|$, gdzie C jest stałą zależącą od własności mieszających procesu. Dla ustalonego procesu, w sytuacji niezdegenerowanej, kiedy wariancja asymptotyczna jest niezerowa, może być ona zastąpiona przez σ_n^2, co jednak kosztuje wprowadzenie dodatkowych stałych multyplikatywnych, zależących od łańcucha i od f.

Według naszej najlepszej wiedzy twierdzenie 3.4 jest zatem pierwszą w literaturze nierównością ogonową, którą można otrzymać dla ogólnego geometrycznie ergodycznego (niekoniecznie silnie aperiodycznego) łańcuchu Markowa i która (z dokładnością do stałych uniwersalnych) jest odbiciem prawidłowego granicznego zachowania gaussowskiego funkcjonałów addytywnych. Problem otrzymania nierówności tego typu został oficjalnie postawiony w pracy [3]. Zauważmy, że badanie ilościowe naszych problemów, związane z CTG dla ogólnych aperiodycznych łańcuchów Markowa, wydaje się być istotnie trudniejsze niż dla łańcuchów, które są silnie aperiodyczne. Dla przykładu rezultaty typu optymalnej silnej aproksymacji są znane tylko w tym drugim przypadku [42].

4 Entropia i odzyskiwanie sygnału

Niech $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ będą zmiennymi losowymi, gdzie $\mathcal{X}$ jest przestrzenią dyskretną (co najwyżej przeliczalną). W tej części szeroko używamy pojęcia (warunkowej) entropii Shannona $\mathbf{H}(X)$ ($\mathbf{H}(X \mid Y)$), która intuicyjnie może być opisana jako miara (jest ona zawsze nieujemna) nieuporządkowania X . Innymi słowy, im większa entropia $\mathbf{H}(X)$, tym bardziej jest niepewny wynik obserwacji zmiennej X. Oczywiście ekstremalnym przypadkiem jest tu $\mathbf{H}(X) = 0$, równoważny z X będącą stałą (jesteśmy więc pewni realizacji zmiennej losowej X). Ponadto zachodzi $\mathbf{H}(X) \leq \log |\mathcal{X}|$, z równością (dla przestrzeni skończonych) wtedy i tylko wtedy, gdy X ma jednostajny rozkład na $\mathcal{X}$ (nie można przewidzieć wyniku X, wszystkie zdarzenia mają to samo prawdopodobieństwo). Krótko mówiąc, gdy $\mathbf{H}(X)$ rośnie, rozkład zmiennej X staje się coraz bardziej jednostajny. Formalnie,

$$\mathbf{H}(X) = -\sum_{x \in \mathcal{X}} \mathbb{P}(X = x) \log \mathbb{P}(X = x) \tag{4.1}$$

oraz

$$\mathbf{H}(X \mid Y) = \mathbb{E}_Y \mathbf{H}\left(X^{(Y)}\right), \qquad X^{(y)} \sim p_{X|Y}(\cdot|y),$$

gdzie $\mathbb{E}_Y$ oznacza całkowanie względem Y, a $p_{X|Y}$ oznacza regularny rozkład warunkowy zmiennej X, mając dane Y. Jeśli X jest dowolna (niekoniecznie dyskretna), to można rozszerzyć definicję entropii Shannona do

$$\mathbf{H}(X) = \sup_f \mathbf{H}(f(X)), \tag{4.2}$$

gdzie supremum bierzemy po zbiorze wszystkich (mierzalnych) funkcji zdefiniowanych na $\mathcal{X}$ oraz przyjmujących **skończenie wiele wartości**. Jednakże takie rozszerzenie ma swoje wady, z których najbardziej istotna pochodzi z faktu, że $\mathbf{H}(X)$ jest wtedy prawie zawsze nieskończona. Żeby to zobaczyć, weźmy proces stacjonarny $X = \mathbf{X} = (X_i)_{i \in \mathbb{Z}}$. Wówczas, na mocy definicji (4.2), dla dowolnego $n \in \mathbb{N}$,

$$\mathbf{H}(\mathbf{X}) \geq \mathbf{H}\left(X_{[1,n]}\right). \tag{4.3}$$

Jednakże jest intuicyjnie jasne (ze względu na stacjonarność procesu $\mathbf{X}$), że $\mathbf{H}\left(X_{[1,n]}\right)$ jest rzędu n (formalnie, odwzorowanie $n \to \mathbf{H}\left(X_{[1,n]}\right)$ jest podaddytywne) i stąd $\mathbf{H}(\mathbf{X}) \geq n$ dla dowolnego n, a więc $\mathbf{H}(\mathbf{X}) = \infty$. Aby ominąć ten problem, dla każdego procesu stacjonarnego $\mathbf{X}$ takiego, że X_0 jest dyskretna, wprowadzamy pojęcie ***entropii procesu*** $\mathbf{X}$ (ang. entropy rate),

$$\mathbf{H}(\mathbf{X}) = \lim_{n \to \infty} \frac{1}{n} \mathbf{H}\left(X_{[1,n]}\right) = \mathbf{H}\left(X_0 \mid X_{(-\infty,-1]}\right). \tag{4.4}$$

Zauważmy, że $\mathbf{H}(\mathbf{X}) \leq \mathbf{H}(X_0)$ i stąd entropia $\mathbf{H}(\mathbf{X})$ jest skończona, o ile $\mathbf{H}(X_0)$ jest skończona.

Uwaga 4.1. Używamy tego samego symbolu zarówno w (4.2), jak i (4.4), ale ponieważ istotna jest jedynie definicja (4.4), więc $\mathbf{H}(\mathbf{X})$ będzie zawsze dotyczyło (4.4).

Podobnie, dla dowolnego procesu stacjonarnego $(\mathbf{X}, \mathbf{Y}) = (X_i, Y_i)_{i \in \mathbb{Z}}$, gdzie X_0, Y_0 są dyskretne oraz $\mathbf{H}(X_0 \mid Y_0) < \infty$, definiujemy ***entropię relatywną*** jako

$$\mathbf{H}(\mathbf{X}|\mathbf{Y}) = \lim_{n \to \infty} \frac{1}{n} \mathbf{H}\left(X_{[1,n]} \mid Y_{[1,n]}\right) = \mathbf{H}\left(X_0 \mid X_{(-\infty,-1]}, \mathbf{Y}\right).$$

Przypomnijmy teraz pojęcie ***entropii topologicznej*** (która opisuje złożoność topologiczną przestrzeni). Dla podshiftu $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}}$ jego entropia topologiczna jest dana przez

$$\mathbf{H}_{\mathscr{X}} = \lim_{n\to\infty} \frac{1}{n} |\mathcal{L}(n)|, \tag{4.5}$$

gdzie $\mathcal{L}^{(n)}$ jest rodziną wszystkich słów długości n, która pojawiają się w $\mathscr{X}$. Można pokazać pewną wersję zasady wariacyjnej dla $\mathbf{H}_{\mathscr{X}}$, a mianowicie,

$$\mathbf{H}_{\mathscr{X}} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \mathbf{H}\left(\mathbf{X}\right). \tag{4.6}$$

Dzięki górnej półciągłości entropii procesu, powyższe supremum jest zawsze osiągane przez pewien proces $\mathbf{X}$. Każdy taki proces $\mathbf{X}$ (w istocie rzeczy chodzi o jego rozkład) jest nazywany ***miarą o maksymalnej entropii***. Jeśli istnieje dokładnie jedna miara o maksymalnej entropii, to mówimy, że $\mathscr{X}$ jest ***wewnętrznie ergodyczny***.

Wszystkie te „typy” entropii mają wiele interpretacji i zastosowań. Wspomnijmy kilka z nich, żeby uświadomić sobie, dlaczego to pojęcie jest tak fundamentalne i uniwersalne.

- (Struktura miar produktowych). Załóżmy, że zmienne losowe $X_i \in \mathcal{X}$ przyjmują skończenie wiele wartości oraz $\mathbf{X}$ jest ergodyczny. Rozważmy rozkład zmiennej $X = X_{[1,n]} \in \mathcal{X}^n$. Chcielibyśmy (w przybliżeniu) opisać nośnik zmiennej X (przez nośnik rozumiemy tutaj dowolny podzbiór S taki, że $\mathbb{P}\left(X \in S\right) = 1$). Jak wygląda rozkład μ zmiennej X na S? Okazuje się, że w zasadzie możemy wziąć za S zbiór **punktów typowych entropijnie**,

$$S = T_{ent,\varepsilon} = \{x \in \mathcal{X}^n \mid 2^{(n-\varepsilon)\mathbf{H}(\mathbf{X})} \le \mathbb{P}\left(X = x\right) \le 2^{(n+\varepsilon)\mathbf{H}(\mathbf{X})}\} \tag{4.7}$$

  (ten fakt jest znany pod nazwą ***asymptotycznego równomiernego rozłożenia***, patrz Theorem I.7.1 w [52]). Intuicyjnie, o X (dla dostatecznie dużych n) można myśleć jako o **równomiernym rozkładzie** na zbiorze S, który ma w przybliżeniu $2^{n\mathbf{H}(\mathbf{X})}$ elementów.

- (Kompresja sygnału). Przypuśćmy, że mamy dany proces i.i.d. (zwykle binarny) $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$ taki, że $X_i \in \mathcal{X}$ oraz zbiór $\mathcal{X}$ jest skończony. Dla danego $n \in \mathbb{N}$ chcemy skompresować $X = X_{[1,n]}$, tzn. znaleźć **koder** $f\colon \mathcal{X}^n \to \mathcal{Y}$ oraz **dekoder** $D\colon \mathcal{Y} \to \mathcal{X}^n$ w taki sposób, aby **prawdopodobieństwo błędu** $\mathbb{P}\left(D(E(X)) \ne X\right)$ było dowolnie małe. Oczywiście w tym problemie chcemy, aby $\mathcal{Y}$ był możliwie mały. Okazuje się, że (z grubsza) można dokonać kompresji X, o ile $|\mathcal{Y}| > n\mathbf{H}\left(X_1\right)$. Z drugiej strony nie można uzyskać kompresji X, gdy $|\mathcal{Y}| < n\mathbf{H}\left(X_1\right)$. W tym sensie $n\mathbf{H}\left(X_1\right)$ jest naturalną barierą dla problemu kompresji.

