

INSTYTUT PODSTAW INFORMATYKI POLSKIEJ AKADEMII NAUK

Łukasz Dębowski

**Własności entropii nadwyżkowej
dla procesów stochastycznych
nad różnymi alfabetami**

rozprawa doktorska
przygotowana pod kierunkiem doc. dr. hab. Jana Mielniczuka



WARSZAWA 2005

Praca finansowana ze środków budżetowych na naukę w roku 2005 jako projekt badawczy
nr 1 P03A 045 28 pt. *Entropia nadwyżkowa jako teorioinformacyjna miara zależności.*

Spis treści

Wstęp	5
Niektóre oznaczenia matematyczne	9
I. Miary informacji	
Rozdział 1. Teoriomiarowe definicje miar informacji	13
Rozdział 2. Graniczne miary informacji procesu	31
Rozdział 3. Całkowe przedstawienia miar granicznych	37
Rozdział 4. Rozkład ergodyczny	43
II. Procesy dyskretne	
Rozdział 5. Procesy nieprzeliczalnego opisu	51
Rozdział 6. Oczekiwana nadwyżkowa długość kodu	55
Rozdział 7. Przeliterowanie i segmentacja	63
III. Procesy gaussowskie	
Rozdział 8. Własności podstawowe	75
Rozdział 9. Niektóre własności graniczne	81
Rozdział 10. Przedstawienia kauzalne i odwracalne	89
Rozdział 11. Sumy szeregów funkcji autokorelacji	95
Rozdział 12. Wzajemne ograniczenia zaniku ACF i PACF	105
Rozdział 13. Przykłady procesów gaussowskich	109
Zakończenie	123
Dodatek A. Podstawy probabilistyczne	125
A.1. Miary prawdopodobieństwa i procesy stochastyczne	125
A.2. Definiowanie miar przez rozkłady	129
A.3. Kodowanie i mierzalność funkcji rzeczywistych	130
A.4. Równość prawie na pewno	133
A.5. Prawdopodobieństwo warunkowe jako miara i martyngał	134
A.6. Warunkowa niezależność i warunkowa miara produktowa	138
A.7. Definiowanie miar przez częstości w ciągach	143
A.8. Procesy Markowa i ukryte procesy Markowa	144
Bibliografia	147
English summary	151

Wstęp

Celem niniejszej pracy jest dyskusja związków pomiędzy asymptotyczną wartością entropii nadwyżkowej a niektórymi innymi własnościami dyskretnych i gaussowskich procesów stacjonarnych.

Entropia nadwyżkowa to pewna miara globalnej zależności stacjonarnego procesu stochastycznego, której koncepcja wywodzi się z teorii informacji.¹ Jak pokazuję w niniejszej pracy, entropię nadwyżkową można określić dla dowolnego procesu stacjonarnego jako informację wzajemną między σ -ciałem półnieskończonej przeszłości a σ -ciałem półnieskończonej przyszłości. W szczególności entropia nadwyżkowa wydaje się być naturalną miarą zależności dla procesów o wartościach nominalnych, dla których opis zależności w tradycyjnych terminach funkcji autokorelacji i liniowych predyktorów jest nieefektywny.

Entropia nadwyżkowa jest wielkością nieujemną, lecz w zależności od procesu może być skończona lub nieskończona. Procesy, dla których entropia nadwyżkowa jest nieskończona, nazywa się niefinitarnymi (Crutchfield i Feldman, 2003). Dla dwóch szerokich podklas procesów dyskretnych i procesów gaussowskich procesy niefinitarne nie mogą być przedstawione jako ukryte łańcuchy Markowa. W przypadku procesów gaussowskich, proces niefinitarny nie może być nieosobliwym procesem ARMA (Finch, 1960). W przypadku procesów o wartościach dyskretnych, proces niefinitarny nie może być ukrytym łańcuchem Markowa o skończonej entropii blokowej.

Do wyboru tematyki rozprawy przyczyniło się moje równoległe zaangażowanie w zastosowania modeli probabilistycznych w lingwistyce komputerowej (przetwarzaniu języka naturalnego). Powszechnie używane w tej dziedzinie ukryte łańcuchy Markowa cechują się skończoną entropią nadwyżkową. Z drugiej strony istotne przesłanki eksperymentalne i teoretyczne wskazują, że bardziej adekwatne modele języka naturalnego powinny być niefinitarne (Hilberg, 1990). O niektórych z tych przesłanek szerzej piszę w osobnym artykule dla lingwistów (Dębowski, 2005). Wspominam o nich pokrótce także w tej pracy, przedstawiając własności procesów dyskretnych.

Niniejsza rozprawa obejmuje rozważania z pogranicza teorii informacji, teorii kodowania, teorii miary i teorii szeregów czasowych. Podzieliłem ją na trzy części o następującej zawartości:

1. Teoriomiarowe ujęcie teorii informacji. Główne wyniki to:
 - a) uogólnione definicje i podstawowe własności warunkowej informacji wzajemnej i entropii warunkowej jako funkcji (prawie) dowolnych σ -ciał (por. Gelfand *et al.*, 1956);
 - b) równoważne definicje entropii nadwyżkowej jako informacji wzajemnej między przeszłością a przyszłością oraz jako sumy pewnego szeregu warunkowych informacji wzajemnych;

¹ Gwoli ścisłości, po raz pierwszy o entropii nadwyżkowej usłyszałem na wykładzie nt. powierzchniowych przemian fazowych prowadzonym przez prof. dr. hab. Marka Napiórkowskiego, promotora mojej pracy magisterskiej na Wydziale Fizyki Uniwersytetu Warszawskiego.

- c) ograniczenia wartości entropii nadwyżkowej wynikające z rozkładu ergodycznego.

2. Własności stacjonarnych procesów dyskretnych. Udowadniam następujące fakty:
 - a) proces jest niefinitarny, jeżeli istnieje nieskończenie wiele niezależnych zmiennych binarnych, których wartości można przewidzieć dowolnie dobrze na podstawie dowolnego dostatecznie długiego fragmentu procesu;
 - b) dla dowolnego kodu uniwersalnego entropie nadwyżkowe skończonego rzędu są nieskończenie często nie większe niż nadwyżkowe długości kodu;
 - c) dla ustalonego kodu uniwersalnego i prawie wszystkich procesów stacjonarnych nadwyżkowe długości kodu dążą do nieskończoności;
 - d) dowolną miarę procesu o skończonej liczbie wartości można przekształcić na miarę procesu o wartościach binarnych z zachowaniem wszystkich wartości entropii blokowej z dokładnością do pewnej stałej;
 - e) niektóre miary procesów niefinitarnych o nieskończonej liczbie wartości można przekształcić na miary procesów o skończonej liczbie wartości — być może przy zachowaniu wartości entropii nadwyżkowej.
3. Własności stacjonarnych procesów gaussowskich. Podstawowe rezultaty to:
 - a) algebraiczne powiązanie funkcji informacji wzajemnej i warunkowej informacji wzajemnej z funkcjami autokorelacji i autokorelacji częściowej;
 - b) kilka warunków dostatecznych istnienia przedstawień kauzalnych i odwracalnych;
 - c) prostszy dowód klasycznego twierdzenia Grenandera i Szegö o zerach wielomianów ortogonalnych na okręgu jednostkowym;
 - d) wykazanie czystego niedeterminizmu (a więc i ergodyczności) procesów finitarnych i kwazifinitarnych;
 - e) wzór wyrażający sumę autokorelacji jako iloczyn autokorelacji częściowych;
 - f) ograniczenie modułu autokorelacji częściowej przez moduł autokorelacji dla dostatecznie małych wartości autokorelacji;
 - g) wskazanie kilku przykładów procesów niefinitarnych o bardzo szybko zanikającej funkcji autokorelacji, w tym procesów ergodycznych.

W zakończeniu pracy umieściłem listę problemów otwartych. Integralną częścią rozprawy jest dodatek, który postronnemu czytelnikowi może służyć jako zwięzłe wprowadzenie do teorii miary w rachunku prawdopodobieństwa. Oprócz materiału podręcznikowego w dodatku przedstawiam kilka specyficznych idei i konwencji notacyjnych, do których odwołuję się w rozprawie.

Chciałbym podziękować promotorowi Janowi Mielniczkowi, a także swojej rodzinie.

Niektóre oznaczenia matematyczne

$\llbracket \phi \rrbracket$	—	1, jeżeli ϕ jest prawdziwe; 0 w przeciwnym wypadku
I_A	—	funkcja charakterystyczna zbioru A ; $I_A(x) := \llbracket x \in A \rrbracket$
$[x]_+$	—	część dodatnia x ; $[x]_+ := \llbracket x > 0 \rrbracket \cdot x$
$[x]_-$	—	część ujemna x ; $[x]_- := -\llbracket x < 0 \rrbracket \cdot x$
$\lfloor x \rfloor$	—	najmniejsza liczba całkowita nie mniejsza niż x
$\lceil x \rceil$	—	największa liczba całkowita nie większa niż x
z^*	—	liczba zespolona sprzężona do z
$ z $	—	moduł liczby zespolonej z
\log	—	logarytm naturalny
$\text{card } \mathbb{V}$	—	moc zbioru \mathbb{V}
$A \ominus B$	—	różnica symetryczna zbiorów; $A \ominus B := A \cup B \setminus A \cap B$
\mathbb{N}	—	zbiór liczb naturalnych (bez zera)
\mathbb{Z}	—	zbiór liczb całkowitych
\mathbb{Q}	—	zbiór liczb wymiernych
\mathbb{R}	—	zbiór liczb rzeczywistych
\mathbb{C}	—	zbiór liczb zespolonych
\times	—	operator iloczynu kartezjańskiego, konkatenacji słów lub produktu miar
λ	—	słowo puste lub ciąg pusty (element neutralny operacji \times)
$\mathbb{V}^k := \underbrace{\mathbb{V} \times \dots \times \mathbb{V}}_{k \text{ razy}}$	—	k -ta potęga kartezjańska zbioru \mathbb{V} , zbiór konkatenacji k elementów zbioru \mathbb{V}
$\mathbb{V}^+ := \bigcup_{n \in \mathbb{N}} \mathbb{V}^n$	—	suma skończonych potęg kartezjańskich zbioru \mathbb{V} , zbiór konkatenacji skończenie wielu elementów zbioru \mathbb{V}
$\mathbb{V}^* := \{\lambda\} \cup \mathbb{V}^+$	—	domknięcie Kleenego zbioru \mathbb{V}
$x_{n:m}$	—	skrótowy zapis słowa lub ciągu $(x_n, x_{n+1}, \dots, x_m)$; $x_{n:m} := \lambda$, jeżeli $m < n$
$\text{len } w$	—	długość słowa lub ciągu $w \in \mathbb{V}^*$; $\text{len } \lambda := 0$; $\text{len } w := k$ dla $w = (x_1, x_2, \dots, x_k)$, $x_i \in \mathbb{V}$
$w' \stackrel{P}{\sqsubset} w$	—	słowo w' jest przedrostkiem słowa w
$w' \stackrel{S}{\sqsubset} w$	—	słowo w' jest przyrostkiem słowa w

(Ω, \mathcal{J}, P)	—	domyślna przestrzeń probabilistyczna
$\sigma(\mathcal{G})$	—	σ -ciało generowane przez klasę zbiorów \mathcal{G}
$\mathcal{A}(\mathcal{G})$	—	zbiór atomów σ -ciała dyskretnego \mathcal{G}
$\mathcal{A}(X)$	—	alfabet zmiennej X ; $\mathcal{A}(\mathcal{G}) \supset X(\Omega)$
$\sigma(X)$	—	σ -ciało przeciwobrazów zmiennej X
$\tilde{\sigma}(X)$	—	σ -ciało obrazów zmiennej X
\mathcal{R}	—	σ -ciało borelowskie na prostej \mathbb{R}
$L^q(\mathcal{G}, \mu)$	—	przestrzeń funkcji f mierzalnych \mathcal{G}/\mathcal{R} i spełniających $\int f ^q d\mu < \infty$
$\nu \ll \mu$	—	miara ν jest absolutnie ciągła względem miary μ
$\nu \perp \mu$	—	miary ν i μ są wzajemnie osobliwe
$d\nu/d\mu$	—	pochodna miary ν względem miary μ
$P(A B)$	—	prawdopodobieństwo warunkowe zdarzenia A względem zdarzenia lub σ -ciała B
$A \perp\!\!\!\perp B C$	—	A i B są niezależne względem C ; A , B i C mogą być zdarzeniami, σ -ciałami lub zmiennymi
P_X	—	przeniesiona miara prawdopodobieństwa zmiennej X
μ_X	—	miara pomiarowa zmiennej X
$\langle X \mathcal{G} \rangle$	—	warunkowa wartość oczekiwana zmiennej X względem σ -ciała \mathcal{G} i miary P ; $\langle X \rangle := \langle X \{\emptyset, \Omega\} \rangle$
$\text{Cov}(X; Y)$	—	kowariancja zmiennych zespolonych X i Y ; $\text{Cov}(X; Y) := \langle X^*Y \rangle - \langle X^* \rangle \langle Y \rangle$
$\ X\ $	—	norma zmiennej zespolonej X ; $\ X\ := \sqrt{\text{Cov}(X; X)}$

Część I

Miary informacji

Rozdział 1

Teoriomiarowe definicje miar informacji

W niniejszym rozdziale przedstawiamy podstawowe definicje i tożsamości z teorii informacji. Kilka ogólnie znanych twierdzeń przeformułujemy do postaci znacznie mocniejszej niż ich wersje spotykane w elementarnych podręcznikach (Yeung, 2002; Cover i Thomas, 1991). Celem przewodnim będzie rozszerzenie definicji podstawowych pojęć teorii informacji w terminach teorii miary (por. Gelfand *et al.*, 1956, a także Billingsley, 1965, sekcja 12). Nieco inne analogie pomiędzy miarami informacji a teorią rzeczywistej miary addytywnej zostały dostrzeżone przez Hu (1962) — są one związane także z teorią grup skończonych (Yeung, 2002, rozdział 16). Nas będzie interesował kierunek badań bliższy klasycznym zastosowaniom teorii (nieujemnej i σ -addytywnej) miary w probabilistyce. Wstęp do używanego przez nas aparatu teoriomiarowego zamieściliśmy w dodatkach A.1–A.8.

Definicja 1.1. (wartość oczekiwana, Billingsley, 1979, sekcja 21) *Wartość oczekiwana $\langle X \rangle$ rzeczywistej zmiennej losowej X zdefiniowana jest wzorem*

$$\langle X \rangle = \int X dP, \quad (1.1)$$

gdzie P oznacza miarę prawdopodobieństwa P na przestrzeni probabilistycznej (Ω, \mathcal{J}, P) , a całkowanie w sensie Lebesgue'a odbywa się po całej przestrzeni zdarzeń Ω .

Istnieją zmienne X , dla których wartość oczekiwana jest nieskończona bądź nieokreślona. W szczególności wartość oczekiwana jest określona jako wielkość $\langle X \rangle \in [0, \infty]$ dla dowolnej nieujemnej zmiennej losowej. Jeżeli zmienna X przyjmuje dowolne wartości rzeczywiste, to $\langle X \rangle$ określa się jako $\langle X \rangle = \langle [X]_+ \rangle - \langle [X]_- \rangle$ wtedy i tylko wtedy, gdy przynajmniej jedna z wielkości $\langle [X]_+ \rangle$ i $\langle [X]_- \rangle$ jest skończona.

Wartość oczekiwaną można określić tylko dla zmiennej losowej o wartościach liczbowych (lub wektorowych) i nie jest ona niezmiennicza ze względu na bijekcje zmiennych losowych. Miary informacji, czyli kombinacje liniowe entropii, są pewnymi charakterystykami zmiennych losowych, które nie mają jednej, a niekiedy obu wymienionych cech. W podejściu wyabstrahowanym przez nas z podejścia elementarnego (Cover i Thomas, 1991, sekcje 2.1 i 9.1) entropię i inne miary informacji konstruuje się w oparciu o dwa typy miar pomocniczych — miary przeniesione i miary pomiarowe (te miary są już miarami w sensie teorii miary).

Definicja 1.2. (miara przeniesiona) *Dla zmiennej X przeniesioną miarą prawdopodobieństwa na σ -ciele obrazów nazywamy miarę $P_X = P \circ X^{-1} : \tilde{\sigma}(X) \ni B \mapsto P(X \in B)$.*

Funkcję f mierzalną $\tilde{\sigma}(X)/\mathcal{R}$ można całkować przez podstawienie $\langle f(X) \rangle = \int f dP_X$, jeżeli oczywiście $\int [f(X)]_+ dP < \infty$ lub $\int [f(X)]_- dP < \infty$.

Definicja 1.3. (miara pomiarowa) *Miara pomiarowa μ_X zmiennej losowej X to dowolna niezerowa i σ -skończona miara na σ -ciele obrazów zmiennej X , $\mu_X : \tilde{\sigma}(X) \rightarrow \mathbb{R}$.*

Miary pomiarowe są niejawnymi parametrami określonej dalej entropii i innych miar informacji. Zwykle używać będziemy następującego układu miar pomiarowych:

- dla dyskretnej zmiennej losowej X — miary liczącej $\mu_X : 2^{\mathbb{V}} \ni B \mapsto \text{card } B$,
- dla zmiennej rzeczywistej Y o nieprzeliczalnym $Y(\Omega)$ — miary Lebesgue'a

$$\mu_Y : \mathcal{R} \ni B \mapsto m(B),$$

- dla zmiennej wielowymiarowej $X_{1:k}$ — miary produktowej $\mu_{X_{1:k}} = \mu_{X_1} \times \dots \times \mu_{X_k}$.
- Układ tak określonych miar nazywać będziemy układem standardowych miar pomiarowych. Napis $X_{1:k} := X_1 \times \dots \times X_k$ oznacza wielowymiarową zmienną losową powstałą przez skonkatowanie wartości zmiennych X_1, \dots, X_k i odpowiedni dobór σ -ciał obrazów i przeciwobrazów, zob. dodatek A.1.

Definicja 1.4. (gęstość prawdopodobieństwa) *Jeżeli $P_{X_1 \times \dots \times X_k} \ll \mu_{X_1} \times \dots \times \mu_{X_k}$ (symbol \ll oznacza relację ciągłości absolutnej — definicja A.21), to definiujemy gęstość prawdopodobieństwa ρ_{X_1, \dots, X_k} zmiennych X_1, \dots, X_k jako pochodną Radona-Nikodyma (zob. twierdzenie A.22)*

$$\rho_{X_1, \dots, X_k} := dP_{X_1 \times \dots \times X_k} / d(\mu_{X_1} \times \dots \times \mu_{X_k}). \quad (1.2)$$

Zauważmy, że jeżeli istnieje gęstość $\rho_{X,Y}$, to istnieje także gęstość ρ_X .

Gęstość prawdopodobieństwa zależy od przyjętych miar pomiarowych. Dla standardowego układu miar pomiarowych mamy $\rho_{X_{1:k}} = \rho_{X_1, \dots, X_k}$, gdyż dla owego układu zachodzi $\mu_{X_{1:k}} = \mu_{X_1} \times \dots \times \mu_{X_k}$. Dla innych układów miar pomiarowych możemy mieć jednak $\mu_{X_{1:k}} \neq \mu_{X_1} \times \dots \times \mu_{X_k}$, a w konsekwencji $\rho_{X_{1:k}} \neq \rho_{X_1, \dots, X_k}$. Podkreślmy, że dla standardowych miar pomiarowych istnieją zmienne X takie, że gęstość ρ_X nie istnieje. Przykładowo, gęstość $\rho_{Y \times Z}$ nie istnieje dla pary zmiennych rzeczywistych Y i Z takich, że $Z = f(Y)$.¹ Dla zmiennej dyskretnej X zachodzi natomiast $\rho_X(a) = P(X = a)$.

Definicja 1.5. (entropia) *Entropią zmiennych X_1, \dots, X_k nazywa się wielkość*

$$\begin{aligned} H(X_1, \dots, X_k) &:= - \langle \log \rho_{X_1, \dots, X_k}(X_1 \times \dots \times X_k) \rangle = - \int \log \rho_{X_1, \dots, X_k} dP_{X_1 \times \dots \times X_k} \\ &= - \int \rho_{X_1, \dots, X_k} \log \rho_{X_1, \dots, X_k} d(\mu_{X_1} \times \dots \times \mu_{X_k}), \end{aligned} \quad (1.3)$$

gdzie funkcja \log to logarytm naturalny, zaś μ_{X_i} są przyjętymi miarami pomiarowymi (por. Cover i Thomas, 1991, sekcje 2.1 i 9.1).

Uwaga 1.6. *Entropia jest niejawną funkcją miary przeniesionej i miar pomiarowych. W niektórych przypadkach wygodnie jest postrzegać miarę przeniesioną jako parametr. Wprowadzamy zatem dodatkowe oznaczenie, w którym $P_{X_1 \times \dots \times X_k}$ pojawia się jawnie, $H(X_1, \dots, X_k; P_{X_1 \times \dots \times X_k}) := H(X_1, \dots, X_k)$.*

Załóżmy chwilowo, że μ_X dla dowolnej rozważanej zmiennej X jest standardową miarą pomiarową (odpowiednim produktem miar liczących i miar Lebesgue'a). W przeciwieństwie do wartości oczekiwanej $\langle X \rangle$, entropia $H(X)$ jest niezmiennicza względem dość szerokiej klasy bijekcji zmiennej X :

- $H(X_{1:k}) = H(X_1, \dots, X_k)$, gdyż $\rho_{X_{1:k}} = \rho_{X_1, \dots, X_k}$;
- $H(X_{\pi(1)}, \dots, X_{\pi(k)}) = H(X_1, \dots, X_k)$ dla dowolnej permutacji $\pi : \{1, \dots, k\} \rightarrow \{1, \dots, k\}$;
- $H(f(X)) = H(X)$ dla dowolnej zmiennej dyskretnej X i dyskretnej iniekcji f .

¹ Formalnie można uogólnić funkcje rzeczywiste do dystrybucji i zdefiniować $\rho_{Y \times Z}(y, z) = \delta(z - f(y))\rho_Y(y)$ dla $Z = f(Y)$, gdzie operator δ zwany jest deltą Diraca. Niestety na dystrybucjach nie można konsystently określić operacji nieliniowych takich, jak podnoszenie do potęgi bądź obliczanie logarytmu.

Zależność $H(f(X)) = H(X)$ nie jest jednak prawdziwa dla dowolnej zmiennej rzeczywistej X i dowolnej iniekcji $f : \mathbb{R} \rightarrow \mathbb{R}$. Na przykład, $H(bX) = H(X) + \log(b)$ dla każdego $b \in \mathbb{R}$.

Entropia jest także pewną miarą rozproszenia rozkładu zmiennej. Jak w przypadku wielu innych miar rozproszenia, nie zawsze można przypisać entropii rzeczywistą wartość liczbową. Dla zmiennych prostych (czyli zmiennych o skończonej liczbie wartości) nie ma takich problemów, ale istnieją zmienne dyskretne o nieskończonym alfabcie, których entropia jest nieskończona.

Przykład 1.7. (zmienna dyskretna o nieskończonej entropii) *Określmy ciąg $r_n = n^{-1}(\log n)^{-\beta}$. Zachodzi $\sum_{m=1}^{\infty} r_m = \infty$ dla $\beta \leq 1$ oraz $\sum_{m=1}^{\infty} r_m < \infty$ dla $\beta > 1$ (Knopp, 1956, rozdział 14, uwagi 2-5).² Zatem $-\sum_{m=1}^{\infty} r_m \log r_m = \infty$ dla $1 < \beta \leq 2$, gdyż*

$$-r_m \log r_m = \frac{1}{n(\log n)^{\beta-1}} + \frac{\beta \log \log n}{n(\log n)^{\beta}}.$$

W szczególności, zmienna naturalna X o rozkładzie $P(X = n) = p_n := r_n / \sum_{m=1}^{\infty} r_m$ ma entropię $H(X) = -\sum_{m=1}^{\infty} p_m \log p_m = \infty$.

Dla zmiennej o wartościach rzeczywistych entropia może być dowolną liczbą ujemną, jeżeli dowolnie mała masa prawdopodobieństwa koncentruje się dostatecznie mocno w otoczeniu przeliczalnego zbioru wartości.

Przykład 1.8. (zmienna rzeczywista o dowolnie małej entropii) *Niech*

$$\rho_X(x) = \begin{cases} p_{\lfloor x \rfloor} b_{\lfloor x \rfloor}, & \text{jeśli } x - \lfloor x \rfloor \leq 1/b_{\lfloor x \rfloor}, \\ 0, & \text{jeśli } x - \lfloor x \rfloor > 1/b_{\lfloor x \rfloor}, \end{cases}$$

gdzie $\sum_{m=1}^{\infty} p_m = 1$ oraz $p_m \geq 0$, $b_m \geq 1$ dla wszystkich $m \in \mathbb{N}$. Wyrazy b_m możemy przedstawić jako $b_m = \exp(w_m/p_m)/p_m$, gdzie $b_m \geq 1 \iff w_m \geq p_m \log p_m$, natomiast $p_m \log p_m \leq 0$. W szczególności wyrazy w_m mogą być dowolnymi nieujemnymi liczbami rzeczywistymi. Mamy

$$H(X) = -\langle \log \rho_X(X) \rangle = -\sum_{m=1}^{\infty} p_m \log(p_m b_m) = -\sum_{m=1}^{\infty} w_m.$$

Jeżeli $w_n = 1$ dla każdego n , to $H(X) = -\infty$.

Przykład 1.8 wskazuje, że istnieją także zmienne rzeczywiste X o nieokreślonym $H(X)$. Aby skonstruować takie X , wystarczy wziąć rozkład o entropii $-\sum_{m=1}^{\infty} p_m \log p_m = \infty$ oraz wyrazy w_m takie, że $\sum_{m=1}^{\infty} [w_m]_+ = \infty$ oraz $\sum_{m=1}^{\infty} [w_m]_- = \infty$.

Miarami informacji nazywa się dowolne kombinacje liniowe entropii skończonej liczby zmiennych losowych. Najczęściej dyskutowanymi miarami informacji są następujące wielkości:

Definicja 1.9. (miary informacji) *Niech X, Y, Z będą zmiennymi takimi, że określone i skończone są ich entropie dla standardowego układu miar pomiarowych. Entropią warunkową zmiennej X przy danej zmiennej Y nazywa się wielkość*

$$H(X|Y) = H(X \times Y) - H(Y), \quad (1.4)$$

² Zbieżność dla $\beta > 1$ wynika także z nierówności Krafta (6.1) i istnienia bezprzedrostkowej binarnej reprezentacji ω zbioru liczb naturalnych (Elias, 1975), zob. też rozdział 6.

por. *Cover i Thomas (1991, sekcje 2.2 i 9.2)*. Informacją wzajemną między zmiennymi X i Y nazywa się wielkość

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= -H(X \times Y) + H(X) + H(Y), \end{aligned} \quad (1.5)$$

por. *Cover i Thomas (1991, sekcje 2.3 i 9.3)*. Warunkową informacją wzajemną między zmiennymi X i Y przy danej zmiennej Z nazywa się wielkość

$$\begin{aligned} I(X; Y|Z) &= H(X|Z) - H(X|Y \times Z) \\ &= -H(X \times Y \times Z) + H(X \times Z) + H(Y \times Z) - H(Z), \end{aligned} \quad (1.6)$$

por. *Cover i Thomas (1991, sekcja 2.5)*. Potrójną informacją wzajemną typu I między zmiennymi X , Y i Z nazywa się wielkość

$$V(X; Y; Z) = -H(X \times Y \times Z) + H(X) + H(Y) + H(Z). \quad (1.7)$$

Potrójną informacją wzajemną typu II między zmiennymi X , Y i Z nazywa się wielkość

$$\begin{aligned} I(X; Y; Z) &= I(X; Y) - I(X; Y|Z) \\ &= +H(X \times Y \times Z) \\ &\quad - H(X \times Y) - H(Y \times Z) - H(X \times Z) \\ &\quad + H(X) + H(Y) + H(Z), \end{aligned} \quad (1.8)$$

por. *Cover i Thomas (1991, problem 2.12)*.

W szczególności dla standardowego układu miar pomiarowych mamy tożsamości

$$I(X; Y; Z) = -V(X; Y; Z) + I(X; Y) + I(Y; Z) + I(X; Z), \quad (1.9)$$

$$V(X; Y; Z) = I(X; Y \times Z) + I(Y; Z) = I(X \times Z; Y) + I(X; Z), \quad (1.10)$$

$$\begin{aligned} I(X; Y \times Z|W) &= I(X; Y|W) + I(X; Z|Y \times W) \\ &= I(X; Z|W) + I(X; Y|Z \times W), \end{aligned} \quad (1.11)$$

jeżeli określone są wielkości po prawej stronie.

Zauważmy, że $H(Z) = H(Z|1)$, gdy określona jest którakolwiek z tych dwóch wielkości, gdyż $\rho_{Z \times 1}(z \times i) = \llbracket i = 1 \rrbracket \rho_Z(z)$. Podobnie $I(X; Y) = I(X; Y|1)$. Wielkości $H(X)$, $H(X|Y)$, $I(X; Y)$, $I(X; Y|Z)$ nazywane są miarami informacji Shannona (*Yeung, 2002, rozdział 2*). Dla zmiennych dyskretnych o skończonej entropii dowolną z miar informacji Shannona można wyrazić jako pewną warunkową informację wzajemną: $H(X) = I(X; X|1)$, $H(X|Y) = I(X; X|Y)$, $I(X; Y) = I(X; Y|1)$.

Wiadomo, że wielkości $I(X; Y|Z)$ i $V(X; Y; Z)$ zdefiniowane jako kombinacje liniowe entropii dla standardowych miar pomiarowych są zawsze nieujemne (*Cover i Thomas, 1991*). Niestety, wielkości $H(X|Y)$, $I(X; Y)$, $I(X; Y|Z)$, $V(X; Y; Z)$, i $I(X; Y; Z)$ nie można określić za pomocą entropii bezwarunkowych, gdy te nie są określone i skończone. W następnych akapitach spróbujemy uogólnić definicje tych obiektów tak, by były one określone dla większego podzbioru zmiennych losowych i spełniały niektóre z tożsamości zachodzących, gdy entropie bezwarunkowe są określone.

Zauważmy, że standardowy układ miar pomiarowych nie jest wystarczający. Przykładowo nie wiemy, jakie są miary pomiarowe dla procesów bądź dla zmiennych mierzalnych względem σ -ciał ogonowych, zob. dodatek [A.1](#). Aby móc zdefiniować miary informacji

Shannona dla dowolnych zmiennych losowych, musimy albo rozszerzyć układ miar pomiarowych na dowolne zmienne, albo definiować miary informacji w oderwaniu od gęstości prawdopodobieństwa.

W pierwszej kolejności zaargumentujemy, że układ standardowych miar pomiarowych nie jest trywialnie rozszerzalny. Zauważmy, że miary te spełniają dwa postulaty:

- (i) Dla dowolnych zmiennych X i Y zachodzi równość $\mu_X \times \mu_Y = \mu_{X \times Y}$.
 - (ii) Dla zmiennych dyskretnych X i $Y = f(X)$, gdzie f jest iniekcją, zachodzi $\mu_Y \circ f = \mu_X$.
- Postulaty (i) i (ii) są dostatecznymi warunkami decydującymi o tym, że w ujęciu elementarnym entropia i pewne inne miary informacji są zawsze nieujemne (po uogólnieniu — zawsze określone dla jakichkolwiek zmiennych) bądź niezmiennicze względem szerokiej klasy (po uogólnieniu — wszystkich) mierzalnych bijekcji zmiennych losowych.

Aby określić miary pomiarowe dla jakichkolwiek procesów o wartościach dyskretnych lub rzeczywistych, przez analogię do postulatu (i) moglibyśmy przyjąć trzeci postulat:

- (iii) Dla dowolnego procesu $X_{\mathbb{T}} = (X_t)_{t \in \mathbb{T}}$ zachodzi $\mu_{X_{\mathbb{T}}} = \times_{t \in \mathbb{T}} \mu_{X_t}$, gdzie miara $\times_{t \in \mathbb{T}} \mu_{X_t}$ jest miarą produktową na produkcie nieskończonym (Pap, 2002, sekcja 5.3).

Niestety, gdy miary μ_{X_t} spełniają $\mu_{X_t}(X_t(\Omega)) > 1$ (względnie $\mu_{X_t}(X_t(\Omega)) < 1$), a zbiór \mathbb{T} jest nieskończony, to miara $\times_{t \in \mathbb{T}} \mu_{X_t}$ jest nieskończona (względnie zerowa) na wszystkich niepustych zbiorach. Aby spełnić postulat (iii) w pewnym układzie σ -skończonych miar pomiarowych, trzeba przyjąć, że dla każdej zmiennej X miara μ_X jest miarą prawdopodobieństwa. Niewątpliwie nie może być to standardowy układ miar. Niestety nie istnieją także takie miary prawdopodobieństwa μ_X , że postulat (ii) jest spełniony. W istocie postulat (ii) głosi, że miara μ_X dla zmiennej dyskretnej X musi być stała na atomach zbioru $X(\Omega)$. Jeżeli $\text{card } X(\Omega) < \infty$, to $\mu_X = [\text{card } X(\Omega)]^{-1} \cdot \text{card}$. Jeżeli $\text{card } X(\Omega) = \infty$, to każda niezerowa miara stała na atomach zbioru $X(\Omega)$ jest nieskończona.

Jeżeli nie można skonstruować uniwersalnych miar pomiarowych dla wszystkich typów zmiennych losowych, nie łamiąc postulatów (i) i (ii), to jedynym wyjściem pozostaje konsekwentne definiowanie miar informacji w oderwaniu od pojęcia miary pomiarowej.

Pokażemy, że warunkową informację wzajemną można zdefiniować jako jednoznacznie określoną wielkość z przedziału $[0, \infty]$ nie tylko dla trzech zmiennych losowych bez określania ich miar pomiarowych, ale także dla (prawie) dowolnych trzech σ -ciał. Nasza konstrukcja jest pewnym rozwinięciem podejść Gelfanda *et al.* (1956), którzy definiowali bezwarunkową informację wzajemną między σ -ciałami, a także Billingsleya (1965, sekcja 12), który określił entropię warunkową jako funkcję dwóch σ -ciał.

Określmy najpierw kilka pojęć pomocniczych.

Definicja 1.10. (niewłaściwa pochodna miary) Niech ν będzie σ -skończoną miarą na (Ω, \mathcal{G}_1) , a μ — σ -skończoną miarą na (Ω, \mathcal{G}_2) spełniającą $\mathcal{G}_1 \subset \mathcal{G}_2$. Określamy niewłaściwą pochodną Radona-Nikodyma

$$\frac{D\nu}{D\mu} = \frac{d\nu}{d(\mu|_{\mathcal{G}_1} + \nu)} : \frac{d\mu|_{\mathcal{G}_1}}{d(\mu|_{\mathcal{G}_1} + \nu)} = \frac{d\nu}{d(\mu|_{\mathcal{G}_1} + \nu)} : \left(1 - \frac{d\nu}{d(\mu|_{\mathcal{G}_1} + \nu)}\right), \quad (1.12)$$

gdzie $r/0 := \infty$ dla $r > 0$.

Oznaczmy $S = (\mu|_{\mathcal{G}_1} + \nu)$. Podczas gdy $d\nu/d\mu$ może być dowolnie redefiniowana/niezdefiniowana na dowolnym zbiorze z klasy zbiorów μ -miary 0, to $D\nu/D\mu$ może być dowolnie redefiniowana/niezdefiniowana wyłącznie na dowolnym zbiorze z mniejszej klasy zbiorów S -miary 0. Pomimo tego, $D\nu/D\mu$ jako odpowiednia klasa abstrakcji funkcji o wartościach $[0, \infty]$ jest określona zawsze, gdyż $\mu|_{\mathcal{G}_1}, \nu \ll S$.

Twierdzenie 1.11.

- (i) Jeżeli $\nu, \mu \ll \lambda$, to $D\nu/D\mu = (d\nu/d\lambda) : (d\mu/d\lambda)$.

(ii) Jeżeli $\nu \ll \mu$, to $D\nu/D\mu = d\nu/d\mu$.

Dowód:

- (i) Równość $D\nu/D\mu = (d\nu/dS \cdot dS/d\lambda) : (d\mu/dS \cdot dS/d\lambda) = (d\nu/d\lambda) : (d\mu/d\lambda)$ zachodzi na zbiorze A zawierającym wszystkie punkty, na których $dS/d\lambda > 0$. Z definicji A wynika, że $S(\Omega \setminus A) = 0$. Możemy zatem utożsamić $(d\nu/d\lambda) : (d\mu/d\lambda)$ z pewną wersją $D\nu/D\mu$.
- (ii) Ponieważ $S \ll \mu|_{\mathcal{G}_1}$, to $[d\mu|_{\mathcal{G}_1}/dS]^{-1} = dS/d\mu|_{\mathcal{G}_1}$, a zatem $D\nu/D\mu = d\nu/dS \cdot dS/d\mu|_{\mathcal{G}_1} = d\nu/d\mu|_{\mathcal{G}_1} = d\nu/d\mu$.

□

Własność (i) stanowi o użyteczności niewłaściwej pochodnej Radona-Nikodyma. Wynika z niej, że $(d\nu/d\lambda) : (d\mu/d\lambda) = (d\nu/d\kappa) : (d\mu/d\kappa)$ dla dowolnych $\nu, \mu \ll \lambda, \kappa$, a samą wielkość $(d\nu/d\lambda) : (d\mu/d\lambda)$ można zdefiniować w oderwaniu od λ . Określenie $D\nu/D\mu$ pozwala na eleganckie wybrnięcie z nieprzejrzystego rozpatrywania implikacji poszczególnych przypadków absolutnej ciągłości i nieciągłości pewnych miar względem innych.

Definicja 1.12. (dywergencja Kullbacka-Leiblera) Niech ν będzie skończoną miarą na (Ω, \mathcal{G}_1) , a μ — skończoną miarą na (Ω, \mathcal{G}_2) , przy czym $\mathcal{G}_1 \subset \mathcal{G}_2$, $\nu(\Omega) = \mu(\Omega)$. Uogólniamy dywergencję Kullbacka-Leiblera (Kullback i Leibler, 1951) jako

$$D_{KL}(\nu||\mu) := \int \log \left(\frac{D\nu}{D\mu} \right) d\nu, \quad (1.13)$$

gdzie $\log 0 := -\infty$, $\log \infty := \infty$.

Całka (1.13) jest jednoznacznie określona nawet wtedy, gdy funkcja $\log \left(\frac{D\nu}{D\mu} \right)$ jest niecałkowalna w standardowym sensie (tzn. gdy $\int \left| \log \left(\frac{D\nu}{D\mu} \right) \right| d\mu = \infty$, Billingsley, 1979, sekcja 13). Dzieje się tak, gdyż $\int \left[\log \left(\frac{D\nu}{D\mu} \right) \right]_- d\nu < \infty$.

Twierdzenie 1.13. Dla miar ν i μ jak powyżej prawdziwe są następujące stwierdzenia:

- (i) Wielkość $D_{KL}(\nu||\mu)$ jest określona i nieujemna.
(ii) $D_{KL}(\nu||\mu) = \infty$, jeżeli $\nu \not\ll \mu$.
(iii) $D_{KL}(\nu||\mu) = 0$ wtedy i tylko wtedy, gdy $\nu = \mu|_{\mathcal{G}_1}$.
(iv) $D_{KL}(\nu|_{\mathcal{G}}||\mu) \leq D_{KL}(\nu||\mu)$ dla $\mathcal{G} \subset \mathcal{G}_1$.

Dowód:

- (i) Oznaczmy ponownie $S = (\mu|_{\mathcal{G}_1} + \nu)$, $h := d\nu/dS$. Niech i_{reg} oznacza funkcję

$$i_{\text{reg}}(x) := x \log x - x \log(1-x) - 2x + 1 \geq 0,$$

dookreśloną jako $i_{\text{reg}}(0) := 1$, $i_{\text{reg}}(1) := \infty$. Całka

$$D_{KL}(\nu||\mu) = \int \log \left(\frac{h}{1-h} \right) d\nu = \int \log \left(\frac{h}{1-h} \right) h dS = \int i_{\text{reg}}(h) dS$$

jest określona i nieujemna, gdyż $\int (1-2h) dS = \int dS - 2 \int d\nu = \mu(\Omega) + \nu(\Omega) - 2\nu(\Omega) = 0$ zaś $\int i_{\text{reg}}(h) dS$ jest określone, gdyż i_{reg} jest nieujemna.

- (ii) $D_{KL}(\nu||\mu) = \infty$, jeżeli $\nu \not\ll \mu$, gdyż $h = 1$ i $i_{\text{reg}}(h) = \infty$ na zbiorze dodatniej S -miary.
(iii) Równość $i_{\text{reg}}(x) = 0$ zachodzi wyłącznie dla $x = 1/2$, a zatem $D_{KL}(\nu||\mu) = 0$ wtedy i tylko wtedy, gdy $d\nu/dS = d\mu|_{\mathcal{G}_1}/dS$ S -prawie wszędzie. Ten warunek jest zaś równoważny $\nu = \mu|_{\mathcal{G}_1}$.

(iv) Jeżeli $D_{\text{KL}}(\nu|\mu) = \infty$, to $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) \leq D_{\text{KL}}(\nu|\mu)$ zachodzi automatycznie.

Jeżeli $D_{\text{KL}}(\nu|\mu) < \infty$, to $\nu \ll \mu$ na mocy punktu (ii). Możemy wówczas napisać $D_{\text{KL}}(\nu|\mu) = \int f(d\nu/d\mu)d\mu$ oraz $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) = \int f(d\nu|_{\mathcal{G}}/d\mu)d\mu$, gdzie $f(x) := x \log x$ dla $x > 0$ i $f(0) := 0$. Jeżeli A jest podzbiorem wszystkich punktów, dla których $d\nu|_{\mathcal{G}}/d\mu = 0$, to $d\nu/d\mu = 0$ na A także. A zatem $D_{\text{KL}}(\nu|\mu) = \int_{\Omega'} f(d\nu/d\mu)d\mu$ oraz $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) = \int_{\Omega'} f(d\nu|_{\mathcal{G}}/d\mu)d\mu$ dla $\Omega' := \Omega \setminus A$. Oznaczmy $g(x_0, x) := x \log x_0 + x - x_0$ dla $x_0 > 0$ i $g(0, x) = 0$. Funkcja f jest wypukła, więc dla $x \geq 0$, $x_0 > 0$ i pochodnej $f'(x_0) = \log x_0 + 1$ mamy $g(x_0, x) = f(x_0) + f'(x_0)(x - x_0) \leq f(x)$. Nierówność $g(x_0, x) \leq f(x)$ zachodzi także dla $x_0 = 0$. Ponieważ $\int_{\Omega'} \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} - \frac{d\nu}{d\mu} \right) d\mu = \nu(\Omega) - \nu(\Omega) = 0$, to zachodzi $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) = \int_{\Omega'} \log \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\nu|_{\mathcal{G}} = \int_{\Omega'} \log \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\nu = \int_{\Omega'} \frac{d\nu}{d\mu} \log \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\mu = \int_{\Omega'} g \left(\frac{d\nu|_{\mathcal{G}}}{d\mu}, \frac{d\nu}{d\mu} \right) \leq \int_{\Omega'} f \left(\frac{d\nu}{d\mu} \right) d\mu = D_{\text{KL}}(\nu|\mu)$. \square

Dowód nierówności $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) \leq D_{\text{KL}}(\nu|\mu)$ w zasadzie bazuje na dowodzie nierówności Jensena (Billingsley, 1979, sekcja 33), który uzupełniliśmy o wykazanie określoności całki $\int f' \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} \right) \left(\frac{d\nu}{d\mu} - \frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\mu$, a właściwie prostszej całki $\int \left(\frac{d\nu}{d\mu} - \frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\mu$. (Określoność całki $\int f' \left(\frac{d\nu|_{\mathcal{G}}}{d\mu} \right) \left(\frac{d\nu}{d\mu} - \frac{d\nu|_{\mathcal{G}}}{d\mu} \right) d\mu$ była milcząco założona w oryginale.)

Definicję $D_{\text{KL}}(\nu|\mu)$ jako $\int \log \left(\frac{d\nu}{d\mu} \right) d\nu$ dla $\nu \ll \mu$ oraz $D_{\text{KL}}(\nu|\mu) = \infty$ dla $\nu \not\ll \mu$ podaje np. Devroye (1987). Definicja ta, choć równoważna naszej, jest mniej przejrzysta. Po pierwsze, w definicji Devroye'a nie wiadomo, dlaczego należy przyjąć $D_{\text{KL}}(\nu|\mu) = \infty$ dla $\nu \not\ll \mu$. Po drugie, w niektórych przypadkach wielkość $\frac{D\nu}{D\mu}$ można wyrazić przez inne obiekty bez jawnego rozpatrywania, czy zachodzi $\nu \ll \mu$ czy też $\nu \not\ll \mu$.

W dodatku A.6 przedstawiamy definicje miary diagonalnej $P^{(n)}$ oraz warunkowej promiary produktowej $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$. Omawiamy też (dość eklektyczne) warunki, przy których skończenie addytywna funkcja $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ na $\mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$ daje się rozszerzyć do miary prawdopodobieństwa na σ -ciele produktowym $\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3$ (twierdzenie A.41). Jeżeli istotnie $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ daje się rozszerzyć do miary, to możemy zdefiniować warunkową informację między σ -ciałami $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$.

Definicja 1.14. (informacja między σ -ciałami) Jeżeli określona jest warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$, to warunkową informację wzajemną między σ -ciałami \mathcal{G}_1 i \mathcal{G}_2 względem \mathcal{G}_3 definiujemy jako

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) := D_{\text{KL}}(P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} | P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}) \geq 0. \quad (1.14)$$

Konstrukcja (1.14) jak najbardziej nawiązuje do poglądu wyrażanego przez Billingsleya (1965, sekcja 12; 1979, sekcja 33), że σ -ciała same w sobie reprezentują częściową informację o wyniku eksperymentu losowego.