- (Powtarzanie wzorca). Niech $\mathbf{X}$ będzie ergodycznym procesem stacjonarnym. Dla danego $n \in \mathbb{N}$, jak długo musimy czekać, aby słowo $X_{[1,n]}$ powtórzyło się w całym ciągu $\mathbf{X} = (X_i)_{i\in\mathbb{N}}$? Można pokazać (patrz Theorem II.5.1 w [52]), że dla

$$R_n = \inf\{k \in \mathbb{N} \mid X_{[k+1,k+n]} = X_{[1,n]}\}$$

  mamy $\lim\limits_{n\to\infty} \frac{1}{n}\log_2 R_n = \mathbf{H}\left(\mathbf{X}\right)$, tzn. $R_n \sim 2^{n\mathbf{H}(\mathbf{X})}$ (o ile $\mathbf{H}\left(\mathbf{X}\right) > 0$). Ta asymptotyka ma swoją następującą intuicyjną konsekwencję: im mniejsza entropia procesu $\mathbf{X}$ (a więc, im mniej jest „skomplikowany” proces $\mathbf{X}$), „tym częściej” będziemy widzieć powtórki wzorców w $\mathbf{X}$. Dodajmy, że powyższe twierdzenie jest wybitnie nieefektywne dla procesów o zerowej entropii. W przypadku ekstremalnym, w którym $\mathbf{X}$ jest okresowy, tzn. $S^p\mathbf{X} = \mathbf{X}$ (co w oczywisty sposób implikuje $\mathbf{H}\left(\mathbf{X}\right) = 0$), stowarzyszony proces R_n jest ograniczony przez p jednostajnie dla wszystkich n!

- (Izomorfizm procesów). Słynne twierdzenie Ornsteina i Friedmana [20] mówi, że dwa układy, które są słabo Bernoulli i mają tę samą entropię, są izomorficzne. W terminach procesów oznacza to, że jeśli procesy $\mathbf{X}$ oraz $\mathbf{Y}$ są stacjonarne, słabo Bernoulli (równoważnie, β-mieszające) oraz $\mathbf{H}\left(\mathbf{X}\right) = \mathbf{H}\left(\mathbf{Y}\right)$, to są one izomorficzne. Zauważmy, że w szczególności otrzymujemy raczej nietrywialny wniosek: każdy mieszający łańcuch Markowa (na skończonej przestrzeni stanów) jest izomorficzny z pewnym procesem i.i.d.!

- (Fizyczne własności miar o maksymalnej entropii) Niech $\mathscr{X}$ będzie podshiftem, a $\mathbf{X}$ miarą o maksymalnej entropii. Okazuje się, że często $\mathbf{X}$ ma „dobre" dodatkowe własności fizyczne. Dla przykładu, jeśli proces $\mathbf{X}$ (tzn. jego rozkład) jest jedyny, to $\mathbf{X}$ musi być ergodyczny. Ponadto, raczej w sposób naturalny, oczekujemy, że $\mathbf{X}$ jest miarą Gibbsa, tzn.

$$c2^{-n\mathbf{H}_{\mathscr{X}}} \leq \mathbb{P}\left(X_{[1,n]} = x_{[1,n]}\right) \leq C2^{-n\mathbf{H}_{\mathscr{X}}} \tag{4.8}$$

dla pewnych stałych numerycznych c, C oraz wszystkich $x_{[1,n]} \in \mathcal{X}^n$, dla których powyższe prawdopodobieństwo jest dodatnie. (Por. (4.8) z (4.7).) Jeśli myślimy o innych konsekwencjach, to (4.8) implikuje np., że proces $\mathbf{X}$ musi być ***quasi-Bernoulli***, tzn.

$$k\mathbb{P}\left(X_{[1,n]} = x_{[1,n]}\right) \leq \mathbb{P}_{X_{[-m,0]}=x_{[-m,0]}}\left(X_{[1,n]} = x_{[1,n]}\right) \leq K\mathbb{P}\left(X_{[1,n]} = x_{[1,n]}\right) \tag{4.9}$$

dla pewnych stałych k, K zależących od c, C z (4.8).

Przypomnijmy, że jeśli rozpatrujemy jedynie ograniczenie dolne w (4.8), to otrzymujemy definicję tzw. ***własności Gibbsa***. Naturalna miara o maksymalnej entropii dla układów $\mathscr{B}$-wolnych nie spełnia własności Gibbsa. W szczególności otrzymujemy szeroką klasę podshiftów wewnętrznie ergodycznych, dla których miara o maksymalnej entropii **nie jest Gibbsa**.

4.1 Entropia iloczynu procesów

Niech $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ i $\mathbf{Y} = (Y_i)_{i\in\mathbb{Z}}$ będą rzeczywistymi procesami o skończenie wielu wartościach, przy czym proces $(\mathbf{X},\mathbf{Y}) = (X_i, Y_i)_i$ jest stacjonarny. Załóżmy dodatkowo, że $Y_i \in \{0,1\}$ dla $i \in \mathbb{Z}$ oraz $0 < \mathbb{P}(Y_0 = 1) < 1$. W tej części badamy entropię relatywną iloczynu procesów $\mathbf{X}$ and $\mathbf{Y}$, przy zadanym procesie $\mathbf{Y}$, a więc wielkość $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$.

Uwaga 4.2. Zauważmy, że jeśli $\mathbf{H}(\mathbf{Y}) = 0$, co zachodzi dla podshiftów $\mathscr{B}$-wolnych, których dynamika jest dla nas główną motywacją badania problemu entropii iloczynu, to $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) = \mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$. Jednakże wydaje się, że znacznie łatwiej jest poradzić sobie z relatywną wersją entropii procesu, $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$, niż z jej wersją bezwarunkową $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y})$.

W rozprawie odpowiadamy na następujące pytania postawione w nieco słabszej wersji w artykule [33] (patrz Question 1 tamże):

1. Czy istnieje ogólny wzór dla entropii $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y})$?
2. Czy zawsze mamy $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) > 0$, o ile $\mathbf{H}(\mathbf{X}) > 0$?
3. Czy możemy mieć $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) = \mathbf{H}(\mathbf{X}) > 0$?

Poniżej, dla dowolnego $A \subset \Omega$ takiego, że $\mathbb{P}(A) > 0$, będziemy pisać $\mathbf{H}_A(X)$ dla entropii Shannona na A, tzn. zastępujemy prawdopodobieństwo $\mathbb{P}$ jego wersją warunkową $\mathbb{P}_A(\cdot) = \mathbb{P}(\cdot \cap A)/\mathbb{P}(A)$ i rozważamy X na $(\Omega, \mathcal{F}, \mathbb{P}_A)$. Innymi słowy, po prostu zastępujemy $\mathbb{P}$ przez $\mathbb{P}_A$ w (4.1). Ta sama umowa obowiązuje w przypadku wersji warunkowej entropii Shannona.

Niech $\mathbf{R} = \mathbf{R}(\mathbf{Y}) = (R_i)_{i\in\mathbb{Z}}$ będzie procesem kolejnych ***czasów wizyt*** procesu $\mathbf{Y}$ do stanu 1 (aby uzyskać dokładny wzór na $\mathbf{R}$, należy zastąpić A przez $\{1\}$ w (1.1)). Proces ten jest dobrze zdefiniowany (tzn. $-\infty < R_i < \infty$ dla wszystkich i) na $\{Y_0 = 1\}$ (dzięki twierdzeniu Poincarégo o powracaniu) i na całej przestrzeni Ω, o ile $\mathbf{Y}$ jest ergodyczny.

Twierdzenie 4.3 (Odpowiedź na pytanie 1)**.** *Przy naszych założeniach,*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1)\,\mathbf{H}_{Y_0=1}\left(X_0 \mid X_{\{R_{-1},R_{-2},\cdots\}}, \mathbf{Y}\right). \tag{4.10}$$

Jeśli dodatkowo $\mathbf{X} \amalg \mathbf{Y}$*, to*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) = \mathbb{P}(Y_0 = 1)\,\mathbb{E}_{Y_0=1}\mathbf{H}\left(X_0 \mid X_{\{r_{-1},r_{-2},\cdots\}}\right)|_{r_{-i}=R_{-i}}. \tag{4.11}$$

Stosując łatwe oszacowania do (4.11), otrzymujemy następujący rezultat:

Wniosek 4.4 (odpowiedź na pytanie 2)**.** *Przy naszych założeniach, jeśli dodatkowo* $0 < \mathbf{H}(\mathbf{X})$ *oraz* $\mathbf{X} \amalg \mathbf{Y}$*, to* $\mathbb{P}(Y_0 = 1)\, \mathbf{H}(\mathbf{X}) \leq \mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) \leq \mathbb{P}(Y_0 = 1)\, \mathbf{H}(X_0)$*. W szczególności* $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y} \mid \mathbf{Y}) > 0$*, o ile* $\mathbf{H}(\mathbf{X}) > 0$.