Przypomnijmy, że w szczególności warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ istnieje dla $\mathcal{G}_3 = \{\emptyset, \Omega\}$ oraz dowolnych $\mathcal{G}_1, \mathcal{G}_2$. Oznacza to, że bezwarunkowa informacja wzajemna między \mathcal{G}_1 a \mathcal{G}_2 zdefiniowana jako $\bar{I}(\mathcal{G}_1; \mathcal{G}_2) := \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \{\emptyset, \Omega\}) = D_{\text{KL}}(P^{(2)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2} | P \times P)$ jest określona dla dowolnych σ -ciał \mathcal{G}_1 i \mathcal{G}_2 . Gelfand *et al.* (1956) pokazali, że wielkość $\bar{I}(\mathcal{G}_1; \mathcal{G}_2)$ można zdefiniować w jeszcze inny sposób.

Przez analogię do zmiennych dyskretnych określimy dyskretne σ -ciała.

Definicja 1.15. (σ -ciała dyskretne) Mówimy, że σ -ciało \mathcal{B} jest dyskretne, jeżeli $\mathcal{B} = \sigma(\mathcal{A}(\mathcal{B}))$, gdzie $\mathcal{A}(\mathcal{B})$ jest pewną przeliczalną klasą rozłącznych niepustych zdarzeń takich, że $\bigcup_{A \in \mathcal{A}(\mathcal{B})} A = \Omega$. Odpowiednio σ -ciało dyskretne \mathcal{B} nazywamy skończonym, jeżeli zbiór $\mathcal{A}(\mathcal{B})$ jest skończony.

Dla każdego σ -ciała dyskretnego \mathcal{B} zbiór $\mathcal{A}(\mathcal{B})$ dany jest jednoznacznie i nazywa się go zbiorem atomów (Billingsley, 1965, sekcja 12). Dla odróżnienia, podobne oznaczenie $\mathcal{A}(X)$ dla zmiennej losowej X oznacza alfabet tej zmiennej, czyli pewien zbiór $\mathcal{A}(X) \supset X(\Omega)$. (definicja A.5).

Zauważmy, że σ -ciała generowane przez zmienne losowe są dyskretnie wtedy i tylko wtedy, gdy zmienne te są dyskretnie. Co więcej, jakiegokolwiek dyskretnie σ -ciała można postrzegać jako σ -ciała generowane przez pewne abstrakcyjne dyskretnie zmienne losowe (istnienie takich zmiennych wynika z pewnika wyboru). Skończone σ -ciała są z kolei generowane przez zmienne o skończonym zbiorze wartości.

Wróćmy do miar informacji. Zachodzi tożsamość

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2) = \sup_{\mathcal{B}_1 \subset \mathcal{G}_1, \mathcal{B}_2 \subset \mathcal{G}_2} \sum_{A_1 \in \mathcal{A}(\mathcal{B}_1), A_2 \in \mathcal{A}(\mathcal{B}_2)} P(A_1 \cap A_2) \log \frac{P(A_1 \cap A_2)}{P(A_1)P(A_2)}, \quad (1.15)$$

gdzie \mathcal{B}_1 i \mathcal{B}_2 są dyskretnymi σ -ciałami zawartymi odpowiednio w \mathcal{G}_1 i \mathcal{G}_2 (Gelfand *et al.*, 1956, twierdzenie 4). Być może $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3)$ spełnia tożsamość analogiczną do (1.15), w której rolę $P(\cdot)$ pełni prawdopodobieństwo warunkowe $P(\cdot | \mathcal{G}_3)$. Taka hipotetyczna tożsamość mogłaby posłużyć do dalszego uogólnienia $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3)$ na przypadek, gdy nie istnieje warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$. Wzór (1.15) można stosować jako uogólnioną definicję informacji wzajemnej również w przypadku, gdy \mathcal{G}_1 i \mathcal{G}_2 są ciałami, a nie σ -ciałami (Gelfand *et al.*, 1956).

Gelfand *et al.* (1956, własności 1–4 oraz I–IV) podają bez dowodu kilka algebraicznych własności bezwarunkowej informacji wzajemnej $\bar{I}(\mathcal{G}_1; \mathcal{G}_2)$. Na pracę tę natknęliśmy się tuż przed złożeniem niniejszej rozprawy. Niezależnie wykazaliśmy wcześniej kilka analogicznych własności dla uogólnionej informacji wzajemnej warunkowej.

Twierdzenie 1.16. *Dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4$ prawdziwe są następujące stwierdzenia:*

- (i) *Równość $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) = 0$ zachodzi wtedy i tylko wtedy, gdy $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3$.*
- (ii) *Zachodzi równość $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_2; \mathcal{G}_1 | \mathcal{G}_3)$.*
- (iii) *Jeżeli $\mathcal{G}_2 \subset \mathcal{G}_3$, to zachodzi nierówność*

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4). \quad (1.16)$$

- (iv) *Nierówność (1.16) zachodzi także, jeżeli $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3 \oplus \mathcal{G}_4$ (definicja działania \oplus w dodatku A.6) i jednocześnie*

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 \oplus \mathcal{G}_3 | \mathcal{G}_4) = \bar{I}(\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4) + \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3 \oplus \mathcal{G}_4). \quad (1.17)$$

Uwaga 1. Jak się okaże dalej, hipotetyczna addytywność (1.17) jest pewnym uogólnieniem równości (1.11).

Uwaga 2. Warunek $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3 \oplus \mathcal{G}_4$ zachodzi w szczególności, gdy $\mathcal{G}_2 \subset \mathcal{G}_3$.

- (v) *Jeżeli addytywność (1.17) zachodzi dla $\mathcal{G}_3 = \mathcal{G}_1$, to zachodzi nierówność*

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_4). \quad (1.18)$$

Uwaga 3. Warunek (1.18) jest mocniejszym ograniczeniem wielkości $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4)$ niż ograniczenie $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4) \leq \bar{I}(\mathcal{G}_1 \oplus \mathcal{G}_2; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_4)$ wynikające bezpośrednio z pkt. (iii).

Dowód:

- (i) Z twierdzenia 1.13 pkt. (iii) wynika, że $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) = 0$ wtedy i tylko wtedy, gdy $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} = P^{(3)} |_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3}$, co jest równoważne $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3$ na mocy twierdzenia A.40.

- (ii) Funkcja podcałkowa w (1.14) jest symetryczna względem przestawienia $\mathcal{G}_1 \leftrightarrow \mathcal{G}_2$.
 (iii) Tezę wynika z nierówności $D_{\text{KL}}(\nu|_{\mathcal{G}}|\mu) \leq D_{\text{KL}}(\nu|\mu)$ i związku

$$\frac{dP^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_4}}{dP^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4}} = \frac{(dP^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_3 \otimes \mathcal{G}_4})|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_4}}{dP^{\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_4}} = \frac{(dP^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_3 \otimes \mathcal{G}_4})|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_4}}{dP^{\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4}}.$$

(iv) Mamy

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_2 \oplus \mathcal{G}_3 | \mathcal{G}_4) = \bar{I}(\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4) + \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3 \oplus \mathcal{G}_4) = \bar{I}(\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_4),$$

gdź $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3 \oplus \mathcal{G}_4) = 0$ na mocy punktu (i).

(v) Na mocy punktu (iii) zachodzi $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_4) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_2 \oplus \mathcal{G}_1 | \mathcal{G}_4)$, natomiast

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 \oplus \mathcal{G}_1 | \mathcal{G}_4) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_4) + \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_1 \oplus \mathcal{G}_4) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_4),$$

gdź $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2 | \mathcal{G}_1 \oplus \mathcal{G}_4$ implikuje $I(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_1 \oplus \mathcal{G}_4) = 0$.

□

Powyższe twierdzenie można prosto przełożyć na pewne twierdzenie o zmiennych losowych. Dla zmiennych losowych X, Y, W oznaczamy

$$\bar{I}(X; Y | W) := \bar{I}(\sigma(X); \sigma(Y) | \sigma(W)), \quad (1.19)$$

a także $\bar{I}(X; Y) := \bar{I}(\sigma(X); \sigma(Y))$, jeżeli wielkości po prawej stronie są określone. Numeracja punktów poniższego twierdzenia odwzorowuje numerację twierdzenia 1.16.

Twierdzenie 1.17. *Dla zmiennych losowych X, Y, Z, W zachodzą następujące fakty:*

- (0) Jeżeli $Y = f(Z)$, gdzie f jest bijektywnie mierzalne $\tilde{\sigma}(Z)/\tilde{\sigma}(Y)$, to $\sigma(Y) = \sigma(Z)$ (por. nieco ogólniejsze twierdzenie A.12), czyli $\bar{I}(X; Y | W) = \bar{I}(X; Z | W)$.
 (i) Równość $\bar{I}(X; Y | Z) = 0$ zachodzi wtedy i tylko wtedy, gdy $X \perp\!\!\!\perp Y | Z$.
 (ii) Zachodzi równość $\bar{I}(X; Y | Z) = \bar{I}(Y; X | Z)$.
 (iii) Jeżeli $Y = f(Z)$, gdzie f jest mierzalne $\tilde{\sigma}(Z)/\tilde{\sigma}(Y)$, to zachodzi nierówność

$$\bar{I}(X; Y | W) \leq \bar{I}(X; Z | W). \quad (1.20)$$

(iv) Nierówność (1.20) zachodzi także, jeżeli $X \perp\!\!\!\perp Y | Z \times W$ i jednocześnie

$$\bar{I}(X; Y \times Z | W) = \bar{I}(X; Z | W) + \bar{I}(X; Y | Z \times W). \quad (1.21)$$

(v) Jeżeli (1.21) zachodzi dla $Z = X$, to zachodzi nierówność

$$\bar{I}(X; Y | W) \leq \bar{I}(X; X | W). \quad (1.22)$$

W powyższym stwierdzeniu pkt. (0) i pkt. (iii) wzięły się z obserwacji, iż $Y = f(Z)$ implikuje $\sigma(Y) \subset \sigma(Z)$, jeżeli f jest mierzalne $\tilde{\sigma}(Z)/\tilde{\sigma}(Y)$, por. twierdzenie A.12.

Założmy chwilowo, że $I(X; Y | W) = \bar{I}(X; Y | W)$, jeżeli $I(X; Y | W)$ jest określone. W elementarnych podręcznikach nierówność (1.20), nazywana nierównością przetwarzania danych, jest dowodzona wyłącznie dla zmiennych dyskretnych w przypadku pkt. (iv) (Cover i Thomas, 1991, twierdzenie 2.8.1) za pomocą addytywności (1.11), odpowiednika (1.21). Dla tychże zmiennych dyskretnych przypadek pkt. (iii) zawiera się w przypadku pkt. (iv). Uogólnienie nierówności przetwarzania danych (1.20) na przypadek dowolnych zmiennych rzeczywistych z elementarnego punktu widzenia wygląda na nietrywialne. Próba wykazania tej nierówności przez dyskretyzację zmiennych i przejście z nią do granicy

wikłaby nas w wyrażenie całki Lebesgue'a całką Riemanna i konieczność czynienia nieprzejrzyistych założeń o odpowiedniej regularności funkcji podcałkowych.

Posłużenie się teoriomiarową definicją warunkowej miary produktowej pozwala bez większych kłopotów uzyskać bardzo mocne twierdzenia 1.16 i 1.17. W przeciwieństwie do $I(X; Y)$ wielkość $\bar{I}(X; Y)$ jest określona także wtedy, gdy X i Y są procesami stochastycznymi i nie istnieją standardowe miary pomiarowe μ_X, μ_Y .

W odróżnieniu od elementarnej teorii informacji, addytywność (1.21) jest w znacznej mierze hipotetyczna. Jest to ciekawe, gdyż dla zmiennych dyskretnych równość (1.11) wraz z własnością $I(X; Y|Z) = 0 \iff X \perp\!\!\!\perp Y|Z$ jest silnym narzędziem we wnioskowaniu o własnościach niezależności warunkowej (Yeung, 2002, sekcja 13.5). Możliwość niezachodzenia teoriomiarowej addytywności (1.21) nie wydaje się jednak zaskakująca. Niektóre własności warunkowej niezależności zależą od tego, czy dla danej miary przeniesionej istnieją nietrywialne zdarzenia o zerowym prawdopodobieństwie. Przykładowo, dla zmiennych dyskretnych takich, że $P(X = \cdot, Y = \cdot, Z = \cdot, W = \cdot) > 0$, zachodzi implikacja $X \perp\!\!\!\perp Y|W \times Z \wedge X \perp\!\!\!\perp Z|W \times Y \implies X \perp\!\!\!\perp Y \times Z|W$. Z drugiej strony implikacja ta nie zachodzi dla $Y = Z = (X, W)$ (Yeung, 2002, sekcja 2.1).

Wykażemy teraz, że istotnie $\bar{I}(X; Y|W)$ jest rozszerzeniem $I(X; Y|W)$. Przy okazji dowiedzimy, że $\bar{I}(X; Y|W)$ jest określone, gdy istnieje gęstość $\rho_{X,Y,W}$ względem pewnych miar pomiarowych, zaś addytywność (1.21) zachodzi, gdy istnieje gęstość $\rho_{X,Y,Z,W}$.

Przypomnijmy pojęcie nadreprezentacji, którym posługują się Kowalczyk *et al.* (2004). Dla standardowego układu miar pomiarowych i pary zmiennych X i Y nadreprezentacją (bezwarunkową) nazywa się funkcję $g_{X;Y}(x \times y) := \rho_{X \times Y}(x \times y) / (\rho_X(x)\rho_Y(y))$. Nadreprezentacją warunkową zmiennych X i Y względem zmiennej Z nazwiemy funkcję analogiczną, w której gęstości $\rho_W(w)$ dla $W = X, Y, X \times Y$ zastępuje się przez gęstości warunkowe $\rho_{W|Z}(w \times z) := \rho_{W \times Z}(w \times z) / \rho_Z(z)$, por. wzór (A.29).

Definicja 1.18. (nadreprezentacja warunkowa) *Ustalmy dowolny układ miar pomiarowych. Niech X, Y, Z będą dowolnymi zmiennymi losowymi, dla których gęstość $\rho_{X,Y,Z}$ jest określona. Warunkową nadreprezentacją nazywamy funkcję*

$$g_{X;Y|Z}(x \times y \times z) = \frac{\rho_{X,Y,Z}(x \times y \times z)\rho_Z(z)}{\rho_{X,Z}(x \times z)\rho_{Y,Z}(y \times z)}. \quad (1.23)$$

Wzięcie w definicji (1.23) gęstości $\rho_{X,Y,Z}$, $\rho_{X,Z}$ i $\rho_{Y,Z}$ zamiast $\rho_{X \times Y \times Z}$, $\rho_{X \times Z}$ i $\rho_{Y \times Z}$ skutkuje tym, że nadreprezentacja jest niezależna od przyjętych miar pomiarowych.

Twierdzenie 1.19. *Rozpatrzmy dowolny układ miar pomiarowych i zmienne X, Y, Z, W .*

(i) *Jeżeli istnieje gęstość $\rho_{X,Y,Z}$, to istnieje warunkowa miara produktowa $P^{\sigma(X); \sigma(Y)|\sigma(Z)}$ i zachodzą równości*

$$\frac{DP^{(3)}|_{\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)}}{DP^{\sigma(X); \sigma(Y)|\sigma(Z)}}(\omega_1 \times \omega_2 \times \omega_3) = g_{X;Y|Z}(X(\omega_1) \times Y(\omega_2) \times Z(\omega_3)), \quad (1.24)$$

$$\bar{I}(X; Y|Z) = \int \log g_{X;Y|Z} dP_{X \times Y \times Z} = \langle \log g_{X;Y|Z}(X \times Y \times Z) \rangle. \quad (1.25)$$

(ii) *Jeżeli dla standardowego układu miar pomiarowych określone jest $I(X; Y|Z)$, to $I(X; Y|Z) = \bar{I}(X; Y|Z)$.*

(iii) *Jeżeli $\mu_{Z \times W} = \mu_Z \times \mu_W$ oraz istnieje gęstość $\rho_{X,Y,Z,W}$, to zachodzi addytywność (1.21).*

Dowód:

(i) Zdefiniujmy zmienną tensorową

$$X \otimes Y \otimes Z : \Omega \times \Omega \times \Omega \ni (\omega_1 \times \omega_2 \times \omega_3) \mapsto (X(\omega_1) \times Y(\omega_2) \times Z(\omega_3)).$$

Zmienna ta jest izomorfizmem σ -ciał $\tilde{\sigma}(X) \otimes \tilde{\sigma}(Y) \otimes \tilde{\sigma}(Z)$ i $\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)$,

$$\tilde{\sigma}(X) \otimes \tilde{\sigma}(Y) \otimes \tilde{\sigma}(Z) \ni A' \mapsto (X \otimes Y \otimes Z)^{-1}(A') \in \sigma(X) \otimes \sigma(Y) \otimes \sigma(Z).$$

Wraz z σ -ciałami odwzorowywane są izomorficznie miary, np.

$$\begin{aligned} P^{(3)}|_{\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)} &= P_{X \times Y \times Z} \circ (X \otimes Y \otimes Z), \\ P_{X \times Y \times Z} &= P^{(3)}|_{\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)} \circ (X \otimes Y \otimes Z)^{-1}. \end{aligned}$$

Izomorficzne odwzorowanie dotyczy także relacji między miarami (absolutna ciągłość, wyrażalność miar przez produkty miar), wartości pochodnych Radona-Nikodyma oraz wartości całek funkcji odpowiednio mierzalnych (całkowanie przez podstawienie).

Dla $W_1 := X$, $W_2 := Y$ i $W_3 := Z$ oznaczmy $\mu_i := \mu_{W_i} \circ W_i$ i $\mu_{i_1, \dots, i_n} := \mu_{i_1} \times \dots \times \mu_{i_n}$, gdzie i_k są różne. Ponieważ $P^{(3)}|_{\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)} \ll \mu_{X,Y,Z}$, to z twierdzeń A.42 i A.41 pkt. (iv) wynika, że istnieje warunkowa miara produktowa $P^{\sigma(X); \sigma(Y) | \sigma(Z)}$. Oznaczywszy $\rho_{i_1, \dots, i_n} := dP^{(n)}|_{\sigma(W_{i_1}) \otimes \dots \otimes \sigma(W_{i_n})} / d\mu_{i_1, \dots, i_n}$, z pkt. (ii) twierdzenia A.42 mamy także

$$\begin{aligned} \frac{DP^{(3)}|_{\sigma(X) \otimes \sigma(Y) \otimes \sigma(Z)}}{DP^{\sigma(X); \sigma(Y) | \sigma(Z)}}(\omega_1 \times \omega_2 \times \omega_3) &= \frac{\rho_{1,2,3}(\omega_1 \times \omega_2 \times \omega_3) \rho_3(\omega_3)}{\rho_{1,3}(\omega_1 \times \omega_3) \rho_{2,3}(\omega_2 \times \omega_3)} \\ &= g_{X;Y|Z}(X(\omega_1) \times Y(\omega_2) \times Z(\omega_3)), \end{aligned}$$

gdyż $\rho_{i_1, \dots, i_n}(\omega_{i_1} \times \dots \times \omega_{i_n}) = \rho_{W_{i_1}, \dots, W_{i_n}}(W_{i_1}(\omega_{i_1}) \times \dots \times W_{i_n}(\omega_{i_n}))$. Wzór (1.25) jest konsekwencją (1.24) i definicji warunkowej informacji wzajemnej między σ -ciałami.

(ii) Niech μ_X , μ_Y , μ_Z będą standardowymi miarami pomiarowymi. Mamy $\rho_{X,Y,Z} = \rho_{X \times Y \times Z}$, $\rho_{X,Z} = \rho_{X \times Z}$, $\rho_{Y,Z} = \rho_{Y \times Z}$ oraz

$$\begin{aligned} I(X; Y|Z) &= \\ &\langle \log \rho_{X,Y,Z}(X \times Y \times Z) \rangle + \langle \log \rho_Z(Z) \rangle - \langle \log \rho_{X,Z}(X \times Z) \rangle - \langle \log \rho_{Y,Z}(Y \times Z) \rangle, \\ &= \langle \log \rho_{X,Y,Z}(X \times Y \times Z) + \log \rho_Z(Z) - \log \rho_{X,Z}(X \times Z) - \log \rho_{Y,Z}(Y \times Z) \rangle \\ &= \langle \log g_{X;Y|Z}(X \times Y \times Z) \rangle = \bar{I}(X; Y|Z). \end{aligned}$$

(iii) Z $\mu_{Z \times W} = \mu_Z \times \mu_W$ mamy $\rho_{U,Z \times W} = \rho_{U,Z,W}$ dla dowolnej zmiennej U . Zatem

$$\begin{aligned} &g_{X;Z|W}(x \times z \times w) g_{X;Y|Z \times W}(x \times y \times z \times w) \\ &= \frac{\rho_{X,Z,W}(x \times z \times w) \rho_W(w)}{\rho_{X,W}(x \times w) \rho_{Z,W}(z \times w)} \cdot \frac{\rho_{X,Y,Z \times W}(x \times y \times z \times w) \rho_{Z \times W}(z \times w)}{\rho_{X,Z \times W}(x \times z \times w) \rho_{Y,Z \times W}(y \times z \times w)} \\ &= g_{X;Y \times Z|W}(x \times y \times z \times w). \end{aligned}$$

Stąd

$$\begin{aligned} &\bar{I}(X; Z|W) + \bar{I}(X; Y|Z \times W) \\ &= \langle \log g_{X;Z|W}(X \times Z \times W) \rangle + \langle \log g_{X;Y|Z \times W}(X \times Y \times Z \times W) \rangle \\ &= \langle \log g_{X;Z|W}(X \times Z \times W) + \log g_{X;Y|Z \times W}(X \times Y \times Z \times W) \rangle \\ &= \langle \log g_{X;Y \times Z|W}(X \times Y \times Z \times W) \rangle = \bar{I}(X; Y \times Z|W). \end{aligned}$$

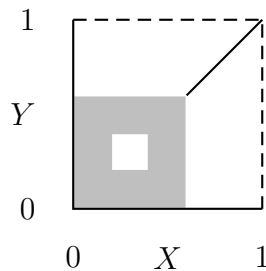
□

Na mocy twierdzenia 1.19 nadreprezentacja $g_{X;Y}(x \times y) := g_{X;Y|1}(x \times y \times 1) = \rho_{X;Y}(x \times y) / (\rho_X(x) \rho_Y(y))$ równa jest niewłaściwej pochodnej Radona-Nikodyma

$$\frac{DP_{X \times Y}}{D(P_X \times P_Y)} = \frac{dP_{X \times Y}}{d(\mu_X \times \mu_Y)} : \frac{d(P_X \times P_Y)}{d(\mu_X \times \mu_Y)} = g_{X;Y}, \quad (1.26)$$

gdyż $P_{X \times Y}, P_X \times P_Y \ll \mu_X \times \mu_Y$. Z drugiej strony nie możemy w ogólności napisać $g_{X;Y} = dP_{X \times Y} / d(P_X \times P_Y)$, gdyż nie możemy zapewnić $P_{X \times Y} \ll P_X \times P_Y$. Dla pary zmiennych dyskretnych X i Y relacja $P_{X \times Y} \ll P_X \times P_Y$ zachodzi zawsze, ale dla innych zmiennych może być inaczej.

Przykład 1.20. Prosty przykład zmiennych, dla których nie zachodzi $P_X \times P_Y \ll P_{X \times Y}$ ani $P_{X \times Y} \ll P_X \times P_Y$, to zmienne rzeczywiste $X, Y \in [0, 1]$ o mierze $P_{X \times Y}$:



Szary obszar to miara proporcjonalna do miary Lebesgue'a na płaszczyźnie, zaś czarna linia to miara proporcjonalna do miary Lebesgue'a na przekątnej. W białym kwadracie zachodzi $P_X \times P_Y \ll P_{X \times Y}$, zaś na czarnej linii mamy $P_{X \times Y} \ll P_X \times P_Y$.

Istotne dla dowodu addytywności (1.21) w twierdzeniu 1.19 było założenie, że istnieją takie miary $\mu_X, \mu_Y, \mu_Z, \mu_W$, że $P_{X \times Y \times Z \times W} \ll \mu_X \times \mu_Y \times \mu_Z \times \mu_W$. Twierdzenie 1.19 możemy bowiem łatwo przełożyć na język σ -ciał.

Twierdzenie 1.21.

- (i) Warunkowa informacja wzajemna $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3)$ jest określona, jeżeli istnieją miary μ_1, μ_2, μ_3 takie, że $P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} \ll \mu_1 \times \mu_2 \times \mu_3$.
- (ii) Addytywność (1.17) zachodzi dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3, \mathcal{G}_4$, jeżeli istnieją miary $\mu_1, \mu_2, \mu_3, \mu_4$ takie, że $P^{(4)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3 \otimes \mathcal{G}_4} \ll \mu_1 \times \mu_2 \times \mu_3 \times \mu_4$.

Dowód: Skonstruujemy zmienne $X_i : \Omega \ni \omega \mapsto \omega$, gdzie $\sigma(X_i) = \mathcal{G}_i$, i obierzmy miary pomiarowe $\mu_{X_i} = \mu_i$ oraz $\mu_{X_i \times X_j} = \mu_{X_i} \times \mu_{X_j}$. Teza wynika z pkt. (i) i (iii) twierdzenia 1.19. \square

Twierdzenie 1.22. Dla σ -ciał $\mathcal{G}_i, 1 \leq i \leq n$, istnieją miary μ_i takie, że $P^{(n)}|_{\mathcal{G}_1 \otimes \dots \otimes \mathcal{G}_n} \ll \mu_1 \times \dots \times \mu_n$, jeżeli co najwyżej jedno \mathcal{G}_i nie jest σ -ciałem dyskretnym.

Dowód: Bez zmniejszania ogólności załóżmy, że dyskretne są \mathcal{G}_i , gdzie $1 \leq i \leq n-1$. Dla tychże \mathcal{G}_i mamy $\mathcal{A}(\mathcal{G}_i) = \{A_{ik} : k \in \mathbb{N}\}$, gdzie A_{ik} są rozłączne. Ponieważ $\Omega = \bigcup_{k \in \mathbb{N}} A_{ik}$, to dowolny zbiór $A \in \mathcal{G}_1 \otimes \dots \otimes \mathcal{G}_{n-1} \otimes \mathcal{G}_n$ ma przedstawienie w terminach przeliczalnej sumy rozłącznych zbiorów

$$\begin{aligned} A &= A \cap (\Omega \times \dots \times \Omega) \\ &= A \cap \bigcup_{k_1, \dots, k_{n-1} \in \mathbb{N}} A_{1, k_1} \times \dots \times A_{n-1, k_{n-1}} \times \Omega \\ &= \bigcup_{k_1, \dots, k_{n-1} \in \mathbb{N}} A_{1, k_1} \times \dots \times A_{n-1, k_{n-1}} \times B_{k_1, \dots, k_{n-1}}. \end{aligned}$$

Jako miary pomiarowe obierzemy $\mu_i = P|_{\mathcal{G}_i}$. Ponieważ prawdopodobieństwo jest przeliczalnie addytywne, aby wykazać $P^{(n)}|_{\mathcal{G}_1 \otimes \dots \otimes \mathcal{G}_n} \ll \mu_1 \times \dots \times \mu_n$ wystarczy pokazać, że z $P(A_{1,k_1}) \cdot \dots \cdot P(A_{n-1,k_{n-1}}) \cdot P(B_{k_1, \dots, k_{n-1}}) = 0$ wynika $P(C_{k_1, \dots, k_{n-1}}) = 0$ dla $C_{k_1, \dots, k_{n-1}} = A_{1,k_1} \cap \dots \cap A_{n-1,k_{n-1}} \cap B_{k_1, \dots, k_{n-1}}$. To zaś jest oczywiste, gdyż $P(C_{k_1, \dots, k_{n-1}}) \leq P(A_{i,k_i})$ oraz $P(C_{k_1, \dots, k_{n-1}}) \leq P(B_{k_1, \dots, k_{n-1}})$, a zatem $P(A_{1,k_1}) \cdot \dots \cdot P(A_{n-1,k_{n-1}}) \cdot P(B_{k_1, \dots, k_{n-1}}) \geq [P(C_{k_1, \dots, k_{n-1}})]^n$. \square

Z twierdzeń 1.21 i 1.22 wynika określoność entropii warunkowych dla procesów dyskretnych. Najpierw jednak zdefiniujemy teoriomiarową entropię warunkową w przypadku ogólnym.

Definicja 1.23. (entropia σ -ciała) *Jeżeli określona jest warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2}$ dla σ -ciała $\mathcal{G}_1, \mathcal{G}_2$, to warunkową entropię σ -ciała \mathcal{G}_1 względem \mathcal{G}_2 definiujemy jako*

$$\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) := \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2) = D_{KL}(P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_1 \otimes \mathcal{G}_2} || P^{\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2}) \geq 0. \quad (1.27)$$

W szczególności warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2}$ istnieje dla $\mathcal{G}_2 = \{\emptyset, \Omega\}$ oraz dowolnego \mathcal{G}_1 . Zatem bezwarunkowa entropia teoriomiarowa zdefiniowana jako $\bar{H}(\mathcal{G}_1) := \bar{H}(\mathcal{G}_1 | \{\emptyset, \Omega\})$ jest określona dla dowolnego σ -ciała \mathcal{G}_1 . Definicja postaci (1.27) jest słabo znana w literaturze z powodu pewnych pozornie paradoksalnych jej własności. Nawet Billingsley (1965, sekcja 12), który uogólnia entropię warunkową $H(\mathcal{G}_1 | \mathcal{G}_2)$ na przypadek σ -ciała dyskretnego \mathcal{G}_1 i dowolnego σ -ciała \mathcal{G}_2 , odwołuje się do zupełnie innej konstrukcji. W dalszej dyskusji pokażemy, że definicja (1.27) jest uogólnieniem definicji zaproponowanej przez Billingsleya.

Dla zmiennych losowych X, Y oznaczamy

$$\bar{H}(X | Y) := \bar{H}(\sigma(X) | \sigma(Y)), \quad (1.28)$$

a także $\bar{H}(X) := \bar{H}(\sigma(X))$. W przeciwieństwie do informacji wzajemnej zachodzenie równości $\bar{H}(X) = H(X)$ dla entropii zależy istotnie od tego, czy miara przeniesiona P_X jest bezatomowa.

Definicja 1.24. (ciągła zmienna rzeczywista) *Rzeczywistą zmienną losową Y nazywamy zmienną ciągłą, jeżeli zachodzi $P_Y \ll m$, gdzie m jest miarą Lebesgue'a na \mathcal{R} .*

Twierdzenie 1.25. *Dla dowolnej ciągłej zmiennej rzeczywistej Y zachodzi $\bar{H}(Y) := \infty$.*

Dowód: Oznaczmy przekątną $D = \{(r, r) : r \in \mathbb{R}\}$. Zauważmy, że $P_Y^{(2)}(D) = 1$ natomiast $(P_Y \times P_Y)(D) = 0$, gdyż $(m \times m)(D) = 0$ zaś $P_Y \times P_Y \ll m \times m$ z założenia. Ponieważ $P_Y^{(2)} \not\ll P_Y \times P_Y$, to $\bar{H}(Y) = \bar{I}(Y; Y) = D_{KL}(P_Y^{(2)} || P_Y \times P_Y) = \infty$. \square

Twierdzenie 1.26. *Niech X będzie dowolną zmienną dyskretną zaś W dowolną zmienną. Zachodzą następujące fakty:*

- (i) *Określona jest teoriomiarowa entropia warunkowa $\bar{H}(X | W)$.*
- (ii) *Jeżeli dla standardowego układu miar pomiarowych określone jest $H(X | W)$, to zachodzi równość*

$$H(X | W) = \bar{H}(X | W). \quad (1.29)$$

Dowód:

- (i) Teza wynika z twierdzenia 1.22 i pkt. (i) twierdzenia 1.21.

(ii) Równość wynika z równości

$$\begin{aligned} g_{X;X|W}(x \times x' \times w) &= \rho_{X,X,W}(x \times x' \times w) \rho_W(w) / \rho_{X,W}(x \times w) \rho_{X,W}(x \times w) \\ &= \llbracket x = x' \rrbracket \rho_W(w) / \rho_{X,W}(x \times w), \end{aligned}$$

gdyż $\rho_{X,X,W}(x \times x' \times w) = \llbracket x = x' \rrbracket \rho_{X \times W}(x \times w)$. □

Z powodu tożsamości (1.29) niekiedy entropię określa się mianem *self-information* (po angielsku). Intrzygujące jest, że teoriomiarowa informacja wzajemna $\bar{I}(X; X|W)$ w szczególności przez równość $H(X|W) = \bar{I}(X; X|W)$ wyróżnia miarę liczącą $\mu_X = \text{card}$. Trzeba podkreślić, że równość $H(X|W) = \bar{I}(X; X|W)$ zachodzi wyłącznie dla zmiennej dyskretnej, ale za to również wtedy, gdy $H(X|W)$ jest nieskończone i gdy z tego powodu $I(X; X|W)$ nie istnieje.

Cover *et al.* (1989, strona 848, *epsilon entropy*) wspominają o własności dowiedzionej w twierdzeniu 1.25. Ich zdaniem własność ta przesądza o bezużyteczności pojęcia entropii teoriomiarowej w teorii kodowania. W dalszych rozdziałach pokażemy, że jest to niczym nieuzasadnione uprzedzenie. Twierdzenia 4.9, 5.1 i 6.12 posiłkują się pojęciem entropii teoriomiarowej w sposób istotny.

Własność $\bar{H}(Y) = \infty$ dla ciągłej zmiennej rzeczywistej Y nie wyda się paradoksalna, jeżeli uświadomimy sobie, że zmienna taka jest w intuicyjnym sensie nieskończonym źródłem informacji. W szczególności istnieje ciąg niezależnych zmiennych losowych o identycznym rozkładzie $Z_{\mathbb{N}}$ takich, że $Y = f(Z_{\mathbb{N}})$ dla pewnej funkcji f bijektywnie mierzalnej (por. dodatek A.3).

Definicja 1.27. (proces Bernoulliego) *Procesem Bernoulliego $Z_{\mathbb{N}}$ nazywamy ciąg niezależnych zmiennych losowych takich, że $Z_i(\Omega) = \{0, 1\}$, $P(Z_i = 1) = P(Z_i = 0) = 1/2$.*

Twierdzenie 1.28. *Dla ciągłej rzeczywistej zmiennej losowej Y istnieje proces Bernoulliego $Z_{\mathbb{N}}$ taki, że $Y = f(Z_{\mathbb{N}})$ prawie na pewno dla pewnej funkcji f bijektywnie mierzalnej.*

Dowód: Zdefiniujmy dystrybuantę zmiennej rzeczywistej Y jako $F_Y(y) := P(Y \leq y) = P_Y(\{r \in \mathbb{R} : r \leq y\})$. Jeżeli $P_Y \ll m$, to F_Y ma pochodną $F'_Y(y) = (dP_Y/dm)(y)$ (Rudin, 1986, sekcje 8.1–8.6), czyli jest ciągła. Stąd $F_Y(\mathbb{R}) \supset (0, 1)$. Określmy funkcję kwantylową $F^{-1} : (0, 1) \ni p \mapsto \inf \{y : F_Y(Y) = p\}$. Funkcję odwrotną do F^{-1} oznaczymy jako F . Istnieje zmienna rzeczywista \bar{Y} taka, że $Y = F^{-1}(\bar{Y})$ i $\bar{Y} = F(Y) = F_Y(Y)$ prawie na pewno. Zauważmy, że $P_{\bar{Y}}((a, b)) = P_Y(F_Y^{-1}((a, b))) = b - a = m((a, b))$ dla odcinków $(a, b) \subset (0, 1)$, czyli miara przeniesiona \bar{Y} równa jest mierze Lebesgue'a $P_{\bar{Y}} = m|_{\mathcal{R}_{(0,1)}}$. Z twierdzenia A.16 wynika także, że istnieje proces $Z_{\mathbb{N}}$ dany mierzalną bijekcją $\bar{Y} = g_2(Z_{\mathbb{N}}) := \sum_{k=1}^{\infty} 2^{-k} Z_k$, gdzie $Z_i \in \{0, 1\}$, $P(\limsup_{n \rightarrow \infty} Z_n = 0) = 0$, $P(\limsup_{n \rightarrow \infty} Z_n = 1) = 0$. Ponieważ $P_{\bar{Y}} = m|_{\mathcal{R}_{(0,1)}}$, to łatwo można sprawdzić, że $Z_{\mathbb{N}}$ jest ciągiem niezależnych zmiennych losowych takich, że $P(Z_i = 1) = P(Z_i = 0) = 1/2$. Mamy także równość $Y = f(Z_{\mathbb{N}})$ prawie na pewno dla $f = F^{-1} \circ g_2$. □

Z twierdzeń 1.28 i 1.26 pkt. (ii) wynika alternatywny dowód nieskończoności teoriomiarowej entropii dla ciągłej rzeczywistej zmiennej Y . Z nierówności przetwarzania danych mamy bowiem

$$\bar{H}(Y) = \bar{H}(Z_{\mathbb{N}}) \geq \sup_{n \in \mathbb{N}} \bar{H}(Z_{1:n}) = \sup_{n \in \mathbb{N}} H(Z_{1:n}) = \sup_{n \in \mathbb{N}} n \log 2 = \infty. \quad (1.30)$$

Pomimo tego, że teoriomiarowe entropie zmiennych ciągłych są nieskończone, informacje wzajemne pomiędzy różnymi zmiennymi ciągłymi mogą być skończone. Tak dzieje się w przypadku procesów gaussowskich.

Niech \mathcal{G}^P będzie uzupełnieniem σ -ciała \mathcal{G} względem miary P (definicja A.18). Można pokazać, że zachodzi równość

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1^P; \mathcal{G}_2^P | \mathcal{G}_3^P). \quad (1.31)$$

Zarys rozumowania wyprowadzającego (1.31) przedstawiamy w dodatku A.6. Mamy także następujące twierdzenie.

Twierdzenie 1.29. *Równość $\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) = 0$ zachodzi wtedy i tylko wtedy, gdy $\mathcal{G}_1^P \subset \mathcal{G}_2^P$.*

Dowód: Równość $\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2) = 0$ jest równoważna $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_1 | \mathcal{G}_2$. Relacja $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_1 | \mathcal{G}_2$ zachodzi z kolei wtedy i tylko wtedy, gdy $P(P(A_1 | \mathcal{G}_2) P(A_1 | \mathcal{G}_2) = P(A_1 | \mathcal{G}_2)) = 1$, czyli gdy $P(P(A_1 | \mathcal{G}_2) \in \{0, 1\}) = 1$ dla każdego $A_1 \in \mathcal{G}_1$. Z twierdzenia A.31 $P(P(A_1 | \mathcal{G}_2) \in \{0, 1\}) = 1$ zachodzi wtedy i tylko wtedy, gdy istnieje $A_2 \in \mathcal{G}_2$ takie, że $P(A_1 \ominus A_2) = 0$. W istocie takie A_2 istnieje wtedy i tylko wtedy, gdy $\mathcal{G}_1^P \subset \mathcal{G}_2^P$. \square

Twierdzenie 1.29 ma kilka ciekawych konsekwencji:

- (i) Równość $\bar{H}(\mathcal{G}) = 0$ zachodzi wtedy i tylko wtedy, gdy $P(A) \in \{0, 1\}$ dla każdego $A \in \mathcal{G}$.
- (ii) Z warunków $\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) = 0$ i $\bar{H}(\mathcal{G}_2 | \mathcal{G}_3) = 0$ wynika $\bar{H}(\mathcal{G}_1 | \mathcal{G}_3) = 0$, gdyż relacja \subset jest przechodnia. Ponadto dla dowolnej zmiennej Y mamy $\bar{H}(Y | Y) = 0$, nawet gdy Y jest ciągłą zmienną rzeczywistą.
- (iii) Z twierdzenia A.20 dla dowolnej zmiennej dyskretnej X równość $\bar{H}(X | Y) = 0$ zachodzi wtedy i tylko wtedy, gdy istnieje funkcja f mierzalna $\bar{\sigma}(Y)/\bar{\sigma}(X)$ taka, że $X = f(Y)$ prawie na pewno. Fakt ten można uogólnić na przypadek, gdy X jest rzeczywistą bądź borelowską zmienną losową.

Twierdzenie 1.30. (por. Billingsley, 1965, sekcja 12) *Dla σ -ciał dyskretnych $\mathcal{G}_1, \mathcal{G}_2$ oraz dowolnego σ -ciała \mathcal{G}_3 zachodzą następujące związki:*

$$\bar{H}(\mathcal{G}_1 | \mathcal{G}_3) \leq \bar{H}(\mathcal{G}_2 | \mathcal{G}_3), \quad \text{jeżeli } \mathcal{G}_1 \subset \mathcal{G}_2, \quad (1.32)$$

$$\bar{H}(\mathcal{G}_1 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) + \bar{H}(\mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3), \quad (1.33)$$

$$\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) = \bar{I}(\mathcal{G}_1; \mathcal{G}_3 | \mathcal{G}_2) + \bar{H}(\mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3), \quad (1.34)$$

$$\bar{H}(\mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) = \bar{H}(\mathcal{G}_2 | \mathcal{G}_3) + \bar{H}(\mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3), \quad (1.35)$$

$$\bar{H}(\mathcal{G}_1 | \mathcal{G}_2) \geq \bar{H}(\mathcal{G}_1 | \mathcal{G}_3), \quad \text{jeżeli } \mathcal{G}_2 \subset \mathcal{G}_3. \quad (1.36)$$

Dowód: (1.32) wynika z nierówności przetwarzania danych (twierdzenie 1.16, pkt. (iii)), $\bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_3) \leq \bar{I}(\mathcal{G}_2; \mathcal{G}_2 | \mathcal{G}_3)$.

Z twierdzeń 1.22, 1.21 i 1.16 pkt. (iii) i (v), mamy $\bar{I}(\mathcal{G}_1; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) \geq \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_3)$ i $\bar{I}(\mathcal{G}_1; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_3)$, czyli

$$\bar{I}(\mathcal{G}_1; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_3). \quad (1.37)$$

A zatem $\bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3) + \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_3)$, czyli zachodzi (1.33). Zamieniając w rozumowaniu \mathcal{G}_2 i \mathcal{G}_3 dowodzimy (1.34). Mamy także $\bar{I}(\mathcal{G}_1 \oplus \mathcal{G}_2; \mathcal{G}_1 \oplus \mathcal{G}_2 | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1 \oplus \mathcal{G}_2; \mathcal{G}_2 | \mathcal{G}_3) + \bar{I}(\mathcal{G}_1 \oplus \mathcal{G}_2; \mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3)$, czyli (1.35).

Nierówność (1.36) wynika z (1.34), gdyż $\bar{H}(\mathcal{G}_1 | \mathcal{G}_2 \oplus \mathcal{G}_3) = \bar{H}(\mathcal{G}_1 | \mathcal{G}_3)$ dla $\mathcal{G}_2 \subset \mathcal{G}_3$. \square

Dla σ -ciała skończonego \mathcal{G}_1 teoriomiarowa entropia $H(\mathcal{G}_1|\mathcal{G}_2)$ pokrywa się z wielkością $\left\langle -\sum_{A \in \mathcal{A}(\mathcal{G}_1)} P(A|\mathcal{G}_2) \log P(A|\mathcal{G}_2) \right\rangle$ zaproponowaną przez Billingsleya (1965, sekcja 12) jako uogólnienie entropii warunkowej dla σ -ciał. Mamy bowiem następujące twierdzenie uogólniające elementarne własności entropii warunkowej, por. Cover i Thomas (1991, lemat 2.1.1, twierdzenie 2.6.4):

Twierdzenie 1.31. *Niech \mathcal{G}_1 będzie σ -ciałem dyskretnym zaś \mathcal{G}_2 dowolnym σ -ciałem. Zachodzą relacje*

$$\bar{H}(\mathcal{G}_1|\mathcal{G}_2) = \int \left[-\sum_{A \in \mathcal{A}(\mathcal{G}_1)} P(A|\mathcal{G}_2) \log P(A|\mathcal{G}_2) \right] dP, \quad (1.38)$$

$$\bar{H}(\mathcal{G}_1|\mathcal{G}_2) \leq \log \text{card } \mathcal{A}(\mathcal{G}_1). \quad (1.39)$$

Dowód: Dowolny zbiór $B \in \mathcal{G}_1 \otimes \mathcal{G}_1 \otimes \mathcal{G}_2$ ma przedstawienie w terminach przeliczalnej sumy rozłącznych zbiorów

$$B = \bigcup_{A, A' \in \mathcal{A}(\mathcal{G}_1)} A \times A' \times B_{A, A'}, \quad B_{A, A'} \in \mathcal{G}_2.$$

Oznaczmy $I_A(\omega) := \llbracket \omega \in A \rrbracket$. Równość $P^{\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2}(A \times A' \times B_{A, A'}) = \int_{B_{A, A'}} P(A|\mathcal{G}_2) I_{A'} dP$ zachodzi na mocy twierdzenia A.32, gdyż funkcja $P(A'|\mathcal{G}_2)$ jest mierzalna $\mathcal{G}_2/\mathcal{R}$. Ponieważ $P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_1 \otimes \mathcal{G}_2}(A \times A' \times B_{A, A'}) = \int_{B_{A, A'}} I_A I_{A'} dP$, więc dla pewnych wersji pochodnych Radona-Nikodyma mamy

$$\frac{DP^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_1 \otimes \mathcal{G}_2}}{DP^{\mathcal{G}_1; \mathcal{G}_1 | \mathcal{G}_2}}(\omega_1 \times \omega_2 \times \omega_3) = \sum_{A \in \mathcal{A}(\mathcal{G}_1)} \frac{I_A(\omega_3)}{P(I_A|\mathcal{G}_2)(\omega_3)}.$$

Z definicji entropii warunkowej (1.29) wnioskujemy, że

$$\begin{aligned} \bar{H}(\mathcal{G}_1|\mathcal{G}_2) &= \int \log \left(\sum_{A \in \mathcal{A}(\mathcal{G}_1)} \frac{I_A(\omega_3)}{P(I_A|\mathcal{G}_2)(\omega_3)} \right) dP^{(3)}(\omega_1 \times \omega_2 \times \omega_3) \\ &= \int \left[-\sum_{A \in \mathcal{A}(\mathcal{G}_1)} I_A \log P(A|\mathcal{G}_2) \right] dP. \end{aligned} \quad (1.40)$$

Przejsie od (1.40) do (1.38) wynika z twierdzenia A.32 oraz twierdzenia o zbieżności monotonicznej (przeliczalne sumowania w obu wzorach można wyciągnąć przed całki), jako że funkcje sumowane w obu wzorach są mierzalne $\mathcal{G}_2/\mathcal{R}$ i nieujemne. Cover i Thomas (1991, twierdzenie 2.6.4) dowodzą, że dla dowolnego rozkładu dyskretnego p spełniającego $\sum_{x \in G} p(x) = 1$, $p(x) \geq 0$, zachodzi $-\sum_{x \in G} p(x) \log p(x) \leq \log \text{card } G$. Stąd (1.38) implikuje (1.39). \square

Z twierdzenia 1.31 wynika uogólnienie nierówności (1.36) na przypadek dowolnego \mathcal{G}_2 , por. Billingsley (1965, sekcja 12).