Sytuacja zmienia się diametralnie, jeśli pozwolimy na pewną zależność procesów $\mathbf{X}$ i $\mathbf{Y}$ – wówczas może się zdarzyć, że

$$\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) = 0 = \mathbf{H}(\mathbf{Y}) < \mathbf{H}(\mathbf{X}). \tag{4.12}$$

Przykład 4.5 (odpowiedź na pytanie 2)**.** Weźmy proces stacjonarny $(\mathbf{X}, \mathbf{Y})$. Heurystycznie, gdy $\mathbf{X}$ jest przemnożony przez $\mathbf{Y}$, część procesu $\mathbf{X}$ na nośniku procesu $\mathbf{Y}$ pozostaje niezmieniona, gdy tymczasem część procesu $\mathbf{X}$ na nośniku procesu $\mathbf{1} - \mathbf{Y}$ znika. Zatem, jeśli $\mathbf{X}$ jest taki, że jego entropia jest zero na nośniku procesu $\mathbf{Y}$ i dodatnia na nośniku procesu $\mathbf{1} - \mathbf{Y}$, to otrzymujemy (4.12). Bardziej formalnie, dla dowolnego połączenia (stacjonarnego couplingu) $(\mathbf{Z}, \mathbf{W}, \mathbf{U})$, gdzie $\mathbf{U}$ jest binary, rozważmy

$$\mathbf{A} = \mathbf{U} \cdot \mathbf{Z} + (\mathbf{1} - \mathbf{U})\mathbf{W}.$$

Zauważmy, że $(\mathbf{A}, \mathbf{U})$ jest stacjonarny i każdy stacjonarny proces $(\mathbf{X}, \mathbf{Y})$ może być zrealizowany jako $(\mathbf{A}, \mathbf{U})$ poprzez wzięcie $\mathbf{W} = \mathbf{Z} = \mathbf{X}$, $\mathbf{U} = \mathbf{Y}$. I teraz, jeśli założymy, że $\mathbf{W} \amalg \mathbf{U} \amalg \mathbf{Z}$ oraz $\mathbf{H}(\mathbf{Z}) = \mathbf{H}(\mathbf{U}) = 0 < \mathbf{H}(\mathbf{W})$ i $\mathbf{U} \neq \mathbf{1}$, to ze względu na podaddytywność entropii mamy

$$\mathbf{H}(\mathbf{A} \cdot \mathbf{U}) = \mathbf{H}(\mathbf{U} \cdot \mathbf{Z}) \leq \mathbf{H}(\mathbf{U}) + \mathbf{H}(\mathbf{Z}) = 0$$

i z wniosku 4.4 wynika, że

$$\mathbf{H}(\mathbf{A}) = \mathbf{H}(\mathbf{A} \mid \mathbf{Z}, \mathbf{U}) = \mathbf{H}((\mathbf{1} - \mathbf{U})\mathbf{W} \mid \mathbf{Z}, \mathbf{U}) = \mathbf{H}((\mathbf{1} - \mathbf{U})\mathbf{W} \mid \mathbf{U}) \overset{\text{wniosek 4.4}}{\geq} \mathbb{P}(U_0 = 0)\, \mathbf{H}(\mathbf{W}) > 0$$

(ażeby otrzymać drugą równość, trzeba użyć definicji entropii warunkowej i zastosować zasadę reetykietowania).

W końcu podajemy przykład, w którym $\mathbf{X}$ oraz $\mathbf{Y}$ mogą być zrekonstruowane z procesu $\mathbf{X} \cdot \mathbf{Y}$ (w szczególności mamy $\mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{X} \cdot \mathbf{Y})$).

Przykład 4.6 (odpowiedź na pytanie 3)**.** Niech $(\xi_i)_{i \in \mathbb{Z}}$ będzie ciągiem i.i.d. zmiennych losowych takim, że $\mathbb{P}(\xi_0 = 0) = \mathbb{P}(\xi_0 = 1) = \frac{1}{2}$, niech $F \colon \{0,1\}^2 \to \{1,2,3,4\}$ będzie dowolną funkcją 1-1 (tzn. mamy do czynienia z reetykietowaniem) i połóżmy $X_i = F(\xi_i, \xi_{i+1})$. Dalej, niech $\mathbf{Y}$ będzie niezależny od procesu $\mathbf{X}$ oraz $\mathbf{Y} \sim \frac{1}{2}(\delta_{\mathbf{x}} + \delta_{S\mathbf{x}})$, gdzie $x_{2i} = 0 = 1 - x_{2i+1}$ dla $i \in \mathbb{Z}$. Ponieważ $\mathbf{X}$ jest łańcuchem Markowa i F jest 1-1, więc mamy $\mathbf{H}(\mathbf{X}) = \mathbf{H}(X_1 \mid X_0) = \mathbf{H}(\xi_1, \xi_2 \mid \xi_0, \xi_1) = \mathbf{H}(\xi_2 \mid \xi_0, \xi_1) = \mathbf{H}(\xi_2) = \log 2$. Ponadto $\mathbb{P}_{Y_0 = 1}(R_{-1} = 2) = 1$ i dlatego ze względu na (4.11), $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) = \frac{1}{2}\mathbf{H}(X_0 \mid X_{-2}) = \frac{1}{2}\mathbf{H}(X_0) = \log 2$, gdzie użyliśmy niezależności X_0 od X_{-2}. Podsumowując,

$$\mathbf{H}(\mathbf{X}) = \mathbf{H}(\mathbf{X} \cdot \mathbf{Y}). \tag{4.13}$$

Możemy nawet wzmocnić (4.13). Zauważmy, że ponieważ F nie osiąga wartości 0, więc możemy odzyskać **zarówno** $\mathbf{X}$, jak i $\mathbf{Y}$ z $\mathbf{X} \cdot \mathbf{Y}$. Rzeczywiście, każda zerowa współrzędna w $\mathbf{X} \cdot \mathbf{Y}$ natychmiast wyznacza $\mathbf{Y}$. Ponadto, z samej definicji procesu $\boldsymbol{\xi}$, $\boldsymbol{\xi}$ (i stąd $\mathbf{X}$) można zrekonstruować, o ile znamy nieparzyste, bądź parzyste współrzędne w $\mathbf{X}$. Stąd, jeśli tylko dostaniemy proces $\mathbf{Y}$ z procesu $\mathbf{X} \cdot \mathbf{Y}$, to łatwo znaleźć proces $\mathbf{X}$.

4.2 Odzyskiwanie sygnału

Przypomnijmy słynny problem Furstenberga (z 1967 r.) odszumiania (zaszumionego) sygnału. Podstawowe pytanie zadane w [21] brzmiało, kiedy można odzyskać sygnał $\mathbf{Z}$ z perturbacji postaci $\mathbf{Z} + \mathbf{W}$, gdzie $\mathbf{Z}$ i $\mathbf{W}$ są rzeczywistymi procesami stacjonarnymi. Aby rozwiązać ten problem, Furstenberg wprowadził pojęcie **absolutnej rozłączności procesów** (znacznie silniejsze niż zwykła niezależność) i udowodnił, że ta własność wystarcza do odszumienia procesu $\mathbf{Z}$ z $\mathbf{Z} + \mathbf{W}$ (dodatkowe założenie całkowalności użyte w [21], jak wykazał znacznie później Garbit [24], okazało się niepotrzebne). W tym samym duchu można interpretować $\mathbf{X} \cdot \mathbf{Y}$ jako stracony sygnał (przypomnijmy, że $Y_i \in \{0,1\}$) i zapytać, czy mimo wszystko można (bądź nie można) odzyskać $\mathbf{X}$ z $\mathbf{X} \cdot \mathbf{Y}$. Oczywiście, jeśli $\mathbf{H}(\mathbf{X} \cdot \mathbf{Y}) < \mathbf{H}(\mathbf{X})$, to nie ma szans na odzyskanie procesu $\mathbf{X}$, stąd naturalnym jest przeformułowanie pytania 3 w następujący sposób:

3'. Czy istnieje naturalne kryterium na $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) < \mathbf{H}(\mathbf{X})$ przy założeniu, że $\mathbf{H}(\mathbf{X}) > 0$?

Uwaga 4.7. Zauważmy, że gdyby przestrzenie stanów procesów $\mathbf{X}$ i $\mathbf{Y}$ były zawarte w zbiorze dodatnich liczb rzeczywistych, to nasz problem zredukowałby się do problemu odszumiania Furstenberga $\mathbf{Z}+\mathbf{W}$, gdzie $\mathbf{Z} = \log \mathbf{X}$ oraz $\mathbf{W} = \log \mathbf{Y}$. W tym sensie jest ważne, że zmienne losowe Y_i przyjmują wartość zero.

Uwaga 4.8. W klasycznym problemie Furstenberga sygnał $\mathbf{X}$ jest reprezentowany przez proces o zerowej entropii (tzn. proces deterministyczny). W naszym problemie jest odwrotnie. To proces $\mathbf{X}$ nie jest deterministyczny i jest on zaburzany przez proces deterministyczny $\mathbf{Y}$, więc interpretacja poprzez klasyczną sytuację sumy nie jest zbyt precyzyjna. Niemniej, jest jasna analogia problemu, a problem ubytku entropii jest sam w sobie interesujący (patrz również następna uwaga).