Twierdzenie 1.32. *Dla σ -ciała dyskretnego \mathcal{G}_1 oraz dowolnych σ -ciał $\mathcal{G}_2, \mathcal{G}_3$ mamy*

$$\bar{H}(\mathcal{G}_1|\mathcal{G}_2) \geq \bar{H}(\mathcal{G}_1|\mathcal{G}_3), \quad \text{jeżeli } \mathcal{G}_2 \subset \mathcal{G}_3. \quad (1.41)$$

Dowód: Oznaczmy $f(x) := x \log x$ dla $x > 0$ i $f(0) := 0$, a także Oznaczmy $g(x_0, x) := x \log x_0 + x - x_0$ dla $x_0 > 0$ i $g(0, x) = 0$. Funkcja f jest wypukła, więc dla $x \geq 0$, $x_0 > 0$ i pochodnej $f'(x_0) = \log x_0 + 1$ mamy $g(x_0, x) = f(x_0) + f'(x_0)(x - x_0) \leq f(x)$. Nierówność $g(x_0, x) \leq f(x)$ zachodzi także dla $x_0 = 0$.

Jako że $\int \sum_{A \in \mathcal{A}(\mathcal{G}_1)} [P(A|\mathcal{G}_3) - P(A|\mathcal{G}_2)] dP = P(\Omega|\mathcal{G}_3) - P(\Omega|\mathcal{G}_2) = 0$, mamy

$$\begin{aligned} \bar{H}(\mathcal{G}_1|\mathcal{G}_2) &= \int \left[- \sum_{A \in \mathcal{A}(\mathcal{G}_1)} I_A \log P(A|\mathcal{G}_2) \right] dP = \int \left[- \sum_{A \in \mathcal{A}(\mathcal{G}_1)} P(A|\mathcal{G}_3) \log P(A|\mathcal{G}_2) \right] dP \\ &= \int \left[- \sum_{A \in \mathcal{A}(\mathcal{G}_1)} g(P(A|\mathcal{G}_2), P(A|\mathcal{G}_3)) \right] dP \leq \int \left[- \sum_{A \in \mathcal{A}(\mathcal{G}_1)} f(P(A|\mathcal{G}_3)) \right] dP \\ &= H(\mathcal{G}_1|\mathcal{G}_2). \end{aligned}$$

W przekształceniach wykorzystaliśmy fakt, że jeżeli $\log P(A|\mathcal{G}_2)$ jest mierzalna $\mathcal{G}_2/\mathcal{R}$, to jest także mierzalna $\mathcal{G}_3/\mathcal{R}$. \square

W dowodzie twierdzenia 1.32 śledziliśmy wywód nierówności Jensena, aby pokazać, że nie ma kłopotów z całkowalnością — analogicznie jak w przypadku twierdzenia 1.13. Jest dobrym pytaniem, czy nierówność (1.41) można uogólnić na przypadek dowolnego \mathcal{G}_1 .

Kończąc rozważania o podstawowych miarach informacji, zauważmy jeszcze, że analogicznie do informacji wzajemnej i entropii warunkowej definicję potrójnej informacji wzajemnej $V(X; Y; Z)$ można rozszerzyć jako

$$\bar{V}(\mathcal{G}_1; \mathcal{G}_2; \mathcal{G}_3) := D_{\text{KL}}(P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} || P \times P \times P) \geq 0. \quad (1.42)$$

(Z własności tej nie będziemy korzystać, więc dowód opuszczamy.)

W przeciwieństwie do potrójnej informacji wzajemnej typu I, potrójna informacja wzajemna typu II nie ma ustalonego znaku i przez to w ogólnym przypadku może być nieokreślona. W przypadku elementarnym przykładowo mamy $I(X; Y; Z) = I(X; Y) - I(X; Y|Z) = -I(X; Y|Z) \leq 0$ dla $X \perp\!\!\!\perp Y$ i $Z = g(X, Y)$, natomiast $I(X; Y; Z) = I(X; Y) - I(X; Y|Z) = I(X; Y) \geq 0$ dla $X = f(Z)$ i $Y = h(Z)$, por. Yeung (2002, sekcja 6.4). Oznacza to, że dla warunkowej informacji wzajemnej nie ma analogonu nierówności (1.41) typu $\bar{I}(\mathcal{G}_1; \mathcal{G}_2|\mathcal{G}_3) \geq \bar{I}(\mathcal{G}_1; \mathcal{G}_2|\mathcal{G}_4)$, gdy $\mathcal{G}_3 \subset \mathcal{G}_4$ oraz $\mathcal{G}_1 \neq \mathcal{G}_2$.

Rozdział 2

Graniczne miary informacji procesu

W tym rozdziale będziemy rozważać procesy stacjonarne $X_{\mathbb{Z}}$ takie, że dla każdej wielowymiarowej zmiennej $X_{m:n} = X_m \times \dots \times X_n$, $n \geq m$, $n, m \in \mathbb{Z}$:

- (i) określona jest standardowa miara pomiarowa $\mu_{X_{m:n}} = \mu_{X_m} \times \dots \times \mu_{X_n}$,
- (ii) istnieje gęstość prawdopodobieństwa $\rho_{X_{m:n}}$,
- (iii) entropia $H(X_{m:n})$ są określona i skończona.

Z rozdziału 1 wiemy, że dla takich procesów $I(X_{m:n}; X_{k:l}|X_{i:j}) = \bar{I}(X_{m:n}; X_{k:l}|X_{i:j})$, a gdy zmienne X_i są dyskretne, mamy także $H(X_{m:n}|X_{k:l}) = \bar{H}(X_{m:n}|X_{k:l})$. Przypomnijmy też, że dla procesów stacjonarnych rozkłady prawdopodobieństwa skończonych podzbiorów zmiennych losowych są niezmiennicze ze względu na translacje. Własność ta implikuje pewne szczególne własności entropii.

Definicja 2.1. (entropia blokowa) Dla procesu stacjonarnego $X_{\mathbb{Z}}$ entropią blokową nazywa się funkcję

$$H_X(n) := \begin{cases} 0, & n = 0; \\ H(X_{i+1:i+n}), & n \geq 1, \end{cases} \quad (2.1)$$

gdzie $i \in \mathbb{Z}$ jest dowolną liczbą całkowitą.

Uwaga 2.2. Dla funkcjonatu entropii jawnie sparametryzowanego miarą przeniesioną $P_{X_{\mathbb{Z}}}$ oznaczamy $H(n; P_{X_{\mathbb{Z}}}) := H_X(n)$. Analogicznie oznaczamy wielkości $h(P_{X_{\mathbb{Z}}}) := h_X$, $h(P_{X_{\mathbb{Z}}}) := h_X$, $E(P_{X_{\mathbb{Z}}}) := E_X$, $E(n; P_{X_{\mathbb{Z}}}) := E_X(n)$ itd. dla wszystkich funkcjonatów entropii blokowej definiowanych w tym rozdziale. Indeks X w oznaczeniach H_X , h_X i E_X należy rozumieć jako skrótowe oznaczeniem procesu $X := X_{\mathbb{Z}}$, do którego odnoszą się te wielkości.

Niech Δ oznacza operator różnicowy, $\Delta F(n) := F(n) - F(n-1)$.

Twierdzenie 2.3. (por. Crutchfield i Feldman, 2003) Entropia blokowa spełnia następujące zależności

$$\Delta H_X(n) = H(X_n|X_{1:n-1}). \quad (2.2)$$

$$\Delta^2 H_X(n) = -I(X_1; X_n|X_{2:n-1}), \quad (2.3)$$

gdzie $H(X_1|X_{1:0}) := H(X_1)$ oraz $I(X_1; X_2|X_{2:1}) := I(X_1; X_2)$. Zatem dla dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ entropia blokowa jest funkcją wklęsłą ($\Delta^2 H_X(n) \leq 0$). Dla procesu $X_{\mathbb{Z}}$ o wartościach dyskretnych entropia blokowa jest także funkcją dodatnią ($H_X(n) \geq 0$) i niemalejącą ($\Delta H_X(n) \geq 0$).

Dowód:

$$\begin{aligned}
H(X_n|X_{1:n-1}) &= H(X_{1:n}) - H(X_{1:n-1}) = \\
&= H_X(n) - H_X(n-1) = \Delta H_X(n). \\
-I(X_1; X_n|X_{2:n-1}) &= H(X_{1:n}) - H(X_{1:n-1}) - H(X_{2:n}) + H(X_{2:n-1}) \\
&= H_X(n) - 2H_X(n-1) + H_X(n-2) = \Delta^2 H_X(n).
\end{aligned}$$

□

Przy okazji omawiania własności różnic entropii blokowej warto zwrócić uwagę na następujące przedstawienie:

$$\begin{aligned}
H_X(n) &= H_X(1) + \sum_{k=2}^n \Delta H_X(k) \\
&= H_X(1) + \sum_{k=2}^n \left[H_X(1) + \sum_{l=2}^k \Delta^2 H_X(l) \right] \\
&= nH_X(1) + \sum_{k=2}^n (n-k+1) \Delta^2 H_X(k). \tag{2.4}
\end{aligned}$$

Wprowadzimy teraz dwie wielkości graniczne — intensywność entropii (*entropy rate*) i entropię nadwyżkową (*excess entropy*).

Definicja 2.4. (intensywność entropii) *Intensywnością entropii procesu stacjonarnego $X_{\mathbb{Z}}$ nazywa się wielkość*

$$h_X = \lim_{n \rightarrow \infty} \Delta H_X(n) = H_X(1) + \sum_{n=2}^{\infty} \Delta^2 H_X(n). \tag{2.5}$$

Jeżeli entropia blokowa jest określona, to intensywność entropii h_X jest jednoznacznie określona jako wielkość z przedziału $[-\infty, H_X(1)]$, gdyż $\Delta^2 H_X(n) \leq 0$. Dla procesu $X_{\mathbb{Z}}$ o wartościach dyskretnych zachodzi $h_X \in [0, H_X(1)]$.

Uzasadnieniem nazwy „intensywność entropii” jest następujące twierdzenie, por. [Yeung \(2002, sekcja 2.9\)](#):

Twierdzenie 2.5. *Intensywność entropii spełnia równość*

$$h_X = \lim_{n \rightarrow \infty} \frac{H_X(n)}{n}. \tag{2.6}$$

Dowód: $\Delta H_X(\cdot)$ jest funkcją malejącą. Zatem $H_X(n) = H_X(m) + \sum_{k=m+1}^n \Delta H_X(k)$ spełnia nierówność

$$H_X(m) + (n-m) \cdot \Delta H_X(n) \leq H_X(n) \leq H_X(m) + (n-m) \cdot \Delta H_X(m). \tag{2.7}$$

Kładąc $m = 0$ w lewej nierówności, otrzymujemy $\Delta H_X(n) \leq H_X(n)/n$. Kładąc $m = n-1$ w prawej nierówności, otrzymujemy $H_X(n) \leq H_X(n-1) + \Delta H_X(n-1) \leq H_X(n-1) + H_X(n-1)/(n-1)$, czyli $H_X(n)/n \leq H_X(n-1)/(n-1)$. Ponieważ $n \mapsto H_X(n)/n$ jest funkcją malejącą, granica $h'_X := \lim_{n \rightarrow \infty} H_X(n)/n$ istnieje i spełnia nierówność $h'_X \geq h_X$. Kładąc $n = 2m$ w prawej nierówności (2.7) i dzieląc obie strony przez m , w granicy $m \rightarrow \infty$ otrzymujemy $2h'_X \leq h'_X + h_X$, czyli $h'_X = h_X$. □

Definicja 2.6. (entropia nadwyżkowa) Entropią nadwyżkową procesu stacjonarnego $X_{\mathbb{Z}}$ nazywa się wielkość $E_X := \lim_{n \rightarrow \infty} E_X(n)$, gdzie

$$E_X(n) = I(X_{-n+1:0}; X_{1:n}). \quad (2.8)$$

Dla każdego procesu $X_{\mathbb{Z}}$, entropia nadwyżkowa E_X jest jednoznacznie określona jako wielkość z przedziału $[0, \infty]$, gdyż $E_X(\cdot)$ jest funkcją niemalejącą. Nierówność $E_X(n) \geq E_X(n-1)$ zachodzi na mocy nierówności przetwarzania danych (1.16).

Uzasadnieniem nazwy „entropia nadwyżkowa” jest następujące twierdzenie, por. Crutchfield i Feldman (2003):

Twierdzenie 2.7. Dla funkcji zdefiniowanych wzorami

$$\bar{E}_X(n) = H_X(n) - n\Delta H_X(n), \quad (2.9)$$

$$\tilde{E}_X(n) = H_X(n) - nh_X \quad (2.10)$$

zachodzą nierówności

$$0 \leq \bar{E}_X(n) \leq E_X(n) \leq \bar{E}_X(2n) \leq E_X, \quad (2.11)$$

$$E_X(n) \leq \tilde{E}_X(n) \leq E_X, \quad (2.12)$$

przy czym $\bar{E}_X(\cdot)$ i $\tilde{E}_X(\cdot)$ są funkcjami niemalejącymi. A zatem

$$E_X = \lim_{n \rightarrow \infty} E_X(n) = \lim_{n \rightarrow \infty} \bar{E}_X(n) = \lim_{n \rightarrow \infty} \tilde{E}_X(n). \quad (2.13)$$

NB. Funkcje $\bar{E}_X(\cdot)$, $E_X(\cdot)$, $\tilde{E}_X(\cdot)$ nazywać będziemy entropiami nadwyżkowymi skończonego rzędu.

Dowód: Mamy

$$\begin{aligned} E_X(n) &= H(X_{1:n}) - H(X_{1:n}|X_{-n+1:0}) = \sum_{i=1}^n [H(X_i|X_{1:i-1}) - H(X_i|X_{-n+1:i-1})] \\ &= \sum_{i=1}^n [\Delta H_X(i) - \Delta H_X(i+n)] = - \sum_{i=1}^n \sum_{j=1}^n \Delta^2 H_X(i+j) \\ &= - \sum_{k=2}^n (k-1)\Delta^2 H_X(k) - \sum_{k=n+1}^{2n} (2n-k+1)\Delta^2 H_X(k), \end{aligned} \quad (2.14)$$

$$\begin{aligned} \bar{E}_X(n) &= \sum_{i=1}^n [\Delta H_X(i) - \Delta H_X(n)] = - \sum_{i=1}^n \sum_{j=i+1}^n \Delta^2 H_X(j) \\ &= - \sum_{k=2}^n (k-1)\Delta^2 H_X(k). \end{aligned} \quad (2.15)$$

Każda z wielkości $\bar{E}_X(n)$, $E_X(n)$, $\bar{E}_X(2n)$ jest sumą nieujemnych wielkości $-\Delta^2 H_X(k)$ mnożonych przez nieujemne współczynniki. Współczynniki mnożące ustaloną wielkość $-\Delta^2 H_X(k)$ rosną odpowiednio dla kolejnych wielkości $\bar{E}_X(n)$, $E_X(n)$, $\bar{E}_X(2n)$. Stąd wynika $0 \leq \bar{E}_X(n) \leq E_X(n) \leq \bar{E}_X(2n)$. Na mocy podobnego argumentu $\bar{E}_X(n) \leq \bar{E}_X(n +$

1), czyli $\bar{E}_X(\cdot)$ jest funkcją rosnącą. Zatem granica $\lim_{n \rightarrow \infty} \bar{E}_X(n)$ istnieje i jest równa $\lim_{n \rightarrow \infty} E_X(n)$.

Dla $m > n$ zdefiniujemy

$$\begin{aligned} \tilde{E}_X(n; m) &= H_X(n) - n\Delta H_X(m) = \bar{E}_X(m) + n[\Delta H_X(n) - \Delta H_X(m)] \\ &= -\sum_{k=2}^n (k-1)\Delta^2 H_X(k) - \sum_{k=n+1}^m n\Delta^2 H_X(k). \end{aligned} \quad (2.16)$$

Z (2.16) wynika $\tilde{E}_X(n; m) \leq \tilde{E}_X(n; m+1)$. Zatem granica $\lim_{m \rightarrow \infty} \tilde{E}_X(n; m)$ istnieje i jest równa $\bar{E}_X(n)$. Porównując współczynniki przy $-\Delta^2 H_X(k)$, wnioskujemy, że $E_X(n) \leq \tilde{E}_X(n; m) \leq \bar{E}_X(m)$ dla $m \geq 2n$. Na mocy podobnego argumentu $\tilde{E}(n; m) \leq \tilde{E}(n+1; m)$. Dla $m \rightarrow \infty$ te dwa zestawy nierówności implikują $E_X(n) \leq \bar{E}_X(n) \leq E_X$ oraz $\bar{E}(n)_X \leq \tilde{E}_X(n+1)$, a zatem $\lim_{n \rightarrow \infty} \bar{E}_X(n) = \lim_{n \rightarrow \infty} \tilde{E}_X(n)$, zaś $\tilde{E}_X(\cdot)$ jest funkcją rosnącą. \square

W świetle (2.11) wartość $\bar{E}_X(2n)$ jest zwykle lepszym niż $E_X(n)$ przybliżeniem entropii nadwyżkowej E_X , które można obliczyć z $2n$ -wymiarowego rozkładu zmiennych $X_{1:2n}$. Pomimo tego warto zauważyć, że dla procesu stacjonarnego $E_X(n) := I(X_{1:n}; X_{n+1:2n})$ maksymalizuje wartość informacji wzajemnej $I(X_{1:k}; X_{k+1:2n})$.

Twierdzenie 2.8. *Dla dowolnego n i $k \leq n/2$ zachodzi nierówność*

$$I(X_{1:k-1}; X_{k:n}) \leq I(X_{1:k}; X_{k+1:n}). \quad (2.17)$$

Dowód: Mamy

$$\begin{aligned} I(X_{1:k-1}; X_{k:n}) - I(X_{1:k}; X_{k+1:n}) &= H_X(k-1) + H_X(n-k+1) - H_X(k) - H_X(n-k) \\ &= \Delta H_X(n-k+1) - \Delta H_X(k) = \sum_{l=k+1}^{n-k+1} \Delta^2 H_X(l) \leq 0. \end{aligned}$$

\square

Chociaż $E_X(\cdot)$, $\bar{E}_X(\cdot)$ i $\tilde{E}_X(\cdot)$ są funkcjami niemalejącymi, nie są one zwykle funkcjami wklęsłymi. W szczególności $\Delta^2 E_X(n)$ wyraża się jako kombinacja liniowa stałej liczby różnic $\Delta^2 H_X(k)$,

$$\Delta E_X(n) = -\sum_{k=n+1}^{2n-1} 2\Delta^2 H_X(k) - \Delta^2 H_X(2n), \quad (2.18)$$

$$\Delta^2 E_X(n) = 2\Delta^2 H_X(n) - \Delta^2 H_X(2n-2) - 2\Delta^2 H_X(2n-1) - \Delta^2 H_X(2n). \quad (2.19)$$

Omówmy teraz niektóre elementarne własności entropii nadwyżkowej.

Definicja 2.9. (proces finitarny, Crutchfield i Feldman, 2003) *Mówimy, że proces stacjonarny $X_{\mathbb{Z}}$ jest finitarny, jeżeli $E_X < \infty$.*

Twierdzenie 2.10. (Shalizi, 2001) *Zachodzi nierówność*

$$H_X(1) - h_X \leq E_X. \quad (2.20)$$

Dowód: Mamy $H_X(1) - h_X = -\sum_{k=2}^{\infty} \Delta^2 H_X(k) \leq -\sum_{k=2}^{\infty} (k-1)\Delta^2 H_X(k) = E_X$. \square

Na mocy twierdzenia 2.10 każdy proces taki, że $h_X = -\infty$, jest nefinitarny. (Przykładem procesów spełniających równość $h_X = -\infty$ są gaussowskie procesy deterministyczne.)

W pewnych przypadkach wartość entropii nadwyżkowej wiąże się z tym, czy dany proces można przedstawić jako ukryty łańcuch Markowa (definicja A.46).

Twierdzenie 2.11. *Proces stacjonarny $X_{\mathbb{Z}}$ jest finitarny, jeżeli $X_{\mathbb{Z}}$ jest ukrytym łańcuchem Markowa z procesem ukrytym $Y_{\mathbb{Z}}$ takim, że $\bar{Y}_{\mathbb{Z}} := (X_i \times Y_i)_{i \in \mathbb{Z}}$ jest stacjonarny, określona jest entropia blokowa $H_{\bar{Y}}$ i $I(\bar{Y}_i; \bar{Y}_{i+1}) < \infty$.*

Dowód: Na mocy nierówności przetwarzania danych (1.20) zachodzi $I(X_{-n+1:0}; X_{1:n}) \leq I(\bar{Y}_{-n+1:0}; \bar{Y}_{1:n})$, a stąd $E_X \leq E_{\bar{Y}}$. Ponieważ $\bar{Y}_{\mathbb{Z}}$ jest stacjonarnym łańcuchem Markowa, to mamy

$$E_{\bar{Y}} = \sum_{n=2}^{\infty} (n-1) I(\bar{Y}_1; \bar{Y}_n | \bar{Y}_{2:n-1}) = I(\bar{Y}_1; \bar{Y}_2) < \infty,$$

gdzie $I(\bar{Y}_1; \bar{Y}_n | \bar{Y}_{2:n-1}) = 0$ dla $n > 2$ na mocy twierdzenia 1.17, pkt. (iii). \square

Przykład 2.12. *Jeżeli $X_{\mathbb{Z}}$ i $Y_{\mathbb{Z}}$ są procesami dyskretnymi o skończonej liczbie wartości, to $I(\bar{Y}_i; \bar{Y}_{i+1}) \leq H(\bar{Y}_i) \leq H(X_i) + H(Y_i) < \infty$. A zatem wszystkie takie ukryte procesy Markowa są finitarne.*

Założenie $I(\bar{Y}_i; \bar{Y}_{i+1}) < \infty$ w twierdzeniu 2.11 jest istotne. Istnieją bowiem przykłady dyskretnych nefinitarnych procesów Markowa o nieskończonej liczbie wartości.

Przykład 2.13. (nefinitarny łańcuch Markowa) *Niech proces $X_{\mathbb{Z}}$, gdzie $X_i(\Omega) \subset \mathbb{N}$, ma rozkład $P(X_{1:k} = n_{1:k}) = p_{n_1} \prod_{i=2}^k p_{n_{i-1}, n_i} / p_{n_{i-1}}$, gdzie $p_n = \sum_{m=1}^{\infty} p_{nm}$. Macierz $(p_{nm})_{n, m \in \mathbb{N}}$ definiujemy jako macierz symetryczną*

$$p_{nm} = pa_n \mathbb{I}[n = m] + (1-p)a_n a_m,$$

gdzie $0 < p, a_n < 1$, $\sum_{j=1}^{\infty} a_n = 1$, $-\sum_{j=1}^{\infty} a_n \log a_n = \infty$.

Pokażemy teraz, że dowolny proces $X_{\mathbb{Z}}$ tej postaci jest nefinitarny. Po pierwsze zauważmy, że $p_n = a_n$. Niech $\kappa := (1-p)/p$. Wówczas

$$\begin{aligned} \bar{I}(X_j; X_{j+1}) &= \sum_{n, m} p_{nm} \log \frac{p_{nm}}{p_n p_m} \\ &= \sum_n (pa_n + (1-p)a_n^2) \log \frac{pa_n + (1-p)a_n^2}{a_n^2} + \sum_{n \neq m} (1-p)a_n a_m \log(1-p) \\ &= \sum_n p(a_n + \kappa a_n^2) (\log p + \log(a_n^{-1} + \kappa)) + \left(1 - \sum_n a_n^2\right) (1-p) \log(1-p) \\ &\geq p \left(-\sum_{j=1}^{\infty} a_n \log a_n\right) + p \log p + (1-p) \log(1-p) = \infty. \end{aligned}$$

Proces z przykładu 2.13 jest jednoklasowym stacjonarnym łańcuchem Markowa, a zatem jest ergodyczny (Hernandez-Lerma i Lasserre, 2003; Kemeny *et al.*, 1966).

W ogólnym przypadku, obliczenie wielkości $\Delta H_X(n)$ i $E_X(n)$ wymaga całkowania po n -wymiarowych rozkładach prawdopodobieństwa. Pełna informacja o wysokowymiarowych rozkładach rzadko bywa dostępna przy modelowaniu danych empirycznych. Z tego

względnie warto badać jakość takich oszacowań intensywności entropii i entropii nadwyżkowej, które wyrażają się przez funkcje wartości oczekiwanych zmiennych niskowymiarowych. Ponieważ $\Delta H_X(n)$ i $E_X(n)$ są funkcjonalami ciągu $(I(X_1; X_n | X_{2:n-1}))_{n \in \mathbb{N}}$, jednym z interesujących problemów szczegółowych jest badanie związków pomiędzy informacją wzajemną $I(X_1; X_n)$ a warunkową informacją wzajemną $I(X_1; X_n | X_{2:n-1})$. Różnica $I(X_1; X_n; X_{2:n-1}) = I(X_1; X_n) - I(X_1; X_n | X_{2:n-1})$ może mieć znak dowolny, ale pewne ograniczenia istnieją, por. Yeung (2002, rozdziały 6, 12–14, 16).

W pierwszej kolejności warto zwrócić uwagę na ograniczenia najprostsze. Z nierówności przetwarzania danych (1.20) wynika banalna nierówność

$$I(X_0; X_n) \leq I(X_0; X_{1:n}) = \sum_{k=1}^n I(X_0; X_k | X_{1:k-1}). \quad (2.21)$$

Nierówność (2.21) ma pewien nieoczekiwany odpowiednik dla szeregów czasowych — nierówność (11.7), lecz niezależnie od tego jest to nierówność bardzo słaba. Prawa strona nierówności jest zawsze rosnącą funkcją n . Z kolei lewa strona jest malejącą funkcją n dla procesów Markowa, gdyż nierówność $I(X_0; X_n) \leq I(X_0; X_{n-1})$ wynika z własności $X_0 \perp\!\!\!\perp X_n | X_{n-1}$ na mocy twierdzenia 1.17 pkt. (iv). Pomimo tego, istnieją proste procesy Markowa, dla których równość w (2.21) jest osiągnięta dla każdego n .

Przykład 2.14. Niech $X_i := Y \times Z_i$, gdzie $Z_{\mathbb{Z}}$ jest zbiorem niezależnych dyskretnych zmiennych losowych o identycznym rozkładzie, zaś Y jest dyskretną zmienną losową. Wówczas proces $X_{\mathbb{Z}}$ jest stacjonarnym procesem Markowa, $I(X_0, X_n) = I(X_0; X_1)$, $I(X_0, X_k | X_{1:k-1}) = 0$ dla $k > 1$, a zatem równość w (2.21) zachodzi dla wszystkich n .

Nieco inny problem oszacowania wartości intensywności entropii i entropii nadwyżkowej pojawia się, gdy wartości $H_X(n)$ znane są tylko z pewną dokładnością. Zadanie tego typu pojawia się jako naturalna konsekwencja problemu szacowania entropii blokowej języka naturalnego przez zgadywanie (Shannon, 1950).

Twierdzenie 2.15. Niech $X_{\mathbb{Z}}$ będzie dowolnym stacjonarnym procesem stochastycznym. Jeżeli entropia blokowa spełnia $A(n) \leq H_X(n) \leq B(n)$ dla $n \in I \subset \mathbb{N}$, to

$$\frac{A(l) - B(k)}{l - k} \leq \Delta H_X(n) \leq \frac{B(l) - A(k)}{l - k} \quad (2.22)$$

dla $k \leq n \leq l$, $k, l \in I$.

Dowód: Dla dowolnego stacjonarnego procesu stochastycznego, $\Delta H_X(n)$ jest malejącą funkcją n . Stąd $A(l) - B(k) \leq H_X(l) - H_X(k) = \sum_{m=k+1}^l \Delta H_X(m) \leq (l - k) \Delta H_X(n)$ dla $k \leq n \leq l$. Analogicznie $A(k) - B(l) \leq H_X(k) - H_X(l) = -\sum_{m=k+1}^l \Delta H_X(m) \leq -(l - k) \Delta H_X(n)$. \square

Przykład 2.16. Niech $A(n) = \alpha B(n)$, $B(n) = n^\beta$, $\beta \in (0, 1)$, zaś $A(n) \leq H_X(n) \leq B(n)$ zachodzi dla $n \in \mathbb{N}$. Dla $n = k$, $l = rk$ mamy $\Delta H_X(n)/n^{\beta-1} \in [\frac{\alpha r^\beta - 1}{r-1}, \frac{r^\beta - \alpha}{r-1}]$.

Twierdzenie 2.17. Jeżeli zachodzi $A(n) \leq H_X(n) \leq B(n)$ dla $n \in I \subset \mathbb{N}$, to

$$\frac{lA(k) - kB(l)}{l - k} \leq \bar{E}_X(k) \leq \frac{lB(k) - kA(l)}{l - k} \quad (2.23)$$

dla $k, l \in I$.

Dowód: Mamy $\bar{E}_X(n) = H_X(n) - n \Delta H_X(n)$. Stąd $A(n) - n[B(l) - A(k)]/(l - k) \leq \bar{E}_X(n) \leq B(n) - n[A(l) - B(k)]/(l - k) \leq \bar{E}_X(n)$ dla $\min(k, l) \leq n \leq \max(k, l)$. Kładąc $n = k$, otrzymujemy tezę. \square

Rozdział 3

Całkowe przedstawienia miar granicznych

Intensywność entropii i entropia nadwyżkowa, zdefiniowane w rozdziale 2, są pewnymi granicznymi miarami nieprzewidywalności i pamięci stacjonarnego procesu stochastycznego. Rodzi się pytanie, czy wielkości te można zdefiniować bezpośrednio jako warunkowe informacje wzajemne pomiędzy pewnymi σ -ciałami.

Z rozdziału 1 wiemy, że możemy określić wielkość $\bar{H}(X_n|X_{-\infty:n-1})$ dla procesu $X_{\mathbb{Z}}$ o wartościach dyskretnych oraz wielkość $\bar{I}(X_{-\infty:n-1}; X_{n:\infty})$ dla dowolnego procesu. Przez $X_{-\infty:n-1}$ i $X_{n:\infty}$ rozumiemy pewne zmienne losowe mierzalne względem σ -ciał $\sigma(\bigcup_{i \leq n-1} \sigma(X_i))$ oraz $\sigma(\bigcup_{i \geq n} \sigma(X_i))$, określone w dodatku A.1. Interesującym pytaniem jest, w jakich przypadkach dla procesu stacjonarnego $X_{\mathbb{Z}}$ zachodzą równości $h_X = \bar{H}(X_{n+1}|X_{-\infty:n})$ oraz $E_X = \bar{I}(X_{-\infty:n}; X_{n+1:\infty})$. Poniższe twierdzenie orzeka, że pierwsza z tych równości jest prawdziwa dla zmiennych X_i o skończonym zbiorze wartości, a druga z nich jest prawdziwa zawsze.

Twierdzenie 3.1. *Niech $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ będą pewnymi σ -ciałami. Niech ciągi σ -ciał $(\mathcal{G}_j^{(n)})_{j \in \mathbb{N}}$, $j = 1, 2$ spełniają $\mathcal{G}_j^{(n)} \subset \mathcal{G}_j^{(n+1)}$ oraz $\mathcal{G}_j = \sigma(\bigcup_{j \in \mathbb{N}} \mathcal{G}_j^{(n)})$. Wówczas prawdziwe są następujące stwierdzenia:*

(i) *Jeżeli istnieje warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$, to zachodzi równość*

$$\lim_{n \rightarrow \infty} \bar{I}(\mathcal{G}_1^{(n)}; \mathcal{G}_2^{(n)} | \mathcal{G}_3) = \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3). \quad (3.1)$$

(ii) *Jeżeli \mathcal{G}_1 jest σ -ciałem skończonym, to zachodzi równość*

$$\lim_{n \rightarrow \infty} \bar{H}(\mathcal{G}_1 | \mathcal{G}_2^{(n)}) = \bar{H}(\mathcal{G}_1 | \mathcal{G}_2). \quad (3.2)$$

NB. Jeżeli \mathcal{G}_1 jest nieskończonym σ -ciałem dyskretnym, to z nierówności (1.36) w ogólnym przypadku mamy tylko nierówność $\lim_{n \rightarrow \infty} \bar{H}(\mathcal{G}_1 | \mathcal{G}_2^{(n)}) \geq \bar{H}(\mathcal{G}_1 | \mathcal{G}_2)$. Pkt. (ii) powyższego twierdzenia udowodnił Billingsley (1965, twierdzenie 12.1). Gelfand *et al.* (1956, twierdzenie 1) podają bez dowodu wersję pkt. (i) dla bezwarunkowej informacji wzajemnej dla ciągów ciał (niekoniecznie σ -ciał) $(\mathcal{G}_j^{(n)})_{j \in \mathbb{N}}$, gdzie $\mathcal{G}_j = \bigcup_{j \in \mathbb{N}} \mathcal{G}_j^{(n)}$.

Dowód:

(i) Niech $(\mathcal{J}_1, \mathcal{J}_2) = (\mathcal{G}_1^{(n)}, \mathcal{G}_2^{(n)})$ dla pewnego n lub $(\mathcal{J}_1, \mathcal{J}_2) = (\mathcal{G}_1, \mathcal{G}_2)$. Z istnienia miary $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ wynika określoność $\bar{I}(\mathcal{J}_1; \mathcal{J}_2 | \mathcal{G}_3)$. Oznaczmy $S = P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} + P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3}$. Z dowodu twierdzenia 1.13 mamy, że

$$\bar{I}(\mathcal{J}_1; \mathcal{J}_2 | \mathcal{G}_3) = \int i_{\text{reg}}(dP^{(3)}|_{\mathcal{J}_1 \otimes \mathcal{J}_2 \otimes \mathcal{G}_3} / dS) dS,$$

gdzie $i_{\text{reg}}(x) := x \log x - x \log(1-x) - 2x + 1 \geq 0$. Z kolei z twierdzenia A.24 o zbieżności martyngałów wynika, że $\lim_{n \rightarrow \infty} dP^{(3)}|_{\mathcal{G}_1^{(n)} \otimes \mathcal{G}_2^{(n)} \otimes \mathcal{G}_3} / dS = dP^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} / dS$ S -prawie na pewno. Ponieważ funkcja i_{reg} jest ciągła i nieujemna, to z lematu

Fatou ($\int \liminf_{n \rightarrow \infty} h_n d\mu \leq \liminf_{n \rightarrow \infty} \int h_n d\mu$ dla h_n nieujemnych i odpowiednio mierzalnych, Rudin, 1986, sekcja 1.28) otrzymujemy

$$\liminf_{n \rightarrow \infty} \bar{I}(\mathcal{G}_1^{(n)}; \mathcal{G}_2^{(n)} | \mathcal{G}_3) \geq \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3).$$

Ponieważ z nierówności przetwarzania danych (twierdzenie 1.17 pkt. (iii)) wynika także relacja przeciwna

$$\limsup_{n \rightarrow \infty} \bar{I}(\mathcal{G}_1^{(n)}; \mathcal{G}_2^{(n)} | \mathcal{G}_3) \leq \bar{I}(\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3),$$

to wnioskujemy, że zachodzi równość (3.1).

(ii) Niech $A \in \mathcal{A}(\mathcal{G}_1)$ będą atomami σ -ciała \mathcal{G}_1 . Z twierdzenia A.24 o zbieżności martingałów zachodzi zbieżność

$$\lim_{n \rightarrow \infty} P(A | \mathcal{G}_2^{(n)}) = P(A | \mathcal{G}_2) \quad \text{prawie na pewno.}$$

Z kolei z twierdzenia 1.31 mamy

$$\bar{H}(\mathcal{G}_1 | \mathcal{J}_2) = \int \left[- \sum_{A \in \mathcal{A}(\mathcal{G}_1)} P(A | \mathcal{J}_2) \log P(A | \mathcal{J}_2) \right] dP,$$

gdzie $\mathcal{J}_2 = \mathcal{G}_2^{(n)}$ dla pewnego n lub $\mathcal{J}_2 = \mathcal{G}_2$. Ponieważ dla $x \in (0, 1]$ funkcja $f(x) = x \log x \in [0, 1]$ jest ciągła i ograniczona, to (3.2) wynika ze skończoności $\mathcal{A}(\mathcal{G}_1)$ i twierdzenia Lebesgue'a o zbieżności majoryzowanej ($\int \lim_{n \rightarrow \infty} h_n d\mu = \lim_{n \rightarrow \infty} \int h_n d\mu$ dla h_n takich, że $|h_n| \leq g$ i $\int g d\mu < \infty$, Rudin, 1986, sekcja 1.34). □

Zauważmy, że równość (3.1) jest szczególnym przypadkiem równości

$$\lim_{n \rightarrow \infty} \int f(d\nu |_{\mathcal{G}^{(n)}} / d\mu) d\mu = \int f(d\nu |_{\mathcal{G}} / d\mu) d\mu, \quad (3.3)$$

która zachodzi dla $\mathcal{G}^{(n)} \subset \mathcal{G}^{(n+1)}$, $\mathcal{G} = \sigma(\bigcup_{i \in \mathbb{N}} \mathcal{G}^{(i)})$ oraz dowolnej nieujemnej i wypukłej funkcji f takiej, że całka $\int f'(d\nu |_{\mathcal{G}^{(n)}} / d\mu) (d\nu |_{\mathcal{G}} / d\mu - d\nu |_{\mathcal{G}^{(n)}} / d\mu) d\mu$ jest określona. W tym przypadku równość (3.3) zachodzi z powodu dość wyjątkowej interakcji pomiędzy lematem Fatou a nierównością Jensena $\int f(d\nu |_{\mathcal{G}^{(n)}} / d\mu) d\mu \leq \int f(d\nu |_{\mathcal{G}} / d\mu) d\mu$ (Billingsley, 1979, sekcja 33) i aby zapewnić tę równość, nie ma potrzeby czynić żadnych dalszych założeń (np. aby móc skorzystać z twierdzenia o zbieżności majoryzowanej).

Równość (3.1) jest oczywiście znacznie mocniejszym stwierdzeniem niż ciągłość warunkowej informacji wzajemnej $I(X; Y | Z)$ względem rozkładu dla zmiennych X, Y, Z o skończonej liczbie wartości, por. Yeung (2002, sekcja 2.3).

W świetle twierdzenia 3.1 pkt. (i) możemy uogólnić dwa pojęcia z rozdziału 2.

Definicja 3.2. (entropia nadwyżkowa i proces finitarny) *Entropią nadwyżkową dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ nazywa się wielkość*

$$E_X = \bar{I}(X_{-\infty:n}; X_{n+1:\infty}). \quad (3.4)$$

Mówimy, że proces $X_{\mathbb{Z}}$ jest finitarny, jeżeli $E_X < \infty$.

Dla porównania, punktu (ii) twierdzenia 3.1 nie można uogólnić na przypadek $\mathcal{G}_1 := \sigma(\mathcal{F}_1)$, gdzie \mathcal{F}_1 jest dowolną przeliczalną klasą rozłącznych zdarzeń. Istnieją zmienne losowe X o nieskończonym zbiorze wartości takie, że $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) > \bar{H}(X|Y_{\mathbb{N}})$. Przykładowo, możemy dobrać takie X , że $\bar{H}(X|Y_{1:n}) = \infty$ natomiast $\bar{H}(X|Y_{\mathbb{N}}) = 0$, gdyż istnieje funkcja g mierzalna $\tilde{\sigma}(Y_{\mathbb{N}})/\tilde{\sigma}(X)$ taka, że $X = g(Y_{\mathbb{N}})$ prawie na pewno.

Przykład 3.3. (niezbiegające entropie warunkowe) Obierzmy zmienną X o wartościach $X(\Omega) \subset \mathbb{N}$ taką, że $\bar{H}(X) = \infty$. Niech $Y_k = 1$, gdy $X \geq k$, zaś $Y_k = 0$ w przeciwnym wypadku. Mamy wówczas $\bar{H}(X) = \bar{H}(X \times Y_{1:n}) = \bar{H}(X|Y_{1:n}) + \bar{H}(Y_{1:n})$, a zatem $\bar{H}(X|Y_{1:n}) = \infty$, gdyż $\bar{H}(Y_{1:n}) \leq n \log 2$. Z drugiej strony X jest mierzalną funkcją $Y_{\mathbb{N}}$, a zatem $\bar{H}(X|Y_{\mathbb{N}}) = 0$.

Równość $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = \bar{H}(X|Y_{\mathbb{N}})$ udowodniona dla zmiennej X o skończonej liczbie wartości wiąże się z kolei z pewnym interesującym twierdzeniem.

Twierdzenie 3.4. Niech zmienna X przyjmuje wartości ze skończonego zbioru, zaś $Y_{\mathbb{N}}$ będzie dowolnym procesem. Funkcje f_n mierzalne $\tilde{\sigma}(Y_{1:n})/\tilde{\sigma}(X)$ takie, że

$$\lim_{n \rightarrow \infty} P(X = f_n(Y_{1:n})) = 1, \quad (3.5)$$

istnieją wtedy i tylko wtedy, gdy $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0$.

Dowód: Niech \mathbb{V} oznacza zbiór wartości zmiennej X .

(i) Załóżmy, że $\lim_{n \rightarrow \infty} P(X = f_n(Y_{1:n})) = 1$. Zdefiniujmy $\Phi_n := \llbracket X = f_n(Y_{1:n}) \rrbracket$. Ponieważ $\bar{H}(\Phi_n|Y_{1:n} \times X) = 0$, to mamy $\bar{H}(X|Y_{1:n}) = \bar{H}(X|Y_{1:n}) + \bar{H}(\Phi_n|Y_{1:n} \times X) = \bar{H}(X \times \Phi_n|Y_{1:n})$, a więc

$$\begin{aligned} \bar{H}(X|Y_{1:n}) &= \bar{H}(X \times \Phi_n|Y_{1:n}) = \bar{H}(\Phi_n|Y_{1:n}) + \bar{H}(X|\Phi_n \times Y_{1:n}) \\ &\leq \bar{H}(\Phi_n) + \bar{H}(X|\Phi_n \times Y_{1:n}). \end{aligned} \quad (3.6)$$

Mamy także $\bar{H}(\Phi_n) = h_{\text{bin}}(P(\Phi_n = 1))$, gdzie $h_{\text{bin}} : [0, 1] \rightarrow \mathbb{R}$ jest funkcją ciągłą określoną jako $h_{\text{bin}}(p) := -p \log p - (1-p) \log(1-p)$ dla $0 < p < 1$. Funkcja ta spełnia $h_{\text{bin}}(1) = 0$, a zatem $\lim_{n \rightarrow \infty} P(X = f(Y_{1:n})) = 1$ implikuje $\lim_{n \rightarrow \infty} \bar{H}(\Phi_n) = 0$. Ponieważ zachodzi

$$\begin{aligned} 0 &\leq \bar{H}(X|\Phi_n \times Y_{1:n}) \\ &= \int_{\Phi_n=0} \left[- \sum_{x \in \mathbb{V}} P(X = x|\Phi_n \times Y_{1:n}) \log P(X = x|\Phi_n \times Y_{1:n}) \right] dP = \\ &\leq [1 - P(\Phi_n = 1)] \log[\text{card } \mathbb{V} - 1], \end{aligned} \quad (3.7)$$

to (3.6) implikuje $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0$.

(ii) Załóżmy, że $\lim_{n \rightarrow \infty} \bar{H}(Y|X_{1:n}) = 0$. Bez zmniejszania ogólności rozważań załóżmy, że \mathbb{V} jest podzbiorem liczb naturalnych. Niech $f_n(Y_{1:n})$ będzie równe najmniejszej liczbie x takiej, że $P(X = x|Y_{1:n}) \geq P(X = x'|Y_{1:n})$ zachodzi dla każdego $x' \in \mathbb{V}$. Funkcja $f_n(Y_{1:n})$ jest mierzalna $\tilde{\sigma}(Y_{1:n})/\tilde{\sigma}(X)$. Dowiedzmy tego ostatniego. Niech $A(x)$ oznacza zdarzenie, że $P(X = x|Y_{1:n}) \geq P(X = x'|Y_{1:n})$ zachodzi dla każdego $x' \in \mathbb{V}$. Mamy

$$A(x) = \bigcap_{q \in \mathbb{Q}} (P(X = x|Y_{1:n}) \leq q)^c \cup \bigcap_{x' \in \mathbb{V}} (P(X = x'|Y_{1:n}) \leq q) \in \sigma(Y_{1:n}),$$

gdzież $[\forall_{b \in B} a \geq b] \iff \forall_{q \in \mathbb{Q}} [a \leq q \implies \forall_{b \in B} b \leq q]$ dla $a \in \mathbb{R}$, $b \subset \mathbb{R}$. Stąd mamy także $(f_n(Y_{1:n}) = x) = A(x) \cap \bigcap_{x' < x} (A(x'))^c \in \sigma(Y_{1:n})$.

Położmy $\Phi_n := \llbracket X = f_n(Y_{1:n}) \rrbracket$ jak poprzednio. Mamy wówczas $P(\Phi_n = 1 | Y_{1:n}) \geq 1/\text{card } \mathbb{V}$ prawie wszędzie. Funkcja h_{bin} z pkt. (i) jest wklęsła i spełnia $h_{\text{bin}}(1) = 0$, a zatem

$$h_{\text{bin}}(p) \geq h_{\text{bin}}(q) \frac{1-p}{1-q} + h_{\text{bin}}(1) \frac{p-q}{1-q} = h_{\text{bin}}(q) \frac{1-p}{1-q}$$

dla $p \in [q, 1]$, $q \in [0, 1]$. W szczególności

$$\begin{aligned} \bar{H}(X|Y_{1:n}) &= \bar{H}(X \times \Phi_n | Y_{1:n}) \geq \bar{H}(\Phi_n | Y_{1:n}) \\ &= \int h_{\text{bin}}(P(\Phi_n = 1 | Y_{1:n})) dP \\ &\geq \int \frac{h_{\text{bin}}(1/\text{card } \mathbb{V})}{1 - 1/\text{card } \mathbb{V}} \cdot [1 - P(\Phi_n = 1 | Y_{1:n})] dP \\ &= \frac{h_{\text{bin}}(1/\text{card } \mathbb{V})}{1 - 1/\text{card } \mathbb{V}} \cdot [1 - P(\Phi_n = 1)]. \end{aligned}$$

Ponieważ $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0$, to wnioskujemy, że $\lim_{n \rightarrow \infty} P(\Phi_n = 1) = 1$. \square

Nierówność $\bar{H}(X|Y_{1:n}) \leq h_{\text{bin}}(P(\Phi_n = 1)) + [1 - P(\Phi_n = 1)] \log[\text{card } \mathbb{V} - 1]$ jest teoriomiarową wersją nierówności Fano (Yeung, 2002, twierdzenie 2.47).

Z twierdzenia 3.4 wypływa wniosek, który można sformułować bez użycia pojęć z zakresu teorii informacji.

Twierdzenie 3.5. *Niech zmienna X przyjmuje wartości ze skończonego zbioru, zaś $Y_{\mathbb{N}}$ będzie dowolnym procesem. Funkcje f_n mierzalne $\tilde{\sigma}(Y_{1:n})/\tilde{\sigma}(X)$ takie, że zachodzi (3.5), istnieją wtedy i tylko wtedy, gdy istnieje funkcja g mierzalna $\tilde{\sigma}(Y_{\mathbb{N}})/\tilde{\sigma}(X)$ taka, że*

$$X = g(Y_{\mathbb{N}}) \quad \text{prawie na pewno.} \quad (3.8)$$

Dowód: Na mocy twierdzenia 3.4, warunek (3.5) jest równoważny $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0$. Z twierdzenia 3.1 zachodzi równość $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = \bar{H}(X|Y_{\mathbb{N}})$. Z kolei z twierdzeń 1.29 i A.20 warunek $\bar{H}(X|Y_{\mathbb{N}}) = 0$ jest równoważny (3.8). \square

Twierdzeń 3.4 i 3.5 nie można uogólnić na zmienne X o nieskończonym przeliczalnym zbiorze wartości. Można bowiem dobrać takie X i $Y_{\mathbb{N}}$ by zachodziło $\lim_{n \rightarrow \infty} P(X = f_n(Y_{1:n})) = 1$ oraz $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = \infty$. Być może jednak zachodzi implikacja odwrotna

$$\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0 \implies \lim_{n \rightarrow \infty} P(X = f_n(Y_{1:n})) = 1. \quad (3.9)$$

Założmy, że X jest zmienną o nieskończonym przeliczalnym zbiorze wartości. Niewątpliwie, jeżeli $\lim_{n \rightarrow \infty} \bar{H}(X|Y_{1:n}) = 0$, to $\bar{H}(X|Y_{\mathbb{N}}) = 0$. Istnieje zatem funkcja g mierzalna $\tilde{\sigma}(Y_{\mathbb{N}})/\tilde{\sigma}(X)$ taka, że $X = g(Y_{\mathbb{N}})$ prawie na pewno. Jeżeli $g \in L^1(\tilde{\sigma}(Y_{\mathbb{N}}), P_{Y_{\mathbb{N}}})$, to istnieją rzuty ortogonalne $f_n \in L^1(\tilde{\sigma}(Y_{1:n}), P_{Y_{1:n}})$ spełniające warunek $f_n(Y_{1:n}) = \langle g(Y_{\mathbb{N}}) | \sigma(Y_{1:n}) \rangle$. Ciąg $(f_n(Y_{1:n}))_{n \in \mathbb{N}}$ jest martyngałem, więc mamy zbieżność $\lim_{n \rightarrow \infty} f_n(Y_{1:n}) = g(Y_{\mathbb{N}}) = X$ prawie na pewno. Bez zmniejszania ogólności możemy założyć, że $X \in \mathbb{N}$. Zbieżność $\lim_{n \rightarrow \infty} f_n(Y_{1:n}) = X$ prawie na pewno implikuje zbieżność według prawdopodobieństwa $\lim_{n \rightarrow \infty} P(|f_n(Y_{1:n}) - X| \geq \epsilon) = 0$ dla każdego $\epsilon > 0$ (Billingsley, 1979, sekcja 20, zbieżność według prawdopodobieństwa). Dowiedliśmy zatem (3.9) w szczególnym przypadku, gdy $g \in L^1(\tilde{\sigma}(Y_{\mathbb{N}}), P_{Y_{\mathbb{N}}})$. Uogólnienia twierdzeń A.20, 3.4 i 3.5 oraz relacji (3.9) na dowolne zmienne losowe warte są dalszych badań.

Szczególny przypadek równości (3.8) zachodzi, gdy $X = X_j$ zaś $Y_n = X_{j-n}$.

Definicja 3.6. (proces deterministyczny) *Mówimy, że proces stacjonarny $X_{\mathbb{Z}}$ jest deterministyczny, jeżeli istnieje funkcja g mierzalna $\tilde{\sigma}(X_{-\infty:j-1})/\tilde{\sigma}(X_j)$ taka, że*

$$X_j = g(X_{-\infty:j-1}) \quad \text{prawie na pewno.} \quad (3.10)$$

Dla kilku klas procesów stochastycznych warunek determinizmu (3.10) jest równoważny pewnym asymptotycznym zachowaniom entropii warunkowej.

- (i) Niech $X_{\mathbb{Z}}$ będzie procesem stacjonarnym, gdzie zmienne X_i przyjmują wartości ze skończonego zbioru. Z twierdzeń 3.1, 3.5, 1.29 i A.20 równoważne są warunki:
- $X_{\mathbb{Z}}$ jest deterministyczny,
 - $\bar{H}(X_j|X_{-\infty:j-1}) = 0$,
 - $\lim_{n \rightarrow \infty} P(X_j = f_n(X_{j-n:j-1})) = 1$ dla pewnych f_n mierzalnych $\tilde{\sigma}(X_{j-n:j-1})/\tilde{\sigma}(X_j)$,
 - $\lim_{n \rightarrow \infty} \bar{H}(X_j|X_{j-n:j-1}) = 0$.
- (ii) Niech $X_{\mathbb{Z}}$ będzie procesem stacjonarnym, gdzie zmienne X_i przyjmują wartości z nieskończonego zbioru przeliczalnego. Z twierdzeń 1.29 i A.20 równoważne są warunki:
- $X_{\mathbb{Z}}$ jest deterministyczny,
 - $\bar{H}(X_j|X_{-\infty:j-1}) = 0$.
- (iii) Niech $X_{\mathbb{Z}}$ będzie stacjonarnym procesem gaussowskim. Z twierdzeń 9.2 i 9.3 równoważne są warunki:
- $X_{\mathbb{Z}}$ jest deterministyczny,
 - $\lim_{n \rightarrow \infty} \text{Var}(X_j - f_n(X_{j-n:j-1})) = 1$ dla pewnych f_n mierzalnych $\tilde{\sigma}(X_{j-n:j-1})/\tilde{\sigma}(X_j)$,
 - $\lim_{n \rightarrow \infty} H(X_j|X_{j-n:j-1}) = -\infty$.

Zauważmy też, że gdy $X_{\mathbb{Z}}$ jest procesem gaussowskim, wielkości $g(X_{-\infty:j-1})$ i $f_n(X_{j-n:j-1})$ ze wspomnianych wyżej warunków są elementami domkniętych podprzestrzeni Hilberta rozpinanych odpowiednio przez $(X_k)_{k \leq j-1}$ oraz $(X_k)_{j-n \leq k \leq j-1}$.

Rozdział 4

Rozkład ergodyczny

Własność stacjonarności dla procesu stochastycznego $X_{\mathbb{Z}}$ definiujemy w dodatku A.7 w oparciu o pewną bijektywną operację T_n , która przesuwa proces w czasie o odstęp $n \in \mathbb{Z}$. Dla $X_{\mathbb{Z}} = (X_t)_{t \in \mathbb{Z}}$ mamy $T_n X_{\mathbb{Z}} := (X_{t+n})_{t \in \mathbb{Z}}$. Dla $n \in \mathbb{Z}$ możemy wyrazić $T_n = T^n$, gdzie $T := T_1$. Transformację T nazywa się przesunięciem (w lewo).

Określmy σ -ciało niezmiennicze jako $\tilde{\sigma}_X^T := \{A \in \tilde{\sigma}(X_{\mathbb{Z}}) : A = TA\}$. Definiujemy także σ -ciało przeciwobrazów $\sigma_X^T = X_{\mathbb{Z}}^{-1}(\tilde{\sigma}_X^T)$. Zauważmy, że

$$\tilde{\sigma}_X^T \ominus \tilde{\sigma}_X^{-\infty} \neq \emptyset, \quad \tilde{\sigma}_X^T \ominus \tilde{\sigma}_X^{\infty} \neq \emptyset, \quad (4.1)$$

gdzie $\tilde{\sigma}_X^{-\infty}$ i $\tilde{\sigma}_X^{\infty}$ są ogonowymi σ -ciałami obrazów (definicje w dodatku A.7). W szczególności istnieje taka przestrzeń probabilistyczna z niestacjonarnym procesem $X_{\mathbb{Z}}$, że $(\liminf_{n \rightarrow \infty} X_{2n} > 0) \in \sigma_X^{\infty} \setminus \sigma_X^T$ oraz $(\forall_{n \in \mathbb{Z}} X_n > 0) \in \sigma_X^T \setminus \sigma_X^{\infty}$.

Definicja 4.1. (proces ergodyczny i regularny) *Mówimy, że proces $X_{\mathbb{Z}}$ jest ergodyczny, jeżeli $P(A) \in \{0, 1\}$ dla każdego $A \in \sigma_X^T$. Proces $X_{\mathbb{Z}}$ nazywamy regularnym, jeżeli $P(A) \in \{0, 1\}$ dla każdego $A \in \sigma_X^{-\infty}$.*

NB. Miarę przeniesioną $P_{X_{\mathbb{Z}}} = P \circ X_{\mathbb{Z}}^{-1}$ będziemy nazywać odpowiednio ergodyczną, regularną lub stacjonarną, jeżeli proces $X_{\mathbb{Z}}$ jest ergodyczny, regularny lub stacjonarny.

Z twierdzenia 1.29 proces $X_{\mathbb{Z}}$ jest ergodyczny wtedy i tylko wtedy, gdy $\bar{H}(\sigma_X^T) = 0$. Podobnie $X_{\mathbb{Z}}$ jest regularny wtedy i tylko wtedy, gdy $\bar{H}(\sigma_X^{-\infty}) = 0$.

Powszechnie znany jest następujący rezultat (Kornfeld *et al.*, 1982, twierdzenie 8.§1.1; Taniguchi i Kakizawa, 2000, sekcja 1.3; Billingsley, 1979, twierdzenie 4.5):

Twierdzenie 4.2. *Stacjonarny ciąg niezależnych zmiennych losowych $Z_{\mathbb{Z}}$ jest ergodyczny i regularny.*

Twierdzenie o regularności ciągu niezależnych zmiennych losowych znane jest jako prawo zero-jedynkowe. Według dowodu Kornfelda *et al.* ciąg niezależnych zmiennych jest ergodyczny, gdyż jest procesem mieszającym (Kornfeld *et al.*, 1982, definicja 1.§6.2, *mixing*). Dowód alternatywny może wykorzystywać fakt, że ergodyczność wynika z regularności przy dodatkowych warunkach, por. twierdzenie 4.7 dla procesów borelowskich, a w kontekście procesów gaussowskich — twierdzenie 5.22 w książce Pourahmadiego (2001).

Regularność i ergodyczność nie są sobie równoważne. By podkreślić ten fakt, uogólnijmy pewien standardowy wynik (Taniguchi i Kakizawa, 2000, twierdzenie 1.3.3).

Twierdzenie 4.3. *Niech $X_{\mathbb{Z}}$ i $Y_{\mathbb{Z}}$ będą dowolnymi procesami.*

- (i) *Jeżeli $X_{\mathbb{Z}}$ jest stacjonarny i istnieje funkcja f mierzalna $\tilde{\sigma}(X_{\mathbb{Z}})/\tilde{\sigma}(Y_n)$ taka, że $Y_n = f((X_{t+n})_{t \in \mathbb{Z}})$ dla każdego n , to $Y_{\mathbb{Z}}$ jest także stacjonarny.*
- (ii) *Jeżeli $X_{\mathbb{Z}}$ jest ergodyczny i istnieje funkcja f mierzalna $\tilde{\sigma}(X_{\mathbb{Z}})/\tilde{\sigma}(Y_n)$ taka, że $Y_n = f((X_{t+n})_{t \in \mathbb{Z}})$ dla każdego n , to $Y_{\mathbb{Z}}$ jest także ergodyczny.*
- (iii) *Jeżeli $X_{\mathbb{Z}}$ jest regularny i istnieje funkcja f mierzalna $\tilde{\sigma}(X_{-\infty:n})/\tilde{\sigma}(Y_n)$ taka, że $Y_n = f(X_{-\infty:n})$ dla każdego n , to $Y_{\mathbb{Z}}$ jest także regularny.*

Dowód:

- (i) Oznaczmy $f_n(x_{\mathbb{Z}}) := f((x_{t+n})_{t \in \mathbb{Z}})$ oraz $f^*(x_{\mathbb{Z}}) := (f_t(x_{\mathbb{Z}}))_{t \in \mathbb{Z}}$. Mamy $Y_n = f_n(X_{\mathbb{Z}})$ oraz $Y_{\mathbb{Z}} = f^*(X_{\mathbb{Z}})$. Dla dowolnego $A \in \tilde{\sigma}(X_{\mathbb{Z}})$ oznaczmy $A_n = f_n A$ oraz $A^* = f^* A$ (dla skrótu opuszczamy nawiasy). Mamy $A_{n+1} = f_n T A$, a zatem $T A^* = T f^* A = f^* T A$. Zawężmy zakres zmiennej A . Dla dowolnego $A^* \in \tilde{\sigma}(Y_{\mathbb{Z}})$ oznaczmy $A = [f^*]^{-1} A^*$. Mamy nadal $A^* = f^* A$, a zatem $T A^* = T f^* A = f^* T A$. Stąd stacjonarność $X_{\mathbb{Z}}$ implikuje stacjonarność $Y_{\mathbb{Z}}$.
- (ii) Dla dowolnego $A^* \in \tilde{\sigma}(Y_{\mathbb{Z}})$ oznaczmy $A = [f^*]^{-1} A^*$. Z tożsamości $T A^* = T f^* A = f^* T A$ z punktu poprzedniego wynika $A^* \ominus T A^* = (f^* A) \ominus (T f^* A) = (f^* A) \ominus (f^* T A) = f^*(A \ominus T A)$. W szczególności $A^* \in \tilde{\sigma}_Y^T \iff A \in \tilde{\sigma}_X^T$. Dla $A^* \in \tilde{\sigma}_Y^T$ mamy także $Y_{\mathbb{Z}}^{-1} A^* = X_{\mathbb{Z}}^{-1} A$, a zatem $\sigma_Y^T \subset \sigma_X^T$. Stąd ergodyczność $X_{\mathbb{Z}}$ implikuje ergodyczność $Y_{\mathbb{Z}}$.
- (iii) Z założenia wynika, że $\sigma(Y_n) \subset \sigma(X_{-\infty:n})$. Stąd mamy także $\sigma(Y_{-\infty:n}) = \sigma(\bigcup_{j \leq n} \sigma(Y_j)) \subset \sigma(X_{-\infty:n})$. Jako że $\sigma_Y^{-\infty} = \bigcap_n \sigma(Y_{-\infty:n}) \subset \bigcap_n \sigma(X_{-\infty:n}) = \sigma_X^{-\infty}$, to regularność $X_{\mathbb{Z}}$ implikuje regularność $Y_{\mathbb{Z}}$. □

Definicja 4.4. (proces kwaziliniowy) *Proces $Y_{\mathbb{Z}}$ nazywamy procesem kwaziliniowym, jeżeli istnieje stacjonarny ciąg niezależnych zmiennych losowych $Z_{\mathbb{Z}}$ oraz funkcja f mierzalna $\tilde{\sigma}(Z_{\mathbb{Z}})/\tilde{\sigma}(Y_n)$ taka, że $Y_n = f((Z_{t+n})_{t \in \mathbb{Z}})$. Analogicznie proces $Y_{\mathbb{Z}}$ nazywamy procesem jednostronnie kwaziliniowym, jeżeli istnieje stacjonarny ciąg niezależnych zmiennych losowych $Z_{\mathbb{Z}}$ oraz funkcja f mierzalna $\tilde{\sigma}(Z_{-\infty:n})/\tilde{\sigma}(Y_n)$ taka, że $Y_n = f(Z_{-\infty:n})$.*

Procesy kwaziliniowe są uogólnieniem gaussowskich procesów czysto niedeterministycznych, zwanych także procesami jednostronnie liniowymi (*one-sided linear processes*). Procesy kwaziliniowe są ergodyczne na mocy twierdzeń 4.2–4.3. Podobnie procesy jednostronnie kwaziliniowe są regularne. Oprócz procesów kwaziliniowych ważną klasą procesów ergodycznych są stacjonarne jednoklasowe przeliczalne łańcuchy Markowa (Hernandez-Lerma i Lasserre, 2003, sekcje 2.4, 2.5, 3.2 i 3.3; Kemeny *et al.*, 1966; zob. też dodatek A.8 i przykład 2.13). Z twierdzenia 4.3 ergodyczne są także ukryte łańcuchy Markowa z ergodycznym procesem ukrytym (Ephraim i Merhav, 2002).

Znanym rezultatem jest następujące twierdzenie.

Twierdzenie 4.5. (ergodyczne, Kallenberg, 1997, twierdzenie 9.6) *Dla dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ i dowolnej funkcji $f \in L^1(\tilde{\sigma}(X_{\mathbb{Z}}), P_{X_{\mathbb{Z}}})$ zachodzi zbieżność*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k} X_{\mathbb{Z}}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=0}^{n-1} f(T^k X_{\mathbb{Z}}) = \langle f(X_{\mathbb{Z}}) | \sigma_X^T \rangle \quad \text{prawie na pewno.} \quad (4.2)$$

Zauważmy, że jeżeli proces $X_{\mathbb{Z}}$ jest ergodyczny, to mamy także $\langle f(X_{\mathbb{Z}}) | \sigma_X^T \rangle = \langle f(X_{\mathbb{Z}}) \rangle$ (por. własność (iv) prawdopodobieństwa warunkowego, dodatek A.5). Prosty dowód twierdzenia ergodycznego istnieje dla $f \in L^2(\tilde{\sigma}(X_{\mathbb{Z}}), P_{X_{\mathbb{Z}}})$ (Varadhan, 2001, twierdzenie 6.1). Istnieją także uogólnienia twierdzenia ergodycznego na przypadek procesów asymptotycznie średnio stacjonarnych (Gray, 1987, twierdzenie 7.2.1; Gray i Kieffer, 1980).

Niech proces stacjonarny $X_{\mathbb{Z}}$ będzie zmienną borelowską (definicja A.14). Przypomnijmy, że borelowskimi zmiennymi losowymi są wszystkie procesy stacjonarne $X_{\mathbb{Z}}$ o wartościach dyskretnych lub rzeczywistych. Przestrzeń wszystkich (przeniesionych) miar stacjonarnych określonych na przestrzeni mierzalnej $(X_{\mathbb{Z}}(\Omega), \tilde{\sigma}(X_{\mathbb{Z}}))$ oznaczamy będziemy jako $\mathbb{S} \subset \mathbb{R}^{\tilde{\sigma}(X_{\mathbb{Z}})}$. W szczególności $P_{X_{\mathbb{Z}}} \in \mathbb{S}$. Niech $\mathbb{E} \subset \mathbb{S}$ będzie zbiorem stacjonarnych miar ergodycznych. Aby móc rozpatrywać miary losowe o wartościach w \mathbb{S} definiujemy przestrzeń mierzalną $(\mathbb{S}, \mathcal{S})$, gdzie \mathcal{S} jest σ -ciałem miar stacjonarnych. Formalnie $\mathcal{S} := \sigma(\mathcal{R}_0^{\tilde{\sigma}(X_{\mathbb{Z}})})|_{\mathbb{S}}$ jest obcięciem $\sigma(\mathcal{R}_0^{\tilde{\sigma}(X_{\mathbb{Z}})})$ do zbioru \mathbb{S} , gdzie ciało $\mathcal{R}_0^{\tilde{\sigma}(X_{\mathbb{Z}})}$ definiujemy wzorem (A.9). Podob-

ną przestrzeń mierzalną $(\mathbb{E}, \mathcal{E})$, gdzie $\mathcal{E} := \sigma(\mathcal{R}_0^{\tilde{\sigma}(X_{\mathbb{Z}})})|_{\mathbb{E}}$, definiujemy dla losowych miar ergodycznych.

Twierdzenie 4.6. (rozkład ergodyczny) *Niech proces stacjonarny $X_{\mathbb{Z}}$ będzie borelowską zmienną losową na przestrzeni (Ω, \mathcal{J}, P) .*

(i) *Na przestrzeni $(X_{\mathbb{Z}}(\Omega), \tilde{\sigma}(X_{\mathbb{Z}}))$ istnieje funkcja $\phi : X_{\mathbb{Z}}(\Omega) \ni x_{\mathbb{Z}} \rightarrow \phi(\cdot)(x_{\mathbb{Z}}) \in \mathbb{E}$ mierzalna $\tilde{\sigma}(X_{\mathbb{Z}})|_{\mathcal{E}}$ taka, że $F_X = \phi(\cdot)(X_{\mathbb{Z}})$ jest ergodyczną miarą losową ($F_X(\cdot)(\omega) = \phi(\cdot)(X_{\mathbb{Z}}(\omega))$), $F_X(\cdot)(\Omega) \subset \mathbb{E}$, $\tilde{\sigma}(F_X) = \mathcal{E}|_{F_X(\cdot)(\Omega)}$ i równości*

$$F_X(A) = P(X_{\mathbb{Z}} \in A | \sigma_X^T) \quad (4.3)$$

zachodzą prawie na pewno jednocześnie dla wszystkich $A \in \tilde{\sigma}(X_{\mathbb{Z}})$.

(ii) *Dla miar $P_{F_X} = P \circ F_X$ i $P_{X_{\mathbb{Z}}} = P \circ X_{\mathbb{Z}}$ zachodzi*

$$P_{X_{\mathbb{Z}}}(A) = \langle F_X(A) \rangle = \int \mu(A) dP_{F_X}(\mu) \quad (4.4)$$

dla każdego $A \in \tilde{\sigma}(X_{\mathbb{Z}})$. Fakt ten zapisujemy skrótowo jako

$$P_{X_{\mathbb{Z}}} = \langle F_X \rangle = \int \mu dP_{F_X}(\mu). \quad (4.5)$$

(iii) *Dla dowolnej miary prawdopodobieństwa ν na \mathcal{S} zachodzi $\int \mu d\nu(\mu) \in \mathbb{S}$.*

(iv) *Dla dowolnej miary prawdopodobieństwa ν na \mathcal{E} takiej, że $P_{X_{\mathbb{Z}}} = \int \mu d\nu(\mu)$ mamy*

$$\mu(\{x_{\mathbb{Z}} \in X_{\mathbb{Z}}(\Omega) : \phi(\cdot)(x_{\mathbb{Z}}) = \mu\}) = 1 \quad \text{dla } \nu\text{-prawie wszystkich } \mu \in \mathbb{E}. \quad (4.6)$$

(v) *Dla dowolnej miary prawdopodobieństwa ν na \mathcal{E} mamy równoważność*

$$P_{X_{\mathbb{Z}}} = \int \mu d\nu(\mu) \iff \nu|_{\tilde{\sigma}(F_X)} = P_{F_X}. \quad (4.7)$$

(vi) *$P_{X_{\mathbb{Z}}} \in \mathbb{E}$ wtedy i tylko wtedy, gdy $P_{F_X}(\{\mu\}) = 1$ dla pewnego $\mu \in \mathbb{E}$.*

Pkt. (i) pochodzi z twierdzenia 9.10 w książce Kallenberg'a 1997. Pkt. (ii)-(vi) są wnioskami z twierdzenia 9.12 tamże i jego dowodu.

Równość (4.6) jest głęboką konsekwencją twierdzenia ergodycznego. Jest ona kluczowa dla jednoznaczności rozkładu w pkt. (v), a w połączeniu z pkt. (v) — dla rozważań na temat kodów uniwersalnych. Gray i Davisson (1974) w kontekście teorii informacji podkreślają wagę pewnych stwierdzeń zbliżonych do (4.6).

Rozkład ergodyczny zadaje jednoznaczną parametryzację miar stacjonarnych. Zauważmy, że zbiór wszystkich miar stacjonarnych \mathbb{S} jest wypukły. Jeżeli miary $\mu_i \in \mathbb{S}$ są stacjonarne, to miara $\mu = \sum_i c_i \mu_i$ jest stacjonarna dla $c_i \geq 0$, $\sum_i c_i = 1$. W świetle twierdzenia 4.6 ta sama miara $\mu \in \mathbb{S}$ jest nieergodyczna, jeżeli $c_i > 0$ dla więcej niż jednego i , a μ_i są różne. Zbiór miar ergodycznych \mathbb{E} jest zatem zbiorem punktów ekstremalnych zbioru \mathbb{S} .

Pomimo nierówności (4.1) mamy następujące twierdzenie.

Twierdzenie 4.7. *Niech proces stacjonarny $X_{\mathbb{Z}}$ będzie borelowską zmienną losową. Dla uzupełnionych σ -ciał zachodzi relacja*

$$\sigma(F_X)^P \subset (\sigma_X^{-\infty})^P \cap (\sigma_X^T)^P \cap (\sigma_X^{\infty})^P. \quad (4.8)$$

Dowód: Z twierdzenia 4.6 dla każdego $A \in \tilde{\sigma}(X_{\mathbb{Z}})$ mamy $F_X(A) = P(X \in A | \sigma_X^T)$ prawie na pewno, gdzie $\sigma(P(X \in A | \sigma_X^T)) \subset \sigma_X^T$. W świetle twierdzenia A.19 pkt. (ii) mamy zatem $\sigma(F_X(A))^P \subset (\sigma_X^T)^P$. Ponieważ $\sigma(F_X) = \sigma\left(\bigcup_{A \in \tilde{\sigma}(X_{\mathbb{Z}})} \sigma(F_X(A))\right)$, zatem $\sigma(F_X)^P \subset (\sigma_X^T)^P$.

Zdefiniujmy rozszerzoną prostą rzeczywistą $\mathbb{R}^\circ := \mathbb{R} \cup \{-\infty, \infty, \perp\}$, gdzie \perp oznacza wartość niezdefiniowaną. Oznaczmy $\mathcal{R}^\circ := \sigma(\mathcal{R} \cup \{-\infty\} \cup \{\infty\} \cup \{\perp\})$, a także

$$\lim_{n \rightarrow \infty} f_n := \begin{cases} \limsup_{n \rightarrow \infty} f_n & \text{gdy } \limsup_{n \rightarrow \infty} f_n = \liminf_{n \rightarrow \infty} f_n, \\ \perp, & \text{inaczej,} \end{cases} \quad (4.9)$$

gdzie funkcje $\limsup_{n \rightarrow \infty} f_n$ i $\liminf_{n \rightarrow \infty} f_n$ przybierają wartości z zakresu $\mathbb{R} \cup \{-\infty, \infty\}$. Dla dowolnego ciągu $(f_n)_{n \in \mathbb{N}}$ funkcji mierzalnych $\mathcal{G}/\mathcal{R}^\circ$ funkcja $\lim_{n \rightarrow \infty} f_n$ jest także mierzalna $\mathcal{G}/\mathcal{R}^\circ$, gdyż tak samo mierzalne są $\limsup_{n \rightarrow \infty} f_n$ i $\liminf_{n \rightarrow \infty} f_n$ (Billingsley, 1979, twierdzenie 13.4). Oznaczmy $I_A(x) := \llbracket x \in A \rrbracket$. Dla każdego $A_+ \in \tilde{\sigma}(X_{1:\infty})$ zmienna $F_+(A_+) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^{p+n-1} I_{A_+}(X_{k:\infty})$ nie zależy od p i jest mierzalna $\sigma(X_{p:\infty})/\mathcal{R}^\circ$. A zatem $F_+(A_+)$ jest mierzalna $\sigma_X^\infty/\mathcal{R}^\circ$. Podobnie dla każdego $A_- \in \tilde{\sigma}(X_{-\infty:0})$ zmienna $F_-(A_-) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^{p+n-1} I_{A_-}(X_{-\infty:-k})$ nie zależy od p i jest mierzalna $\sigma_X^{-\infty}/\mathcal{R}^\circ$.

Oznaczmy teraz $\mathcal{G}_+ := X_{\mathbb{Z}}(\sigma(X_{1:\infty}))$, $\mathcal{G}_- := X_{\mathbb{Z}}(\sigma(X_{-\infty:0}))$. Z twierdzeń 4.5 i 4.6 dla każdego A_+ i A_- mamy

$$F_+(A_+) = F_X(X_{\mathbb{Z}}(X_{1:\infty}^{-1}(A_+))), \quad F_-(A_-) = F_X(X_{\mathbb{Z}}(X_{-\infty:0}^{-1}(A_-))) \quad \text{prawie na pewno.}$$

Z twierdzenia A.19 pkt. (ii) wynika zatem $\sigma(F_X(A))^P \subset (\sigma_X^\infty)^P$ dla każdego $A \in \mathcal{G}_+$ oraz $\sigma(F_X(A))^P \subset (\sigma_X^{-\infty})^P$ dla każdego $A \in \mathcal{G}_-$. Reasumując, $\sigma(F_X|_{\mathcal{G}_+})^P \subset (\sigma_X^\infty)^P$ oraz $\sigma(F_X|_{\mathcal{G}_-})^P \subset (\sigma_X^{-\infty})^P$. Wiadomo, że dowolna miara stacjonarna $\mu \in \mathbb{S}$ jest unikatowym rozszerzeniem na $\tilde{\sigma}(X_{\mathbb{Z}})$ swoich obcięć $\mu|_{\mathcal{G}_+}$ oraz $\mu|_{\mathcal{G}_-}$ (Kallenberg, 1997, lemat 9.2), więc mamy także $\sigma(F_X)^P \subset (\sigma_X^\infty)^P \cap (\sigma_X^{-\infty})^P$. \square

Z twierdzeń 4.7 i 4.6 pkt. (vi) każdy regularny proces borelowski jest ergodyczny, gdyż zachodzą implikacje $\bar{H}(\sigma_X^{-\infty}) = 0 \implies \bar{H}(F_X) = 0 \implies \bar{H}(\sigma_X^T) = 0$.

Z zawierania (4.8) dla dowolnej zmiennej Z mamy $F_X \perp\!\!\!\perp Z | X_{\mathbb{N}}$. Jeżeli zmienne X_i przybierają skończoną liczbę wartości, to mamy także

$$\begin{aligned} 0 &= \bar{I}(X_0; F_X | X_{\mathbb{N}}) = \lim_{n \rightarrow \infty} \bar{I}(X_0; F_X | X_{1:n}) \\ &= \lim_{n \rightarrow \infty} \bar{H}(X_0 | X_{1:n}) - \lim_{n \rightarrow \infty} \bar{H}(X_0 | F_X \times X_{1:n}) \end{aligned} \quad (4.10)$$

w świetle rezultatów z rozdziałów 1 i 3. Dowodzenie równości (4.10) bez rozwiniętego w tych rozdziałach aparatu jest żmudne, zob. Gray i Davisson (1974, twierdzenie 5.1).

Niech $X_{\mathbb{Z}}$ będzie dyskretnym procesem stacjonarnym. Dla stacjonarnej miary prawdopodobieństwa $\mu \in \mathbb{S}$ niech $H(n; \mu)$ będzie entropią blokową tej miary (definicja w uwadze po definicji 2.1). Dla stacjonarnej miary μ wprowadziliśmy także oznaczenia intensywności entropii $h(\mu)$ i entropii nadwyżkowej $E(\mu)$. Zgodnie z tymi oznaczeniami $H_X(n) = H(n; P_{X_{\mathbb{Z}}})$, $h_X = h(P_{X_{\mathbb{Z}}})$, $E_X = E(P_{X_{\mathbb{Z}}})$. Jeżeli proces $X_{\mathbb{Z}}$ jest ergodyczny, to $F_X = P_{X_{\mathbb{Z}}}$ prawie na pewno i mamy

$$H(n; F_X) = H_X(n), \quad h(F_X) = h_X, \quad E(F_X) = E_X \quad \text{prawie na pewno.}$$

W szczególności, jeżeli zmienna losowa $h(F_X)$ prawie na pewno nie jest stałą (tzn. gdy $\bar{H}(h(F_X)) > 0$), to F_X nie jest stałą, a proces $X_{\mathbb{Z}}$ nie jest ergodyczny.

Twierdzenie 4.8. Niech $X_{\mathbb{Z}}$ będzie procesem stacjonarnym, gdzie zmienne X_i są dyskretne. Dla każdego n zachodzi równość

$$\langle H(n; F_X) \rangle = \bar{H}(X_{1:n}|F_X). \quad (4.11)$$

Dowód: Z twierdzenia A.30 mamy równość prawie na pewno

$$F_X(A) = P(X_{\mathbb{Z}} \in A | \sigma_X^T) = P(X_{\mathbb{Z}} \in A | \sigma(F_X)).$$

Niech $\text{Cyl}_{(1,\dots,n)}^{\mathbb{Z}}(X_1, \dots, X_n)$ będzie dyskretnym σ -ciałem cylindrów (A.6). Oznaczmy zbiór jego atomów $\mathcal{G}_n := \mathcal{A}(\text{Cyl}_{(1,\dots,n)}^{\mathbb{Z}}(X_1, \dots, X_n))$, zob. definicja 1.15. Z definicji entropii zmiennych dyskretnych (miarami pomiarowymi są miary liczące) mamy

$$H(n; F_X) = - \sum_{A \in \mathcal{G}_n} F_X(A) \log F_X(A).$$

Z drugiej strony ze wzoru (1.38) mamy

$$\begin{aligned} \bar{H}(X_{1:n}|F_X) &= \int \left[- \sum_{A \in \mathcal{G}_n} P(X_{\mathbb{Z}} \in A | \sigma(F_X)) \log(X_{\mathbb{Z}} \in A | \sigma(F_X)) \right] dP \\ &= \int \left[- \sum_{A \in \mathcal{G}_n} F_X(A) \log F_X(A) \right] dP = \langle H(n; F_X) \rangle. \end{aligned}$$

□

Twierdzenie 4.9. Niech $X_{\mathbb{Z}}$ będzie procesem stacjonarnym, gdzie zmienne X_i przybierają skończoną liczbę wartości. Zachodzą związki

$$h_X = \langle h(F_X) \rangle, \quad (4.12)$$

$$E_X \geq \bar{H}(F_X) + \langle E(F_X) \rangle. \quad (4.13)$$

Dowód: Niech X_i przybierają M różnych wartości, gdzie $M \in \mathbb{N}$. Oznaczmy $K = \log M$. Mamy $0 \leq K - \Delta H(i; F_X) \leq K - \Delta H(i+1; F_X)$. Z twierdzenia 4.8, równości (3.2) i twierdzenia o zbieżności monotonicznej (Rudin, 1986, sekcja 1.26) wynika

$$\begin{aligned} \langle K - h(F_X) \rangle &= \left\langle \lim_{n \rightarrow \infty} [K - \Delta H(n; F_X)] \right\rangle = \lim_{n \rightarrow \infty} \langle K - \Delta H(n; F_X) \rangle \\ &= \lim_{n \rightarrow \infty} [K - \bar{H}(X_0|F_X \times X_{1:n-1})] = [K - \bar{H}(X_0|F_X \times X_{1:\infty})] \\ &= [K - \bar{H}(X_0|X_{1:\infty})] = [K - h_X], \end{aligned}$$

gdyż $\sigma(F_X)^P \subset \sigma(X_{1:\infty})^P$ (twierdzenie 4.7).

Analogicznie z twierdzenia 4.8, równości (3.1) i twierdzenia o zbieżności monotonicznej mamy

$$\begin{aligned} \langle E(F_X) \rangle &= \left\langle \lim_{n \rightarrow \infty} E(n; F_X) \right\rangle = \lim_{n \rightarrow \infty} \langle E(n; F_X) \rangle \\ &= \lim_{n \rightarrow \infty} \bar{I}(X_{-n+1:0}; X_{1:n}|F_X) = \bar{I}(X_{-\infty:0}; X_{1:\infty}|F_X). \end{aligned}$$

Z nierówności przetwarzania danych i twierdzenia 4.7 wynika, że $\bar{H}(F_X) = \bar{I}(F_X; F_X) \leq \bar{I}(X_{-\infty:0}; F_X)$. Z addytywności $\bar{I}(X_{-n+1:0}; F_X) + \bar{I}(X_{-n+1:0}; X_{1:n}|F_X) = \bar{I}(X_{-n+1:0}; F_X \times X_{1:n})$ (twierdzenia 1.21 i 1.22) i równości (3.1) mamy z kolei

$$\bar{I}(X_{-\infty:0}; F_X) + \bar{I}(X_{-\infty:0}; X_{1:\infty}|F_X) = \bar{I}(X_{-\infty:0}; F_X \times X_{1:\infty}) = \bar{I}(X_{-\infty:0}; X_{1:\infty}) = E_X,$$

skąd otrzymujemy (4.13). □

Na mocy twierdzenia 4.9 wszystkie procesy dyskretne $X_{\mathbb{Z}}$, dla których $\bar{H}(F_X) = \infty$, są niefinitarne. W istocie wszystkie procesy $X_{\mathbb{Z}}$ będące zmiennymi borelowskimi są niefinitarne i nieergodyczne, jeżeli tylko $\bar{H}(F_X) = \infty$. Wynika to z twierdzenia 4.7 oraz z nierówności przetwarzania danych

$$E_X = \bar{I}(X_{-\infty:0}; X_{1:\infty}) \geq \bar{I}(\sigma_X^{-\infty}; \sigma_X^{\infty}) \geq \bar{I}(F_X; F_X) = \bar{H}(F_X). \quad (4.14)$$

Nierówność (4.13) jest jednak znacznie mocniejsza niż (4.14). Prowadzi ona także do pewnych interesujących wniosków w teorii kodów uniwersalnych (twierdzenie 6.12).

Miara $\mu \in \mathbb{S}$ jest nieergodyczna, jeżeli umiemy przedstawić $\mu = \sum_i c_i \mu_i$, gdzie $\mu_i \in \mathbb{S}$ są różne, a $c_i > 0$ dla więcej niż jednego i . Prostsze kryterium nieergodyczności wynika z twierdzenia Shannona-McMillana-Breimanna o ekwipartycji.

Twierdzenie 4.10. (o ekwipartycji) *Niech $X_{\mathbb{Z}}$ będzie procesem stacjonarnym, gdzie zmienne X_i przybierają skończoną liczbę wartości. Oznaczmy zmienne*

$$P(X_{1:n} = X_{1:n}(\cdot)) : \Omega \ni \omega \mapsto P(X_{1:n} = X_{1:n}(\omega)) \quad (4.15)$$

Zachodzi równość

$$-\lim_{n \rightarrow \infty} \frac{1}{n} \log P(X_{1:n} = X_{1:n}(\cdot)) = L_X \quad \text{prawie na pewno,} \quad (4.16)$$

gdzie L_X jest pewną zmienną losową spełniającą $\langle L_X \rangle = h_X$. Ponadto, jeżeli proces $X_{\mathbb{Z}}$ jest ergodyczny, to $L_X = h_X$ prawie na pewno (Billingsley, 1965, twierdzenie 13.1).

NB. Pomimo tego, że zachodzi $\langle L_X \rangle = h_X = \langle h(F_X) \rangle$, twierdzenie o ekwipartycji nie orzeka, że zachodzi równość $L_X = h(F_X)$ prawie na pewno w przypadku ogólnym. Po kilku bezskutecznych próbach dowodu przypuszczamy, że równość ta może nie być prawdziwa.

Część II

Procesy dyskretne

Rozdział 5

Procesy nieprzeliczalnego opisu

W rozdziałach 2–4 zdefiniowaliśmy trzy asymptotyczne własności niektórych procesów stacjonarnych: finitarność (definicje 2.9 i 3.2), niedeterminizm (definicja 3.6) i ergodyczność (definicja 4.1). Są to trzy własności, które są własnościami nie tyle procesu stochastycznego $X_{\mathbb{Z}}$ rozumianego jako funkcja $\Omega \rightarrow X_{\mathbb{Z}}(\Omega)$, co σ -ciała $\sigma(X_{\mathbb{Z}})$ generowanego przez proces. Z tego względu finitarność, niedeterminizm i ergodyczność są niezmiennikami szerokich klas bijektywnych transformacji całych procesów. Dla każdej z tych własności dokładna klasa zachowujących ją bijekcji jest nieco inna.

Warto podkreślić, że finitarność jest własnością istotnie różną od ergodyczności i niedeterminizmu. Problem klasyfikacji różnych własności asymptotycznych omawiamy dość szczegółowo w przypadku procesów gaussowskich (podsumowanie i diagram w rozdziale 13). Dla procesów gaussowskich finitarność implikuje zarówno ergodyczność (twierdzenia 10.3, 10.4, 10.5) jak i niedeterminizm (twierdzenie 9.3 i nierówność 2.20). Z drugiej strony istnieją przykłady procesów gaussowskich, które są: (a) niefinitarne lecz ergodyczne i niedeterministyczne, (b) nieergodyczne lecz niedeterministyczne, (c) deterministyczne lecz ergodyczne.

Dla procesów dyskretnych nie dysponujemy na razie klasyfikacją równie szczegółową jak dla procesów gaussowskich. Nie wiemy, czy pewne kombinacje warunków (nie)finitarności, (nie)ergodyczności i (nie)determinizmu nie są wykluczone. W rozdziale 2 podaliśmy przykład niefinitarnego lecz ergodycznego procesu Markowa o nieskończonej przeliczalnej liczbie wartości (przykład 2.13). Dla porównania, nie znamy żadnego procesu o skończonej liczbie wartości, który byłby jednocześnie niefinitarny i ergodyczny.

Korzystając z rozkładu ergodycznego (twierdzenie 4.9), bardzo łatwo jest konstruować przykłady dyskretnych procesów nieergodycznych, które są finitarne lub niefinitarne. Niech miary $\mu_i \in \mathbb{E}$ będą miarami stacjonarnych jednoklasowych (a więc ergodycznych) procesów Markowa. Jeżeli $\mu = \sum_{i \in J} c_i \mu_i$, to

- (i) μ jest miarą nieergodycznego procesu finitarnego, jeżeli J jest zbiorem skończonym,
- (ii) μ jest miarą nieergodycznego procesu niefinitarnego, jeżeli $-\sum_{i \in J} c_i \log c_i = \infty$.

Pkt. (i) wynika z twierdzenia 2.11, pkt. (ii) — z nierówności (4.14).

W niniejszym rozdziale przedstawimy pewną szeroką klasę procesów, nazwanych przez nas procesami nieprzeliczalnego opisu, które są jednocześnie niefinitarne i nieergodyczne na mocy rozumowania zbliżonego do nierówności (4.14). Rozumowanie to nie korzysta z rozkładu ergodycznego i w przypadku procesów dyskretnych pozwala uzyskać mocne twierdzenie nie odwołujące się jawnie do aparatu teoriomiarowego.

Z równości (3.1) wynika prosty warunek dostateczny niefinitarności procesu.

Twierdzenie 5.1. *Proces stacjonarny $X_{\mathbb{Z}}$ jest niefinitarny, jeżeli istnieje zmienna losowa Y o entropii teoriomiarowej $\bar{H}(Y) = \infty$, liczba $n \in \mathbb{Z}$ oraz funkcje f_- mierzalna $\bar{\sigma}(X_{-\infty:n-1})/\bar{\sigma}(Y)$ i f_+ mierzalna $\bar{\sigma}(X_{n:\infty})/\bar{\sigma}(Y)$ takie, że*

$$f_-(X_{-\infty:n-1}) = Y = f_+(X_{n:\infty}) \quad \text{prawie na pewno.} \quad (5.1)$$

NB. Równość $\bar{H}(Y) = \infty$ zachodzi w dwóch przypadkach szczególnych

- (i) jeżeli Y jest zmienną dyskretną i $H(Y) = \infty$, por. tożsamość (1.29),
- (ii) jeżeli Y jest ciągłą zmienną rzeczywistą, por. twierdzenie 1.25.

Dowód: Zauważmy, że $\sigma(Y)^P \subset \sigma(X_{-\infty:n-1})^P \cap \sigma(X_{n:\infty})^P$. Z równości (3.1) i nierówności przetwarzania danych $E_X = \bar{I}(X_{-\infty:n-1}; X_{n:\infty}) \geq \bar{I}(Y; Y) = \bar{H}(Y) = \infty$, czyli $X_{\mathbb{Z}}$ jest niefinitarny. \square

Gdy Y jest ciągłą zmienną rzeczywistą, twierdzenie 5.1 jest intuicyjne. Losowa liczba rzeczywista jest ciągiem nieskończenie wielu niezależnych bitów, a zatem powinna nieść nieskończoną informację. Jeżeli możemy ustalić wartości wszystkich tych bitów, obserwując przeszłość i przyszłość procesu niezależnie, to informacja wzajemna między przeszłością a przyszłością musi być nieskończona. Pomimo intuicyjności nieformalnego rozumowania, ścisły dowód musi odwoływać się do dość wyrafinowanych rozważań z teorii miary. W szczególności trzeba założyć, że Y nie jest zmienną o skończonej liczbie wartości.