Uwaga 4.9. Podobny (a tak naprawdę znacznie ogólniejszy) problem odszumiania sygnału był badany przez Furtenberga, Peresa i Weissa w [22]. Niech $\mathbf{X}^{(i)} = \left(X_j^{(i)}\right)_{j\in\mathbb{Z}}$, gdzie $i \in \mathbb{N}$, będzie rodziną procesów i niech $\mathbf{U}$ będzie procesem przyjmującym wartości w $\mathbb{N}$. Zakładamy, że wszystkie te procesy są wspólnie stacjonarne. Kładziemy

$$\mathbf{X}^{(\mathbf{U})} = \left(X_i^{(U_i)}\right)_{i\in\mathbb{Z}} \tag{4.14}$$

(mówiąc nieformalnie, proces $\mathbf{U}$ określa wybór z rodziny procesów). Aby pokazać możliwość odszumienia procesu $\mathbf{U}$ z $\mathbf{X}^{(\mathbf{U})}$, autorzy pracy [22] wprowadzili pojęcie **podwójnej rozłączności** procesów. Mówimy, że proces $\mathbf{A}$ jest ***podwójnie rozłączny (PR) z*** procesem $\mathbf{B}$, gdy każde samopołączenie procesu $\mathbf{A}$ jest absolutnie rozłączne z procesem $\mathbf{B}$. Innymi słowy, jeśli $(\mathbf{A}', \mathbf{A}'', \mathbf{B}')$ jest procesem stacjonarnym takim, że $\mathbf{A}', \mathbf{A}'' \sim \mathbf{A}$ oraz $\mathbf{B}' \sim \mathbf{B}$, to $(\mathbf{A}', \mathbf{A}'') \amalg \mathbf{B}'$. Najbardziej podstawowy przykład PR procesów otrzymujemy rozpatrując $\mathbf{A}$ deterministyczny (wówczas każde samopołączenie jest również o zerowej entropii), a $\mathbf{B}$ ma trywialną σ-algebrę ogonową (zauważmy jednocześnie, że jeśli $\mathbf{A}$ jest PR z procesem $\mathbf{B}$, to **koniecznie** $\mathbf{H}(\mathbf{A}) = 0$ oraz $\mathbf{B}$ jest ergodyczny). Główny rezultat pracy [22] można podsumować (z grubsza) następująco. Przypuśćmy, że $\mathbf{X}^{(i)}$ dla $i \in \mathbb{N}$ oraz $\mathbf{U}$ są wspólnie stacjonarne. Jeśli $\mathbf{U}$ jest PR z każdym z procesów $\mathbf{X}^{(i)}$ dla $i \in \mathbb{N}$, to można odzyskać proces $\mathbf{U}$ z procesu $\mathbf{X}^{(\mathbf{U})}$.

Wyjaśnijmy, jak można użyć tego twierdzenia w naszym kontekście mnożenia procesów i pytania 3'. Rozważmy dwa procesy $\mathbf{X}^{(i)}$, dla $i \in \{0,1\}$, gdzie

$$X_j^{(i)} = iX_j \tag{4.15}$$

i weźmy $\mathbf{U} = \mathbf{Y}$. Wówczas $\mathbf{X}^{(\mathbf{U})} = \mathbf{X}\cdot\mathbf{Y}$ i twierdzenie powyższe mówi nam, że możemy odzyskać $\mathbf{Y}$ z procesu $\mathbf{X}\cdot\mathbf{Y}$, o ile $\mathbf{Y}$ jest PR z $\mathbf{X}$. Zauważmy, że ponieważ zakładamy, że $\mathbf{H}(\mathbf{X}) > 0$, więc nie możemy zamienić rolami $\mathbf{X}$ i $\mathbf{Y}$ i stąd problem, który rozważamy w rozprawie (np. ubytku entropii dla iloczynu), jest uzupełnieniem problemu odszumiania badanego w [22].

Uwaga 4.10. Byłoby interesujące znaleźć ogólne kryterium gwarantujące, że $\mathbf{X}$ może być odzyskany z iloczynu $\mathbf{X}\cdot\mathbf{Y}$ (jak to widzieliśmy w przykładzie 4.6).

Jak poprzednio, niech $\mathbf{X} = (X_i)_{i\in\mathbb{Z}}$ oraz $\mathbf{Y} = (Y_i)_{i\in\mathbb{Z}}$ będą rzeczywistymi procesami o skończenie wielu wartościach takimi, że $(\mathbf{X}, \mathbf{Y}) = (X_i, Y_i)_i$ jest stacjonarny, przy czym $Y_i \in \{0,1\}$ dla $i \in \mathbb{Z}$ oraz $0 < \mathbb{P}(Y_0 = 1) < 1$. W tej części zakładamy dodatkowo, że $\mathbf{X}$ i $\mathbf{Y}$ są niezależne, $\mathbf{H}(\mathbf{Y}) = 0 < \mathbf{H}(\mathbf{X})$ oraz $\mathbf{Y}$ jest ergodyczny. Nasz główny rezultat podaje kryterium na spadek entropii.

Twierdzenie 4.11 (odpowiedź na pytanie 3'.). *Przy naszych założeniach,*

$$\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) \le \mathbf{H}(\mathbf{X}) - \mathbb{P}(Y_0 = 1)^2\, \mathbb{E}_{Y_0=1}\mathbf{H}\left(X_{[1,r_1)} \mid X_{(-\infty,0]}, X_{\{r_1,r_2,\dots\}}\right)|_{r_i=R_i}.$$

W szczególności, $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y} \mid \mathbf{Y}) \le \mathbf{H}(\mathbf{X}) - \mathbb{P}(Y_0 = 1)^2 \sum_{k=1}^{\infty} \mathbb{P}_{Y_0=1}(R_1 = k)\, \mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0]\cup[k,\infty)}\right).$

Zatem, aby sprawdzić, czy $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) < \mathbf{H}(\mathbf{X})$, wystarcza sprawdzić, kiedy powyższa suma (formalnie) nieskończona jest dodatnia. Niech $\mathcal{T}_{double} = \bigcap_{i\geq 0}\sigma\left(X_{(-\infty,-i]}, X_{[i,\infty)}\right)$ oznacza ***podwójną ogonową σ-algebrę*** (dla $\mathbf{X}$) i zauważmy, że dla dostatecznie dużych (parzystych) $k\in\mathbb{N}$ mamy

$$\begin{aligned}\mathbf{H}\left(X_{[1,k)} \mid X_{(-\infty,0]\cup[k,\infty)}\right) &\geq \mathbf{H}\left(X_{k/2} \mid X_{(-\infty,0]\cup[k,\infty)}\right)\\ &= \mathbf{H}\left(X_0 \mid X_{(-\infty,-k/2]\cup[k/2,\infty)}\right) \approx \mathbf{H}\left(X_0 \mid \mathcal{T}_{double}\right).\end{aligned}$$

Stąd, ponieważ każdy proces słabo Bernoulli (równoważnie, β-mieszający) musi mieć trywialną podwójną σ-algebrę ogonową $\mathcal{T}_{double}$ (w szczególności $\mathbf{H}(X_0 \mid \mathcal{T}_{double}) = \mathbf{H}(X_0)$), więc natychmiast otrzymujemy następujący wniosek:

Wniosek 4.12. *Oprócz naszych dotychczasowych założeń przyjmijmy dodatkowo, że* $\mathbf{X}\amalg\mathbf{Y}$, $\mathbf{H}(\mathbf{X}) > 0$, $\mathbf{X}$ *jest słabo Bernoulli oraz* $\mathbb{P}_{Y_0=1}(R_1 = k) > 0$ *dla nieskończenie wielu* $k\in\mathbb{N}$. *Wówczas* $\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) < \mathbf{H}(\mathbf{X})$.

5 Ciśnienie topologiczne i miary Gibbsa

Ciśnienie topologiczne Ciśnienie topologiczne jest naturalnym uogólnieniem pojęcia entropii topologicznej. Ponadto jest ono jednym z głównych pojęć formalizmu termodynamicznego, który z kolei odgrywa istotną rolę w rozwoju teorii układów dynamicznych. Dodajmy, że zasada wariacyjna dla ciśnienia topologicznego jest naturalnym pomostem łączącym dynamikę topologiczną i miarową. Związki pomiędzy tymi dynamikami są często wykorzystywane w zastosowaniach, m.in. badaniach wykładników Lapunowa, wymiarów fraktalnych, widm multi-fraktalnych, naturalnych miar niezmienniczych (np. miar o maksymalnej entropii, fizycznych, SRB, o zerowej temperaturze, czy miar o maksymalnym wymiarze), i zbiorów obrotu, patrz [7, 30, 39, 47, 49] oraz bibliografia tamże. Dodajmy, że ciśnienie topologiczne ma zastosowanie zarówno w matematyce, jak i pokrewnych dziedzinach nauki, takich jak fizyka statystyczna, czy biologia matematyczna, patrz [4, 16, 17, 15, 25, 54] oraz bibliografia tamże. Przegląd zagadnień związanych z entropią i ciśnieniem czytelnik może znaleźć w książkach i artykułach przeglądowych [7, 29, 39, 49, 56]. Zauważmy, że istnieje gałąź matematyki, która zajmuje się obliczalnością ciśnienia topologicznego. W skrócie chodzi o odpowiedź na następujące pytanie (kluczowe z punktu widzenia zastosowań): dla jakich podshiftów $\mathscr{X}$ istnieje program komputerowy, który jest w stanie znaleźć $\mathscr{P}_{\mathscr{X}}$ (patrz [9] oraz bibliografia tamże).

Niech $(\mathscr{X}, S)$ będzie podshiftem, $\varphi\colon \mathscr{X}\to\mathbb{R}$ zaś funkcją **ciągłą** (zwaną ***potencjałem***). Potencjał nazywamy ***lokalnym***, jeśli zależy on tylko od skończenie wielu współrzędnych. Dla danego podshiftu $\mathscr{X}\subset\mathcal{X}^{\mathbb{Z}}$ ***ciśnienie topologiczne, $\mathscr{P}_{\mathscr{X},\varphi}$, podshiftu $\mathscr{X}$*** jest dane przez

$$\mathscr{P}_{\mathscr{X},\varphi} = \lim_{n\to\infty}\frac{1}{n}\log_2\sum_{\mathbf{A}\in\mathcal{L}^{(n)}} 2^{\sup_{\mathbf{A}} S_n\varphi(\mathbf{x})}, \tag{5.1}$$

gdzie $S_n = \sum_{i=1}^{n} S^i$, a $\mathcal{L}^{(n)}$ oznacza zbiór słów długości n, które pojawiają się w $\mathscr{X}$. Następująca ***zasada wariacyjna (VP)*** jest dobrze znana (patrz [57], Theorem 4.1):

$$\mathscr{P}_{\mathscr{X},\varphi} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}}\left[\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})\right]. \tag{5.2}$$

Jest jasne, że jeśli weźmiemy $\varphi = 0$, to (5.1) i (5.2) redukują się do ich odpowiedników dla entropii topologicznej (przypomnijmy (4.5) oraz (4.6)). Zauważmy, że odwzorowanie $\mathbf{X}\to\mathbf{H}(\mathbf{X}) + \mathbb{E}\varphi(\mathbf{X})$ jest górnie półciągłe (w słabej topologii) i stąd wynika, że zawsze istnieje pewien optymalny proces $\mathbf{X}$ osiągający supremum w (5.2). Proces ten (a raczej jego rozkład) nazywamy ***stanem równowagi***. Jeśli miara ta jest jedyna, to musi być ona ergodyczna (używamy tego samego argumentu co w Theorem 14.25, point 2 w [26]). Ponadto oczekuje się, że zwykle stany równowagi są miarami ***Gibbsa***, tzn., dla pewnych stałych numerycznych c, C oraz dowolnych $\mathbf{x}\in\mathscr{X}$, $n\in\mathbb{N}$, $\mathbf{y}\in[x_{[1,n]}]$ ($[\cdot]$ oznacza zbiór cylindryczny),

$$c \leq \frac{\mathbb{P}\left(X_{[1,n]} = x_{[1,n]}\right)}{2^{\sum_{k=0}^{n-1}\varphi(S^k\mathbf{y}) - n\mathscr{P}}} \leq C. \tag{5.3}$$

Można pokazać, że powyższa stała $\mathscr{P}$ musi być równa ciśnieniu topologicznemu podshiftu $\mathscr{X}$ (z domyślnym potencjałem φ). W szczególności, jeśli $\varphi = 0$, to wracamy do (4.8). Dodatkowo można pokazać, że każda miara Gibbsa musi być quasi-Bernoulli (przypominamy (4.9)).

Dodajmy ponadto, że pojęcie miary Gibbsa pochodzi z fizyki statystycznej [48, 34] i odpowiada ono idei stanu równowagi skomplikowanych układów fizycznych. Miary te okazały się być również interesującym obiektem badań z punktu widzenia dynamiki i odgrywają one ważną rolę w teorii ergodycznej (patrz np. [8, 53]).

Układy $\mathscr{B}$-wolne Badanie układów $\mathscr{B}$-wolnych częściowo wynika z zainteresowania własnościami **funkcji Möbiusa** $\boldsymbol{\mu}\colon \mathbb{Z} \to \{-1, 0, 1\}$. Przypomnijmy, że liczba całkowita $z \in \mathbb{Z}$ nazywa się ***bezkwadratową***, gdy nie dzieli się ona przez żaden kwadrat liczby pierwszej oraz, że $\boldsymbol{\mu}$ jest dana przez

$$\boldsymbol{\mu}(z) = \begin{cases} 1, & \text{gdy } z \text{ jest bezkwadratowa z parzystą liczbą czynników pierwszych,} \\ -1, & \text{gdy } z \text{ jest bezkwadratowa z nieparzystą liczbą czynników pierwszych,} \\ 0, & \text{w przeciwnym wypadku.} \end{cases} \tag{5.4}$$

Arytmetyczna funkcja Möbiusa jest jedną z najbardziej fundamentalnych funkcji w teorii liczb. Wspomnijmy jedynie kilka ważnych faktów związanych z $\boldsymbol{\mu}$:

- Jest ona związana z funkcją Riemanna poprzez szereg Dirichleta, $\frac{1}{\zeta(s)} = \sum_{n\geq 1} \boldsymbol{\mu}(n)/n^s$.
- Twierdzenie o rozkładzie liczb pierwszych (liczba liczb pierwszych pomiędzy 0 i $n \in \mathbb{N}$ jest w przybliżeniu równa $n/\ln n$) jest równoważne warunkowi $\sum_{n\leq N} \boldsymbol{\mu}(n) = \mathrm{o}(N)$.
- Gdybyśmy mogli wzmocnić warunek powyżej do stwierdzenia, że dla dowolnego $\varepsilon > 0$ (wiadomo już, że $\varepsilon = 0$ nie działa) zachodzi $\sum_{n\leq N} \boldsymbol{\mu}(n) = \mathrm{O}(N^{\frac{1}{2}+\varepsilon})$ (a więc, w jakimś sensie, ciekawi nas, czy $\boldsymbol{\mu}$ zachowuje się jak losowy ciąg znaków), to otrzymalibyśmy dowód hipotezy Riemanna (tak naprawdę powyższy warunek prędkości zbieżności jest równoważny hipotezie Riemanna).

W świetle ostatniego punktu zrozumienie asymptotyki funkcji $\boldsymbol{\mu}$ jest jednym z największych wyzwań matematyki. Naturalną koleją rzeczy jest studiowanie trochę prostszego obiektu, mianowicie kwadratu funkcji Möbiusa $\boldsymbol{\mu}^2$. Zauważmy, że $\boldsymbol{\mu}^2 \in \{0, 1\}^{\mathbb{Z}}$ oraz $\boldsymbol{\mu}^2(n) = 1$ wtedy i tylko wtedy, gdy n jest bezkwadratowa, co motywuje badanie **zbioru liczb bezkwadratowych**, $\mathcal{F}_{\boldsymbol{\mu}^2} = \{n \in \mathbb{Z} \mid \boldsymbol{\mu}^2(n) = 1\}$. Zbiory tej postaci były już rozpatrywane, z punktu widzenia teorii liczb, w latach 30 ub. wieku przez Besicovitcha, Erdösa, Chowlę i innych. Podejście dynamiczne zostało zapoczątkowane przez P. Sarnaka w jego nowatorskich wykładach [51] o losowości funkcji Möbiusa. Mówiąc dokładniej, Sarnak analizował podshift wyznaczony przez $\boldsymbol{\mu}^2$,

$$\mathscr{X}_{\boldsymbol{\eta}} = \overline{\{S^n \boldsymbol{\eta} \mid n \in \mathbb{Z}\}} \subset \{0, 1\}^{\mathbb{Z}}, \tag{5.5}$$

gdzie $\boldsymbol{\eta} = \mathbb{1}_{\mathcal{F}_{\boldsymbol{\mu}^2}} \in \{0, 1\}^{\mathbb{Z}}$ jest funkcją charakterystyczną zbioru liczb bezkwadratowych. Krótko mówiąc, kodujemy zbiór $\mathcal{F}_{\boldsymbol{\mu}^2}$ funkcją $\boldsymbol{\eta}$ i rozpatrujemy najmniejszy podzbiór S-niezmienniczy $\mathscr{X}_{\boldsymbol{\eta}}$ zawierający $\boldsymbol{\eta}$. Można łatwo rozszerzyć tę konstrukcję do następującej: startujemy z podzbioru liczb naturalnych $\mathscr{B} \subset \mathbb{N} \setminus \{1\}$ i określamy ***zbiór liczb $\mathscr{B}$-wolnych*** $\mathcal{F}_{\mathscr{B}} = \mathbb{Z} \setminus \bigcup_{b\in\mathscr{B}} b\mathbb{Z}$, który następnie wyrażamy w $\{0, 1\}^{\mathbb{Z}}$, używając $\boldsymbol{\eta} = \mathbb{1}_{\mathcal{F}_{\mathscr{B}}}$ (oczywiście $\mathcal{F}_{\mathscr{P}^2} = \mathcal{F}_{\boldsymbol{\mu}^2}$, gdzie $\mathscr{P}^2$ jest zbiorem kwadratów liczb pierwszych). Podshift $\mathscr{B}$-wolny $\mathscr{X}_{\boldsymbol{\eta}}$ jest następnie dany przez (5.5).

Powiedzmy teraz kilka słów o ogólnych podshiftach ***$\mathscr{B}$-wolnych*** $\mathscr{X}_{\boldsymbol{\eta}}$. Dla prostoty ograniczymy się do tzw. przypadku Erdösa, w którym zbiór $\mathscr{B}$ składa się z liczb parami względnie pierwszych oraz szereg ich odwrotności jest zbieżny, $\sum_{b\in\mathscr{B}} 1/b < \infty$ (zauważmy, że to obejmuje przypadek bezkwadratowy). Dodajmy, że przy tych założeniach podshift $\mathscr{X}_{\boldsymbol{\eta}}$ jest ***dziedziczny***, tzn. jest on zamknięty na zastępowanie jedynek przez zera na dowolnej współrzędnej punktu $\mathbf{x} \in \mathscr{X}_{\boldsymbol{\eta}}$. Ponadto $\boldsymbol{\eta}$ **jest punktem generującym dla *miary Mirsky'ego*** $\nu_{\boldsymbol{\eta}} \in \mathcal{M}_{\mathscr{X}_{\boldsymbol{\eta}}}$. Innymi słowy,

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{S^i \boldsymbol{\eta}} \quad \Rightarrow \quad \nu_{\boldsymbol{\eta}}. \tag{5.6}$$