Interesującą kwestią jest, czy w twierdzeniu 5.1 słowo “jeżeli” można zastąpić przez frazę “wtedy i tylko wtedy, gdy”. W rozdziale 10 pokazujemy, że niefinitarne są wszystkie procesy gaussowskie, które nie są czysto niedeterministyczne. Dowód polega na jawnej konstrukcji zmiennej Y przy wykorzystaniu jednoznaczności rozkładu Wolda i symetrii odbicia czasu. W rozdziale 11 udowadniamy z kolei, że czysto niedeterministyczne są wszystkie tak zwane procesy kwazifinitarne. Ponieważ istnieją kwazifinitarne procesy niefinitarne, oznacza to, że istnieją gaussowskie procesy niefinitarne $X_{\mathbb{Z}}$, dla których nie znamy jawnej konstrukcji zmiennej Y , a nawet nie wiemy, czy taka zmienna istnieje.

Delikatnie modyfikując twierdzenie 5.1 otrzymujemy dwa podobne stwierdzenia.

Twierdzenie 5.2. *Proces stacjonarny $X_{\mathbb{Z}}$ jest nieergodyczny, jeżeli istnieje zmienna Y o entropii teoriomiarowej $\bar{H}(Y) > 0$ oraz funkcja f mierzalna $\tilde{\sigma}(X_{\mathbb{Z}})/\tilde{\sigma}(Y)$ taka, że dla każdego $n \in \mathbb{Z}$ zachodzi*

$$f((X_{t+n})_{t \in \mathbb{Z}}) = Y \quad \text{prawie na pewno.} \quad (5.2)$$

Dowód: Weźmy dowolne $A \in \tilde{\sigma}(Y)$. Zauważmy, że dla każdego n różnica $(Y \in A) \ominus (X_{\mathbb{Z}} \in T^n f^{-1}(A))$ jest zbiorem P -miary 0. Oznacza to, że różnice symetryczne dowolnych dwóch spośród zbiorów $T^n f^{-1}(A)$, $n \in \mathbb{Z}$, są także zbiorami $P_{X_{\mathbb{Z}}}$ -miary 0. Ponieważ $\{A \in \tilde{\sigma}(X_{\mathbb{Z}}) : P(X_{\mathbb{Z}} \in A \ominus TA) = 0\} = (\tilde{\sigma}_X^T)^{P_{X_{\mathbb{Z}}}}$ (Kallenberg, 1997, lemat 9.4), to $f^{-1}(A) \in (\tilde{\sigma}_X^T)^{P_{X_{\mathbb{Z}}}}$. W rezultacie $\sigma(Y) \subset (\sigma_X^T)^P$. Stąd mamy $\bar{H}(\sigma_X^T) \geq \bar{H}(Y) > 0$, czyli $X_{\mathbb{Z}}$ jest nieergodyczny. \square

Twierdzenie 5.3. *Proces stacjonarny $X_{\mathbb{Z}}$ jest nieergodyczny i niefinitarny, jeżeli istnieje zmienna Y o entropii teoriomiarowej $\bar{H}(Y) = \infty$ oraz funkcje f_{-} mierzalna $\tilde{\sigma}(X_{-\infty:n-1})/\tilde{\sigma}(Y)$ i f_{+} mierzalna $\tilde{\sigma}(X_{n:\infty})/\tilde{\sigma}(Y)$ takie, że dla każdego $n \in \mathbb{Z}$ zachodzi*

$$f_{-}(X_{-\infty:n-1}) = Y = f_{+}(X_{n:\infty}) \quad \text{prawie na pewno.} \quad (5.3)$$

NB. Jedyna różnica między twierdzeniami 5.1 i 5.3 polega na tym, że w pierwszym pojawia się sformułowanie „dla pewnego n ”, a w drugim — „dla każdego n ”.

Dowód: Teza wynika z zastosowania twierdzeń 5.1 i 5.2. \square

Twierdzenie 5.3 wygodnie jest ograniczyć do postaci, która nie operuje pojęciami teoriomiarowymi. Postać ta jest szczególnie przydatna, gdy procesy $X_{\mathbb{Z}}$ są dyskretne. Wprowadźmy następującą klasę procesów:

Definicja 5.4. (proces opisu) *Proces stacjonarny $X_{\mathbb{Z}}$ nazywamy procesem nieprzeliczalnego opisu, jeżeli istnieje proces Bernoulliego $Z_{\mathbb{N}}$ (definicja 1.27) i funkcje f_{nk} takie, że prawdopodobieństwo $P(f_{nk}(X_{i+1:i+n}) = Z_k)$ nie zależy od $i \in \mathbb{Z}$ oraz*

$$\lim_{n \rightarrow \infty} P(f_{nk}(X_{1:n}) = Z_k) = 1. \quad (5.4)$$

W definicji procesu nieprzeliczalnego opisu istotne jest, że $P(f_{nk}(X_{i+1:i+n}) = Z_k)$ nie zależy od i . Z tego powodu nie może zachodzić $X_i = Z_i$, a proces Bernoulliego sam nie może być procesem nieprzeliczalnego opisu.

Nazwę “proces nieprzeliczalnego opisu” wybraliśmy kierując się skojarzeniem, że wartości zmiennych procesu $X_{\mathbb{Z}}$ w systematyczny sposób opisują (determinują) jedną z nieprzeliczalnie wielu wartości procesu Bernoulliego $Z_{\mathbb{N}}$. Dla niektórych procesów $X_{\mathbb{Z}}$ pojedyncza zmienna X_i może mieć za małą liczbę wartości, aby móc determinować stan całego procesu $Z_{\mathbb{N}}$. Jednakże nawet, gdy liczba wartości X_i jest mniejsza niż liczba wartości $Z_{\mathbb{N}}$, można skonstruować proces stacjonarny $X_{\mathbb{Z}}$, dla którego dowolnie wiele wartości Z_k można przewidywać dowolnie dobrze na podstawie dostatecznie długiego ciągu zmiennych X_i .

Twierdzenie 5.5. *Każdy proces nieprzeliczalnego opisu jest niefinitarny i nieergodyczny.*

Dowód: Załóżmy, że $X_{\mathbb{Z}}$ jest procesem nieprzeliczalnego opisu. Z założeń wynika, że

$$\lim_{n \rightarrow \infty} P(f_{nk}(X_{m:m+n-1}) = Z_k) = \lim_{n \rightarrow \infty} P(f_{nk}(X_{m-n:m-1}) = Z_k) = 1,$$

gdyż $P(f_{nk}(X_{i+1:i+n}) = Z_k)$ nie zależy od $i \in \mathbb{Z}$. Z twierdzenia 3.5 wynika zatem, że istnieją funkcje f_k^- mierzalna $\tilde{\sigma}(X_{-\infty:n-1})/\tilde{\sigma}(Z_k)$ i f_k^+ mierzalna $\tilde{\sigma}(X_{n:\infty})/\tilde{\sigma}(Z_k)$ takie, że $f_k^-(X_{-\infty:n-1}) = Z_k = f_k^+(X_{n:\infty})$ dla każdego $n \in \mathbb{Z}$.

Z twierdzenia A.16 wynika, że istnieje zmienna rzeczywista $Y = \sum_{k=1}^{\infty} 2^{-k} Z_k$ i funkcje f_- mierzalna $\tilde{\sigma}(X_{-\infty:m-1})/\mathcal{R}$ i f_+ mierzalna $\tilde{\sigma}(X_{m:\infty})/\mathcal{R}$ spełniające równość 5.1. Ponieważ $Z_{\mathbb{N}}$ jest ciągiem niezależnych zmiennych losowych takich, że $P(Z_i = 1) = P(Z_i = 0) = 1/2$, to łatwo można sprawdzić, że $P_Y = m|_{\mathcal{R}_{(0,1)}}$, zob. Billingsley (1979, sekcje 1 i 5). Z równości $P_Y = m|_{\mathcal{R}_{(0,1)}}$ zmienna Y jest zmienną ciągłą, czyli $\bar{H}(Y) = \infty$. Na mocy twierdzenia 5.3 proces $X_{\mathbb{Z}}$ jest nieergodyczny i niefinitarny. \square

Nazwę “proces przeliczalnego opisu” chcielibyśmy zarezerwować na przypadek procesu $X_{\mathbb{Z}}$ spełniającego warunki twierdzenia 5.3 przy dodatkowym założeniu, że Y jest zmienną dyskretną o nieskończonej entropii. Formalnej definicji nie podajemy, gdyż zależałoby nam na sformułowaniu niekorzystającym z funkcji nieskończonych przeszłości i przyszłości procesu. Badania potrzebnych analogonów twierdzeń 3.5 i 5.5 odkładamy na przyszłość.

Podajmy teraz przykład dyskretnego procesu nieprzeliczalnego opisu.

Przykład 5.6. (prosty proces opisu) *Niech $Z_{\mathbb{N}}$ będzie procesem Bernoulliego. Niech proces $N_{\mathbb{Z}}$, niezależny od $Z_{\mathbb{N}}$, będzie stacjonarnym ciągiem niezależnych zmiennych N_i , gdzie $N_i(\Omega) \subset \mathbb{N}$. Niech $P(N_i = n) = p_n$, gdzie $p_n > 0$ dla wszystkich $n \in \mathbb{N}$. Wartości zmiennych procesu X_i , gdzie $X_i(\Omega) \subset \{0, 1\}^+$, definiujemy jako słowa o losowej długości*

$$X_i := Z_1 \times Z_2 \times \dots \times Z_{N_i}. \quad (5.5)$$

Proces $X_{\mathbb{Z}}$ jest procesem nieprzeliczalnego opisu. Oznaczmy bowiem $f_{1k}(X_i) := Z_k$ dla $N_i \geq k$ oraz $f_{1k}(X_i) := -1$ w przeciwnym przypadku. Dla $f_{nk}(x_{1:n}) := \max_{i \in \{1, \dots, n\}} f_{1k}(x_i)$ zachodzi (5.4), gdyż $P(f_{nk}(X_{1:n}) = Z_k) = P(\max_{i \in \{1, \dots, n\}} N_i \geq k) = 1 - P(N_1 < k, \dots, N_n < k) = 1 - (\sum_{m=1}^{k-1} p_m)^n$.

W istocie można podać wiele przykładów procesów nieprzeliczalnego opisu podobnych do przykładu 5.6. Istnieje kilka potencjalnych kierunków modyfikacji:

- (i) Zamiast odsłaniać wartości wszystkich bitów Z_k , $k \in \{1, \dots, N_i\}$, możemy ujawniać wartość tylko N_i -tego z nich, jeżeli uczynimy jawną samą wartość N_i . Połóżmy zatem $X_i^{(1)} := (N_i, Z_{N_i})$.
- (ii) Nawet, gdy odsłaniamy ciągi bitów, a nie pojedyncze bity, nie musimy wypisywać wszystkich bitów od początku. Niech proces $M_{\mathbb{Z}}$, niezależny od $Z_{\mathbb{N}}$ i $N_{\mathbb{Z}}$, będzie stacjonarnym ciągiem niezależnych zmiennych M_i , gdzie $M_i(\Omega) \subset \mathbb{N}$. Niech $P(M_i = n) = q_n$, gdzie $q_n > 0$ dla wszystkich $n \in \mathbb{N}$. Połóżmy $X_i^{(2)} := (M_i, Z_{M_i: M_i + N_i})$.
- (iii) Możemy wprowadzić losowe przekłamania, które utrudnią rozstrzygnięcie, czy w wartościach zmiennych $X_{1:n}$ pojawia się prawidłowa wartość Z_k . Niech proces $U_{\mathbb{Z}}$, niezależny od $Z_{\mathbb{N}}$ i $N_{\mathbb{Z}}$, będzie stacjonarnym ciągiem niezależnych zmiennych U_i , gdzie $U_i(\Omega) \subset \{0, 1\}$. Niech $P(U_i = 1) > 1/2$. Połóżmy $X_i^{(3)} := (N_i, Z_{N_i})$ dla $U_i = 1$ oraz $X_i := (N_i, 1 - Z_{N_i})$ dla $U_i = 0$.

Łatwo wykazać, że procesy $X_{\mathbb{Z}}^{(i)}$ określone w punktach (i)–(iii) są procesami nieprzeliczalnego opisu. W myśl twierdzenia 5.5 są one nefinitarne, czyli dla każdego z nich zachodzi równość $\lim_{n \rightarrow \infty} E_{X^{(i)}}(n) = \infty$. Wykazanie tychże równości rachunkiem elementarnym jest dość żmudne i nie odsłania przed rachującym ogólnej prawidłowości.

Warto zwrócić uwagę na pewne związki pomiędzy ściśle matematyczną teorią procesów nieprzeliczalnego opisu a rozwijanymi przez niektórych przyrodników empirycznymi teoriami układów złożonych. Koncepcja procesów nefinitarnych w kształcie definicji 2.9 pojawiła się w kontekście prób probabilistycznego modelowania tekstów w języku naturalnym (Crutchfield i Feldman, 2003). Załóżmy, że wartościami zmiennej $Y_{\mathbb{Z}}$ są nieskończone teksty, przy czym dla każdego n wartością Y_n jest n -ta litera odpowiedniego tekstu. Niektórzy badacze w oparciu o próbki skończonych tekstów w języku angielskim próbowali oszacować wartości entropii blokowej $H_Y(n)$ dla możliwie wielu n (Shannon, 1950). Według niektórych publikacji wielkość $E_Y(n) = I(Y_{-n+1:0}; Y_{1:n})$ spełnia przybliżoną zależność $E_Y(n) \approx \text{const} \cdot \sqrt{n}$ dla $n \leq 100$ (Hilberg, 1990; Ebeling i Pöschel, 1994). Przypuszczamy zatem, że pewne intuicje i inspiracje lingwistyczne mogą być przydatne w czysto matematycznych rozważaniach na temat procesów nefinitarnych (Dębowski, 2005, 2004a).

W świetle twierdzenia 5.5 nefinitarność hipotetycznego procesu stacjonarnego $Y_{\mathbb{Z}}$ modelującego teksty w języku naturalnym nie wydaje się zaskakująca. Jeżeli każdy tekst opisuje w sposób losowy pewien losowy stan nieskończonej fikcyjnej rzeczywistości (świata przedstawionego), to proces $Y_{\mathbb{Z}}$ powinien być procesem nieprzeliczalnego opisu, czyli procesem nefinitarnym.

Przypomnijmy proces $X_{\mathbb{Z}}^{(1)}$ z przykładu (i). Przedstawienie postaci $X_i^{(1)} := (N_i, Z_{N_i})$ jest bliskie koncepcji struktury tematyczno-rematycznej rozwijanej przez niektórych lingwistów (Polański, 1999, hasło „Aktualne rozczłonkowanie zdania”). Według tejże koncepcji każde zdanie w języku naturalnym można podzielić na dwie części: temat, który wskazuje pewien wyróżniony obiekt, oraz reumat, który (w pewnym przybliżeniu) określa stan owego obiektu. Analogicznie dowolną wartość $y_i^{(1)} := (n_i, b_i)$ zmiennej $X_i^{(1)}$ możemy podzielić na temat n_i , który wskazuje indeks zmiennej Z_{n_i} , oraz reumat b_i , który opisuje wartość Z_{n_i} . Aby uczynić z procesu $X_{\mathbb{Z}}^{(1)}$ bardziej realistyczny model tekstu w języku naturalnym prawdopodobnie należałoby rozluźnić warunek niezależności zmiennych N_i oraz wprowadzić pewne losowe przekłamania jak w przykładzie (iii).

Rozdział 6

Oczekiwana nadwyżkowa długość kodu

Niech \mathbb{V} i \mathbb{W} będą pewnymi przeliczalnymi zbiorami, których elementy nazwiemy znakami. Same zbiory \mathbb{V} i \mathbb{W} nazwiemy alfabetami. Określamy też zbiory słów $\mathbb{V}^* = \{\lambda\} \cup \mathbb{V}^+$ i $\mathbb{W}^* = \{\lambda\} \cup \mathbb{W}^+$, gdzie $\mathbb{V}^+ = \bigcup_{n \in \mathbb{N}} \mathbb{V}^n$, $\mathbb{W}^+ = \bigcup_{n \in \mathbb{N}} \mathbb{W}^n$, zaś λ , nazywane słowem pustym, jest elementem neutralnym łącznego działania konkatencji \times . Niech $\text{len } w$ oznacza długość słowa $w \in \mathbb{V}^*$ lub $w \in \mathbb{W}^*$. Oznaczamy następujące relacje:

- (i) $w' \overset{P}{\sqsubset} w$ — w' jest przedrostkiem słowa w , tzn. $w = w' \times y$ dla pewnego słowa y ,
- (ii) $w' \overset{S}{\sqsubset} w$ — w' jest przyrostkiem słowa w , tzn. $w = u \times w'$ dla pewnego słowa u .

Definicja 6.1. (zbiory bezrostkowe) Mówimy, że zbiór słów $U \subset \mathbb{W}^*$ jest zbiorem bezprzedrostkowym (prefix-free), jeżeli dla dowolnych $w, w' \in U$ warunek $w' \overset{P}{\sqsubset} w$ implikuje $w' = w$. Analogicznie zbiór $U \subset \mathbb{W}^*$ nazywamy bezprzyrostkowym (suffix-free), gdy dla dowolnych $w, w' \in U$ warunek $w' \overset{S}{\sqsubset} w$ implikuje $w' = w$. Mówimy, że U jest bezrostkowy (fix-free, [Gillman i Rivest, 1995](#)), gdy jest zarazem bezprzedrostkowy i bezprzyrostkowy.

Twierdzenie 6.2. (Krafta) Niech $U \subset \mathbb{W}^*$ będzie dowolnym zbiorem bezprzedrostkowym lub bezprzyrostkowym. Zachodzi nierówność

$$\sum_{w \in U} (\text{card } \mathbb{W})^{-\text{len } w} \leq 1 \quad (6.1)$$

(zob. [Cover i Thomas, 1991](#), twierdzenie 5.2.2).

Definicja 6.3. (zupełne zbiory bezrostkowe) Niech U będzie zbiorem bezprzedrostkowym lub bezprzyrostkowym. Mówimy, że U jest zupełny, gdy w nierówności Krafta (6.1) zachodzi równość.

Pojęcie zupełnego zbioru bezrostkowego okaże się ważne w rozdziale 7.

Definicja 6.4. (kod) Kodem nazywamy dowolną iniekcję $C : \mathbb{V}^* \rightarrow \mathbb{W}^*$, czyli funkcję taką, że dla słów $v, v' \in \mathbb{V}^*$ zachodzi wynikanie $v \neq v' \implies C(v) \neq C(v')$. Kod C nazywamy bezprzedrostkowym, jeżeli zbiór $C(\mathbb{V}^*)$ jest bezprzedrostkowy. Dla $\text{card } \mathbb{W} < \infty$ oznaczamy unormowaną długość kodu

$$l_C(v) := \log(\text{card } \mathbb{W}) \cdot \text{len } C(v). \quad (6.2)$$

Konsekwencją nierówności Krafta (6.1) jest następujące twierdzenie.

Twierdzenie 6.5. (Shannona) Dla każdego kodu bezprzedrostkowego $C : \mathbb{V}^* \rightarrow \mathbb{W}^*$ i zmiennej losowej X , gdzie $\text{card } \mathbb{W} < \infty$ i $X(\Omega) \subset \mathbb{V}^*$, zachodzi nierówność

$$\langle l_C(X) \rangle \geq H(X) \quad (6.3)$$

(zob. [Cover i Thomas, 1991](#), twierdzenie 5.3.1).

W twierdzeniu 6.5 zmienna X jest dyskretna, jej miarą pomiarową jest miara licząca, a $H(X) = \bar{H}(X)$.

Z nierówności (6.3) oczekiwane długości kodów bezprzedrostkowych są górnymi ograniczeniami entropii. Można zadać pytanie, czy istnieją kody bezprzedrostkowe o długościach asymptotycznie równych entropii.

Definicja 6.6. (kod uniwersalny, por. Weissman, 2004) Niech $X_{\mathbb{Z}}$ oznacza stacjonarny proces stochastyczny, gdzie $X_i(\Omega) \subset \mathbb{V}$. Kod bezprzedrostkowy $C : \mathbb{V}^* \rightarrow \mathbb{W}^*$ nazywamy kodem (prawie na pewno) uniwersalnym, jeżeli dla każdego procesu ergodycznego $X_{\mathbb{Z}}$ zachodzi

$$\limsup_{n \rightarrow \infty} l_C(X_{1:n})/n \leq h_X \quad \text{prawie na pewno.} \quad (6.4)$$

Kod bezprzedrostkowy C nazywamy kodem uniwersalnym w L^1 , jeżeli dla każdego procesu $X_{\mathbb{Z}}$ zachodzi

$$\lim_{n \rightarrow \infty} \langle l_C(X_{1:n}) \rangle / n = h_X. \quad (6.5)$$

Dla procesu dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ i kodu bezprzedrostkowego C nierówność (6.4) jest równoważna równości $\limsup_{n \rightarrow \infty} l_C(X_{1:n})/n = h_X$ prawie na pewno, jak wykażemy w następnym twierdzeniu. Pokażemy też, że każdy bezprzedrostkowy kod uniwersalny prawie na pewno o odpowiednio majoryzowalnej długości kodu jest uniwersalny w L^1 .

Twierdzenie 6.7. Niech $C : \mathbb{V}^* \rightarrow \mathbb{W}^*$ będzie bezprzedrostkowym kodem uniwersalnym prawie na pewno. Jeżeli istnieje stała K taka, że

$$l_C(v) \leq K \text{ len } v \quad (6.6)$$

dla każdego słowa $v \in \mathbb{V}^*$, to kod C jest uniwersalny w L^1 , a dla każdego procesu stacjonarnego $X_{\mathbb{Z}}$, gdzie $X_i(\Omega) \subset \mathbb{V}$, zachodzi także

$$\limsup_{n \rightarrow \infty} l_C(X_{1:n})/n = h(F_X) \quad \text{prawie na pewno,} \quad (6.7)$$

gdzie F_X jest miarą losową określoną w twierdzeniu 4.6.

NB. Dowód twierdzenia 6.7 podobny do naszego podaje Weissman (2004), jakkolwiek niezbyt formalnie odwołuje się on do subtelności rozkładu ergodycznego. Weissman podaje też przykład kodu, który jest uniwersalny w L^1 , ale nie jest uniwersalny prawie na pewno. Co więcej, z (6.5) i (6.7) prawdopodobnie nie wynika $\lim_{n \rightarrow \infty} l_C(X_{1:n})/n = h(F_X)$.¹

Dowód: Niech $X_{\mathbb{Z}}$ będzie dowolnym procesem stacjonarnym. W pierwszej kolejności wykażemy nierówność \leq w równości (6.7). W tym celu skorzystamy z twierdzenia 4.6 (rozkład ergodyczny) i jego konwencji notacyjnych. Oznaczmy zdarzenie $A := (\limsup_{n \rightarrow \infty} l_C(X_{1:n})/n \leq h(F_X))$. W istocie $A = X_{\mathbb{Z}}^{-1}(\tilde{A})$ dla

$$\tilde{A} = \left\{ x_{\mathbb{Z}} \in X_{\mathbb{Z}}(\Omega) : \limsup_{n \rightarrow \infty} l_C(x_{1:n})/n \leq h(\phi(\cdot)(x_{\mathbb{Z}})) \right\} \in \tilde{\sigma}(X_{\mathbb{Z}}).$$

¹ Warunki $\limsup_{n \rightarrow \infty} Y_n \leq 0$ prawie na pewno i $\lim_{n \rightarrow \infty} \langle Y_n \rangle = 0$ nie implikują $\liminf_{n \rightarrow \infty} Y_n \geq 0$ prawie na pewno. Przykład: Obierzmy $(\Omega, \mathcal{J}, P) := ((0, 1], \mathcal{R}|_{(0,1]}, m)$, gdzie m jest miara Lebesgue'a. Dla $w_n := \sum_{k=1}^n 1/k$ i $K < 0$ określmy $Y_n(\omega) := K \cdot \sum_{k=0}^{\infty} \mathbb{1}[w_n < \omega + k \leq w_{n+1}]$. Wówczas $\liminf_{n \rightarrow \infty} Y_n = K$, $\limsup_{n \rightarrow \infty} Y_n = 0$ i $\lim_{n \rightarrow \infty} \langle Y_n \rangle = 0$.

Z twierdzenia 4.6 pkt. (iv)-(v) mamy $\mu(\{x_{\mathbb{Z}} \in X_{\mathbb{Z}}(\Omega) : h(\phi(\cdot)(x_{\mathbb{Z}})) = h(\mu)\}) = 1$ dla P_{F_X} -prawie wszystkich $\mu \in \mathbb{E}$, a zatem z definicji kodu prawie na pewno uniwersalnego (6.4) zachodzi $\mu(\tilde{A}) = 1$ dla tychże μ . Korzystając ze wzoru (4.4), otrzymujemy

$$P(A) = P_{X_{\mathbb{Z}}}(\tilde{A}) = \int \mu(\tilde{A}) dP_{F_X}(\mu) = \int 1 dP_{F_X}(\mu) = 1.$$

Mamy zatem $\limsup_{n \rightarrow \infty} l_C(X_{1:n})/n \leq h(F_X)$ prawie na pewno.

Oznaczmy $g_n := K - l_C(X_{1:n})/n \geq 0$. Korzystając z lematu Fatou (zob. dowód twierdzenia 3.1), otrzymujemy

$$\begin{aligned} K - \limsup_{n \rightarrow \infty} \langle l_C(X_{1:n}) \rangle / n &= \liminf_{n \rightarrow \infty} \langle g_n \rangle \geq \left\langle \liminf_{n \rightarrow \infty} g_n \right\rangle = K - \left\langle \limsup_{n \rightarrow \infty} l_C(X_{1:n})/n \right\rangle \\ &\geq K - \langle h(F_X) \rangle = K - h_X, \end{aligned}$$

gdzie ostatnia równość wynika z równości (4.12). Zatem $\limsup_{n \rightarrow \infty} \langle l_C(X_{1:n}) \rangle / n \leq h_X$. Z drugiej strony z twierdzenia 6.5 mamy $\langle l_C(X_{1:n}) \rangle \geq H_X(n)$, skąd

$$\liminf_{n \rightarrow \infty} \langle l_C(X_{1:n}) \rangle / n \geq \liminf_{n \rightarrow \infty} H_X(n)/n = h_X,$$

a zatem zachodzi (6.5).

Oznaczmy $Y := \limsup_{n \rightarrow \infty} l_C(X_{1:n})/n - h(F_X)$. Z poprzednich przekształceń mamy $Y \leq 0$ prawie na pewno oraz $\langle Y \rangle \geq \limsup_{n \rightarrow \infty} \langle l_C(X_{1:n}) \rangle / n - h_X = 0$. Gdyby $P(Y \leq -b) > a$ dla pewnych $a, b > 0$, to w świetle pierwszej relacji mielibyśmy $\langle Y \rangle \leq \langle -b \cdot \mathbb{I}[Y \leq -b] + 0 \cdot \mathbb{I}[Y > -b] \rangle = -ba < 0$. Jako, że $\langle Y \rangle \geq 0$, to musimy mieć $P(Y \leq -b) = 0$ dla każdego $b > 0$, czyli $Y \geq 0$ prawie na pewno. Zatem $Y = 0$ prawie na pewno, czyli mamy relację (6.7). \square

Zauważmy, że twierdzenie 6.7 odnosi się wyłącznie do przypadku, gdy zarówno \mathbb{V} jak i \mathbb{W} są zbiorami skończonymi. Po pierwsze, zbiór \mathbb{W} musi być skończony, aby funkcja l_C była określona. Gdyby wówczas \mathbb{V} był nieskończony, to mielibyśmy $\sup_{v \in \mathbb{V}} l_C(v) = \infty > K = \sup_{v \in \mathbb{V}} K \text{ len } v$ (bo C jest iniekcją w \mathbb{W}^*), czyli dla żadnego skończonego K nierówność (6.6) nie mogłaby zachodzić.

Kody uniwersalne istnieją. Pewnym prostym przykładem kodu uniwersalnego prawie na pewno i w L^1 jest kod Lempela-Ziva (Ziv i Lempel, 1977, 1978). W podstawowej konstrukcji, kod Lempela-Ziva LZ jest funkcją przekształcającą słowa binarne na słowa binarne, $LZ : \{0, 1\}^* \rightarrow \{0, 1\}^*$, ale można skonstruować analogiczne kody D -arne $\{0, 1, \dots, D-1\}^* \rightarrow \{0, 1, \dots, D-1\}^*$. Kod Lempela-Ziva ma małą czasową złożoność obliczeniową, więc liczne jego modyfikacje są powszechnie stosowane przez programy kompresujące dane (m.in. linuksowy program `gzip`). Warto jednak podkreślić, że kody obliczane de facto przez popularne programy kompresujące zwykle nie są kodami uniwersalnymi.

W literaturze znany jest dowód faktu, że kod LZ jest uniwersalny prawie na pewno. Uzupełnijmy zatem dowód uniwersalności w L^1 . Definicja kodu LZ wykorzystuje pomocniczą funkcję zróżnicowanego parsowania (Cover i Thomas, 1991, sekcja 12.10).

Definicja 6.8. (funkcja zróżnicowanego parsowania) *Funkcją zróżnicowanego parsowania (distinct parsing) nazywamy dowolną funkcję $f : \{0, 1\}^* \rightarrow (\{0, 1\}^+)^*$, która przekształca słowo binarne w na ciąg słów binarnych $f(w) = (w_1, \dots, w_{c(w;f)})$ takich, że $w = w_1 \times \dots \times w_{c(w;f)}$, $w_k \neq \lambda$ dla $1 \leq k \leq c(w;f)$, a także $w_i \neq w_j$ dla $i \neq j$, gdzie $1 \leq i, j < c(w;f)$.*

NB. Od funkcji zróżnicowanego parsowania nie wymagamy, aby ostatnie słowo $w_{c(w;f)}$ było różne od wszystkich poprzednich w_j .

Pewną nieskomplikowaną funkcję zróżnicowanego parsowania f_{LZ} możemy skonstruować w sposób następujący: Poczynając od lewego krańca słowa w odcinamy od niego kolejne w_i tak, aby każde słowo w_i z wyjątkiem pierwszego i ostatniego było konkatencją pewnego poprzednio odciętego słowa $w_{j_i(w)}$, $j_i(w) \in \{1, 2, \dots, i-1\}$, i pewnego bitu $a_i(w) \in \{0, 1\}$. Indeks $j_i(w)$ definiujemy jako taki, że dla $w = w_1 \times w_2 \times \dots \times w_{i-1} \times u$ słowo $w_{j_i(w)}$ jest najdłuższym przedrostkiem słowa u spośród elementów zbioru $(w_j)_{j < i}$. Dla takiego indeksu $j_i(w)$ słowo $w_i = w_{j_i(w)} \times a_i(w)$ jest na pewno różne od $(w_j)_{j < i}$.

Formalnie funkcja f_{LZ} to taka funkcja zróżnicowanego parsowania, że dla każdego słowa w i $f_{LZ}(w) = (w_1, \dots, w_{c(w)})$ istnieją ciągi $j(w) = (j_i(w))_{i \in \{2, 3, \dots, c(w)\}}$, $j_i(w) \in \{1, 2, \dots, i-1\}$, oraz $a(w) = (a_i(w))_{i \in \{0, 1, \dots, c(w)\}}$, $a_i(w) \in \{0, 1\}$, gdzie

$$w_i = \begin{cases} a_i(w), & i = 1, \\ w_{j_i(w)} \times a_i(w), & 1 < i < c(w), \\ w_{j_i(w)} & i = c(w), a_0(w) = 0, \\ w_{j_i(w)} \times a_i(w), & i = c(w), a_0(w) = 1, \end{cases} \quad (6.8)$$

zaś $w_{j_i(w)}$ jest najdłuższym przedrostkiem słowa $w_i \times \dots \times w_{c(w)}$ spośród słów $(w_j)_{j < i}$. Przy dodatkowej definicji $a_{c(w)}(w) := 0$ dla $a_0(w) = 0$ — trójka $(c(w), a(w), j(w))$ jest dana jednoznacznie dla każdego słowa w . Każda trójka $(c(w), a(w), j(w))$ jest także obrazem dokładnie jednego słowa w . W istocie kod Lempel-Ziva reprezentuje każde słowo w za pomocą binarnej reprezentacji takiejże trójki.

Niech $C_B(\cdot; N) : \{1, 2, \dots, N\} \rightarrow \{0, 1\}^{\lceil \log_2 N \rceil}$ będzie bezprzedrostkowym binarnym kodem zbioru liczb $\{1, 2, \dots, N\}$ określonym jako $C_H(k; N) = u \times v$, gdzie v jest rozwinięciem binarnym liczby $k-1$, zaś u jest ciągiem $\lceil \log_2 N \rceil - \text{len } v$ zer. Z kolei $C_E : \mathbb{N} \rightarrow \{0, 1\}^*$ niech będzie bezprzedrostkową ω -reprezentacją zbioru liczb naturalnych (Elias, 1975).²

Definicja 6.9. (kod Lempel-Ziva) Dla każdego słowa $w \in \{0, 1\}^*$ definiujemy wartość kodu Lempel-Ziva $LZ(w) \in \{0, 1\}^*$ jako

$$\begin{aligned} LZ(w) := & C_E(c(w)) \times a_0(w) \times a_1(w) \\ & \times C_B(j_2(w); 1) \times a_2(w) \\ & \times C_B(j_3(w); 2) \times a_3(w) \\ & \times \dots \\ & \times C_B(j_{c(w)}(w); c(w) - 1) \times a_{c(w)}(w). \end{aligned} \quad (6.9)$$

Łatwo pokazać, że zbiór $LZ(\{0, 1\}^*)$ jest bezprzedrostkowy. Ciąg binarny $C_E(c(w))$ reprezentuje liczbę podłów $c(w)$. Ciągi $C_B(j_i(w); i-1)$ reprezentują indeksy $j_i(w)$.

² Formalnie wielkość C_E definiuje się jako $C_E(k) = \bar{C}_E(k) \times 0$, gdzie $\bar{C}_E(1) := \lambda$ zaś dla $k > 1$ mamy $\bar{C}_E(k) := \bar{C}_E(\text{len } v - 1) \times v$, gdzie v jest rozwinięciem binarnym liczby k . Wartości C_E dla kilku pierwszych liczb naturalnych to

$$\begin{array}{ll} C_E(1) = 0, & C_E(5) = 10 \times 101 \times 0, \\ C_E(2) = 10 \times 0, & C_E(6) = 10 \times 110 \times 0, \\ C_E(3) = 11 \times 0, & C_E(7) = 10 \times 111 \times 0, \\ C_E(4) = 10 \times 100 \times 0, & C_E(8) = 11 \times 1000 \times 0. \end{array}$$

Długości kodów C_B i C_E szacują się jako $\text{len } C_B(k; N) \leq \log_2 N + 1$ oraz $\text{len } C_E(N) \leq \log_2^* N + 1$, gdzie \log_2^* oznacza logarytm iterowany³, a zatem istnieje stała M_1 taka, że

$$l_{\text{LZ}}(w) \leq c(w) \log c(w) + M_1 c(w). \quad (6.11)$$

Aby wykazać, że kod Lempel-Ziva jest prawie na pewno uniwersalny, wystarczy dowieść, że $\lim_{n \rightarrow \infty} c(X_{1:n})/n = 0$ oraz $\lim_{n \rightarrow \infty} c(X_{1:n}) \log c(X_{1:n})/n \leq h_X$ prawie na pewno dla każdego ergodycznego procesu stacjonarnego $X_{\mathbb{Z}}$. Pierwsza z relacji wynika z nierówności Lempel-Ziva

$$c(w; f) \log \text{len } w \leq M_2 \text{len } w, \quad (6.12)$$

gdzie f jest dowolną funkcją zróżnicowanego parsowania zaś M_2 jest stałą niezależną od f (Cover i Thomas, 1991, lemat 12.10.1). Dla dowolnej funkcji zróżnicowanego parsowania f i ergodycznego procesu stacjonarnego $X_{\mathbb{Z}}$ nierówność

$$\lim_{n \rightarrow \infty} c(X_{1:n}; f) \log c(X_{1:n}; f)/n \leq h_X \quad \text{prawie na pewno} \quad (6.13)$$

jest także prawdziwa (Cover i Thomas, 1991, twierdzenie 12.10.1). Nierówność (6.13) wynika z nierówności (6.12), nierówności Ziva dla dowolnego procesu stacjonarnego (Cover i Thomas, 1991, lemat 12.10.3) oraz twierdzenia ergodycznego (twierdzenie 4.5) dla prawdopodobieństw warunkowych. W rezultacie otrzymujemy następujące twierdzenie.

Twierdzenie 6.10. *Kod Lempel-Ziva LZ jest uniwersalny prawie na pewno i w L^1 .*

Dowód: Dowód uniwersalności prawie na pewno podają Cover i Thomas (1991, twierdzenie 12.10.1). Z twierdzenia 6.7 i nierówności (6.11), aby dowieść uniwersalności w L^1 , wystarczy pokazać, że nierówność

$$c(w) \log c(w) + M_1 c(w) \leq K \text{len } w \quad (6.14)$$

zachodzi dla pewnej stałej K . Z definicji $c(w)$ mamy $c(w) \leq \text{len } w$. Z nierówności (6.12) mamy także $c(w) \log c(w) \leq c(w) \log \text{len } w \leq M_2 \text{len } w$, więc nierówność (6.14) jest prawdziwa dla $K = M_1 + M_2$. \square

Uniwersalność kodu Lempel-Ziva wynika z dość ogólnych własności funkcji zróżnicowanego parsowania. Zbieżność ilorazów $\langle l_{\text{LZ}}(X_{1:n}) \rangle / n$ do granicznej wartości h_X nie jest zatem szybka. Empiryczne badania kompresji (m. in. dla tekstów w języku naturalnym) wskazują, że istnieje wiele takich kodów C , dla których obserwowane długości l_C są istotnie mniejsze niż l_{LZ} (Bell *et al.*, 1990; Nevill-Manning i Witten, 1997; de Marcken, 1996). Uniwersalności niektórych z tych kodów dowieść jest trudno. W ostatnich latach wykazano jednak uniwersalność szerokiej podklasy kodów opartych na gramatykach, które kompresują dane empiryczne istotnie lepiej niż kod Lempel-Ziva (Kieffer i Yang, 2000, 2002; Charikar *et al.*, 2002; Lehman i Shelat, 2002; Dębowski, 2005).

³ Dla $k \geq 1$ określamy $\log_2^* k := \sum_{n=1}^{\infty} \log_2^{(n)} k$, gdzie

$$\log_2^{(n)} k := \begin{cases} \log_2 k, & k = 1, \\ \log_2 \log_2^{(n-1)} k, & k > 1, \log_2^{(n-1)} k \geq 1, \\ 0, & k > 1, \log_2^{(n-1)} k < 1. \end{cases} \quad (6.10)$$

Niewątpliwie warto poszukiwać takich kodów C , dla których unormowane długości $l_C(X_{1:n})$ są istotnie mniejsze niż $l_{LZ}(X_{1:n})$. Wskażmy jednak na nieuchronne ograniczenia wynikające z rozważań teorii informacyjnych, w szczególności na ograniczenia narzucające przez wartości entropii nadwyżkowych skończonego rzędu. Ograniczenia te dotyczą także kodu uniwersalnego, jakim jest bezprzedrostkowa złożoność Kołmogorowa (zob. [Li i Vitányi, 1993](#), twierdzenie 8.1.1, a także [Cover et al., 1989](#)).

Dla procesu stacjonarnego $X_{\mathbb{Z}}$ i kodu C określmy funkcję oczekiwanej długości kodu

$$H_X^C(n) := \langle l_C(X_{i+1:i+n}) \rangle. \quad (6.15)$$

Jeżeli C jest kodem uniwersalnym w L^1 , to dla entropii blokowej H_X mamy

$$H_X^C(n) \geq H_X(n), \quad (6.16)$$

$$\lim_{n \rightarrow \infty} H_X^C(n)/n = h_X = \lim_{n \rightarrow \infty} H_X(n)/n. \quad (6.17)$$

Przez analogię do entropii nadwyżkowej skończonego rzędu $E_X(n) := I(X_{-n+1:0}; X_{1:n}) = 2H_X(n) - H_X(2n)$ określmy funkcję oczekiwanej nadwyżkowej długości kodu

$$E_X^C(n) := 2H_X^C(n) - H_X^C(2n). \quad (6.18)$$

Funkcja $E_X^C(n)$ nie musi być nieujemna i rosnąca, gdyż funkcja H_X^C nie musi być subaddytywna. Dla wygody możemy jednak zdefiniować funkcję rosnącą $E_X^{m,C}(n) = \max_{k \leq m} E_X^C(k)$.

Twierdzenie 6.11. *Dla dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ i kodu C uniwersalnego w L^1 zachodzi nierówność*

$$\limsup_{n \rightarrow \infty} [E_X^C(n) - E_X(n)] \geq 0. \quad (6.19)$$

Dowód: Dla każdej funkcji $f : \mathbb{N} \rightarrow \mathbb{R}$ i $m \in \mathbb{R}$ zachodzi tożsamość

$$\sum_{k=0}^{m-1} [2f(2^k n) - f(2^{k+1} n)] \cdot \frac{1}{2^{k+1}} = f(n) - \frac{f(2^m n)}{2^m n} \cdot n.$$

Z założenia mamy (6.17), a zatem

$$\begin{aligned} H_X(n) - nh_X &= \sum_{k=0}^{\infty} [2H_X(2^k n) - H_X(2^{k+1} n)] \cdot \frac{1}{2^{k+1}} = \sum_{k=0}^{\infty} \frac{E_X(2^k n)}{2^{k+1}}, \\ H_X^C(n) - nh_X &= \sum_{k=0}^{\infty} [2H_X^C(2^k n) - H_X^C(2^{k+1} n)] \cdot \frac{1}{2^{k+1}} = \sum_{k=0}^{\infty} \frac{E_X^C(2^k n)}{2^{k+1}}. \end{aligned}$$

Korzystając z (6.16) i kładąc $n = 2^p M$ dla dowolnego $p \in \mathbb{N}$ i ustalonego $M \in \mathbb{N}$, otrzymujemy

$$\sum_{k=p}^{\infty} \frac{E_X^C(2^k M) - E_X(2^k M)}{2^{k+1}} \geq 0. \quad (6.20)$$

Gdyby $E_X^C(2^k M) - E_X(2^k M) \geq 0$ zachodziło tylko dla skończonego wielu k , to dla pewnego p i wszystkich $k \geq p$ mielibyśmy $E_X^C(2^k M) - E_X(2^k M) < 0$, co stoi w sprzeczności z (6.20). A zatem zachodzi nierówność (6.19). \square

Zauważmy, że z nierówności (6.19) wynika nierówność $E_X^C(n) \geq 0$ dla nieskończenie wielu n , chociaż nie zakładaliśmy subaddytywności funkcji H_X^C . Z (6.19) mamy także nierówność $\limsup_{n \rightarrow \infty} [E_X^{m,C}(n) - E_X(n)] \geq 0$.

Skojarzmy nierówność (6.19) z rozkładem ergodycznym. Oznaczmy funkcjonal oczekiwanej długości kodu jawnie sparametryzowany miarą przeniesioną $P_{X_{\mathbb{Z}}}$ jako $H^C(n; P_{X_{\mathbb{Z}}}) := H_X^C(n) = \int l_C(\xi_{i+1:i+n}) dP_{X_{\mathbb{Z}}}$, gdzie $\xi_n(X_{\mathbb{Z}}) := X_n$. Analogicznie oznaczamy wielkości $E^C(n; P_{X_{\mathbb{Z}}}) := E_X^C(n)$. Niech F_X będzie miarą losową określoną w twierdzeniu 4.6. Dla ograniczonej funkcji f mierzalnej $\tilde{\sigma}(X_{\mathbb{Z}})/\mathcal{R}$ zachodzi tożsamość

$$\int f dP_{X_{\mathbb{Z}}} = \int f d \left(\int m dP_{F_X}(m) \right) = \int \left(\int f dm \right) dP_{F_X}(m) \quad (6.21)$$

(Billingsley, 1979, ćwiczenie 18.19). Kładąc $f = 2l_C(\xi_{i+1:i+n}) - l_C(\xi_{i+1:i+2n})$, otrzymujemy $E_X^C(n) = \langle E^C(n; F_X) \rangle$. Z nierówności (6.19) mamy $\limsup_{n \rightarrow \infty} E^C(n; F_X) \geq E(F_X)$, a z (4.13) także

$$\left\langle \limsup_{n \rightarrow \infty} E^C(n; F_X) \right\rangle \geq \langle E(F_X) \rangle, \quad (6.22)$$

$$\limsup_{n \rightarrow \infty} \langle E^C(n; F_X) \rangle \geq \bar{H}(F_X) + \langle E(F_X) \rangle. \quad (6.23)$$

Warto zauważyć, że istnieje tylko przeliczalnie wiele ergodycznych miar stacjonarnych m o skończonej granicy $\limsup_{n \rightarrow \infty} E^C(n; m)$.

Twierdzenie 6.12. *Dla dowolnego kodu C uniwersalnego w L^1 niech $N^C(K)$ oznacza moc zbioru stacjonarnych miar ergodycznych m takich, że $\limsup_{n \rightarrow \infty} E^C(n; m) \leq K$. Dla dowolnego $K \in \mathbb{R}$ zachodzi nierówność*

$$\log N^C(K) \leq K. \quad (6.24)$$

Dowód: Weźmy dowolne $M \in \mathbb{N}$ spełniające $M \leq N^C(K)$. Niech A będzie zbiorem dowolnych M stacjonarnych miar ergodycznych m takich, że $\limsup_{n \rightarrow \infty} E^C(n; m) \leq K$. Skonstruujmy przestrzeń probabilistyczną z procesem stacjonarnym $X_{\mathbb{Z}}$ takim, że $P_{X_{\mathbb{Z}}} = [\text{card } A]^{-1} \sum_{m \in A} m$. Z jednoznaczności rozkładu ergodycznego miara losowa F_X jest dyskretną zmienną losową przyjmującą dowolną wartość $m \in A$ z prawdopodobieństwem $P(F_X = m) = 1/\text{card } A$. Mamy zatem $\bar{H}(F_X) = \log M$. Dla ustalonego $\epsilon > 0$ zmienne losowe $K + \epsilon - E^C(n; F_X)$, $n \in \mathbb{N}$, są prawie na pewno nieujemne dla prawie wszystkich n . Z lematu Fatou mamy zatem $K + \epsilon - \langle \limsup_{n \rightarrow \infty} E^C(n; F_X) \rangle \leq K + \epsilon - \limsup_{n \rightarrow \infty} \langle E^C(n; F_X) \rangle$, czyli $\limsup_{n \rightarrow \infty} \langle E^C(n; F_X) \rangle \leq \langle \limsup_{n \rightarrow \infty} E^C(n; F_X) \rangle \leq \langle K \rangle = K$. Z nierówności (6.23) mamy $\log M = \bar{H}(F_X) \leq \limsup_{n \rightarrow \infty} \langle E^C(n; F_X) \rangle \leq K$. Założyliśmy, że M jest dowolną liczbą naturalną spełniającą $M \leq N^C(K)$, czyli z nierówności $\log M \leq K$ wynika (6.24). \square

Z twierdzenia 6.12 wynika, że rozbieżność $\langle \limsup_{n \rightarrow \infty} E^C(n; F_X) \rangle = \infty$ dla procesów stacjonarnych $X_{\mathbb{Z}}$ jest powszechna. Nie wyklucza to jednak możliwości, że ciąg $E_X^{m,C}(n)$ rośnie z szybkością różną dla różnych procesów $X_{\mathbb{Z}}$ i kodów C przy zachowaniu nierówności (6.19). Ponieważ istnieje nieprzeliczalnie wiele finitarnych i ergodycznych ukrytych łańcuchów Markowa, to na mocy twierdzenia 6.12 nie istnieje kod C taki, że $\limsup_{n \rightarrow \infty} [E_X^C(n) - E_X(n)] < \infty$ dla dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$. Być może jednak istnieje kod C taki, że $\lim_{n \rightarrow \infty} [E_X^C(n) - E_X(n)]/n^\alpha = 0$ dla każdego procesu $X_{\mathbb{Z}}$

i $\alpha > 0$ spełniającego $\lim_{n \rightarrow \infty} E_X(n)/n^\alpha < \infty$. Kod taki mógłby być używany do testowania istnienia silnej zależności w dyskretnych szeregach czasowych oraz do określania rzędu wzrostu entropii nadwyżkowych $E_X(n)$.

Nierówność (6.19) ma też ciekawą interpretację w kontekście badań języka naturalnego. Szerzej piszemy o tym w odrębnym artykule skierowanym do lingwistów (Dębowski, 2005), zreferujemy jednak tutaj niektóre idee. W świetle twierdzenia 5.5 można przypuszczać, że stacjonarny proces stochastyczny modelujący język naturalny powinien być niefinitarny. Jak wspominaliśmy, według niektórych publikacji (Hilberg, 1990; Ebeling i Pöschel, 1994) entropia nadwyżkowa n kolejnych liter tekstu w języku angielskim spełnia przybliżoną zależność $E_X(n) \approx \text{const} \cdot \sqrt{n}$ dla $n \leq 100$. Gdyby zależność ta zachodziła dla wszystkich $n \in \mathbb{N}$, to dla dowolnego kodu uniwersalnego C i dla nieskończenie wielu n mielibyśmy $E_X^C(n) \geq \text{const} \cdot \sqrt{n}$.

W istocie lingwistom znana jest prawidłowość zbliżona. Blisko związane z prawem Zipfa-Mandelbrota (Mandelbrot, 1983) empiryczne prawo Guirauda (Guiraud, 1954) głosi, że liczba V różnych słów w tekście liczącym ogółem N słów i n liter spełnia przybliżoną zależność $V \geq \text{const} \cdot \sqrt{N}$, gdzie $N \approx \text{const} \cdot n$. Na podstawie kilku badań empirycznych (Wolff, 1980; Nevill-Manning, 1996; Nevill-Manning i Witten, 1997; de Marcken, 1996) sądzimy (Dębowski, 2005), że V jest w przybliżeniu proporcjonalne do liczby różnych jednostek nieterminalnych definiowanych dla danego tekstu przez kod MDL określony jako najkrótszy kod uniwersalny oparty na gramatyce (Charikar *et al.*, 2002). Przedstawiliśmy heurystyczną argumentację (Dębowski, 2005), że dla kodu MDL i stacjonarnego losowego tekstu $X_{1:2n}$ oczekiwana długość definicji wszystkich jednostek nieterminalnych oprócz jednostki początkowej nie może być mniejsza niż nadwyżkowa długość kodu $E_X^{\text{MDL}}(n)$. Reasumując, prawo Guirauda prawdopodobnie można by wywieść z hipotezy o rzędzie wzrostu entropii nadwyżkowej $E_X(n)$. Lepsze uzasadnienie naszych przypuszczeń wymaga jednak dalszych badań kodów opartych na gramatykach.

Rozdział 7

Przeliterowanie i segmentacja

W rozdziale 6 dowiedliśmy, że dla procesów stacjonarnych o skończonej liczbie wartości nadwyżkowe długości dowolnego kodu uniwersalnego są nieskończenie często nie mniejsze niż entropie nadwyżkowe skończonego rzędu (twierdzenie 6.11). Warto byłoby móc porównywać rzędy wzrostu nadwyżkowej długości kodu i przybliżeń entropii nadwyżkowej na konkretnych przykładach procesów, dla których $E_X(n)$ rośnie dostatecznie szybko (np. potęgowo). W tym celu należałoby najpierw takie przykłady skonstruować.

Konstrukcja dyskretnych procesów niefinitarnych o skończonej liczbie wartości jest znacznie trudniejsza niż konstrukcja niefinitarnych procesów gaussowskich. W szczególności znacznie trudniejszy jest opis przestrzeni osiągalnych wartości entropii blokowej $H_X(n)$ (por. Yeung, 2002, rozdział 12 i 14). Dla procesów dyskretnych nie ma bowiem prostego analogonu twierdzenia 8.9. Owo twierdzenie głosi, że warunki $\Delta^2 H_X(k) \leq 0$ są jedynymi warunkami, które muszą spełniać entropie blokowe procesów gaussowskich. Niewątpliwie dla procesu dyskretnego $X_{\mathbb{Z}}$ pewne dodatkowe więzy narzucane są przez nierówność $\Delta H_X(k) = H_X(1) + \sum_{l=2}^k \Delta^2 H_X(l) \geq 0$. Nie są to jednak wszystkie więzy.

Przygotowując niniejszą pracę, prowadziliśmy pewne wstępne badania parametryzacji miar procesów stacjonarnych o wartościach binarnych, $X_i(\Omega) \subset \{0, 1\}$, za pomocą obiektu zbliżonego koncepcyjnie do funkcji częściowej autokorelacji dla procesów gaussowskich. Przy okazji zauważyliśmy, że dla procesów binarnych wartości $\Delta^2 H_X(l)$ muszą spełniać mocniejszy warunek $\Delta^2 H_X(k) \in [-C(k), 0]$, gdzie liczba $C(k) \leq \Delta H_X(k-1)$ zależy od rozkładu zmiennych $X_{1:k-1}$. Z tego powodu sam fakt istnienia stacjonarnych procesów binarnych o potęgowo rosnących entropiach nadwyżkowych $E_X(n)$ jest nietrywialny.

Wiemy jednak, że pewne procesy niefinitarne o wartościach binarnych istnieją. Fakt ten wynika ze wzoru (4.13). Ponieważ istnieje nieskończenie wiele miar stacjonarnych i ergodycznych procesów binarnych, to można dobrać taką miarę P_{F_X} , że $\bar{H}(F_X) = \infty$. Wówczas dla $P_{X_{\mathbb{Z}}} = \int m dP_{F_X}(m)$ proces $X_{\mathbb{Z}}$ jest niefinitarny i nieergodyczny.

Kilka konkretnych przykładów niefinitarnych i nieergodycznych procesów binarnych uzyskuje się przez przekształcenie częstościowe (A.35) zastosowane do pewnych ciągów automatycznych (Allouche i Shallit, 2003). Aby przywołać definicje tychże ciągów, pochyńmy kilka uwag notacyjnych.

Przypomnijmy, że $\mathbb{V}^+ = \bigcup_{n \in \mathbb{N}} \mathbb{V}^n = \{x_{1:k} : x_i \in \mathbb{V}, k \in \mathbb{N}\}$ oznacza zbiór wszystkich niepustych słów, których znaki są elementami \mathbb{V} . Zbiory nieskończonych ciągów będziemy oznaczać jako $\mathbb{V}^{\rightarrow} := \{x_{1:\infty} : x_i \in \mathbb{V}\}$ oraz $\mathbb{V}^{\leftarrow} := \{x_{-\infty:0} : x_i \in \mathbb{V}\}$. (Oznaczenia $x_{1:\infty}$ używamy zamiennie z $x_{\mathbb{N}}$.) Dla dowolnej funkcji $f : \mathbb{V} \rightarrow \mathbb{W}^+$ przekształcającej znaki ze zbioru \mathbb{V} w niepuste słowa ze zbioru \mathbb{W}^+ oznaczymy funkcje $f^+ : \mathbb{V}^+ \rightarrow \mathbb{W}^+$, $f^+ : \mathbb{V}^{\rightarrow} \rightarrow \mathbb{W}^{\rightarrow}$ oraz $f^+ : \mathbb{V}^{\leftarrow} \rightarrow \mathbb{W}^{\leftarrow}$ jako

$$f^+(x_{1:k}) := f(x_1) \times f(x_2) \times \dots \times f(x_k), \quad (7.1)$$

$$f^+(x_{1:\infty}) := f(x_1) \times f(x_2) \times f(x_3) \times \dots, \quad (7.2)$$

$$f^+(x_{-\infty:0}) := \dots \times f(x_{-2}) \times f(x_{-1}) \times f(x_0), \quad (7.3)$$

gdzie $x_i \in \mathbb{V}$. Dla $f : \mathbb{V} \rightarrow \mathbb{V}^+$ oznaczmy iteracje f^+ jako $f^1 := f^+$ oraz $f^{n+1} := f^+ \circ f^n$ dla $n \geq 1$. Jeżeli dla pewnego $x \in \mathbb{V}$ i każdego $k \in \mathbb{N}$ ciąg $f^k(x)$ jest przedrostkiem $f^{k+1}(x)$, to niech $f^\infty(x) \in \mathbb{V}^+$ oznacza taki nieskończony ciąg, że wszystkie $f^n(x)$ są przedrostkami $f^\infty(x)$.

Korzystając z powyższych oznaczeń definiuje się cztery ciągi automatyczne:

- (i) ciąg Fibonacciego $a_{\mathbb{N}}^F := f_F^\infty(0)$, gdzie $f_F(0) = 01$, $f_F(1) = 0$ (Gramss, 1994);
- (ii) ciąg Thue-Morse'a $a_{\mathbb{N}}^T := f_T^\infty(0)$, gdzie $f_T(0) = 01$, $f_T(1) = 10$ (Berthé, 1994);
- (iii) ciąg Rudina-Shapiro $a_{\mathbb{N}}^R := p_R^+ \circ f_R^\infty(0)$, gdzie $f_R(a) = ab$, $f_R(b) = ac$, $f_R(c) = db$, $f_R(d) = dc$, $p_R(a) = p^R(b) = 0$, $p_R(c) = p^R(d) = 1$ (Berthé, 1994);
- (iv) ciąg *paperfolding* $a_{\mathbb{N}}^P := p_P^+ \circ f_P^\infty(0)$, gdzie $f_P(a) = ab$, $f_P(b) = cb$, $f_P(c) = ad$, $f_P(d) = cd$, $p_P(a) = p^R(b) = 1$, $p_P(c) = p^R(d) = 0$ (Berthé, 1994).

Konstruuje się także procesy $X_{\mathbb{Z}}^u$, gdzie $u = F, T, R, P$, zaś

$$P_1(X_{1:p}^u = x_{1:p}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbb{1}[a_{i:i+p-1}^u = x_{1:p}]. \quad (7.4)$$

dla każdego $x_{1:p} \in \{0,1\}^+$. Dla każdego z $X_{\mathbb{Z}}^u$ liczba różnych słów $x_{1:p} \in \mathbb{V}^p$ takich, że $P_1(X_{1:p}^u = x_{1:p}) > 0$, jest ograniczona przez $K^u p$ dla pewnych $K^u > 0$ (Berthé, 1994; Gramss, 1994). Stąd $H_{X^u}(n) \leq \log K^u + \log n$. Dokładniejsze obliczenia pokazują, że zmienne $\lim_{n \rightarrow \infty} \frac{1}{n} \log P(X_{1:n}^u = X_{1:n}^u(\cdot))$ prawie na pewno nie są stałymi zaś $|H_{X^u}(n) - \log n| \leq \log K^u$ (Berthé, 1994; Gramss, 1994). Zatem wszystkie cztery procesy są niefinitarne, a w świetle twierdzenia 4.10 także nieergodyczne.

Dla każdego z procesów $X_{\mathbb{Z}}^u$ entropia nadwyżkowa $E_{X^u}(n)$ rośnie logarytmicznie z n . Rozpatrzmy zatem inny potencjalny sposób konstrukcji procesów o potęgowo rozbiegającej entropii nadwyżkowej skończonego rzędu.

W rozdziale 5 pokazaliśmy, jak łatwo można konstruować dyskretne niefinitarne procesy nieprzeliczalnego opisu $X_{\mathbb{Z}}$, gdy liczba wartości zmiennych X_i jest nieskończona. Nie wiemy jednak niestety, jak konstruować procesy nieprzeliczalnego opisu o skończonej liczbie wartości — do których twierdzenie 6.11 mogłoby znaleźć zastosowanie. Pewnym ogólnym rozwiązaniem byłoby wskazanie klasy przekształceń, które przekształcają miary dyskretnych procesów stacjonarnych o nieskończonej liczbie wartości na miary procesów stacjonarnych o skończonej liczbie wartości z zachowaniem interesujących nas własności teorii informacyjnych.

Niech zmienne X_i , $i \in \mathbb{Z}$, na przestrzeni probabilistycznej $(\Omega_1, \mathcal{J}_1, P_1)$ spełniają $X_i(\Omega_1) \subset \mathbb{V}$, gdzie \mathbb{V} jest dowolnym przeliczalnym zbiorem znaków. Dla $f : \mathbb{V} \rightarrow \mathbb{W}^+$ niech f^+ będzie dane wzorami (7.2) i (7.3). Wówczas możemy skonstruować proces $Y'_{\mathbb{Z}}$, gdzie $Y'_{1:\infty} = f^+(X_{1:\infty})$, $Y'_{-\infty:0} = f^+(X_{-\infty:0})$, a zmienne Y'_i przyjmują wartości w \mathbb{W} . Niestety sama stacjonarność procesu $X_{\mathbb{Z}}$ zwykle nie zapewnia stacjonarności procesu $Y'_{\mathbb{Z}}$. Proces $Y'_{\mathbb{Z}}$ jest niestacjonarny na przykład dla $f(x) = g(x) \times h(x)$, gdzie $g(\mathbb{V}) \cup h(\mathbb{V}) \subset \mathbb{W}$ zaś $g(\mathbb{V}) \cap h(\mathbb{V}) = \emptyset$.

Zaproponujmy konstrukcję procesu o wartościach w \mathbb{W} , która produkuje wyłącznie procesy stacjonarne. Dla prostoty notacji będziemy rozpatrywać procesy stacjonarne $X_{\mathbb{N}}$ i $Y_{\mathbb{N}}$, gdzie zmienne X_i i Y_i indeksowane są liczbami naturalnymi. W razie potrzeby procesy te można zawsze rozszerzyć do $X_{\mathbb{Z}}$ i $Y_{\mathbb{Z}}$ (Kallenberg, 1997, lemat 9.2).

Przynajmniej niektóre miary stacjonarne $P_1 \circ X_{\mathbb{N}}^{-1}$ mają interpretację częstościową (zob. dodatek A.7). Istnieje wówczas taki ciąg $\bar{x}_{\mathbb{N}}$, $\bar{x}_i \in \mathbb{V}$, że prawdopodobieństwa wartości

$X_{1:p}$ są granicznymi częstościami p -literowych słów w $\bar{x}_\mathbb{N}$,

$$P_1(X_{1:p} = x_{1:p}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \llbracket \bar{x}_{i:i+p-1} = x_{1:p} \rrbracket. \quad (7.5)$$

Przy pewnych dodatkowych założeniach, na pewnej przestrzeni probabilistycznej $(\Omega_2, \mathcal{J}_2, P_2)$ istnieją zmienne Y_i takie, że $Y_i(\Omega_2) \subset \mathbb{W}$, zaś prawdopodobieństwa wartości $Y_{1:q}$ są granicznymi częstościami q -literowych słów w ciągu $\bar{y}_\mathbb{N} := f^+(\bar{x}_\mathbb{N})$,

$$P_2(Y_{1:q} = y_{1:q}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \llbracket \bar{y}_{i:i+q-1} = y_{1:q} \rrbracket. \quad (7.6)$$

Kierując się bardzo ogólnymi intuicjami, przypuszczamy, że dla pewnych funkcji f własności entropii blokowej H_Y procesu stacjonarnego $Y_\mathbb{N}$ są podobne do własności entropii blokowej H_X procesu $X_\mathbb{N}$. Pokażemy, że tak jest istotnie, gdy $f(\mathbb{V})$ jest zbiorem słów o stałej długości, a f jest iniekcją.

Jest dobrym pytaniem, czy dla dowolnej miary stacjonarnej $P_1 \circ X_\mathbb{N}^{-1}$ można skonstruować ciąg $\bar{x}_\mathbb{N}$ spełniający (7.5), ale tym zagadnieniem nie będziemy zajmować. Aby przekształcić miarę $P_1 \circ X_\mathbb{N}^{-1}$ na miarę $P_2 \circ Y_\mathbb{N}^{-1}$ nie trzeba bowiem konstruować ciągu $\bar{x}_\mathbb{N}$.

Definicja 7.1. (przeliterowanie) *Rozpatrzmy funkcję $f : \mathbb{V} \rightarrow \mathbb{W}^+$ i proces $X_\mathbb{N}$ na przestrzeni probabilistycznej $(\Omega_1, \mathcal{J}_1, P_1)$, gdzie $X_i(\Omega_1) \subset \mathbb{V}$ oraz*

$$\langle \text{len } f(X_1) \rangle < \infty. \quad (7.7)$$

Przeliterowaniem procesu $X_\mathbb{N}$ względem f nazywamy proces $(N, \bar{X}_\mathbb{N}, Y_\mathbb{N})$ na pewnej przestrzeni probabilistycznej $(\Omega_2, \mathcal{J}_2, P_2)$ taki, że $\bar{X}_i(\Omega_2) \subset \mathbb{V}$, $N(\Omega_2) \subset \mathbb{N}$, $Y_i(\Omega_2) \subset \mathbb{W}$, zaś

$$P_2(\bar{X}_1 = x_1) = \frac{P_1(X_1 = x_1) \text{len } f(x_1)}{\langle \text{len } f(X_1) \rangle}, \quad (7.8)$$

$$P_2(\bar{X}_{i:j} = x_{i:j} | \bar{X}_1 = x_1) = P_1(X_{i:j} = x_{i:j} | X_1 = x_1), \quad i, j \in \mathbb{N}, \quad (7.9)$$

$$N \perp\!\!\!\perp \bar{X}_\mathbb{N} | \bar{X}_1, \quad (7.10)$$

$$P_2(N = n | \bar{X}_1 = x_1) = \llbracket n \leq \text{len } f(x_1) \rrbracket / \text{len } f(x_1), \quad (7.11)$$

$$Y_\mathbb{N} = [f^+(\bar{X}_\mathbb{N})]_{N:\infty}, \quad (7.12)$$

gdzie $[y_{1:\infty}]_{k:l} := y_{k:l}$. Proces $Y_\mathbb{N}$ nazywamy procesem przeliterowanym.

Zwróćmy uwagę, że jeżeli $\text{len } f(x_1) = K$ dla każdego $x_1 \in \mathbb{V}$, to $N \perp\!\!\!\perp \bar{X}_\mathbb{N}$, $P_2(N = n) = \llbracket n \leq K \rrbracket / K$ oraz $P_2(\bar{X}_{1:m} = x_{1:m}) = P_1(X_{1:m} = x_{1:m})$. Można więc założyć, że $(\Omega_1, \mathcal{J}_1, P_1) = (\Omega_2, \mathcal{J}_2, P_2)$ zaś $\bar{X}_\mathbb{N} = X_\mathbb{N}$. Jeżeli długości $\text{len } f(x_1)$ nie są stałe, to zależność między $X_\mathbb{N}$ a jego przeliterowaniem jest znacznie bardziej skomplikowana. W ogólności proces $\bar{X}_\mathbb{N}$ nie musi być stacjonarny nawet, gdy $X_\mathbb{N}$ jest stacjonarny.

Dla uproszczenia oznaczmy $P_{X_\mathbb{N}} := P_1 \circ X_\mathbb{N}^{-1}$. Rozkład procesu $Y_\mathbb{N}$ ma postać

$$P_2(Y_{1:k} = y_{1:k}) = \int \left(\sum_{i=1}^{\text{len } f(x_1)} F(i, x_\mathbb{N}, y_{1:k}) \right) dP_{X_\mathbb{N}}(x_\mathbb{N}), \quad (7.13)$$

$$F(i, x_\mathbb{N}, y_{1:k}) := \frac{\llbracket [f^+(x_\mathbb{N})]_{i:i+k-1} = y_{1:k} \rrbracket}{\langle \text{len } f(X_1) \rangle}. \quad (7.14)$$

Mało przejrzysta definicja 7.1 pokrywa się ze wcześniej wspomnianą interpretacją częstościową, jeżeli tylko interpretacja taka istnieje.

Twierdzenie 7.2. Niech (N, \bar{X}_N, Y_N) będzie przeliterowaniem procesu X_N względem f . Jeżeli istnieje taki ciąg \bar{x}_N , że zachodzi (7.5), to dla $\bar{y}_N := f^+(\bar{x}_N)$ zachodzi (7.6).

Dowód: Podstawiając (7.5) otrzymujemy

$$\begin{aligned} \langle \text{len } f(X_1) \rangle &= \sum_{x_1 \in \mathbb{V}} P_1(X_1 = x_1) \text{len } f(x_1) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \sum_{x_1 \in \mathbb{V}} \llbracket \bar{x}_j = x_1 \rrbracket \text{len } f(x_1) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \text{len } f(\bar{x}_j) =: L. \end{aligned} \quad (7.15)$$

Niech $p \geq k$. Mamy wówczas

$$\begin{aligned} P_2(Y_{1:k} = y_{1:k}) &= \frac{1}{L} \int \sum_{i=1}^{\text{len } f(x_1)} \llbracket [f^+(x_N)]_{i:i+k-1} = y_{1:k} \rrbracket dP_{X_N}(x_N) \\ &= \frac{1}{L} \sum_{x_{1:p} \in \mathbb{V}^p} \sum_{i=1}^{\text{len } f(x_1)} \llbracket [f^+(x_{1:p})]_{i:i+k-1} = y_{1:k} \rrbracket P_1(X_{1:p} = x_{1:p}) \\ &= \frac{1}{L} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^{\text{len } f(\bar{x}_j)} \llbracket [f^+(\bar{x}_{j:j+p-1})]_{i:i+k-1} = y_{1:k} \rrbracket \\ &= \lim_{n \rightarrow \infty} \frac{\sum_{j=1}^n \sum_{i=1}^{\text{len } f(\bar{x}_j)} \llbracket [f^+(\bar{x}_{j:\infty})]_{i:i+k-1} = y_{1:k} \rrbracket}{\sum_{j=1}^n \text{len } f(\bar{x}_j)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{M(n)} \sum_{i=1}^{M(n)} \llbracket [f^+(\bar{x}_N)]_{i:i+k-1} = y_{1:k} \rrbracket, \end{aligned}$$

gdzie $M(n) := \sum_{j=1}^n \text{len } f(\bar{x}_j) \geq n$.

Oznaczmy $B(m) := \sum_{i=1}^m \llbracket [f^+(\bar{x}_N)]_{i:i+k-1} = y_{1:k} \rrbracket$. Ponieważ $B(m)$ jest rosnącą funkcją m , to dla każdego m spełniającego $M(n) \leq m \leq M(n+1)$ mamy

$$\frac{B(M(n))}{M(n+1)} \leq \frac{B(m)}{m} \leq \frac{B(M(n+1))}{M(n)}.$$

Aby dowieść, że $P_2(Y_{1:k} = y_{1:k}) = \lim_{n \rightarrow \infty} B(M(n))/M(n) = \lim_{m \rightarrow \infty} B(m)/m$, wystarczy pokazać, że

$$\lim_{n \rightarrow \infty} \frac{B(M(n-1))}{M(n)} = \lim_{n \rightarrow \infty} \frac{B(M(n+1))}{M(n)} = \lim_{n \rightarrow \infty} \frac{B(M(n))}{M(n)}. \quad (7.16)$$

Z (7.15) mamy, że $Ln - \epsilon(n)n \leq M(n) \leq Ln + \epsilon(n)n$ dla pewnego $\epsilon(n)$ malejącego z n do 0. Stąd $B(M(n+1)) \leq B(M(n)) + [M(n+1) - M(n)] \leq B(M(n)) + L + 2\epsilon(n)(n+1)$. A więc $\left| \frac{B(M(n+1))}{M(n)} - \frac{B(M(n))}{M(n)} \right| \leq \frac{L}{M(n)} + 2\epsilon(n) \frac{n+1}{M(n)}$. Ponieważ prawa strona nierówności dąży do 0 dla $n \rightarrow \infty$, to istotnie zachodzi (7.16). A zatem zachodzi także (7.6). \square

Niezależnie od tego, czy istnieje interpretacja częstościowa procesu X_N , mamy następujące twierdzenie.

Twierdzenie 7.3. Dla dowolnego przeliterowania (N, \bar{X}_N, Y_N) procesu stacjonarnego X_N względem f proces Y_N jest stacjonarny.

Dowód: Zgodnie z warunkiem (A.34), aby dowieść stacjonarności procesu $Y_{\mathbb{N}}$ wystarczy wykazać, że

$$\sum_{y_0 \in \mathbb{W}} P_2(Y_{1:k+1} = y_{0:k}) = P_2(Y_{1:k} = y_{1:k}) = \sum_{y_{k+1} \in \mathbb{W}} P_2(Y_{1:k+1} = y_{1:k+1}). \quad (7.17)$$

W istocie mamy

$$\begin{aligned} \sum_{y_0 \in \mathbb{W}} P_2(Y_{1:k+1} = y_{0:k}) &= \int \left(\sum_{i=1}^{\text{len } f(x_1)} F(i+1, x_{1:\infty}, y_{1:k}) \right) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) \\ &= \int \left(\sum_{i=2}^{\text{len } f(x_1)} F(i, x_{1:\infty}, y_{1:k}) \right) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) + \int F(1, x_{2:\infty}, y_{1:k}) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) \\ &= \int \left(\sum_{i=1}^{\text{len } f(x_1)} F(i, x_{1:\infty}, y_{1:k}) \right) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) = P_2(Y_{1:k} = y_{1:k}), \end{aligned}$$

korzystając w drugiej równości ze stacjonarności $X_{\mathbb{N}}$. Z drugiej strony

$$\sum_{y_{k+1} \in \mathbb{W}} P_2(Y_{1:k+1} = y_{1:k+1}) = \int \left(\sum_{i=1}^{\text{len } f(x_1)} F(i, x_{1:\infty}, y_{1:k}) \right) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) = P_2(Y_{1:k} = y_{1:k}).$$

□

Zauważmy, że dla dowolnego procesu stacjonarnego o skończonej liczbie wartości istnieje proces stacjonarny o wartościach binarnych i o zbliżonej funkcji entropii blokowej.

Twierdzenie 7.4. Niech $(N, \bar{X}_{\mathbb{N}}, Y_{\mathbb{N}})$ będzie przeliterowaniem procesu stacjonarnego $X_{\mathbb{N}}$ względem $f : \mathbb{V} \rightarrow \mathbb{W}^K$, gdzie \mathbb{V} jest skończone, a f jest iniekcją. Istnieje wówczas stała C taka, że dla wszystkich $n \in \mathbb{N}$ zachodzi

$$|H(Y_{1:nK}) - H(X_{1:n})| \leq C. \quad (7.18)$$

NB. Z ograniczenia (7.18) wynika $h_Y = h_X/K$ oraz $|E_Y(nK) - E_X(n)| \leq 3C$.

Dowód: Ponieważ dla $f : \mathbb{V} \rightarrow \mathbb{W}^K$ zachodzi $P_2(\bar{X}_{1:m} = x_{1:m}) = P_1(X_{1:m} = x_{1:m})$, to wystarczy pokazać, że $|H(Y_{1:nK}) - H(\bar{X}_{1:n})| \leq C$. Jest to łatwiejsze, gdyż zmienne $\bar{X}_{\mathbb{N}}$ i $Y_{\mathbb{N}}$ określone są na tej samej przestrzeni probabilistycznej. Mamy też $1 \leq N \leq K$.

Ponieważ $N + nK - 1 \leq (n+1)K$, to mamy $Y_{1:nK} = [f^+(\bar{X}_{\mathbb{N}})]_{N:N+nK-1} = [f^+(\bar{X}_{1:(n+1)})]_{N:N+nK-1}$. Z nierówności przetwarzania danych

$$\begin{aligned} H(Y_{1:nK}) &\leq H(\bar{X}_{1:n+1} \times N) \leq H(\bar{X}_{1:n}) + H(\bar{X}_{n+1}) + H(N) \\ &\leq H(\bar{X}_{1:n}) + \log \text{card } \mathbb{V} + \log K. \end{aligned}$$

Z drugiej strony mamy $Y_{K-N+2:(n+1)K-N+1} = [f^+(\bar{X}_{\mathbb{N}})]_{K+1:(n+1)K}$. Ponieważ $K - N + 2 \geq 2$ oraz $(n+1)K - N + 1 \leq (n+1)K$, to na mocy iniektwności f istnieje taka funkcja g , że $\bar{X}_{2:n+1} = g(Y_{2:(n+1)K}, N)$. Z nierówności przetwarzania danych wynika

$$\begin{aligned} H(\bar{X}_{1:n}) &= H(\bar{X}_{2:n+1}) \leq H(Y_{2:(n+1)K} \times N) \\ &\leq H(Y_{2:nK+1}) + H(Y_{nK+2:(n+1)K}) + H(N) \\ &\leq H(Y_{1:nK}) + H(Y_{1:K}) + \log K \\ &\leq H(Y_{1:nK}) + (K+1) \log \text{card } \mathbb{V} + 2 \log K. \end{aligned}$$

Jako stałą w (7.18) możemy zatem wziąć $C = (K+1) \log \text{card } \mathbb{V} + 2 \log K$.

□

W powyższym twierdzeniu kluczowe było założenie o iniektywności funkcji f . Uogólnienie twierdzenia 7.4 na przypadek, gdy $X_{\mathbb{N}}$ jest procesem o nieskończonej liczbie wartości, jest niestety nietrywialne. Jest to przypadek, który byłby potrzebny do przeliterowania na procesy binarne przykładów procesów nieprzeliczalnego opisu z rozdziału 5. Mimo to dobrą wiadomością jest, że dla szerokiej klasy rozkładów $(p_n)_{n \in \mathbb{N}}$ i iniekcji $f : \mathbb{N} \rightarrow \{0, 1\}^+$ proces $X_{\mathbb{N}}$ z przykładu 5.6 spełnia warunek (7.7), a więc można go przeliterować na odpowiedni proces binarny $Y_{\mathbb{N}}$.

Twierdzenie 7.5. *Niech \mathbb{W} będzie zbiorem skończonym. Dla dowolnej stałej $A > 1$ i zmiennej W , gdzie $W(\Omega_1) \subset \mathbb{W}^+$, zachodzi*

$$H(W) \leq \left[\langle \text{len } W \rangle \log(A \text{ card } \mathbb{W}) - \frac{A-2}{A-1} \right]. \quad (7.19)$$

Nierówność (7.19) zachodzi także, gdy $H(W)$ lub $\langle \text{len } W \rangle$ są nieskończone.

Dowód: Niech $p(w)$ i $q(w)$ spełniają $0 \leq p(w) \leq 1$, $0 < q(w) \leq 1$, $\sum_{w \in \mathbb{W}^+} p(w) < \infty$, $\sum_{w \in \mathbb{W}^+} q(w) < \infty$. Z nierówności $x - 1 - \log x \geq 0$ dla $x = q(w)/p(w) > 0$ mamy

$$- \sum_{w \in \mathbb{W}^+ : p(w) > 0} p(w) \log p(w) \leq - \sum_{w \in \mathbb{W}^+ : p(w) > 0} p(w) \log q(w) + \sum_{w \in \mathbb{W}^+} [q(w) - p(w)]$$

również, gdy dwa pierwsze szeregi są nieskończone. Kładąc $p(w) := P(W = w)$ i $q(w) := (A \text{ card } \mathbb{W})^{-\text{len } w}$, otrzymujemy

$$\begin{aligned} H(W) &\leq - \sum_{w \in \mathbb{W}^+ : p(w) > 0} p(w) \log q(w) + \sum_{w \in \mathbb{W}^+} [q(w) - p(w)] \\ &= \langle \text{len } W \rangle \log(A \text{ card } \mathbb{W}) + \sum_{n=1}^{\infty} A^{-n} - 1. \end{aligned}$$

□

Twierdzenie 7.6. *Jeżeli $H(X_1) = \infty$ dla zmiennej dyskretnej X_1 , to dla żadnego zbioru skończonego \mathbb{W} nie istnieje iniekcja $f : X_1(\Omega_1) \rightarrow \mathbb{W}^+$ taka, że $\langle \text{len } f(X_1) \rangle < \infty$.*

Dowód: Oznaczmy $W = f(X_1)$. Jeżeli f jest iniekcją, to $H(W) = H(X_1) = \infty$. Z nierówności (7.19) mamy zatem $\langle \text{len } W \rangle = \infty$. □

W świetle twierdzenia 7.6 istnieją procesy dyskretne $X_{\mathbb{N}}$ o nieskończonej liczbie wartości, których nie można przeliterować względem żadnej iniekcji f na proces $Y_{\mathbb{N}}$ o skończonej liczbie wartości, gdyż nie można spełnić warunku (7.7).

Wiemy, że dla procesów dyskretnych zachodzi nierówność $H(X_j) = \bar{H}(X_j) \geq \bar{I}(X_j; X_{j+1})$. Jeżeli $X_{\mathbb{Z}}$ jest dyskretnym niefinitarnym łańcuchem Markowa, to zachodzi

$$H(X_j) = \bar{H}(X_j) \geq \bar{I}(X_j; X_{j+1}) = \bar{I}(X_{-\infty:j}; X_{j+1:\infty}) = \infty. \quad (7.20)$$

Zatem żaden niefinitarny łańcuch Markowa nie może być przeliterowany na proces binarny względem iniektywnej funkcji f . Jest to wynik intrygujący, gdyż znamy przykłady dyskretnych ergodycznych niefinitarnych procesów Markowa (zob. przykład 2.13), natomiast nie znamy żadnego ergodycznego procesu niefinitarnego o skończonej liczbie wartości.

Pewnych pomysłów na uogólnienie twierdzenia 7.4 może przysporzyć rozpatrzenie transformacji w zamyśle odwrotnej do przeliterowania. Dla pewnych funkcji $f : \mathbb{V} \rightarrow \mathbb{W}^+$ i dowolnego procesu stacjonarnego $Y_{\mathbb{Z}}$ istnieje dokładnie jeden proces stacjonarny $X_{\mathbb{Z}}$ na tej samej przestrzeni probabilistycznej spełniający warunek $Y_{1:\infty} = f^+(X_{1:\infty})$ oraz $Y_{-\infty:0} = f^+(X_{-\infty:0})$.

Twierdzenie 7.7. Niech $f : \mathbb{V} \rightarrow \mathbb{W}^+$ będzie iniekcją taką, że $f(\mathbb{V})$ jest zbiorem bezprzedrostkowym. Wówczas $f^+ : \mathbb{V}^\rightarrow \rightarrow \mathbb{W}^\rightarrow$ jest iniekcją. Jeżeli dodatkowo zbiór $f(\mathbb{V})$ jest zupełny i skończony, to $f^+ : \mathbb{V}^\rightarrow \rightarrow \mathbb{W}^\rightarrow$ jest bijekcją.

NB. Analogiczne twierdzenie zachodzi dla $f^+ : \mathbb{V}^\leftarrow \rightarrow \mathbb{W}^\leftarrow$, gdy $f(\mathbb{V})$ jest zbiorem bezprzyrostkowym.

Dowód: Załóżmy, że ciąg $y_{\mathbb{N}} = f^+(x_{\mathbb{N}}) = f^+(x'_{\mathbb{N}})$ jest obrazem dwóch ciągów $x_{\mathbb{N}}, x'_{\mathbb{N}} \in \mathbb{V}^\rightarrow$. Ponieważ $f(\mathbb{V})$ jest bezprzedrostkowy, to istnieje dokładnie jedno słowo $u^1 \in f(\mathbb{V})$ i ciąg $y_{\mathbb{N}}^1 \in \mathbb{W}^\rightarrow$ takie, że $y_{\mathbb{N}} = u^1 \times y_{\mathbb{N}}^1$. Stąd wynika, że $f(x_1) = u^1 = f(x'_1)$ oraz $f^+(x'_{2:\infty}) = y^1 = f^+(x'_{2:\infty})$. Przez indukcję można pokazać, że dla każdego n istnieje dokładnie jeden ciąg słów $u^1, \dots, u^n \in \mathbb{W}^+$ i ciąg $y_{\mathbb{N}}^n \in \mathbb{W}^\rightarrow$ takie, że $f(x_i) = u^i = f(x'_i)$ dla $1 \leq i \leq n$ oraz $f^+(x'_{n+1:\infty}) = y_{\mathbb{N}}^n = f^+(x'_{n+1:\infty})$. Ponieważ f jest iniekcją, stąd wynika, że $x_i = x'_i$ dla każdego $i \in \mathbb{N}$. Zatem $x_{\mathbb{N}} = x'_{\mathbb{N}}$ a f^+ jest iniekcją.

Założmy teraz, że zbiór $f(\mathbb{V})$ jest skończony i że istnieje ciąg $y_{\mathbb{N}} \in \mathbb{W}^\rightarrow$, który nie jest obrazem żadnego $x_{\mathbb{N}} \in \mathbb{V}^\rightarrow$. Wówczas musi istnieć ciąg $y'_{\mathbb{N}} \in \mathbb{W}^\rightarrow$ taki, że żadne słowo z $f(\mathbb{V})$ nie jest przedrostkiem $y'_{\mathbb{N}}$. (Ciąg $y'_{\mathbb{N}}$ jest wynikiem odcięcia od $y_{\mathbb{N}}$ pewnej liczby przedrostków z $f(\mathbb{V})$. Liczba ta musi być skończona, gdyż założyliśmy, że $y_{\mathbb{N}}$ nie jest obrazem żadnego $x_{\mathbb{N}} \in \mathbb{V}^\rightarrow$.) Niech l będzie długością najdłuższego słowa z $f(\mathbb{V})$, a $w' \in \mathbb{W}^l$ niech będzie przedrostkiem ciągu $y'_{\mathbb{N}}$ długości l . Zachodzi $w' \notin f(\mathbb{V})$, a zbiór $\{w'\} \cup f(\mathbb{V})$ jest bezprzedrostkowy. Z nierówności Krafta (6.1) mamy więc

$$\sum_{w \in \{w'\} \cup f(\mathbb{V})} (\text{card } \mathbb{W})^{-\text{len } w} = \sum_{w \in f(\mathbb{V})} (\text{card } \mathbb{W})^{-\text{len } w} + (\text{card } \mathbb{W})^{-l} \leq 1, \quad (7.21)$$

czyli $f(\mathbb{V})$ nie jest zupełny.

Zatem jeżeli $f(\mathbb{V})$ jest zupełny i skończony, to $f^+ : \mathbb{V}^\rightarrow \rightarrow \mathbb{W}^\rightarrow$ jest bijekcją. \square

Przypomnijmy, że zbiór nazywamy bezrostkowym, gdy jest jednocześnie bezprzedrostkowy i bezprzyrostkowy. W przeciwieństwie do zbiorów bezprzedrostkowych i bezprzyrostkowych, zbiory bezrostkowe badane są od niedawna (Gillman i Rivest, 1995). Oprócz trywialnych przykładów zbiorów zupełnych istnieją przykłady mniej trywialne.

Przykład 7.8. (zbiór słów o stałej długości) Dla dowolnego $n \in \mathbb{N}$ zbiór $U = \mathbb{W}^n$ jest zupełnym zbiorem bezrostkowym.

Przykład 7.9. (zbiór słów o zmiennej długości) Dla $\mathbb{W} = \{a, b\}$ zbiór

$$U = \{ab, aaa, baa, bba, bbb, aaba, aabb, baba, babb\} \quad (7.22)$$

jest zupełnym zbiorem bezrostkowym (zob. Gillman i Rivest, 1995).

Skończone zupełne zbiory bezrostkowe o dość dowolnie zróżnicowanych długościach słów można definiować rekurencyjnie (Gillman i Rivest, 1995; Ahlswede et al., 1996).

Zdefiniujemy segmentację procesu stochastycznego.

Definicja 7.10. (segmentacja) Rozpatrzmy iniekcję $f : \mathbb{V} \rightarrow \mathbb{W}^+$ taką, że $f(\mathbb{V})$ jest zupełnym i skończonym zbiorem bezrostkowym, oraz dowolny proces $Y_{\mathbb{Z}}$, gdzie $Y_i(\Omega) \subset \mathbb{W}$. Segmentacją procesu $Y_{\mathbb{Z}}$ względem f nazywamy proces $X_{\mathbb{Z}}$ spełniający $X_i(\Omega) \subset \mathbb{V}$, $Y_{1:\infty} = f^+(X_{1:\infty})$ oraz $Y_{-\infty:0} = f^+(X_{-\infty:0})$.

Segmentacja $X_{\mathbb{Z}}$ dowolnego procesu $Y_{\mathbb{Z}}$ względem funkcji f z definicji 7.10 jest jednoznaczna na mocy twierdzenia 7.7. Konstrukcję prawie jednoznacznej segmentacji $X_{\mathbb{Z}}$ można by uogólnić na przypadek, gdy $f(\mathbb{V})$ jest dowolnym zbiorem bezrostkowym, a $P(Y_{1:\infty} \in f^+(\mathbb{V}^\rightarrow)) = P(Y_{-\infty:0} \in f^+(\mathbb{V}^\leftarrow)) = 1$. Uogólnienia tego nie będziemy tutaj rozpatrywać. W dalszych twierdzeniach istotnie wykorzystujemy skończoność i zupełność zbioru $f(\mathbb{V})$.

Aby udowodnić, że segmentacja procesu stacjonarnego jest stacjonarna, potrzebujemy prostego lematu.

Lemat 7.11. *Niech $U \subset \mathbb{W}^*$ będzie skończonym zupełnym zbiorem bezprzedrostkowym. Niech $U_m \subset U$ będzie podzbiorem słów o maksymalnej długości. Wówczas istnieje zbiór U_p taki, że $U_m = U_p \times \mathbb{W}$, $(U \setminus U_m) \cap U_p = \emptyset$, zaś $(U \setminus U_m) \cup U_p$ jest skończonym zupełnym zbiorem bezprzedrostkowym.*

NB. Analogiczne twierdzenie zachodzi dla zbiorów bezprzyrostkowych.

Dowód: Zdefiniujmy $U_p := \{w \in \mathbb{W}^* : \exists y \in \mathbb{W} : w \times y \in U_m\}$. Ponieważ zbiór U jest bezprzedrostkowy, to $(U \setminus U_m) \cap U_p = \emptyset$, zaś $(U \setminus U_m) \cup U_p$ jest także zbiorem bezprzedrostkowym. Mamy $U_m \subset U_p \times \mathbb{W}$, więc jeżeliby zachodziło $U_m \neq U_p \times \mathbb{W}$, to mielibyśmy

$$\begin{aligned} \sum_{w \in U} (\text{card } \mathbb{W})^{-\text{len } w} &< \sum_{w \in U \setminus U_m} (\text{card } \mathbb{W})^{-\text{len } w} + \sum_{w \in U_p \times \mathbb{W}} (\text{card } \mathbb{W})^{-\text{len } w} \\ &= \sum_{w \in U \setminus U_m} (\text{card } \mathbb{W})^{-\text{len } w} + \sum_{w \in U_p} (\text{card } \mathbb{W})^{-\text{len } w} \leq 1, \end{aligned}$$

a więc zbiór U nie byłby zupełny. Stąd dla zupełnego zbioru U musi zachodzić $U_m = U_p \times \mathbb{W}$, a zbiór $(U \setminus U_m) \cup U_p$ także musi być zupełny. \square

Twierdzenie 7.12. *Niech $X_{\mathbb{Z}}$ będzie segmentacją procesu $Y_{\mathbb{Z}}$ względem $f : \mathbb{V} \rightarrow \mathbb{W}^+$. Jeżeli $Y_{\mathbb{Z}}$ jest stacjonarny, to $X_{\mathbb{Z}}$ jest także stacjonarny.*

Dowód: Oznaczmy $\mathbf{P}(w) := P(Y_{i+1:i+\text{len } w} = w)$ dla $w \in \mathbb{W}^+$. Określmy długość $\text{len } U$ zbioru $U \subset \mathbb{W}^*$ jako maksymalną długość słowa należącego do U . Ze stacjonarności $Y_{\mathbb{Z}}$ i z lematu 7.11 dla dowolnego skończonego zupełnego zbioru bezprzedrostkowego U mamy

$$\begin{aligned} \sum_{w \in U} \mathbf{P}(w' \times w) &= \sum_{w \in U \setminus U_m} \mathbf{P}(w' \times w) + \sum_{w \in U_p \times \mathbb{W}} \mathbf{P}(w' \times w) \\ &= \sum_{w \in U \setminus U_m} \mathbf{P}(w' \times w) + \sum_{w \in U_p} \mathbf{P}(w' \times w) = \sum_{w \in U'} \mathbf{P}(w' \times w), \end{aligned} \quad (7.23)$$

gdzie $U' := (U \setminus U_m) \cup U_p$ jest zupełnym zbiorem bezprzedrostkowym o $\text{len } U' = \text{len } U - 1$. Ponieważ $\mathbf{P}(w' \times \lambda) = \mathbf{P}(w')$, a $\{\lambda\}$ jest zupełnym zbiorem bezprzedrostkowym o najmniejszej długości, to równość

$$\sum_{w \in U} \mathbf{P}(w' \times w) = \mathbf{P}(w') \quad (7.24)$$

zachodzi dla dowolnego skończonego zupełnego bezprzedrostkowego zbioru U .