Niech $\mathbf{Y} \sim \nu_{\boldsymbol{\eta}}$. Intuicyjnie, (5.6) oznacza, że $\nu_{\boldsymbol{\eta}}$ **liczy frekwencję bloków** pojawiających się na $\boldsymbol{\eta}$. Dla przykładu $\nu_{\boldsymbol{\eta}}(1) = \mathbb{P}(Y_0 = 1) = \lim_n \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\eta_i=1}$. Oprócz naturalności tej definicji okazuje się ponadto, że $\nu_{\boldsymbol{\eta}}$ ma wiele interesujących własności. Po pierwsze podshift $\mathbf{Y}$ jest ergodyczny oraz $\mathbf{H}(\mathbf{Y}) = 0$. Po drugie $\mathbf{Y}$ ma **największą gęstość jedynek** pośród wszystkich procesów stacjonarnych na $\mathscr{X}_{\boldsymbol{\eta}}$, tzn. $\mathbb{P}(Y_0 = 1) = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}_{\boldsymbol{\eta}}}} \mathbb{P}(X_0 = 1)$. Ponadto **miara o maksymalnej entropii** dla $\mathscr{X}_{\boldsymbol{\eta}}$ jest jedyna i jest ona dana przez $\mathbf{B} \cdot \mathbf{Y}$, gdzie $\mathbf{B}$ jest i.i.d. $\mathbb{P}(B_0 = 0) = 1 - \mathbb{P}(B_0 = 1) = 1/2$ oraz $\mathbf{B} \amalg \mathbf{Y}$. W końcu miara Mirsky'ego wyjaśnia strukturę elementów w $\mathcal{M}_{\mathscr{X}_{\eta}}$. Mówiąc dokładniej, okazuje się, że każdy proces stacjonarny $\mathbf{Z} \in \mathscr{X}_{\boldsymbol{\eta}}$ (tzn. jego rozkład) jest postaci $\mathbf{X} \cdot \mathbf{Y}$ dla pewnego procesu $\mathbf{X} \in \mathcal{M}_{\{0,1\}^{\mathbb{Z}}}$ takiego, że $(\mathbf{X}, \mathbf{Y})$ jest stacjonarny.

Jednym z otwartych pytań zadanych przez Sarnaka był problem wewnętrznej ergodyczności układu bezkwadratowego. Problem ten został (pozytywnie) rozwiązany przez Pecknera w [46] i następnie rozwiązany w przypadku ogólnym w [33, 19]. Powstało naturalne pytanie, czy miara o maksymalnej entropii ma własność Gibbsa (jest tak w wielu naturalnych sytuacjach, wspomnijmy przypadek układów soficznych [58], tzn. faktorów topologicznych łańcuchów Markowa). Odpowiedź na to pytanie okazała się negatywna w przypadku bezkwadratowym (Peckner w [46]), co było pewną niespodzianką. Dowód Pecknera używał nietrywialnych faktów teorio-liczbowych dotyczących liczb pierwszych (i bazował na dokładnym wzorze liczącym miarę Mirsky'ego danego bloku) i z tego powodu zapytał on, czy jego rezultat rozszerza się na wszystkie układy $\mathscr{B}$-wolne. Nasz główny rezultat daje odpowiedź pozytywną na pytanie Pecknera. W rzeczywistości, prezentujemy ogólniejsze kryterium (stosowalne poza światem układów $\mathscr{B}$-wolnych), bazujące na pojęciach entropii topologicznej i gęstości (topologicznej) jedynek, i które gwarantuje brak własności Gibbsa (patrz twierdzenie 6.1). Warto wspomnieć, że uogólniając nasze argumenty, autorzy w [37] uzyskali ostatnio (grudzień 2020) rozszerzenie naszego rezultatu do klasy (niezerowych) potencjałów spełniających pewne zależności arytmetyczne.

Wprowadźmy teraz kilka pojęć pomocniczych. Niech $\mathscr{X}$ będzie dowolnym podshiftem. Wówczas, dla dowolnego $\mathbf{X} \in \mathcal{M}_{\mathscr{X}}$, określamy

$$\mathbf{D} = \mathbf{D}_{\mathscr{X}} = \lim_{n\to\infty} \frac{1}{n} \max_{W\in\mathcal{L}^{(n)}} \#_1 W, \qquad \mathbf{D}_{\mathbf{X}} = \lim_{n\to\infty} \frac{1}{n} \max_{W\in\mathcal{L}^{(n)},\, \mathbb{P}\left(X_{[1,n]}\in W\right)>0} \#_1 W$$

oraz

$$\mathbf{d} = \mathbf{d}_{\mathscr{X}} = \sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \mathbb{P}(X_0 = 1), \qquad \mathbf{d}_{\mathbf{X}} = \mathbb{P}(X_0 = 1).$$

Można pokazać, że $\mathbf{d}_{\mathbf{X}} \le \mathbf{D}_{\mathbf{X}} \le \mathbf{D} = \mathbf{d}$. W szczególności $\sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \mathbf{D}_{\mathbf{X}} = \mathbf{D}_{\mathscr{X}}$ (każda miara, która osiąga supremum, jest nazywana miarą ***jedynkowo-wysyconą***) i każda miara o ***maksymalnej gęstości jedynek*** (tzn. każdy proces $\mathbf{X}$, który osiąga supremum w $\sup_{\mathbf{X}\in\mathcal{M}_{\mathscr{X}}} \mathbf{d}_{\mathbf{X}} = \mathbf{d}$) musi byc jedynkowo-wysycony (stąd miara Mirsky'ego musi byc jedynkowo-wysycona). Dalej, dla układów $\mathscr{B}$-wolnych, zawsze mamy $\mathbf{D}_{\mathscr{X}_{\eta}} = \mathbf{H}_{\mathscr{X}_{\eta}}$ (patrz Proposition K w [19]).

Powiemy, że podshift $\mathscr{X} \subset \mathcal{X}^{\mathbb{Z}} \subset \mathbb{R}^{\mathbb{Z}}$ ma ***własność splotu multyplikatywnego***, gdy istnieje proces stacjonarny $\mathbf{Y} \in \mathcal{M}_{\mathscr{X}}$ spełniający $\mathbf{H}(\mathbf{Y}) = 0$ i taki, że

$$\mathbf{Z} \in \mathcal{M}_{\mathscr{X}} \quad \Leftrightarrow \quad \mathbf{Z} = \mathbf{X} \cdot \mathbf{Y}, \qquad \text{gdzie } (\mathbf{X}, \mathbf{Y}) \in \mathcal{M}_{\mathcal{X}^{\mathbb{Z}}\times\mathscr{X}}. \tag{5.7}$$

Jak już wspomnieliśmy wcześniej, klasa układów $\mathscr{B}$-wolnych ma tę własność. Aby zobaczyć inne przykłady, zauważmy, że jeśli weźmiemy stały proces $\mathbf{Y} = \mathbf{1}$, to jako wynik (5.7) otrzymamy pełny shift $\mathscr{X} = \mathcal{X}^{\mathbb{Z}}$. Ponadto, jeśli $\mathscr{X}$ spełnia (5.7), to zasada wariacyjna dla ciśnienia topologicznego może być napisana jako

$$\mathscr{P}_{\mathscr{X},\varphi} = \sup_{(\mathbf{X},\mathbf{Y})\in\mathcal{M}_{\mathcal{X}^{\mathbb{Z}}\times\mathcal{Y}^{\mathbb{Z}}}} \left[\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y})\right]. \tag{5.8}$$

W szczególności wszystkie twierdzenia związane z (5.8) mogą być bezpośrednio stosowane do układów $\mathscr{B}$-wolnych (poniżej, ze względu na zwięzłość, często omijamy formułowanie takich wniosków).

5.1 Ciśnienie topologiczne w przestrzeniach z własnością splotu multyplikatywnego

W tej części podamy (częściową) odpowiedź na następujące pytanie:

4. Ile wynosi ciśnienie topologiczne układów dynamicznych spełniających (5.7)? Czy odpowiadające stany równowagi są jedyne?

Dokładniej mówiąc, mamy trzy rodzaje twierdzeń związanych z wzorami na ciśnienie topologiczne. Pierwszy rodzaj (twierdzenie 5.1) dotyczy przypadku, w którym φ zależy od tylko jednej współrzędnej. To jest oczywiście najprostsze możliwe rozszerzenie w porównaniu z przypadkiem entropii topologicznej. Drugi rezultat (twierdzenie 5.2) zachodzi dla (wystarczająco) lokalnych potencjałów procesów okresowych. W twierdzeniu 5.3 mamy do czynienia z dowolnym ciągłym potencjałem i procesami, które mogą być aproksymowane w pewien sposób procesami okresowymi. Przypadek, gdy sam proces graniczny jest okresowy, jest tu wykluczony – to jest istotne dla stosowalności naszych metod (zauważmy, że m oraz p w twierdzeniu 5.2 nie są całkiem dowolne). Oprócz powyższych faktów udowodnimy również jedyność stanów równowagi w przypadku potencjałów zależnych od jednej współrzędnej (twierdzenie 5.1).

Od teraz zakładamy, że $\mathcal{X} \subset \mathbb{R}$, $0 \in \mathcal{X}$ i że procesy oznaczane literą $\mathbf{Y}$ są **binarne**, tzn. $\mathcal{Y} = \{0, 1\}$.

Twierdzenie 5.1 (Potencjał zależący od jednej współrzędnej)**.** *Jeśli $\varphi \colon \mathcal{X} \to \mathbb{R}$ zależy tylko od jednej współrzędnej, to*

$$\sup_{(\mathbf{X},\mathbf{Y})\in\mathcal{M}_{\mathcal{X}^{\mathbb{Z}}\times\mathcal{Y}^{\mathbb{Z}}}} [\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y})] = (1-d)\varphi(0) + d\log_2\left(\sum_{x\in\mathcal{X}} 2^{\varphi(x)}\right),$$

gdzie $d = \mathbb{P}(Y_0 = 1)$. Ponadto, jeśli $\mathbf{X}$ osiaga powyższe supremum, to $\mathbf{X}\cdot\mathbf{Y} \sim \mathbf{G}\cdot\mathbf{Y}$, gdzie $\mathbf{G} \amalg \mathbf{Y}$, $\mathbf{G}$ jest procesem i.i.d. oraz $\mathbb{P}(G_i = x)$ jest proporcjonalne do $2^{\varphi(x)}$ (więc G_i jest miarą Gibbsa stowarzyszoną z φ).