Ze stwierdzenia analogicznego do lematu 7.11 wnioskujemy też, że równość

$$\sum_{w \in U} \mathbf{P}(w \times w') = \mathbf{P}(w') \quad (7.25)$$

zachodzi dla dowolnego skończonego zupełnego bezprzyrostkowego zbioru U .

Zwróćmy teraz uwagę, że

$$\begin{aligned} (X_{1:n} = x_{1:n}) &= (Y_{1:\text{len } f^+(x_{1:n})} = f^+(x_{1:n})), \\ (X_{m:0} = x_{m:0}) &= (Y_{-\text{len } f^+(x_{m:0})+1:0} = f^+(x_{m:0})). \end{aligned}$$

Zatem dla dowolnego $m \leq 1$ i $n \geq 0$ mamy

$$\begin{aligned} P(X_{m:n} = x_{m:n}) &= P(Y_{-\text{len } f^+(x_{m:0})+1:\text{len } f^+(x_{1:n})} = f^+(x_{m:n})) \\ &= \mathbf{P}(f^+(x_{m:n})). \end{aligned}$$

Z tożsamości $P(X_{1:n} = x_{1:n}) = \mathbf{P}(f^+(x_{1:n}))$ dla $m, n > 1$ uzyskujemy

$$\begin{aligned} P(X_{m:n} = x_{m:n}) &= \sum_{x_{1:m-1} \in \mathbb{V}^{m-1}} P(X_{1:n} = x_{1:n}) \\ &= \sum_{w \in [f(\mathbb{V})]^{m-1}} \mathbf{P}(w \times f^+(x_{m:n})) = \mathbf{P}(f^+(x_{m:n})), \end{aligned}$$

gdyż $[f(\mathbb{V})]^{m-1}$ jest zupełnym zbiorem bezrostkowym.

Podobnie z $P(X_{m:0} = x_{m:0}) = \mathbf{P}(f^+(x_{m:0}))$ dla $m, n < 0$ mamy

$$\begin{aligned} P(X_{m:n} = x_{m:n}) &= \sum_{x_{n-1:0} \in \mathbb{V}^{-n+2}} P(X_{m:0} = x_{m:0}) \\ &= \sum_{w \in [f(\mathbb{V})]^{-n+2}} \mathbf{P}(f^+(x_{m:n}) \times w) = \mathbf{P}(f^+(x_{m:n})), \end{aligned}$$

gdyż $[f(\mathbb{V})]^{-n+2}$ jest zupełnym zbiorem bezrostkowym.

Reasumując, proces $X_{\mathbb{Z}}$ jest stacjonarny. \square

Mamy też pewien analogon twierdzenia 7.4.

Twierdzenie 7.13. *Niech $X_{\mathbb{Z}}$ będzie segmentacją procesu $Y_{\mathbb{Z}}$ względem $f : \mathbb{V} \rightarrow \mathbb{W}^+$. Zachodzą związki*

$$H(Y_{1:n}) \leq H(X_{1:n}) \leq H(Y_{1:nK}), \quad (7.26)$$

$$\bar{I}(Y_{-m+1:0}; Y_{1:n}) \leq \bar{I}(X_{-m+1:0}; X_{1:n}) \leq \bar{I}(Y_{-mK+1:0}; Y_{1:nK}), \quad (7.27)$$

$$E_Y = \bar{I}(Y_{-\infty:0}; Y_{1:\infty}) = \bar{I}(X_{-\infty:0}; X_{1:\infty}) = E_X, \quad (7.28)$$

gdzie K jest długością najdłuższego słowa w $f(\mathbb{V})$.

Dowód: Ponieważ $Y_{1:\infty} = f^+(X_{1:\infty})$ oraz $Y_{-\infty:0} = f^+(X_{-\infty:0})$, a f jest iniekcją, to istnieją funkcje g_1, \dots, g_4 takie, że $Y_{-m+1:0} = g_1(X_{-m+1:0})$, $X_{-m+1:0} = g_2(Y_{-mK+1:0})$, $Y_{1:n} = g_3(X_{1:n})$ i $X_{1:n} = g_4(Y_{1:nK})$. A zatem nierówności (7.26) i (7.27) wynikają z nierówności przetwarzania danych. Równość (7.28) wynika z (7.27), (3.1) i twierdzenia o trzech ciągach. \square

Nie jest jasne, w jakim zakresie segmentacja jest operacją odwrotną do przeliterowania. Niech proces $X_{\mathbb{Z}}$ będzie segmentacją procesu stacjonarnego $Y_{\mathbb{Z}}$ względem f . Wówczas proces $Y_{\mathbb{Z}}$ jest procesem przeliterowanym uzyskanym z $X_{\mathbb{Z}}$ za pomocą f wtedy i tylko wtedy, gdy zachodzą równości

$$\begin{aligned} P_2(Y_{1:k} = y_{1:k}) &= \int \mathbb{I}[[f^+(x_{\mathbb{N}})]_{1:k} = y_{1:k}] dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}) \\ &= \int \left(\frac{1}{\langle \text{len } f(X_1) \rangle} \sum_{i=1}^{\text{len } f(x_1)} \mathbb{I}[[f^+(x_{\mathbb{N}})]_{i:i+k-1} = y_{1:k}] \right) dP_{X_{\mathbb{N}}}(x_{\mathbb{N}}). \quad (7.29) \end{aligned}$$

Rozstrzygnięcie, czy równość (7.29) jest prawdziwa, pozostawiamy jako problem otwarty.

Część III

Procesy gaussowskie

Rozdział 8

Własności podstawowe

Procesem gaussowskim nazywa się dowolny proces $X_{\mathbb{Z}}$ o wartościach rzeczywistych lub zespolonych taki, że skończone podzbiory zmiennych X_n mają wielowymiarowy rozkład normalny. Dowolny wielowymiarowy rozkład normalny zadany jest wyłącznie przez miary zależności par zmiennych (X_m, X_n) . Dlatego studium procesów gaussowskich wydaje się być naturalnym pierwszym etapem w badaniu związków pomiędzy informacją wzajemną $I(X_1; X_n)$ a warunkową informacją wzajemną $I(X_1; X_n | X_{2:n-1})$ jako funkcjami n .

Dowolny proces gaussowski $X_{\mathbb{Z}}$ można przedstawić jako $X_n = Y_n + \langle X_n \rangle$, gdzie wyrazy $\langle X_n \rangle$, $n \in \mathbb{Z}$, są stałymi zaś $Y_{\mathbb{Z}}$ jest procesem gaussowskim o zerowej wartości oczekiwanej. Dla $K = (k_1, \dots, k_n)$, gdzie $k_i \in \mathbb{Z}$ i $n \in \mathbb{N}$, oznaczajmy $Y_K = (Y_{k_1}, \dots, Y_{k_n})$ oraz $X_K = (X_{k_1}, \dots, X_{k_n})$. Ponieważ $H(Y_K) = H(X_K)$ dla dowolnego K , badając własności teo-rioinformacyjne procesów gaussowskich, wystarczy ograniczyć się do przypadku procesów o zerowej wartości oczekiwanej. Przypomnijmy ich jawną definicję.

Definicja 8.1. (proces gaussowski) *Procesem gaussowskim o zerowej wartości oczekiwanej nazywa się dowolny proces $X_{\mathbb{Z}}$ taki, że zmienne wielowymiarowe X_K mają rozkład normalny*

$$\rho_{X_K}(x_K) = \frac{1}{[(2\pi)^{\text{card } K} \det \Gamma(K)]^{d/2}} \exp \left[-\frac{1}{2d} \sum_{i \in K} \sum_{j \in K} x_i^* [\Gamma(K)^{-1}]_{ij} x_j \right], \quad (8.1)$$

gdzie $d = 1$ dla procesów o wartościach rzeczywistych (X_n są zmiennymi rzeczywistymi), zaś $d = 2$ dla procesów o wartościach zespolonych ($X_n = \text{Re } X_n + i \text{Im } X_n$, gdzie $\text{Re } X_n$ i $\text{Im } X_n$ są zmiennymi rzeczywistymi).

Wzór (8.1) zapisujemy w skrócie jako $\rho_{X_K} \sim N[0, \Gamma(K)]$. Macierz $\Gamma(K)$ jest skończonym minorem o wymiarach $\text{card } K \times \text{card } K$ nieskończonej macierzy kowariancji Γ o elementach $\Gamma(K)_{ij} = \Gamma_{ij} = \text{Cov}(X_i; X_j) := \langle X_i^* X_j \rangle - \langle X_i^* \rangle \langle X_j \rangle = \langle X_i^* X_j \rangle$ dla $i, j \in K \subset \mathbb{Z}$ (Brockwell i Davis, 1987, stwierdzenia 1.6.4, 1.6.6). Dla zmiennych X, Y określmy ponadto normę $\|X\| = \sqrt{\text{Cov}(X; X)}$ oraz współczynnik korelacji $\text{Corr}(X, Y) = \text{Cov}(X; Y) / \|X\| \cdot \|Y\|$, gdzie $|\text{Corr}(X; Z)| \leq 1$ z nierówności Schwarzera.

Twierdzenie 8.2. (Brockwell i Davis, 1987, stwierdzenie 1.6.2) *Proces gaussowski $X_{\mathbb{Z}}$, dla którego skończone podzbiory zmiennych mają rozkład (8.1), istnieje wtedy i tylko wtedy, gdy nieskończona macierz Γ jest hermitowska i nieujemnie określona, tzn. $\Gamma_{ij} = \Gamma_{ji}^*$ oraz*

$$\sum_{i \in K} \sum_{j \in K} a_i^* \Gamma_{ij} a_j \geq 0 \quad (8.2)$$

zachodzi dla każdego ciągu $(a_i)_{i \in K}$.

Warunek (8.2) jest równoważny stwierdzeniu, że $\left\| \sum_{i \in K} a_i X_i \right\|^2 \geq 0$.

Ważnym teoretycznie przypadkiem procesu gaussowskiego jest biały szum.

Definicja 8.3. (biały szum) Standardowym białym szumem gaussowskim nazywamy proces gaussowski $Z_{\mathbb{Z}}$ o zerowej wartości oczekiwanej taki, że $\text{Cov}(Z_i; Z_j) = \llbracket i = j \rrbracket$.

Wielowymiarowe rozkłady normalne mają dwie ważne własności. Po pierwsze dla każdej macierzy B o wymiarach $\text{card } I \times \text{card } K$ i gęstości $\rho_{X_K} \sim N[0, \Gamma(K)]$, kombinacja liniowa $U_I = BX_K$ ma rozkład normalny $\rho_{U_K} \sim N[0, B\Gamma(K)B^T]$, gdzie $B_{ij}^T := B_{ji}^*$. Po drugie, dla każdej gęstości

$$\rho \begin{bmatrix} X_I \\ X_K \end{bmatrix} \sim N \left[0, \Gamma \left(\begin{bmatrix} I \\ K \end{bmatrix} \right) \right], \quad \Gamma \left(\begin{bmatrix} I \\ K \end{bmatrix} \right) = \begin{bmatrix} \Gamma(I) & \Gamma(I, K) \\ \Gamma(I, K)^T & \Gamma(K) \end{bmatrix}, \quad (8.3)$$

zmienne X_I i X_K są niezależne wtedy i tylko wtedy, gdy $\Gamma(I, K) = 0$. Przestrzeń liniowa rozpinana przez zmienne X_i , $i \in K$, jest skończenie wymiarową przestrzenią Hilberta z iloczynem $\text{Cov}(\cdot; \cdot)$. Ortogonalność dowolnych zmiennych z tej przestrzeni jest równoważna ich niezależności.

Dla rozkładu (8.3) kombinacja liniowa $Y_I^K := X_I - \Gamma(I, K)\Gamma(K)^{-1}X_K$ jest niezależna od X_K (Brockwell i Davis, 1987, stwierdzenie 1.6.6). Y_I^K nazywa się innowacją X_I względem X_K . Zmienna $\Phi_I^K := X_I - Y_I^K = \Gamma(I, K)\Gamma(K)^{-1}X_K$ nazywana jest najlepszym liniowym predyktorem X_I względem X_K . Dla $I = (i_1, \dots, i_n)$ mamy $\Phi_I^K = (\Phi_{i_1}^K, \dots, \Phi_{i_n}^K)$, $Y_I^K = (Y_{i_1}^K, \dots, Y_{i_n}^K)$, gdzie

$$\Phi_i^K = \phi_i^K(X_K) := \sum_{j \in K} \phi_{ij}^K X_j, \quad Y_i^K := X_i - \Phi_i^K, \quad (8.4)$$

zaś zmienne Φ_i^K i Y_i^K oraz liczby ϕ_{ij}^K nie zależą od I . Współczynniki ϕ_{ij}^K spełniają równania Yule'a-Walkera

$$\phi_{ij}^K \|X_j\|^2 = \text{Cov}(X_i; X_j) - \sum_{k \in K \setminus \{j\}} \phi_{ik}^K \text{Cov}(X_k; X_j), \quad j \in K. \quad (8.5)$$

Wielkość $\left\| X_i - \sum_{j \in K} a_j X_j \right\|^2$ jest kwadratową funkcją liczb $(a_j)_{j \in K}$ z jedynym minimum lokalnym dla $a_j = \phi_{ij}^K$ (Brockwell i Davis, 1987, twierdzenie 2.3.1). Dla zadanej macierzy kowariancji współczynniki $(\phi_{ij}^K)_{j \in K}$ można obliczyć z zadaną w czasie rzędu $(\text{card } K)^3$ korzystając z algorytmu innowacji, będącego pewną wersją algorytmu ortogonalizacji Grama-Schmidta (Brockwell i Davis, 1987, twierdzenie 5.2.2). Idea algorytmu innowacji zasadza się na iteracyjnym wykorzystaniu tożsamości

$$\Phi_i^{m:n} = \sum_{j=m}^n \frac{\text{Cov}(Y_j^{m:j-1}; \Phi_i^{m:n})}{\|Y_j^{m:j-1}\|^2} \cdot Y_j^{m:j-1}, \quad (8.6)$$

gdzie $\text{Cov}(Y_j^{m:j-1}; \Phi_i^{m:n}) = \text{Cov}(Y_j^{m:j-1}; X_i)$.

Dla stacjonarnych procesów gaussowskich, pojawiają się pewne uproszczenia. Po pierwsze, macierz kowariancji Γ jest macierzą Toeplitza $\Gamma_{ij} = \gamma(i-j)$, gdzie funkcja γ nazywana jest funkcją autokowariancji. Wygodnie jest też zdefiniować następujące obiekty: funkcję

autokorelacji (ACF) $\rho(n)$, współczynniki predyktorów ϕ_{ni} , funkcję częściowej autokorelacji (PACF) $\alpha(n)$ oraz względną wariancję innowacji $v(n)$,

$$\rho(n) := \text{Corr}(X_{k+n+1}; X_{k+1}) = \gamma(n)/\gamma(0), \quad (8.7)$$

$$\phi_{ni} := \phi_{k+n+1, k+n+1-i}^{k+1:k+n} = (\phi_{k, k+i}^{k+1:k+n})^*, \quad (8.8)$$

$$\alpha(n) := \text{Corr}(Y_{k+n+1}^{k+2:k+n}; Y_{k+1}^{k+2:k+n}), \quad (8.9)$$

$$v_n := \frac{\|Y_{k+n+1}^{k+1:k+n}\|^2}{\|X_0\|^2} = \frac{\|Y_k^{k+1:k+n}\|^2}{\|X_0\|^2}. \quad (8.10)$$

Przy danym $(\rho(n))_{n \in \{1, \dots, N\}}$ lub $(\alpha(n))_{n \in \{1, \dots, N\}}$, pozostałe trzy wielkości można obliczyć za pomocą algorytmu Durбина-Levinsona w czasie rzędu N^2 (Durbin, 1960; Brockwell i Davis, 1987, stwierdzenie 5.2.1). W algorytmie tym kładzie się $v_0 = 1$, $\alpha(1) = \rho(1)$, a następnie iteruje

$$v_n = [1 - |\alpha(n)|^2] v_{n-1}, \quad (8.11)$$

$$\phi_{nn} = \alpha^*(n), \quad (8.12)$$

$$\phi_{nj} = \phi_{n-1, j} - \alpha^*(n) \phi_{n-1, n-j}^*, \quad j \in \{1, \dots, n-1\}, \quad (8.13)$$

$$\alpha(n+1) = \left[\rho(n+1) - \sum_{j=1}^n \phi_{nj}^* \rho(n+1-j) \right] / v_n \quad (8.14)$$

dla $n \in \{1, \dots, N-1\}$. Ostatnie równanie można odwrócić, aby otrzymać $\rho(n+1)$ zamiast $\alpha(n+1)$.¹ Zależność pomiędzy ciągami $(\rho(n))_{n \in \{1, \dots, N\}}$ i $(\alpha(n))_{n \in \{1, \dots, N\}}$ jest wzajemnie jednoznaczna (Pourahmadi, 2001, sekcja 7.3). Alternatywnym algorytmem obliczania częściowej autokorelacji jest algorytm Schura, który jest wygodniejszy do implementacji równoległej (Kailath, 1987).

Warunek istnienia procesu dla funkcji częściowej autokorelacji przybiera znacznie prostszą postać niż dla funkcji autokorelacji.

Twierdzenie 8.4. (Ramsey, 1974; por. Schur, 1917) *Stacjonarny proces gaussowski $X_{\mathbb{Z}}$, dla którego skończone podzbiory zmiennych mają rozkład (8.1), istnieje wtedy i tylko wtedy, gdy nieskończona macierz Γ spełnia warunek $\Gamma_{ij} = \gamma(i-j)$, $\gamma(-k) = \gamma(k)^*$, zaś $\alpha(\cdot)$ obliczone z $\rho(\cdot) = \gamma(\cdot)/\gamma(0)$ wzorami (8.11)–(8.14) spełnia dwa warunki:*

(i) $|\alpha(n)| \leq 1$ dla wszystkich $n \geq 1$,

(ii) jeżeli $|\alpha(k)| = 1$ dla pewnego $k > 0$, to $\alpha(n) = 0$ dla wszystkich $n > k$.

Podobne twierdzenie można sformułować dla uogólnionej funkcji częściowej autokorelacji dla procesu niestacjonarnego (Dégerine i Lambert-Lacroix, 2003).

Przejdźmy teraz do własności entropii dla zmiennych o rozkładzie normalnym.

Twierdzenie 8.5. (Cover i Thomas, 1991, twierdzenie 9.4.1) *Dla układu standardowych miar pomiarowych entropia zmiennej X_K o rozkładzie (8.1) wynosi*

$$H(X_K) = \frac{1}{2} [1 + d \log(2\pi)] \text{card } K + \frac{d}{2} \log \det \Gamma(K). \quad (8.16)$$

¹ Zdecydowanie dogodniejsze niż (8.14) do obliczania autokorelacji ρ przy danej częściowej autokorelacji α są równania Yule'a-Walkera (8.5), które dla procesu stacjonarnego przybierają postać

$$\sum_{i=1}^n \phi_{ni}^* \rho(k-i) = \rho(k), \quad n \geq 1, \quad k \in \{1, \dots, n\}. \quad (8.15)$$

Bezpośrednią konsekwencją tego stwierdzenia i równości (1.5) jest następujący wniosek:

Twierdzenie 8.6. *Dla pary (X, Y) zmiennych rzeczywistych lub zespolonych o dwuwymiarowym rozkładzie normalnym zachodzi*

$$I(X; Y) = -\frac{d}{2} \log [1 - |\text{Corr}(X; Y)|^2]. \quad (8.17)$$

Dla zmiennych gaussowskich X i Y istnieje także kąt korelacji $\phi(X; Y) \in [0, \pi]$, który spełnia $e^{-I(X; Y)/d} = \sin \phi(X; Y)$ i $\text{Corr}(X; Y) = \cos \phi(X; Y)$. Funkcja $\sin \phi(X; Y)$ jest metryką na zbiorze unormowanych zmiennych losowych ($\|X\| = \|Y\| = 1$).

Twierdzenie 8.7. *Dla dowolnych zmiennych X_1, X_2, X_3 zachodzi nierówność*

$$\sin \phi(X_2; X_3) \leq \sin \phi(X_2; X_1) + \sin \phi(X_1; X_3), \quad (8.18)$$

gdzie $\sin \phi(X; Y) := \sqrt{1 - |\text{Corr}(X; Y)|^2}$.

Dowód: Bez zmniejszania ogólności połóżmy $\|X_1\| = \|X_2\| = \|X_3\| = 1$. Możemy przedstawić $X_2 = a_2 X_1 + b_2 Y_2$, $X_3 = a_3 X_1 + b_3 Y_3$, gdzie $\text{Cov}(Y_2, X_1) = 0$, $\text{Cov}(Y_3, X_1) = 0$, $\|Y_2\| = \|Y_3\| = 1$ oraz $|a_2|^2 + |b_2|^2 = |a_3|^2 + |b_3|^2 = 1$. Mamy $\sin^2 \phi(X_2; X_1) = |b_2|^2$, $\sin^2 \phi(X_1; X_3) = |b_3|^2$ oraz $\sin^2 \phi(X_2; X_1) = 1 - |a_2^* a_3 + b_2^* b_3 \text{Corr}(Y_2; Y_3)|^2$. Rozpatrzmy teraz dwa przypadki.

(i) Załóżmy, że $|b_2| |b_3| \geq |a_2| |a_3|$. Wówczas

$$0 \leq |b_2|^2 |b_3|^2 - (1 - |b_2|^2)(1 - |b_3|^2) = |b_2|^2 + |b_3|^2 - 1 \leq (|b_2| + |b_3|)^2 - 1.$$

Mamy zatem $\sin \phi(X_2; X_1) + \sin \phi(X_1; X_3) = |b_2| + |b_3| \geq 1 \geq \sin \phi(X_2; X_3)$.

(ii) Załóżmy, że $|b_2| |b_3| \leq |a_2| |a_3|$. Wówczas $|a_2| |a_3| - |b_2| |b_3| \in [0, 1]$ i mamy

$$\begin{aligned} \sin^2 \phi(X_2; X_1) &\leq 1 - (|a_2| |a_3| - |b_2| |b_3|)^2 \\ &= |b_2|^2 + |b_3|^2 + 2(|a_2| |a_3| - |b_2| |b_3|) |b_2| |b_3| \\ &\leq |b_2|^2 + |b_3|^2 + 2|b_2| |b_3| = (\sin \phi(X_2; X_1) + \sin \phi(X_1; X_3))^2. \end{aligned}$$

□

Odpowiednik równości (8.17) zachodzi także dla warunkowej informacji wzajemnej.

Twierdzenie 8.8. *Dla stacjonarnego procesu gaussowskiego X_Z zachodzą równości*

$$I(X_1; X_n) = -\frac{d}{2} \log [1 - |\rho(n-1)|^2], \quad (8.19)$$

$$I(X_1; X_n | X_{2:n-1}) = -\Delta^2 H_X(n) = -\frac{d}{2} \log [1 - |\alpha(n-1)|^2]. \quad (8.20)$$

Dowód: (8.19) wynika bezpośrednio z (8.7) i (8.17). Aby udowodnić równość (8.20) zauważmy, że innowacja $Y_K^I = X_K - \Phi_K^I$ jest niezależna od X_I , a rzut ortogonalny Φ_K^I jako funkcja X_I jest stały przy danym X_I . Ponieważ entropia $H(X_K + A_K(X_I) | X_I)$ jest równa entropii $H(X_K | X_I)$ dla dowolnej funkcji A_K mierzalnej $\tilde{\sigma}(X_I)/\tilde{\sigma}(X_K)$, to

$$H(X_K | X_I) = H(X_K - \Phi_K^I | X_I) = H(Y_K^I | X_I) = H(Y_K^I). \quad (8.21)$$

A zatem

$$\begin{aligned} I(X_1; X_n | X_{2:n-1}) &= H(X_1 | X_{2:n-1}) + H(X_n | X_{2:n-1}) - H(X_1 \times X_n | X_{2:n-1}) \\ &= H(Y_1^{2:n-1}) + H(Y_n^{2:n-1}) - H(Y_1^{2:n-1} \times Y_n^{2:n-1}) \\ &= I(Y_1^{2:n-1}; Y_n^{2:n-1}). \end{aligned}$$

Ponieważ para $(Y_1^{2:n-1}, Y_n^{2:n-1})$ ma rozkład normalny, (8.20) wynika z (8.9) i (8.17). □

Ze wzorów (8.20) i (8.16) wynika, że algorytm Durбина-Levinsona jest efektywnym algorytmem obliczania wyznacznika dla nieujemnie określonej macierzy Toeplitza. W istocie (8.20) i (8.11) implikują

$$\Delta H_X(n) - H_X(1) = \sum_{k=2}^n \Delta^2 H_X(k) = \frac{d}{2} \log v_{n-1} \leq 0, \quad (8.22)$$

$$H_X(n) - nH_X(1) = \sum_{k=1}^n [\Delta H_X(k) - H_X(1)] = \frac{d}{2} \sum_{k=1}^n \log v_{k-1} \leq 0. \quad (8.23)$$

Z drugiej strony, z (8.16) wynika

$$H_X(n) - nH_X(1) = \frac{d}{2} [\log \det \Gamma(\{1, \dots, n\}) - n \log \gamma(0)]. \quad (8.24)$$

A zatem $\det \Gamma(\{1, \dots, n\}) = \gamma(0)^n \prod_{k=1}^n v_{k-1} = \gamma(0)^n \prod_{k=1}^{n-1} [1 - |\alpha(k)|^2]^{n-k}$.

Można zadać pytanie, czy istnieją procesy gaussowskie o dowolnej wartości entropii nadwyżkowej. Odpowiedź jest pozytywna.

Twierdzenie 8.9. *Dla dowolnej rzeczywistej funkcji wklęsłej H_X spełniającej $H_X(0) = 0$ istnieje stacjonarny proces gaussowski X_Z taki, że H_X jest entropią blokową tego procesu.*

Dowód: Stwierdzenie, że funkcja $H_X : \mathbb{N} \ni n \mapsto H_X(n) \in \mathbb{R}$ jest wklęsła, jest równoważne stwierdzeniu, że $\Delta^2 H_X(n) \in (-\infty, 0]$ zachodzi dla wszystkich $n \geq 2$. Relacja $I(X_1; X_n | X_{2:n-1}) = -\Delta^2 H_X(n) \in [0, \infty)$ dla wszystkich $n \geq 2$ implikuje $|\alpha(n-1)| < 1$ dla wszystkich $n \geq 2$. Ten ostatni warunek jest warunkiem dostatecznym istnienia procesu gaussowskiego. \square

Rozdział 9

Niektóre własności graniczne

Rezultaty przedstawione dotychczas odnoszą się wyłącznie do własności skończonych zbiorów zmiennych procesów gaussowskich. W dalszej kolejności porównamy kilka własności asymptotycznych dla procesów stacjonarnych.

Niech $X_{\mathbb{Z}}$ będzie procesem gaussowskim o zerowej wartości oczekiwanej. Dla dowolnego $U \subset \mathbb{Z}$ niech \mathcal{M}_U oznacza domkniętą przestrzeń Hilberta rozpinaną przez kombinacje liniowe $\sum_{j \in B} c_j X_j$ dla skończonych podzbiorów $B \subset U$ (tzn. przestrzeń granic ciągów Cauchy'ego w L^2 , których elementami są sumy $\sum_{j \in B} c_j X_j$ dla skończonych $B \subset U$). Z twierdzenia o rzucie ortogonalnym dla domkniętych podprzestrzeni Hilberta (Rudin, 1986, twierdzenie 4.11) istnieje najlepszy liniowy predyktor Φ_j^U określony jako jedyny element \mathcal{M}_U , dla którego zachodzi $\|X_j - \Phi_j^U\| = \inf_{V \in \mathcal{M}_U} \|X_j - V\|$.

Twierdzenie 9.1. *Niech $X_{\mathbb{Z}}$ będzie procesem gaussowskim o zerowej wartości oczekiwanej. Jeżeli istnieje funkcja f mierzalna $\tilde{\sigma}(X_U)/\tilde{\sigma}(X_j)$ taka, że $X_j = f(X_U)$ dla $U \subset \mathbb{Z}$ i $j \in \mathbb{Z}$, to $f(X_U) = \Phi_j^U$.*

Dowód: Załóżmy, że $Z := \Phi_j^U - X_j = \Phi_j^U - f(X_U)$ jest różne od 0. Zmienne Z i X_U są nieskorelowane i gaussowskie, a zatem $\sigma(Z) \perp \sigma(X_U)$. Ponieważ Z jest mierzalną funkcją X_U , to $\sigma(Z) \subset \sigma(X_U)$. W świetle $\sigma(Z) \perp \sigma(X_U)$ wynika stąd, że $\sigma(Z) \perp \sigma(Z)$. Jako, że $\langle Z \rangle = 0$, z $\sigma(Z) \perp \sigma(Z)$ mamy $Z = 0$ wbrew poczynionemu założeniu. Zatem istotnie zachodzi $f(X_U) = \Phi_j^U$. \square

Rozszerzając konwencję z poprzedniego rozdziału, dla $m, n \in \mathbb{Z} \cup \{-\infty, \infty\}$ oraz $U = \{k \in \mathbb{Z} : m \leq k \leq p\}$ będziemy pisać $\mathcal{M}_U = \mathcal{M}_{m:n}$ oraz $\Phi_j^U = \Phi_j^{m:n}$. Na mocy poniższego twierdzenia predyktory $\Phi_p^{-\infty:p-1}$ są granicami predyktorów skończonych.

Twierdzenie 9.2. *Dla każdego gaussowskiego procesu stacjonarnego zachodzi równość*

$$\lim_{n \rightarrow \infty} \Phi_p^{p-n:p-1} = \Phi_p^{-\infty:p-1}. \quad (9.1)$$

Dowód: Łatwo zauważyć, że $\lim_{n \rightarrow \infty} \Phi_p^{p-n:p-1}$ istnieje i należy do $\mathcal{M}_{-\infty:p-1}$. Ze wzoru 8.6 (por. Brockwell i Davis, 1987, wzór 5.2.6) wynika, że

$$\Phi_p^{p-n:p-1} - \Phi_p^{p-n+1:p-1} = \alpha^*(n) Y_{p-n}^{p-n+1:p-1}. \quad (9.2)$$

Ponieważ $\|\alpha^*(n) Y_{p-n}^{p-n+1:p-1}\|^2 = \|X_p\|^2 |\alpha(n)|^2 v_{n-1} = \|X_p\|^2 [v_{n-1} - v_n]$, otrzymujemy, że

$$\|\Phi_p^{p-m:p-1} - \Phi_p^{p-n:p-1}\|^2 = \|X_p\|^2 \sum_{k=n+1}^m [v_{k-1} - v_k] = \|X_p\|^2 [v_n - v_m].$$

Granica $\lim_{n \rightarrow \infty} v_n$ istnieje zawsze, a więc ciąg $(\Phi_p^{p-m:p-1})_{m \in \mathbb{N}}$ jest ciągiem Cauchy'ego i ma granicę w $\bigcup_{n \in \mathbb{N}} \mathcal{M}_{p-n:p-1} = \mathcal{M}_{-\infty:p-1}$.

Niech $D = \{n \in \mathbb{N} : \|Y_{p-n}^{p-n+1:p-1}\| > 0\}$. Dla $n \in D$ definiujemy zmienne

$$\bar{Z}_n^p := \frac{Y_{p-n}^{p-n+1:p-1}}{\|Y_{p-n}^{p-n+1:p-1}\|}. \quad (9.3)$$

Zmienne $(\bar{Z}_n^p)_{n \in D}$ są ortonormalną bazą przestrzeni $\mathcal{M}_{-\infty:p-1}$, natomiast $(\bar{Z}_n^p)_{n \in D \cap \{1,2,\dots,n\}}$ są ortonormalną bazą przestrzeni $\mathcal{M}_{p-n:p-1}$. Z twierdzenia 2.4.2 w książce Brockwella i Davisa (1987) wynika, że istnieje ciąg $V_{\mathbb{N}}$ taki, że $V_n \in \mathcal{M}_{p-n:p-1}$ oraz $\|\Phi_p^{-\infty:p-1} - V_n\| < \epsilon_n$, gdzie $\epsilon_n \rightarrow 0$. Korzystając z nierówności trójkąta, otrzymujemy

$$\begin{aligned} \|X_p - \Phi_p^{-\infty:p-1}\| &\leq \|X_p - \Phi_p^{p-n:p-1}\| \leq \|X_p - V_n\| \\ &\leq \|X_p - \Phi_p^{-\infty:p-1}\| + \epsilon_n, \end{aligned}$$

czyli $\|X_p - \Phi_p^{-\infty:p-1}\| = \|X_p - \lim_{n \rightarrow \infty} \Phi_p^{p-n:p-1}\|$. Zatem $\lim_{n \rightarrow \infty} \Phi_p^{p-n:p-1} = \Phi_p^{-\infty:p-1}$. \square

Z definicji 3.6 i twierdzenia 9.1 wynika, że stacjonarny proces gaussowski $X_{\mathbb{Z}}$ jest deterministyczny wtedy i tylko wtedy, gdy $\|X_p - \Phi_p^{-\infty:p-1}\| = 0$. Z twierdzenia 9.2 stacjonarny proces gaussowski $X_{\mathbb{Z}}$ jest deterministyczny wtedy i tylko wtedy, gdy

$$\lim_{n \rightarrow \infty} \|X_j - \Phi_j^{j-n:j-1}\| = 0. \quad (9.4)$$

Warunek 9.4 pozwala powiązać determinizm procesu gaussowskiego z wartością intensywności entropii.

Twierdzenie 9.3. *Stacjonarny proces gaussowski $X_{\mathbb{Z}}$ jest deterministyczny wtedy i tylko wtedy, gdy $h_X = -\infty$.*

Dowód: Zauważmy, że $\|X_j - \Phi_j^{j-n:j-1}\|^2 = \|X_0\|^2 v_n$ zaś

$$\sigma^2 := \lim_{n \rightarrow \infty} \|X_0\|^2 v_n = \|X_0\|^2 \exp(-2[H_X(1) - h_X]/d). \quad (9.5)$$

Mamy $\sigma^2 \geq 0$, przy czym $\sigma^2 = 0$ zachodzi wtedy i tylko wtedy, gdy $h_X = -\infty$. \square

Z nierówności (2.20) wynika, że gaussowskie procesy deterministyczne są niefinitarne.

Tradycyjnie wiele własności procesów gaussowskich było dotychczas ujmowane w języku funkcji korelacji a nie miar informacji. Wielkość σ^2 nazywana jest wariancją innowacji (Brockwell i Davis, 1987, *innovation variance*), z kolei analogon entropii nadwyżkowej

$$g := \lim_{n \rightarrow \infty} \det \Gamma(\{1, \dots, n\}) / \sigma^{2n} = \lim_{n \rightarrow \infty} \exp(2[H(X_1^n) - nh]/d) = \exp(2E_X/d) \quad (9.6)$$

nazywany jest wariancją uogólnioną (Finch, 1960; McLeod, 1998, *generalized variance*). Wariancja uogólniona g jest nieskończona wtedy i tylko wtedy, gdy nieskończona jest entropia nadwyżkowa E_X .

Twierdzenie 9.4. *Dla procesów gaussowskich, warunki determinizmu i niefinitarności są równoważne prostym asymptotycznym własnościom funkcji częściowej autokorelacji:*

$$H_X(1) - h_X < \infty \iff \forall_{m \in \mathbb{N}} |\alpha(m)| < 1 \wedge \sum_{k=1}^{\infty} |\alpha(k)|^2 < \infty, \quad (9.7)$$

$$E_X < \infty \iff \forall_{m \in \mathbb{N}} |\alpha(m)| < 1 \wedge \sum_{k=1}^{\infty} k |\alpha(k)|^2 < \infty. \quad (9.8)$$

Dowód: Kiedy $|\alpha(m)| = 1$ zachodzi dla pewnego m to

$$I(X_1; X_{m+1}|X_{2:m}) = -(d/2) \log [1 - |\alpha(m)|^2] = \infty,$$

więc $H_X(1) - h_X = \infty$ i $E_X = \infty$. Teraz rozpatrzmy przypadek, gdy $|\alpha(m)| < 1$ zachodzi dla wszystkich m . Z nierówności $x/2 \leq \log(1+x) \leq x$ dla $0 \leq x \leq 1$ wynika, że $H_X(1) - h_X = \sum_{k=1}^{\infty} \log 1/(1-|\alpha(k)|^2) < \infty$ jest równoważne $\sum_{k=1}^{\infty} |\alpha(k)|^2/(1-|\alpha(k)|^2) < \infty$, to zaś jest równoważne $\sum_{k=1}^{\infty} |\alpha(k)|^2 < \infty$. Tak dowiedliśmy (9.7). Dowód (9.8) jest analogiczny. \square

Warunek (9.7) jako warunek determinizmu procesu jest znany (Ramsey, 1974).

Warto tutaj wspomnieć, że twierdzenia, które formułujemy w niniejszej pracy dla procesów gaussowskich, można przetłumaczyć na lepiej znane twierdzenia dla procesów słabostacjonarnych. Procesem słabostacjonarnym $X_{\mathbb{Z}}$ nazywa się proces o wartościach zespolonych taki, że $\langle X_k \rangle$ nie zależy od k , zaś kowariancje $\text{Cov}(X_m, X_n)$ są określone dla dowolnych m i n oraz są funkcjami $n - m$. Aby przetłumaczyć twierdzenie o procesach gaussowskich na równoważne twierdzenie dla procesów słabostacjonarnych, należy dokonać następujących zastąpień:

- (i) pojęcie gaussowskości zastąpić przez pojęcie słabej stacjonarności,
- (ii) pojęcie niezależności zastąpić przez pojęcie nieskorelowania,
- (iii) gaussowskie funkcje mierzalne zastąpić przez najlepsze liniowe predyktory,
- (iv) determinizm i nefinitarność w sensie definicji 3.6 i 2.9 zastąpić przez warunki po prawych stronach (9.7) i (9.8).

Pod dokonaniu takich zastąpień prawdziwe pozostaje nawet twierdzenie 10.4, którego dowód w istotny sposób odwołuje się do założenia gaussowskości i twierdzenia 5.1.

Determinizm i nefinitarność są jednymi z wielu wyodrębnionych asymptotycznych własności procesów gaussowskich. Ponieważ wiele interesujących warunków asymptotycznych ma postać warunków sumowalności $\sum_{k=1}^{\infty} k^m |b_k|^n < \infty$ dla pewnych wyrazów b_k , warto poczynić prostą uwagę na temat zgrubnego porównywania mocy różnych warunków tej postaci.

Twierdzenie 9.5. Dla $r = -\frac{m+1}{n}$ zachodzą wynikania

$$\limsup_{k \rightarrow \infty} \frac{\log |b_k|}{\log k} < r \implies \sum_{k=1}^{\infty} k^m |b_k|^n < \infty \implies \liminf_{k \rightarrow \infty} \frac{\log |b_k|}{\log k} \leq r. \quad (9.9)$$

Niech ℓ^{Power} będzie zbiorem wszystkich ciągów $b = (b_k)_{k \in \mathbb{N}}$ takich, że $b_n \neq 0$ dla prawie wszystkich n oraz istnieje

$$\text{Power } b := \lim_{k \rightarrow \infty} \frac{\log |b_k|}{\log k}. \quad (9.10)$$

Jeżeli $b \in \ell^{\text{Power}}$ i $r = -\frac{m+1}{n}$, to $\text{Power } b > r$ implikuje $\sum_{k=1}^{\infty} k^m |b_k|^n = \infty$, zaś $\text{Power } b < r$ implikuje $\sum_{k=1}^{\infty} k^m |b_k|^n < \infty$.

Dla ciągów nie należących do ℓ^{Power} , porównywanie mocy warunków $\sum_{k=1}^{\infty} k^m |b_k|^n = \infty$ dla różnych n i m jest trudniejsze:

Twierdzenie 9.6. Dla dowolnego ciągu $(b_k)_{k \in \mathbb{N}}$ zachodzi

$$\sum_{k=1}^{\infty} k^m |b_k|^n < \infty \implies \sum_{k=1}^{\infty} |b_k| < \infty, \quad (9.11)$$

jeżeli $m \geq 0$ i $n \in (0, m+1)$.

Dowód: Oznaczmy $a_k := k^m |b_k|^n$. Suma $\sum_{k=1}^{\infty} a_k$ jest skończona, a zatem istnieje permutacja $f : \mathbb{N} \rightarrow \mathbb{N}$ taka, że $\tilde{a}_{k+1} \leq \tilde{a}_k$ dla wszystkich $\tilde{a}_k := a_{f(k)}$.¹ Mamy $|b_k| = \left(\frac{a_k}{k^m}\right)^{1/n}$.

Oznaczmy analogicznie $\tilde{b}_k := \left(\frac{\tilde{a}_k}{k^m}\right)^{1/n}$. Ponieważ funkcje $k \mapsto k^m$ i $x \mapsto x^{1/n}$ jest monotonicznie rosnąca, to $\sum_{k=1}^{\infty} |b_k| < \sum_{k=1}^{\infty} \tilde{b}_k$.² Dowiedzimy, że $\sum_{k=1}^{\infty} \tilde{b}_k < \infty$. Zauważmy, że ciąg posortowany spełnia $\tilde{a}_k < k^{-1}$ dla prawie wszystkich k , gdyż $\lim_{k \rightarrow \infty} k \tilde{a}_k = 0$ (Knopp, 1956, rozdział 14, pkt. 80). Zatem dla prawie wszystkich k zachodzi

$$\tilde{b}_k < \left(\frac{k^{-1}}{k^m}\right)^{1/n} = k^{-(m+1)/n},$$

czyli $\sum_{k=1}^{\infty} \tilde{b}_k < \infty$, gdyż $(m+1)/n > 1$. \square

Przykładowo mamy $\sum_{k=1}^{\infty} k^{1+\epsilon} |b_k|^2 < \infty \implies \sum_{k=1}^{\infty} |b_k| < \infty$ dla $\epsilon > 0$. Z drugiej strony, zwykle nie zachodzi ani $\sum_{k=1}^{\infty} k |b_k|^2 < \infty \implies \sum_{k=1}^{\infty} |b_k| < \infty$ ani $\sum_{k=1}^{\infty} |b_k| < \infty \implies \sum_{k=1}^{\infty} k |b_k|^2 < \infty$.

Wróćmy do procesów stochastycznych. Pomimo poczynionych w poprzednim akapicie zastrzeżeń, kryterium (9.8) niefinitarności procesu gaussowskiego przypomina znaną z literatury definicję zależności dalekiego zasięgu (*long memory, long-range dependence* — Granger i Joyeux, 1980; Beran, 1994). Rozważmy wariancję estymatora wartości oczekiwanej z próbki dla procesu stacjonarnego o zerowej wartości oczekiwanej,

$$\|\bar{X}_n\|^2 := \left\| \frac{1}{n} \sum_{i=1}^n X_i \right\|^2 = \frac{\|X_0\|^2}{n} \left[\sum_{k=-n+1}^{n-1} \left(1 - \frac{|k|}{n}\right) \rho(k) \right] \leq \|X_0\|^2, \quad (9.12)$$

gdzie ostatnia równość wynika z $|\rho(k)| \leq 1$.

Ograniczmy się do procesów takich, że $\rho(k) \geq 0$ dla wszystkich k . W tym przypadku istnieje granica $\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2$, gdyż $n \|\bar{X}_n\|^2$ rośnie monotonicznie z n dla $\rho(n) \geq 0$. Możemy mieć albo $\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2 \in (0, \infty)$ (jak dla białego szumu) lub $\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2 = \infty$. Niewątpliwie,

$$\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2 = \infty \iff \lim_{n \rightarrow \infty} \sum_{k=-n}^n \left(1 - \frac{|k|}{n}\right) \rho(k) = \infty. \quad (9.13)$$

Lewa strona (9.13) określa procesy o zależności dalekiego zasięgu (Beran, 1994).

Zwróćmy uwagę, że jeżeli warunek $\rho(k) \geq 0$ nie jest spełniony, granica $\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2$ nie musi być określona, a $\lim_{n \rightarrow \infty} n \|\bar{X}_n\|^2 = 0$ zachodzi na przykład dla $X_i = Z_i - Z_{i-1}$, gdzie $Z_{\mathbb{Z}}$ jest standardowym białym szumem.

¹ Aby wyjąć największy wyraz z nieskończonego ciągu $(a_k)_{k \in \mathbb{N}}$ liczb nieujemnych o znanej skończonej sumie $S = \sum_{k=1}^{\infty} a_k$, wystarczy rozpatrzyć n pierwszych wyrazów takich, że $S - \sum_{k=1}^n a_k < \max_{1 \leq k \leq n} a_k$. Wówczas $\max_{1 \leq k \leq n} a_k = \sup_{k \geq 1} a_k$. Liczba naturalna n o powyższej własności istnieje, gdyż ciąg sum częściowych $(\sum_{l=1}^k a_l)_{k \in \mathbb{N}}$ jest rosnący i zbieżny do S . Aby wyjąć P pierwszych największych wyrazów ciągu, procedurę wyjmowania największego wyrazu należy zastosować kolejno P razy do niewybranych jeszcze wyrazów. Pozostaje dowieść, że każdy wyraz dodatni a_K zostanie kiedyś wybrany. W rzeczy samej istnieje skończona liczba N taka, że $S - \sum_{k=1}^N a_k < a_K$. Oznacza to, że istnieje co najwyżej N wyrazów niemniejszych niż a_K i wyraz a_K zostanie wyjęty co najwyżej za N -tym razem.

² Mamy $(l_1/m_1 - l_2/m_2) - (l_1/m_2 - l_2/m_1) = (l_1 - l_2)(m_2 - m_1)/m_1 m_2 \geq 0$ wtedy i tylko wtedy, gdy $l_1 \geq l_2 \iff m_1 \leq m_2$. Sortowanie można wyrazić przez przestawianie par elementów.