Zauważmy, że twierdzenie to rozszerza fakt wewnętrznej ergodyczności udziedziczniеń układów $\mathscr{B}$-wolnych do przypadku, w którym potencjał zależy od jednej współrzędnej. Zupełnie niedawno autorzy już wspomnianej pracy [37] uzyskali troszkę słabszy wynik dla zbiorów $\mathscr{B}$ zawierających 2.[2]

Dla danego podzbioru skończonego liczb rzeczywistych $\mathcal{X}$ (zawierającego zero), ciągłego potencjału $\varphi \colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ i procesu $\mathbf{Y}$, istnieje pewna naturalna operacja (nazywana ***Y-podniesieniem (ang. upgrade) potencjału φ***)

$$\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi : \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$$

konieczna dla sformułowania naszych rezultatów. Dla dowolnego ściśle rosnącego ciągu liczb całkowitych $\mathbf{r}$ określamy $\varphi_{\mathbf{r}}$ poprzez

$$\varphi_{\mathbf{r}}(\mathbf{z}) = \varphi\left(\ldots, 0^{r_{-1}-r_{-2}-1}, \underbrace{z_{-1}}_{r_{-1}-coor.}, 0^{r_0-r_{-1}-1}, \underbrace{z_0}_{r_0-coor.}, 0^{r_1-r_0-1}, \underbrace{z_1}_{r_1-coor.}, 0^{r_2-r_1-1}, \ldots\right).$$

Następnie definiujemy Φ jako

$$\Phi = \mathbb{E}\varphi_{\mathbf{R}}, \tag{5.9}$$

gdzie $\mathbb{E}$ oznacza całkę Bochnera.

Najpierw przedyskutujemy przypadek okresowy procesu $\mathbf{Y}$. Niech $p \in \mathbb{N}$ bedzie ***okresem*** procesu $\mathbf{Y}$, tzn. najmniejszą liczbą naturalną taką, że $S^p\mathbf{Y} = \mathbf{Y}$. Zauważmy, że wówczas $m = \sum_{i=1}^p Y_i$ jest równa ***liczbie jedynek zawartych w okresie***.

Twierdzenie 5.2 (przypadek okresowy)**.** *Niech p będzie okresem procesu $\mathbf{Y}$. Jeśli $\varphi \colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ zależy od co najwyżej p kolejnych współrzędnych, to*

$$\sup_{(\mathbf{X},\mathbf{Y})\in\mathcal{M}_{\mathcal{X}^{\mathbb{Z}}\times\{0,1\}^{\mathbb{Z}}}} [\mathbf{H}(\mathbf{X}\cdot\mathbf{Y}) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y})] = \frac{1}{p}\log_2\left[\sum_{z_{[0,m-1]}\in\mathcal{X}^m} 2^{p\Phi(z_{[0,m-1]})}\right],$$

gdzie $m = \sum_{i=1}^p Y_i$ oraz $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

[2] Na pierwszy rzut oka ich twierdzenie 1.2 wydaje się silniejsze. Jednak jeśli rozpiszemy zasadę wariacyjną i użyjemy założenia $2 \in \mathscr{B}$, to natychmiastowo widzimy, że przypadek rozpatrywany w [37] redukuje się do potencjałów zależących od jednej współrzędnej, a $\mathscr{B}$ nadal musi zawierać 2.

Twierdzenie 5.2 prowadzi do następującego rezultatu:

Twierdzenie 5.3 (słabe granice układów okresowych)**.** *Załóżmy, że* $\left(\mathbf{Y}^{(n)}\right)_n$ *jest ciągiem okresowych procesów binarnych zbiegającym słabo do* $\mathbf{Y}$*. Niech* p_n *oznacza okres procesu* $\mathbf{Y}^{(n)}$*. Jeśli* $\sum_{i=1}^{p_n} Y_i^{(n)} \xrightarrow{n\to\infty} \infty$*, to dla dowolnego ciągłego potencjału* $\varphi\colon \mathcal{X}^{\mathbb{Z}} \to \mathbb{R}$ *mamy*

$$\sup_{(\mathbf{X},\mathbf{Y}_n)\in\mathcal{M}_{\mathcal{X}^{\mathbb{Z}}\times\{0,1\}^{\mathbb{Z}}}} \left[\mathbf{H}\left(\mathbf{X}\cdot\mathbf{Y}\right) + \mathbb{E}\varphi(\mathbf{X}\cdot\mathbf{Y})\right] \xrightarrow{n\to\infty} \mathbb{P}\left(Y_0=1\right)\log|\mathcal{X}| + \sup_{\mathbf{z}\in\mathcal{X}^{\mathbb{Z}}} \Phi(\mathbf{z}),$$

gdzie $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

Powyższe twierdzenie pozwala nam na otrzymanie następującego wniosku, który podaje dokładny wzór na ciśnienie topologiczne (udziedzicznienia) układu $\mathscr{B}$-wolnego.

Twierdzenie 5.4. *Dla dowolnego układu* $\mathscr{B}$*-wolnego* $\mathscr{X}_{\boldsymbol{\eta}}$ *takiego, że odpowiadająca miara Mirsky'ego* $\mathbf{Y} \sim \nu_{\boldsymbol{\eta}}$ ***nie*** *jest okresowa, i dowolnego ciągłego potencjału* $\varphi\colon \{0,1\}^{\mathbb{Z}} \to \mathbb{R}$ *mamy*

$$\mathscr{P}_{\widetilde{\mathscr{X}_{\eta}},\varphi} = \mathbb{P}\left(Y_0=1\right) + \sup_{\mathbf{z}\in\mathcal{X}^{\mathbb{Z}}} \Phi(\mathbf{z}), \tag{5.10}$$

gdzie $\varphi \overset{\mathbf{Y}}{\rightsquigarrow} \Phi$.

6 Układy $\mathscr{B}$-wolne

Na koniec opiszemy pozostałe rezultaty dotyczące układów $\mathscr{B}$-wolnych. Przypomnijmy, że podshift $(\mathscr{X}, S)$, gdzie $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, jest ***dziedziczny***, jeśli dla każdego słowa $W \in \mathcal{L}$ oraz $W' \le W$ mamy $W' \in \mathcal{L}$, gdzie $\mathcal{L} = \mathcal{L}_{\mathscr{X}}$ oznacza język podshiftu $\mathscr{X}$ (zbiór wszystkich skończonych słów pojawiających się w $\mathscr{X}$). Ponadto, dla danego podshiftu $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$, ***udziedzicznienie*** podshiftu $\mathscr{X}$ jest zdefiniowane jako

$$\widetilde{\mathscr{X}} = \{\mathbf{z} \in \{0,1\}^{\mathbb{Z}} : \mathbf{z} \le \mathbf{x} \text{ dla pewnego } \mathbf{x} \in \mathscr{X}\}.$$

Niech $\mathbf{B}$ oznacza symetryczny, i.i.d. proces binarny, zatem $\mathbb{P}\left(B_i=0\right) = \mathbb{P}\left(B_i=1\right) = 1/2$. Nasz główny rezultat jest następujący:

Twierdzenie 6.1. *Ustalmy podshift* $(\mathscr{X}, S)$*, gdzie* $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$*, i załóżmy, że proces* $\mathbf{Y} \in \mathcal{M}^e_{\mathscr{X}}$ *jest jedynkowo-wysycony i bezatomowy. Jeśli* $\mathbf{D}_{\widetilde{\mathscr{X}}} = \mathbf{H}_{\widetilde{\mathscr{X}}}$ *oraz* $\mathbf{Y} \amalg \mathbf{B}$*, to* $\mathbf{B}\cdot\mathbf{Y}$ ***nie ma*** *własności Gibbsa.*

Twierdzenie 6.1 natychmiast przynosi pozytywną odpowiedź na pytanie Pecknera z pracy [46].

Wniosek 6.2. *Niech* $\mathscr{B} \subset \mathbb{N}\setminus\{1\}$*. Załóżmy, że miara Mirsky'ego* ν_η *nie jest okresowa. Wówczas miara o maksymalnej entropii podshiftu* $(\widetilde{\mathscr{X}}_\eta, S)$ ***nie ma*** *własności Gibbsa.*

Jak już wspomnieliśmy, ten wniosek został nieco rozszerzony (do przypadku pewnych punktów równowagi) w pracy [37].

Twierdzenie 6.1 stosuje się poza światem układów $\mathscr{B}$-wolnych. Przypomnijmy, że podshift $\mathscr{X}$ nazywa się ***monoergodyczny***, gdy posiada on dokładnie jedną (stąd ergodyczną) miarę niezmienniczą (ze względu na shift).

Wniosek 6.3. *Jeśli* $(\mathscr{X}, S)$*, gdzie* $\mathscr{X} \subset \{0,1\}^{\mathbb{Z}}$*,* $\mathbf{Y}$ *jest jedyną miarą ergodyczną miara oraz* $\mathbf{H}_{\mathscr{X}} = 0$*, to* $\mathbf{B}\cdot\mathbf{Y}$ *nie ma własności Gibbsa, jeśli tylko* $\mathbf{Y}$ *jest bezatomowa.*

Wniosek 6.4. *Jeśli* $(\widetilde{\mathscr{X}}, S)$ *jest dziedzicznym układem Sturma, to jego miara o maksymalnej entropii nie ma własności Gibbsa.*

Jako pewien „produkt uboczny" naszych rozważań w rozprawie udowodniono kilka nowych rezultatów, wzbogacających wiedzę o samych układach $\mathscr{B}$-wolnych. W szczególności, udowodniono twierdzenie odwrotne do niedawnego rezultatu Kellera [31], otrzymując charakteryzację dynamiczną ważnej własności arytmetycznej naprężenia (ang. tautness) podzbioru liczb całkowitych (z grubsza, naprężenie oznacza, że każdy element zbioru $\mathscr{B}$ odgrywa nietrywialną rolę w zbiorze wielokrotności z punktu widzenia gęstości).

Twierdzenie 6.5. *Niech* $\mathscr{B} \subset \mathbb{N}\setminus\{1\}$*. Jeśli odpowiadająca miara Mirsky'ego* ν_η *ma pełny nośnik topologiczny* $\mathscr{X}_\eta$*, to zbiór* $\mathscr{B}$ *jest naprężony.*

References

[1] J. Aaronson, D. Gilat, and M. Keane. On the structure of 1-dependent Markov chains. *Journal of Theoretical Probability*, 5(3):545–561, 1992.

[2] R. Adamczak. A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electron. J. Probab.*, 13:no. 34, 1000–1034, 2008.

[3] R. Adamczak and W. Bednorz. Exponential concentration inequalities for additive functionals of Markov chains. *ESAIM Probab. Stat.*, 19:440–481, 2015.

[4] L. Arnold, V. M. Gundlach, and L. Demetrius. Evolutionary formalism for products of positive random matrices. *The Annals of Applied Probability*, pages 859–901, 1994.

[5] P. Bertail and G. Ciołek. New Bernstein and Hoeffding type inequalities for regenerative Markov chains. *ALEA Lat. Am. J. Probab. Math. Stat.*, 16(1):259–277, 2019.

[6] P. Bertail and S. Clémencon. Regenerative block bootstrap for Markov chains. *Bernoulli*, 12(4):689–712, 2006.

[7] R. Bowen. Equilibrium states and the ergodic theory of anosov diffeomorphisms. *Springer Lecture Notes in Math*, 470:78–104, 1975.

[8] R. Bowen. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, volume 470 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, revised edition, 2008. With a preface by David Ruelle, Edited by Jean-René Chazottes.

[9] M. Burr, S. Das, C. Wolf, and Y. Yang. Computability of topological pressure on compact shift spaces beyond finite type. *arXiv preprint arXiv:2010.14686*, 2020.

[10] R. M. Burton, M. Denker, and M. Smorodinsky. Finite state bilaterally deterministic strongly mixing processes. *Israel Journal of Mathematics*, 95(1):115–133, 1996.

[11] X. Chen. Limit theorems for functionals of ergodic Markov chains with general state space. *Mem. Amer. Math. Soc.*, 139(664):xiv+203, 1999.

[12] S. J. M. Clémencon. Moment and probability inequalities for sums of bounded additive functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.*, 55(3):227–238, 2001.

[13] V. De Valk. Hilbert space representations of m-dependent processes. *The Annals of Probability*, pages 1550–1570, 1993.

[14] J. Dedecker and S. Gouëzel. Subgaussian concentration inequalities for geometrically ergodic Markov chains. *Electron. Commun. Probab.*, 20:no. 64, 12, 2015.

[15] L. Demetrius and C. Wolf. Evolutionary entropy and the second law of thermodynamics. *arXiv preprint arXiv:2005.10332*, 2020.

[16] L. A. Demetrius. Boltzmann, Darwin and directionality theory. *Physics reports*, 530(1):1–85, 2013.

[17] L. A. Demetrius and V. M. Gundlach. Directionality theory and the entropic principle of natural selection. *Entropy*, 16(10):5428–5522, 2014.

[18] R. Douc, A. Guillin, and E. Moulines. Bounds on regeneration times and limit theorems for subgeometric Markov chains. *Ann. Inst. Henri Poincaré Probab. Stat.*, 44(2):239–257, 2008.

[19] A. Dymek, S. Kasjan, J. Kułaga-Przymus, and M. Lemańczyk. $\mathscr{B}$-free sets and dynamics. *Trans. Amer. Math. Soc.*, 370:5425–5489, 2018.

[20] N. A. Friedman and D. Ornstein. On isomorphism of weak Bernoulli transformations. *Advances in mathematics*, 5(3):365–394, 1970.

[21] H. Furstenberg. Disjointness in ergodic theory, minimal sets, and a problem in Diophantine approximation. *Math. Systems Theory*, 1:1–49, 1967.

[22] H. Furstenberg, Y. Peres, and B. Weiss. Perfect filtering and double disjointness. In *Annales de l'IHP Probabilités et statistiques*, volume 31, pages 453–465, 1995.

[23] F. Gao, A. Guillin, and L. Wu. Bernstein-type concentration inequalities for symmetric Markov processes. *Theory Probab. Appl.*, 58(3):358–382, 2014.

[24] R. Garbit. A note on Furstenberg's filtering problem. *Israel J. Math.*, 182:333–336, 2011.

[25] H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9. Walter de Gruyter, 2011.

[26] E. Glasner. *Ergodic theory via joinings.* American Mathematical Soc., 2003.

[27] A. Holroyd and T. Liggett. Finitely dependent coloring. *Forum of Mathematics, Pi*, 4, 02 2014.

[28] B. Jiang, Q. Sun, and J. Fan. Bernstein's inequality for general Markov chains. *arXiv preprint arXiv:1805.10721*, 2018.

[29] A. Katok. Fifty years of entropy in dynamics: 1958–2007. *Journal of Modern Dynamics*, 1(4):545, 2007.

[30] G. Keller. *Equilibrium states in ergodic theory*, volume 42. Cambridge university press, 1998.

[31] G. Keller. Tautness of sets of multiples and applications to $\mathcal{B}$-free systems. *Studia Math.*, 247(2):205–216, 2019.

[32] A. Kłopotowski and J. B. Robertson. Some new examples of pairwise independent random variables. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 72–88, 1999.

[33] J. Kułaga-Przymus, M. Lemańczyk, and B. Weiss. On invariant measures for $\mathscr{B}$-free systems. *Proc. Lond. Math. Soc. (3)*, 110(6):1435–1474, 2015.

[34] O. E. Lanford. Entropy and equilibrium states in classical statistical mechanics. *Lecture Notes in Physics*, pages 1–113, 1975.

[35] P. Lezaud. Chernoff-type bound for finite Markov chains. *Ann. Appl. Probab.*, 8(3):849–867, 1998.

[36] P. Lezaud. Chernoff and Berry-Esséen inequalities for Markov processes. *ESAIM Probab. Statist.*, 5:183–201, 2001.

[37] Z. Lin and E. Chen. Equilibrium states which are not gibbs measure on hereditary subshifts. *arXiv preprint arXiv:2012.03409*, 2020.

[38] F. Matúš. On two–block–factor sequences and one–dependence. *Proceedings of the American Mathematical Society*, 124(4):1237–1242, 1996.

[39] R. D. Mauldin, M. Urbanski, and M. Urbański. *Graph directed Markov systems: geometry and dynamics of limit sets*, volume 148. Cambridge University Press, 2003.

[40] F. Merlevède, M. Peligrad, and E. Rio. *High dimensional probability V: the Luminy volume*, volume 5. Institute of Mathematical Statistics, Beachwood, OH, 2009. Including papers from the 5th International Conference (HDP V) held in Luminy, May 26–30, 2008.

[41] F. Merlevède, M. Peligrad, and E. Rio. A Bernstein type inequality and moderate deviations for weakly dependent sequences. *Probab. Theory Related Fields*, 151(3-4):435–474, 2011.

[42] F. Merlevède and E. Rio. Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electronic Journal of Probability*, 20, 2015.

[43] S. P. Meyn and R. L. Tweedie. *Markov chains and stochastic stability.* Communications and Control Engineering Series. Springer-Verlag London, Ltd., London, 1993.

[44] E. Nummelin. *General irreducible Markov chains and nonnegative operators*, volume 83 of *Cambridge Tracts in Mathematics.* Cambridge University Press, Cambridge, 1984.

[45] D. Paulin. Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.*, 20:no. 79, 32, 2015.

[46] R. Peckner. Uniqueness of the measure of maximal entropy for the squarefree flow. *Israel J. Math.*, 210(1):335–357, 2015.

[47] Y. B. Pesin. *Dimension theory in dynamical systems: contemporary views and applications.* University of Chicago Press, 2008.

[48] D. Ruelle. *Statistical mechanics.* The Mathematical Gazette, 1999. Rigorous results, Reprint of the 1989 edition.

[49] D. Ruelle. *Thermodynamic formalism: the mathematical structure of equilibrium statistical mechanics.* Cambridge University Press, 2004.

[50] P.-M. Samson. Concentration of measure inequalities for Markov chains and Φ-mixing processes. *The Annals of Probability*, 28(1):416–461, 2000.

[51] P. Sarnak. Three lectures on the Möbius function, randomness and dynamics. unpublished, 2010.

[52] P. C. Shields. *The ergodic theory of discrete sample paths*, volume 13. American Mathematical Soc., 1996.

[53] J. G. Sinaĭ. Gibbs measures in ergodic theory. *Uspehi Mat. Nauk*, 27(4(166)):21–64, 1972.

[54] J. T. Trevors and M. H. Saier Jr. Thermodynamic perspectives on genetic instructions, the laws of biology and diseased states. *Comptes rendus biologies*, 334(1):1–5, 2011.

[55] R. Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press, 2018.

[56] P. Walters. Some results on the classification of non-invertible measure preserving transformations. In *Recent Advances in Topological Dynamics*, pages 266–276. Springer, 1973.

[57] P. Walters. A variational principle for the pressure of continuous transformations. *American Journal of Mathematics*, 97(4):937–971, 1975.

[58] B. Weiss. Subshifts of finite type and sofic systems. *Monatsh. Math.*, 77:462–474, 1973.

[59] O. Wintenberger. Deviation inequalities for sums of weakly dependent time series. *Electron. Commun. Probab.*, 15:489–503, 2010.