Niech $\ell_+^{\text{Power}} = \{(b_n)_{n \in \mathbb{N}} \in \ell^{\text{Power}} : b_n > 0\}$. Zgodnie z (9.13), procesy z funkcją autokorelacji $\rho \in \ell_+^{\text{Power}}$ są procesami o zależności dalekiego zasięgu, jeżeli $\text{Power } \rho > -1$, zaś nie są, jeżeli $\text{Power } \rho < -1$. Zgodnie z (9.8), procesy z funkcją częściowej autokorelacji $\alpha \in \ell_+^{\text{Power}}$ są procesami niefinitarnymi, jeżeli $\text{Power } \alpha > -1$, zaś nie są, jeżeli $\text{Power } \alpha < -1$. Kryteria niefinitarności i zależności dalekiego zasięgu wydają się być prawie analogiczne.

Interesujące jest, w jakim zakresie asymptotyczne zachowania funkcji autokorelacji (obiektu zbliżonego do informacji $I(X_1; X_n)$) powiązane są z asymptotyką funkcji częściowej autokorelacji (obiektu zbliżonego do warunkowej informacji $I(X_1; X_n | X_{2:n-1})$). Jak wspomnieliśmy w rozdziale 2, zdecydowanie łatwiej jest badać empiryczne zależności w rozkładach dwuwymiarowych $X_{(1,n)}$ niż w rozkładach wielowymiarowych $X_{1:n}$. Zależność między funkcją autokorelacji a funkcją częściowej autokorelacji, pomimo swej nieliniowości, jest wzajemnie jednoznaczna. Pewne proste częściowe informacje o asymptotyce jednej z tych funkcji implikują pewne proste częściowe informacje o asymptotyce drugiej funkcji (Baxter, 1962; McLeod, 1998; Inoue, 2005). Katalog takich odpowiedniości wydaje się być bogaty. W następnych rozdziałach przedstawimy pewne nowe rezultaty wiążące zachowanie ACF i PACF, najpierw przypomnijmy jednak kilka klasycznych wyników.

Twierdzenie 9.7. (Herglotza) *Dla dowolnej funkcji autokowariancji γ istnieje dokładnie jedna miara F na $I = (-\pi, \pi]$ taka, że*

$$\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp(ik\omega) dF(\omega). \quad (9.14)$$

Podobnie, dla dowolnej skończonej miary F funkcja γ dana jako (9.14) jest funkcją autokowariancji pewnego procesu (zob. Brockwell i Davis, 1987, twierdzenie 4.3.1).

Miarę F nazywa się miarą spektralną. Dla miary Lebesgue'a m zgodnie z twierdzeniem A.22 istnieje jednoznaczny rozkład

$$F = F_{\perp} + F_{\ll}, \quad F_{\perp} \perp m, \quad F_{\ll} \ll m, \quad f := dF_{\ll}/dm \in L^1(I). \quad (9.15)$$

Jeżeli $F_{\perp} = 0$, to f nazywa się gęstością spektralną. Dowolne $f \in L^1(I)$, gdzie $f \geq 0$, jest gęstością spektralną pewnego procesu. Jeżeli dla dwóch gęstości spektralnych autokowariancje są identyczne, to gęstości te są równe prawie wszędzie. Dla $F_{\perp} = 0$ gęstość $f \in L^2(I)$ możemy wyrazić jako sumę szeregu Fouriera autokowariancji.

Twierdzenie 9.8. (Riesza-Fischera, Rudin, 1986, sekcja 4.26)

(i) *Jeżeli $f \in L^2(I)$, to współczynniki*

$$\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) \exp(ik\omega) d\omega \quad (9.16)$$

spełniają $\sum_{k=-\infty}^{\infty} |\gamma(k)|^2 < \infty$.

(ii) *Jeżeli $\sum_{k=-\infty}^{\infty} |\gamma(k)|^2 < \infty$, to funkcja*

$$f(\omega) = \sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\omega) \quad (9.17)$$

należy do $L^2(I)$

(iii) *Dla $f \in L^2(I)$ i $\gamma(k)$ danego przez (9.16) równość (9.17) zachodzi dla prawie wszystkich ω . Dla $\sum_{k=-\infty}^{\infty} |\gamma(k)|^2 < \infty$ i f danego przez (9.17) równość (9.16) zachodzi dla wszystkich n .*

Dla $\log f \in L^1(I)$, możemy z kolei przedstawić f za pomocą pewnego szeregu $\psi(z)$.

Twierdzenie 9.9. (Grenander i Szegö, 1958, sekcje 1.1, 1.13, 1.14)

(i) Niech pewna funkcja $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$ spełnia $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$. Wówczas szereg $\psi(z)$ jest funkcją analityczną dla $|z| < 1$, dla prawie wszystkich $\omega \in (-\pi, \pi]$ określone są sumy szeregów $\psi(e^{i\omega})$ i mamy $\psi(e^{i\omega}) = \lim_{r \rightarrow 1^-} \psi(re^{i\omega})$. Funkcja $f(\omega) := |\psi(e^{i\omega})|^2$ spełnia $f \in L^1(I)$ i zachodzą równości

$$\gamma(n) = \sum_{j=0}^{\infty} \psi_{j+n}^* \psi_j \quad \text{dla } n \geq 0, \quad (9.18)$$

gdzie γ dane jest przez (9.16). Jeżeli $\psi(0) > 0$, to $\log f \in L^1(I)$, zachodzi też

$$|\psi(0)|^2 = \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega \right], \quad (9.19)$$

oraz prawdziwe jest przedstawienie całkowe

$$\psi(z) = \exp \left[\frac{1}{2} h(z) \right], \quad h(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\omega} + z}{e^{i\omega} - z} \log f(\omega) d\omega. \quad (9.20)$$

(ii) Jeżeli pewna funkcja f spełnia $f(\omega) \geq 0$ dla prawie wszystkich ω , $f \in L^1(I)$ oraz $\log f \in L^1(I)$, to funkcja $\psi(z)$ określona dla $|z| < 1$ za pomocą (9.20) spełnia $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$, $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$, $\psi(0) > 0$ i $f(\omega) = |\psi(e^{i\omega})|^2$ dla prawie wszystkich ω .

(iii) Niech pewna funkcja $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$ spełnia $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$, zaś $f \in L^1(I)$ będzie dowolną funkcją. Jeżeli zachodzą równości (9.18) dla γ danego przez (9.16), to $f(\omega) = |\psi(e^{i\omega})|^2$ dla prawie wszystkich ω .

Powyższe twierdzenie jest niewielkim przeformułowaniem względem oryginału. Odwołaliśmy się m.in. do jawnej definicji zespolonej całki Poissona — we wzorze na funkcję $h(z)$ (Rudin, 1986, sekcje 11.5–11.10). Twierdzenie 9.9 jest twierdzeniem z analizy zespolonej, ale jak się okaże, ma ono interpretację stochastyczną. Przykładowo pkt. (iii) implikuje jednoznaczność rozkładu Wolda dla procesu czysto niedeterministycznego — por. silniejsze twierdzenie 10.3.

Twierdzenie 9.10. (formuła Kołmogorowa) Niech $f := dF_{\ll} / dm$. Wariancja innowacji (9.5) równa się

$$\sigma^2 = \begin{cases} \exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega \right], & \log f \in L^1(I), \\ 0, & \log f \notin L^1(I). \end{cases} \quad (9.21)$$

Grenander i Szegö (1958, sekcja 5.2) wyprowadzili wzór (9.21) dla $f \in [a, b] \subset (0, \infty)$ oraz $F_{\perp} = 0$. Z twierdzenia 10.3 pkt. (iii) proces $X_{\mathbb{Z}}$ jest niedeterministyczny wtedy i tylko wtedy, gdy f dane rozkładem (9.15) spełnia $\log f \in L^1(I)$. Z twierdzenia 11.3 wynika z kolei, że dla $\log f \in L^1(I)$ zachodzi $\exp \left[\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\omega) d\omega \right] = \psi(0) = \lim_{n \rightarrow \infty} \psi_{n0} = \lim_{n \rightarrow \infty} \|X_0\| \sqrt{v_n} = \sigma > 0$. Stąd wzór (9.21) jest prawdziwy także w przypadku ogólnym.

Twierdzenie 9.11. (Baxter, 1962, twierdzenie 2.3) Jeżeli $F_{\perp} = 0$, gęstość spektralna f jest ciągła i ściśle dodatnia oraz $\sum_{k=1}^{\infty} k^l |\rho(k)| < \infty$ dla pewnej liczby całkowitej $l \geq 0$, to $\sum_{k=1}^{\infty} k^l |\alpha(k)| < \infty$.

Warunek ściślejszej dodatniości gęstości spektralnej w twierdzeniu 9.11 jest konieczny. Przykładem procesów spełniających $\sum_{k=1}^{\infty} |\rho(k)| < \infty$ przy $\sum_{k=1}^{\infty} |\alpha(k)| = \infty$ są procesy ARIMA(0, d , 0) dla $d < 0$ (przykład 13.5).

Twierdzenie 9.12. (Grenander i Szegö, 1958, sekcja 5.5) Jeżeli $F_{\perp} = 0$, gęstość spektralna spełnia $f \in [a, b] \subset (0, \infty)$ oraz istnieje pochodna f' gęstości spektralnej i spełnia ona warunek Lipschitza $|f'(\omega_1) - f'(\omega_2)| < \text{const} |\omega_1 - \omega_2|^{\alpha}$ dla $0 < \alpha < 1$, to wariancja uogólniona (9.6) równa się

$$g = \exp \left[\frac{1}{\pi} \int_{|z| < 1} \left| \frac{d}{dz} \log \psi(z) \right|^2 d(\text{Re } z) d(\text{Im } z) \right], \quad (9.22)$$

gdzie funkcja $\psi(z)$ dana jest przez (9.20).

Definicja 9.13. (procesy dualne, McLeod, 1998) Procesy $X_{\mathbb{Z}}^A$ i $X_{\mathbb{Z}}^B$ nazywamy dualnymi, gdy istnieją ich gęstości spektralne $f^A, f^B \in L^1(I)$ i spełniają one $f^A f^B = 1$.

Jeżeli $f^A, f^B \in L^1(I)$, gdzie $f^A f^B = 1$, to $\log f^A, \log f^B \in L^1(I)$. Określone są zatem funkcje $\psi^A(z)$ i $\psi^B(z)$ dane odpowiednio przez (9.20) dla f^A i f^B . Jeżeli gęstość spektralna f^A spełnia warunek Lipschitza oraz warunek $f^A \in [a, b] \subset (0, \infty)$, to gęstość spektralna $X_{\mathbb{Z}}^B$ także spełnia warunek Lipschitza, a $X_{\mathbb{Z}}^A$ i $X_{\mathbb{Z}}^B$ mają równe wariancje uogólnione $g^A = g^B$ i entropie nadwyżkowe $E^A = E^B$ na mocy wzoru (9.22) (McLeod, 1998).

Rozdział 10

Przedstawienia kauzalne i odwracalne

Równość (9.18), gdzie $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$, sugeruje, że dla pewnego standardowego białego szumu gaussowskiego $Z_{\mathbb{Z}}$ możemy przedstawić proces gaussowski $X_{\mathbb{Z}}$ o zerowej wartości oczekiwanej jako szereg

$$X_n = \sum_{k=0}^{\infty} \psi_k Z_{n-k}, \quad (10.1)$$

gdzie ψ_k nie zależą od n . W istocie dla dowolnego niestacjonarnego procesu gaussowskiego $Y_{\mathbb{Z}}$ szereg $\sum_{k=0}^{\infty} \psi_k Y_{n-k}$ definiuje pewną zmienną gaussowską wtedy i tylko wtedy, gdy zbiega on w L^2 , tzn. $\lim_{N \rightarrow \infty} \sup_{M > N} \left\| \sum_{k=N}^M \psi_k Y_{n-k} \right\| = 0$. Dla pewnych ψ_k i $Y_{\mathbb{Z}}$ zbieżność sum $\sum_{k=0}^{\infty} \psi_k Y_{n-k}$ można udowodnić. W szczególności $\sum_{k=0}^{\infty} \psi_k Z_{n-k}$ jest określone, jeżeli $Z_{\mathbb{Z}}$ jest białym szumem, zaś $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$. W tym przypadku (10.1) nazywa się przedstawieniem kauzalnym procesu $X_{\mathbb{Z}}$ lub reprezentacją $MA(\infty)$.

Twierdzenie 10.1. *Dla stacjonarnego procesu gaussowskiego $Y_{\mathbb{Z}}$ suma $\sum_{k=0}^{\infty} \psi_k Y_{n-k}$ zbiega w L^2 , jeżeli $\sum_{k=0}^{\infty} |\psi_k| < \infty$ (por. Brockwell i Davis, 1987, stwierdzenie 3.1.1).*

Dowód: Mamy

$$\lim_{N \rightarrow \infty} \sup_{M > N} \left\| \sum_{k=N}^M \psi_k Y_{n-k} \right\| \leq \lim_{N \rightarrow \infty} \sup_{M > N} \sum_{k=N}^M |\psi_k| \|Y_{n-k}\| = \|Y_i\| \cdot \lim_{N \rightarrow \infty} \sup_{M > N} \sum_{k=N}^M |\psi_k| = 0.$$

□

Twierdzenie 10.2. *Dla białego szumu gaussowskiego $Z_{\mathbb{Z}}$, suma $\sum_{k=0}^{\infty} \psi_k Z_{n-k}$ zbiega w L^2 wtedy i tylko wtedy, gdy $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$.*

Dowód: Mamy $\lim_{N \rightarrow \infty} \sup_{M > N} \left\| \sum_{k=N}^M \psi_k Z_{n-k} \right\|^2 = \lim_{N \rightarrow \infty} \sup_{M > N} \sum_{k=N}^M |\psi_k|^2 = 0$ wtedy i tylko wtedy, gdy $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$. □

Twierdzenie 10.2 implikuje, że dla dowolnych liczb ψ_k spełniających $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$ funkcja $\gamma(n) = \sum_{j=0}^{\infty} \psi_{j+n}^* \psi_j$ dla $n \geq 0$ jest funkcją autokowariancji pewnego procesu.¹ Fakt ten wynika też z twierdzenia 9.9.

¹ Warto zauważyć, że standardowa procedura estymacji funkcji autokowariancji zapewniająca jej nieujemną określoność (Brockwell i Davis, 1987, sekcja 7.2) sprowadza się do położenia

$$\psi_k := \begin{cases} \frac{1}{\sqrt{N}} \left[x_k - \frac{1}{N} \sum_{j=1}^N x_j \right], & 1 \leq k \leq N, \\ 0, & \text{w przeciwnym wypadku,} \end{cases}$$

gdzie x_k , $1 \leq k \leq N$, są zaobserwowanymi wartościami procesu.

Zauważmy, że dla dowolnego niedeterministycznego procesu stacjonarnego $X_{\mathbb{Z}}$ można skonstruować standardowy biały szum $Z_{\mathbb{Z}}^{-\infty}$,

$$Z_p^{-\infty} = \frac{Y_p^{-\infty:p-1}}{\|Y_p^{-\infty:p-1}\|}, \quad (10.2)$$

w oparciu o innowacje $Y_p^{-\infty:p-1} = X_p - \Phi_p^{-\infty:p-1}$, gdyż $\|Y_p^{-\infty:p-1}\| \neq 0$. Wiadomo, że jeżeli szum $Z_{\mathbb{Z}}^{-\infty}$ istnieje, to zadaje on jedyne przedstawienie kauzalne procesu.

Twierdzenie 10.3. (rozkład Wolda) *Dla stacjonarnego gaussowskiego procesu niedeterministycznego $X_{\mathbb{Z}}$ istnieją współczynniki ψ_k oraz procesy stacjonarne $Z_{\mathbb{Z}}$ i $V_{\mathbb{Z}}$ takie, że $Z_{\mathbb{Z}} \perp V_{\mathbb{Z}}$, $Z_{\mathbb{Z}}$ jest standardowym białym szumem gaussowskim, $V_{\mathbb{Z}}$ jest deterministycznym procesem gaussowskim lub spełnia $V_n = 0$, zaś*

$$X_n = \sum_{k=0}^{\infty} \psi_k Z_{n-k} + V_n, \quad \psi_0 > 0. \quad (10.3)$$

Przedstawienie to jest jednoznaczne i jest tożsame z rozkładem Lebesgue'a (9.15):

- (i) Dla dowolnego procesu gaussowskiego $X_{\mathbb{Z}}$ wzór (10.3) zachodzi wtedy i tylko wtedy, gdy $Z_n = Z_n^{-\infty}$ zaś $\psi_k = \text{Cov}(Z_{n-k}^{-\infty}, X_n)$ (Brockwell i Davis, 1987, twierdzenie 5.7.1).
- (ii) Dla dowolnego gaussowskiego procesu niedeterministycznego $X_{\mathbb{Z}}$ o mierze spektralnej F i rozkładzie (9.15) miarą spektralną procesu $Z_{\mathbb{Z}}$ jest F_{\ll} , a miarą spektralną procesu $V_{\mathbb{Z}}$ jest F_{\perp} (Brockwell i Davis, 1987, twierdzenie 5.7.2).
- (iii) Dowolny proces gaussowski $X_{\mathbb{Z}}$ jest niedeterministyczny wtedy i tylko wtedy, gdy f dane rozkładem (9.15) spełnia $\log f \in L^1(I)$.

Pierwszą część tego twierdzenia udowodnił Wold (1938, sekcja 20). Doob (1953, str. 499) pokazał tożsamość rozkładu Wolda z rozkładem (9.15). Pkt. (iii) wynika z pkt. (i) i (ii), jednoznaczności przedstawienia spektralnego (twierdzenie 9.7) oraz twierdzenia 9.9.

Procesy niedeterministyczne $X_{\mathbb{Z}}$, dla których $V_{\mathbb{Z}}$ z rozkładu Wolda spełnia $V_n = 0$ (równoważnie $F_{\perp} = 0$), nazywa się czysto niedeterministycznymi (*purely nondeterministic*). Klasa gaussowskich procesów czysto niedeterministycznych pokrywa się z klasą gaussowskich procesów jednostronnie kwaziliniowych (zob. definicja 4.4).

Twierdzenie 10.4. *Każdy gaussowski proces finitarny jest czysto niedeterministyczny.*

Dowód: Dowolny gaussowski proces deterministyczny jest nefinitarny z nierówności (2.20) i twierdzenia 9.3, o czym wspominaliśmy.

Dowolny gaussowski proces niedeterministyczny $X_{\mathbb{Z}}$, który nie jest czysto niedeterministyczny, można przedstawić jako $X_n = \sum_{k=0}^{\infty} \psi_k Z_{n-k} + V_n$, gdzie $V_n \neq 0$. Proces $X'_{\mathbb{Z}}$ określony jako $X'_n := \sum_{k=0}^{\infty} \psi_k Z_{n-k}$ jest czysto niedeterministyczny.

Niech $F_{X'}$, $F_{\bar{X}'}$, F_V oraz $F_{\bar{V}}$ będą miarami spektralnymi procesów $X'_{\mathbb{Z}}$, $\bar{X}'_{\mathbb{Z}}$, $V_{\mathbb{Z}}$ i $\bar{V}_{\mathbb{Z}}$, gdzie $\bar{X}'_n := X'_{-n}$, $\bar{V}_n := \bar{V}_{-n}$. Mamy $F_{\bar{X}'} = F_{X'} \circ \theta$ oraz $F_{\bar{V}} = F_V \circ \theta$, gdzie $\theta(x) = -x$. Zatem $\bar{X}'_{\mathbb{Z}}$ jest czysto niedeterministyczny, a $\bar{V}_{\mathbb{Z}}$ jest deterministyczny i spełniają one $X_{\mathbb{Z}} \perp V_{\mathbb{Z}}$. Mamy więc rozkład $X'_n = \bar{X}'_{-n} = \sum_{k=0}^{\infty} \psi_k^* \bar{Z}_{-n-k}$ dla pewnego białego szumu $\bar{Z}_{\mathbb{Z}}$, a także $X_n = \sum_{k=0}^{\infty} \psi_k^* \bar{Z}_{-n-k} + V_n$. Oznaczmy $\sigma := Y_n^{n+1:\infty} = Y_n^{-\infty:n-1}$. Z jednoznaczności dwóch rozkładów Wolda dla $X_{\mathbb{Z}}$ mamy

$$V_n = X_n - \sum_{k=0}^{\infty} \psi_k^* Y_{n+k}^{n+k+1:\infty} / \sigma,$$

$$V_p = X_p - \sum_{k=0}^{\infty} \psi_k Y_{p-k}^{-\infty:p-k-1} / \sigma, \quad p < n.$$

gdź $Y_{n+k}^{n+k+1:\infty}/\sigma = \bar{Z}_{-n-k}$ oraz $Y_{p-k}^{-\infty:p-k-1}/\sigma = Z_{n-k}$.

Oznaczmy $\mathcal{G} = \sigma(\mathcal{R}_0^{\mathbb{N}})$. Granice ciągów funkcji mierzalnych \mathcal{G}/\mathcal{R} są mierzalne \mathcal{G}/\mathcal{R} . Stąd z twierdzenia 9.2 dla predyktorów procesów $X_{\mathbb{Z}}$ i $V_{\mathbb{Z}}$ wynika, że istnieją funkcje f_+ , f_- i g_- mierzalne \mathcal{G}/\mathcal{R} takie, że $V_n = f_+(X_{n:\infty})$, $V_p = f_-(X_{-\infty:p})$, $V_n = g_-(V_{-\infty:n-1})$. Ze złożenia funkcji f_- i g_- otrzymujemy funkcję h_- mierzalną \mathcal{G}/\mathcal{R} taką, że $V_n = h_-(X_{-\infty:n-1}) := g_-(f_-(X_{-\infty:p})_{p \leq n-1})$. Zmienna $V_n = h_-(X_{-\infty:n-1}) = f_+(X_{n:\infty})$ jest ciągłą zmienną losową o rozkładzie normalnym, a więc proces $X_{\mathbb{Z}}$ jest nefinitarny z twierdzenia 5.1. \square

Proces gaussowski $X_{\mathbb{Z}}$ taki, że nie istnieje proces gaussowski $(Y, X_{\mathbb{Z}})$, gdzie $Y = h(X_{-\infty:n-1}) = f(X_{n:\infty})$ dla pewnych funkcji mierzalnych h i f , nazywa się zupełnie niedeterministycznym (Bloomfield *et al.*, 1983; Inoue i Kasahara, 2005; Inoue, 2000, *completely nondeterministic*). W istocie dowód twierdzenia 10.4 można rozbić na dowody dwóch twierdzeń:

- (i) Każdy gaussowski proces finitarny jest zupełnie niedeterministyczny.
- (ii) Każdy proces zupełnie niedeterministyczny jest czysto niedeterministyczny.

Pkt. (ii) udowodnili Helson i Sarason (1967).

Interesującą kwestią są związki pomiędzy istnieniem przedstawienia kauzalnego (zerowaniem się składowej deterministycznej) a ergodycznością procesu. Gaussowskie procesy czysto niedeterministyczne są procesami kwaziliniowymi, a zatem są ergodyczne na mocy twierdzenia 4.3. Znane jest stwierdzenie mocniejsze:

Twierdzenie 10.5. *Proces gaussowski jest ergodyczny wtedy i tylko wtedy, gdy jego miara spektralna jest bezatomowa (Kornfeld *et al.*, 1982, wniosek po lemacie 8.§2.2, twierdzenie 14.§2.1).*

Podkreślmy, że bezatomowość miary spektralnej F nie jest równoważna jej absolutnej ciągłości względem miary Lebesgue'a m . Jeżeli $F \ll m$ (tzn. jeżeli proces jest czysto niedeterministyczny), to miara F jest bezatomowa (czyli proces jest ergodyczny w myśl twierdzenia 10.5). Stwierdzenie odwrotne nie jest jednak prawdziwe, gdyż istnieją bezatomowe miary osobliwe względem miary Lebesgue'a.² Warto także zauważyć, że dla procesu gaussowskiego absolutna ciągłość miary spektralnej względem miary Lebesgue'a jest równoważna jego regularności (Pourahmadi, 2001, sekcja 5.6.4 i rozdział 8). Jednokierunkowe implikowanie bezatomowości F przez relację $F \ll m$ jest zatem pewną reminiscencją twierdzenia 4.7.

Przy okazji omawiania związków czystego niedeterminizmu i ergodyczności, wspomnijmy, że proces gaussowski jest mieszający wtedy i tylko wtedy, gdy $\lim_{n \rightarrow \infty} |\rho(n)| = 0$ (Kornfeld *et al.*, 1982, definicja 1.§6.2, twierdzenie 14.§2.2). Własność mieszania implikuje ergodyczność (Kornfeld *et al.*, 1982, sekcja 1.§6). Z kolei z twierdzenia 11.5, czysty niedeterminizm (regularność) implikuje własność mieszania. Oba wynikania zostały zwizualizowane na zbiorczym diagramie w rozdziale 13.

Wróćmy do bardziej elementarnych rozważań. Z twierdzenia 9.2 predyktory dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$ spełniają równość

$$\Phi_n^{-\infty:n-1} = \Phi_n^{n-1:n-1} + \sum_{k=2}^{\infty} (\Phi_n^{n-k:n-1} - \Phi_n^{n-k-1:n-1}), \quad (10.4)$$

² Oznaczmy $B = \{b_{\mathbb{N}} : b_i \in \{0, 1\}, \limsup_{n \rightarrow \infty} b_n > 0\}$. Funkcje $\phi_C(b_{\mathbb{N}}) = \sum_{i=1}^{\infty} 2b_i 3^{-i}$ oraz $\phi(b_{\mathbb{N}}) = \sum_{i=1}^{\infty} b_i 2^{-i}$ są bijekcjami pomiędzy zbiorem B a zbiorem Cantora $C = \phi_C(B)$ oraz odcinkiem $(0, 1] = \phi(B)$. Dla miary Lebesgue'a m mamy też $m(C) = 0$. Definiując $m_C = m \circ \phi \circ \phi_C^{-1}$ otrzymujemy bezatomową miarę m_C , która jest skupiona na zbiorze Cantora C , a zatem spełnia $m_C \perp m$.

gdzie $\Phi_n^{n-k:n-1} = \sum_{j=1}^k \phi_{kj} X_{n-j}$. Gdybyśmy mogli przedstawiać zmienne X_i w szeregu (10.4) bez zmiany jego wartości, to otrzymalibyśmy

$$\Phi_n^{-\infty:n-1} = \sum_{j=1}^{\infty} \phi_j X_{n-j}, \quad (10.5)$$

gdzie $\phi_j := \phi_{1j} + \sum_{k=2}^{\infty} (\phi_{kj} - \phi_{k-1,j}) = \lim_{k \rightarrow \infty} \phi_{kj}$, a $\phi_{kl} = 0$ dla $l > k$. Pomimo tego, że $\Phi_n^{-\infty:n-1}$ jest mierzalną funkcją $X_{-\infty:n-1}$ i, jak pokażemy dalej, zawsze zachodzi zbieżność $\lim_{k \rightarrow \infty} \phi_{kj} = \phi_j$, to wzór (10.5) nie musi być prawdziwy. Równość (10.5) może nie zachodzić z dwóch powodów. Po pierwsze, szereg $\sum_{j=1}^{\infty} \phi_j X_{n-j}$ może nie być zbieżny. Po drugie, suma $\sum_{j=1}^{\infty} \phi_j X_{n-j}$ nawet, gdy jest określona, może się różnić od $\lim_{k \rightarrow \infty} \sum_{j=1}^k \phi_{kj} X_{n-j}$.

Dla procesów czysto niedeterministycznych równość (10.5) jest równoważna równości

$$Z_n = \sum_{k=0}^{\infty} \pi_k X_{n-k}, \quad (10.6)$$

gdzie $\pi_0 := 1/\sigma$, $\pi_k := -\phi_k/\sigma$ dla $k > 0$, zaś $Z_{\mathbb{Z}}$ jest białym szumem gaussowskim z przedstawienia kauzalnego (10.1). W hipotetycznym przypadku ogólnego procesu $X_{\mathbb{Z}}$ i ogólnego standardowego białego szumu gaussowskiego $Z_{\mathbb{Z}}$, wzór (10.6) nazywa się przedstawieniem odwracalnym lub reprezentacją AR(∞). Z kolei mówi się, że proces $X_{\mathbb{Z}}$ ma własność ARR, jeżeli tylko zachodzi zbieżność szeregu $\sum_{k=0}^{\infty} \pi_k X_{n-k}$ (Pourahmadi, 2001, sekcja 6.2.1). Pourahmadi (2001, twierdzenie 6.5) dowodzi także, że każdy niedeterministyczny proces $X_{\mathbb{Z}}$, mający własność ARR, ma przedstawienie dualne do rozkładu Wolda

$$Z_n = \sum_{k=0}^{\infty} \pi_k X_{n-k} + \tilde{V}_n, \quad (10.7)$$

gdzie $Z_{\mathbb{Z}}$ jest procesem ze wzoru (10.3), natomiast $\tilde{V}_n = 0$, jeżeli $V_n = 0$. (Ściśle, należy jeszcze dowieść, że współczynniki π_k w definicjach naszej i Pourahmadiego są sobie równe.)

Przedstawienia kauzalne i odwracalne są ze sobą dość mocno powiązane.

Twierdzenie 10.6. Niech $Z_{\mathbb{Z}}$ będzie białym szumem. Niech w L^2 zbiegają szeregi $X_n = \sum_{k=0}^{\infty} \psi_k Z_{n-k}$ i $\tilde{Z}_n = \sum_{k=0}^{\infty} \pi_k X_{n-k}$. Wówczas współczynniki ψ_k i π_k spełniają równość

$$\sum_{j=0}^k \pi_j \psi_{k-j} = \mathbb{I}[k=0] \quad \text{dla każdego } k \geq 0 \quad (10.8)$$

wtedy i tylko wtedy, gdy $\tilde{Z}_n = Z_n$ dla każdego n .

Dowód: Szereg $\tilde{Z}_n = \sum_{k=0}^{\infty} \pi_k \sum_{k'=0}^{\infty} \psi_{k'} Z_{n-k-k'}$ jest elementem podprzestrzeni Hilberta rozpinanej przez $\{Z_m\}_{m \leq n}$. Możemy go przedstawić jako $\tilde{Z}_n = S_N + R_N$, gdzie

$$S_N := \sum_{k=0}^N \left[\sum_{j=0}^k \pi_j \psi_{k-j} \right] Z_{n-k}, \quad R_N := \sum_{k=0}^{\infty} \pi_k \sum_{k'=\max(0, N+1-k)}^{\infty} \psi_{k'} Z_{n-k-k'}. \quad (10.9)$$

S_N jest elementem podprzestrzeni Hilberta rozpinanej przez $\{Z_m\}_{n-N \leq m \leq n}$ zaś R_N jest elementem podprzestrzeni Hilberta rozpinanej przez $\{Z_m\}_{m \leq n-N-1}$. Jeżeli $\tilde{Z}_n = Z_n$, to $\tilde{Z}_n = S_N$ dla każdego N , a zatem zachodzi (10.8). Odwrotnie, jeżeli zachodzi (10.8), to $S_N = Z_n$, czyli $\tilde{Z}_n = Z_n + R_N$ dla każdego N . To znaczy, że $\tilde{Z}_n - Z_n$ jest elementem przecięcia podprzestrzeni Hilberta rozpinanych przez $\{Z_m\}_{n-N \leq m \leq n}$, gdzie $N \in \mathbb{N}$. Jedynym elementem tego przecięcia jest jednak 0. \square

Równość (10.8) wyznacza współczynniki π_k dla danych współczynników ψ_k i odwrotnie. Jest ona koniecznym ale niedostatecznym warunkiem jednoczesnego zachodzenia wzorów (10.1) i (10.6). Wykazanie istnienia przedstawienia kauzalnego (zerowania się składowej deterministycznej V_n) dla danego przedstawienia odwracalnego wymaga dodatkowych założeń. Pomocne są założenia o sumowalności współczynników (Brockwell i Davis, 1987, sekcja 3.1):

Definicja 10.7. (proces kauzalny i odwracalny) *Mówimy, że proces $X_{\mathbb{Z}}$ jest kauzalny (causal) względem białego szumu $Z_{\mathbb{Z}}$, jeżeli zachodzi równość (10.1) oraz $\sum_{k=0}^{\infty} |\psi_k| < \infty$. Mówimy, że proces $X_{\mathbb{Z}}$ jest odwracalny (invertible) względem białego szumu $Z_{\mathbb{Z}}$, jeżeli zachodzi równość (10.6) oraz $\sum_{k=0}^{\infty} |\pi_k| < \infty$.*

Twierdzenie 10.8. *Załóżmy, że współczynniki ψ_k i π_k spełniają warunki $\sum_{k=0}^{\infty} |\psi_k| < \infty$, $\sum_{k=0}^{\infty} |\pi_k| < \infty$ oraz (10.8). Dla dowolnego stacjonarnego procesu gaussowskiego $X_{\mathbb{Z}}$ i szeregów $Y_n := \sum_{k=0}^{\infty} \pi_k X_{n-k}$ oraz $\tilde{X}_n := \sum_{k=0}^{\infty} \psi_k Y_{n-k}$ zachodzi $\tilde{X}_n = X_n$.*

Dowód: Procesy gaussowskie $Y_{\mathbb{Z}}$ i $\tilde{X}_{\mathbb{Z}}$ są dobrze zdefiniowane na mocy twierdzenia 10.1. Weźmy dowolne $M > 0$. Równania (10.8) implikują tożsamość

$$X_n = \sum_{k=0}^{2M} \sum_{k'=0}^{2M} \psi_k \pi_{k'} X_{n-k-k'} \mathbb{I}[k+k' \leq 2M]. \quad (10.10)$$

Stąd dla $N > 2M$ zachodzi

$$\begin{aligned} X_n - \sum_{k=0}^M \psi_k \sum_{k'=0}^N \pi_{k'} X_{n-k-k'} &= - \sum_{k'=M+1}^N \sum_{k=0}^M \psi_k \pi_{k'} X_{n-k-k'} \mathbb{I}[k+k' > 2M] \\ &\quad + \sum_{k'=0}^M \sum_{k=M+1}^{2M} \psi_k \pi_{k'} X_{n-k-k'} \mathbb{I}[k+k' \leq 2M]. \end{aligned} \quad (10.11)$$

Zatem

$$\begin{aligned} \left\| X_n - \sum_{k=0}^M \psi_k \sum_{k'=0}^N \pi_{k'} X_{n-k-k'} \right\|^2 &\leq \left| \sum_{k'=M+1}^N \sum_{k=0}^M |\psi_k| |\pi_{k'}| + \sum_{k'=0}^M \sum_{k=M+1}^{2M} |\psi_k| |\pi_{k'}| \right|^2 \|X_n\|^2 \\ &\leq \left| \sum_{k'=0}^{\infty} |\pi_{k'}| \sum_{k=0}^{\infty} |\psi_k| - \sum_{k'=0}^M |\pi_{k'}| \sum_{k=0}^M |\psi_k| \right|^2 \|X_n\|^2. \end{aligned}$$

W rezultacie otrzymujemy, że $\lim_{M \rightarrow \infty} \sup_{N > 2M} \left\| X_n - \sum_{k=0}^M \psi_k \sum_{k'=0}^N \pi_{k'} X_{n-k-k'} \right\| = 0$.

Ponieważ $\left\| \sum_{k=0}^M \psi_k \left(Y_{n-k} - \sum_{k'=0}^N \pi_{k'} X_{n-k-k'} \right) \right\| \leq \sum_{k=0}^M |\psi_k| \left\| Y_n - \sum_{k'=0}^N \pi_{k'} X_{n-k'} \right\|$, to

$$\begin{aligned} \|X_n - \tilde{X}_n\| &\leq \left\| X_n - \sum_{k=0}^M \psi_k \sum_{k'=0}^N \pi_{k'} X_{n-k-k'} \right\| + \left\| \tilde{X}_n - \sum_{k=0}^M \psi_k Y_{n-k} \right\| \\ &\quad + \sum_{k=0}^M |\psi_k| \left\| Y_n - \sum_{k'=0}^N \pi_{k'} X_{n-k'} \right\|. \end{aligned}$$

Stosując $\lim_{M \rightarrow \infty} \sup_{N > 2M}$ do prawej strony, otrzymujemy $\|X_n - \tilde{X}_n\| = 0$. \square

Do powiązania asymptotyki współczynników ψ_k i π_k pomocna jest analiza zespolona. Oznaczmy $\psi(z) := \sum_{k=0}^{\infty} \psi_k z^k$ i $\pi(z) := \sum_{k=0}^{\infty} \pi_k z^k$.

Twierdzenie 10.9. *Niech szeregi $\psi(z)$ i $\pi(z)$ zbiegają bezwzględnie dla $|z| < r$, gdzie $r > 0$. Wówczas (10.8) zachodzi wtedy i tylko wtedy, gdy $\pi(z)\psi(z) = 1$ dla $|z| < r$.*

Dowód: Dla $|z| < r$ iloczyn Cauchy'ego $\lambda(z) = \sum_{k=0}^{\infty} \lambda_k z^k$, $\lambda_k = \sum_{j=0}^k \pi_j \psi_{k-j}$, jest zbieżny bezwzględnie, gdyż $\sum_{k=0}^N |\lambda_k| |z|^k \leq \sum_{j=0}^N |\pi_j| |z|^j \sum_{j'=0}^N |\psi_{j'}| |z|^{j'}$. Z twierdzenia Abela o mnożeniu szeregów, $\pi(z)\psi(z) = \lambda(z)$ dla $|z| < r$. Stąd wynika teza, gdyż jedynym szeregiem $\lambda(z)$ spełniającym $\lambda(z) = 1$ dla $|z| < r$ jest szereg o $\lambda_k = \llbracket k = 0 \rrbracket$. \square

Twierdzenie 10.10. *Niech $\pi(z)$ zbiega dla $|z| \leq 1$ i niech $\pi(z) \neq 0$ dla $|z| \leq 1$. Niech ψ_k będą współczynnikami danymi przez (10.8).*

(i) *Niech szereg $\pi(z)$ zbiega bezwzględnie dla $|z| < r$, gdzie $r > 1$. (Jest to warunek mocniejszy niż $\sum_{k=0}^{\infty} |\pi_k| < \infty$.) Wówczas istnieje $r'' \in (1, r)$ takie, że $\pi(z) \neq 0$ dla wszystkich $|z| \leq r''$. Zachodzi też $\sum_{k=0}^{\infty} |\psi_k| < \infty$, a $\psi(z)$ zbiega bezwzględnie do $\psi(z) = 1/\pi(z)$ dla $|z| < r''$.*

(ii) *Niech szereg $\pi(z)$ zbiega do funkcji ciągłej dla $|z| \leq 1$. (Jest to warunek mocniejszy niż $\sum_{k=0}^{\infty} |\pi_k|^2 < \infty$.) Wówczas zachodzi $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$, a $\psi(z)$ zbiega do funkcji ciągłej $\psi(z) = 1/\pi(z)$ dla $|z| < 1$.*

Dowód:

(i) Funkcja $|\pi(z)|$ jest jednostajnie ciągła w kole $|z| \leq r'$ dla każdego $r' \in (0, r)$. Przyjmuje ona swoje infimum na kole jednostkowym, a więc $|\pi(z)| > a$ dla każdego $|z| \leq 1$ i pewnego $a > 0$. Na mocy jednostajnej ciągłości $|\pi(z)|$ i zwartości okręgu, istnieją takie $a > 0$ i $r'' \in (1, r)$, że $|\pi(z)| > a$ dla każdego $|z| < r''$. A zatem funkcja $1/\pi(z)$ jest analityczna dla $|z| < r''$. Na mocy twierdzenia 10.9 mamy wówczas $1/\pi(z) = \psi(z)$. Szereg $\psi(z)$ jest bezwzględnie zbieżny dla $|z| < r''$, a zatem $\sum_{k=0}^{\infty} |\psi_k| < \infty$.

(ii) Ponieważ szereg $\pi(z)$ zbiega dla $|z| \leq 1$, to zbiega on bezwzględnie dla $|z| < 1$. A zatem $\psi(z)$ zbiega bezwzględnie dla $|z| < 1$ i mamy wówczas $\psi(z) = 1/\pi(z)$. Z ciągłości funkcji $\pi(z)$ wynika, że istnieją stałe $a, b > 0$ takie, że dla $|z| \leq 1$ mamy $|\pi(z)| \in (a, b)$ i $|\psi(z)| \in (b, a)$.

Dla funkcji $u(z) = \sum_{k=0}^{\infty} u_k z^k$ analitycznej dla $|z| < 1$ warunek $\sum_{k=0}^{\infty} |u_k|^2 < \infty$ zachodzi wtedy i tylko wtedy, gdy istnieje stała C taka, że $\int_{-\pi}^{\pi} |u(re^{i\omega})|^2 d\omega < C$ dla każdego $r < 1$ (Grenander i Szegö, 1958, sekcja 1.1). Z kolei z twierdzenia 9.9, jeżeli $\sum_{k=0}^{\infty} |u_k|^2 < \infty$, to $\lim_{r \rightarrow 1^-} u(re^{i\omega}) = u(e^{i\omega})$ dla prawie wszystkich ω . W świetle tych wyników oraz faktu, że $|\pi(z)|$ jest ograniczone, zachodzi $\sum_{k=0}^{\infty} |\pi_k|^2 < \infty$, $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$ i $\psi(e^{i\omega}) = \lim_{r \rightarrow 1^-} \psi(re^{i\omega}) = \lim_{r \rightarrow 1^-} 1/\pi(re^{i\omega}) = 1/\pi(e^{i\omega})$, czyli $\psi(z)$ jest ciągła także dla $|z| \leq 1$. \square

Rozdział 11

Sumy szeregów funkcji autokorelacji

W tym rozdziale wyprowadzimy wzór wyrażający przy pewnych warunkach sumy autokorelacji $\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k)$ jako iloczyny funkcji wymiernych częściowych autokorelacji (twierdzenie 11.17). Punktem wyjścia jest następujący wniosek z wcześniejszych rozważań.

Twierdzenie 11.1. *Niech $X_{\mathbb{Z}}$ będzie stacjonarnym procesem gaussowskim. Następujące stwierdzenia są równoważne:*

- (i) $X_l = \sum_{k=0}^{\infty} \psi_k Z_{l-k}^{-\infty}$, tzn. $X_{\mathbb{Z}}$ jest czysto niedeterministyczny;
- (ii) $\gamma(k) = \sum_{j=0}^{\infty} \psi_{j+k}^* \psi_j$ dla $k \geq 0$;
- (iii) $\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^2 \exp(ik\omega) d\omega$, tzn. $F_{\perp} = 0$ zaś gęstość spektralna wynosi $f(\omega) = |\psi(e^{i\omega})|^2$.

Dowód: Jeżeli $X_l = \sum_{k=0}^{\infty} \psi_k Z_{l-k}^{-\infty}$, to zachodzi pkt. (ii) (zob. Brockwell i Davis, 1987, stwierdzenie 3.1.2). Z twierdzenia 9.9 pkt. (i) i jednoznaczności rozkładu spektralnego (twierdzenie 9.7) pkt. (ii) dowodzonego twierdzenia implikuje pkt. (iii). Jeżeli $X_l = \sum_{k=0}^{\infty} \psi_k Z_{l-k}^{-\infty} + V_l$, gdzie $\|V_l\| > 0$, to z rozkładu Wolda procesy $(Z_i^{-\infty})_{i \in \mathbb{Z}}$ i $(V_i)_{i \in \mathbb{Z}}$ są nieskorelowane i mamy $\frac{1}{2\pi} F([-\pi, \pi]) = \|X_l\|^2 > \sum_{k=0}^{\infty} |\psi_k|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^2 d\omega$. Oznacza to, że pkt. (iii) implikuje pkt. (i). \square

Twierdzenie 11.2. *Niech $X_{\mathbb{Z}}$ będzie procesem czysto niedeterministycznym.*

- (i) Zachodzi nierówność $\sum_{k=-\infty}^{\infty} |\gamma(k)| \leq (\sum_{k=0}^{\infty} |\psi_k|)^2$.
- (ii) Zachodzi równość $\sum_{k=-\infty}^{\infty} |\gamma(k)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^4 d\omega$.
- (iii) Jeżeli $\frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^4 d\omega < \infty$, to $\sum_{k=-\infty}^{\infty} \gamma(k) \exp(-ik\omega) = |\psi(e^{i\omega})|^2$.

Dowód: Połóżmy $\psi_j := 0$ dla $j < 0$. Mamy $|\gamma(n)| \leq \sum_{j=-\infty}^{\infty} |\psi_{j+n}| |\psi_j|$ oraz

$$\sum_{k=-\infty}^{\infty} |\gamma(k)| \leq \sum_{k=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\psi_{j+k}| |\psi_j| \leq \left(\sum_{k=0}^{\infty} |\psi_k| \right)^2.$$

Pkt. (ii)–(iii) wynikają z twierdzenia 9.8 (Riesza-Fischera) i tożsamości Parsewała (Rudin, 1986, sekcja 4.26) dla współczynników Fouriera $\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^2 \exp(ik\omega) d\omega$. \square

Przedstawienia kauzalne i odwracalne wydają się być potencjalnie silnym narzędziem do powiązania niektórych własności asymptotycznych autokorelacji i częściowej autokorelacji. Przedstawienie kauzalne jest bliższe algebraicznie funkcji autokorelacji, natomiast przedstawienie odwracalne okazuje się być bliższe algebraicznie częściowej autokorelacji. Pokażemy, że dla pewnych funkcji częściowej autokorelacji przedstawienia te można skonstruować z przedstawień skończonych (por. Pourahmadi, 2001, sekcja 7.6).

Zauważmy, że proces $Z_{\{n+1, n+2, \dots\}}^n$ zdefiniowany przez

$$Z_p^n := \frac{Y_p^{n+1:p-1}}{\|Y_p^{n+1:p-1}\|} \quad (11.1)$$

jest standardowym białym szumem. Spełnia on zależności

$$X_p = \sum_{k=0}^n \psi_{nk} Z_{p-k}^{p-n-1}, \quad (11.2)$$

$$Z_p^{p-n-1} = \sum_{k=0}^n \pi_{nk} X_{p-k}, \quad (11.3)$$

gdzie współczynniki ψ_{nk} i π_{nk} dane są wzorami

$$\pi_{nk} := \frac{1}{\|X_0\| \sqrt{v_n}} \cdot \begin{cases} 1, & k = 0, \\ -\phi_{nk}, & k \in \{1, \dots, n\}, \end{cases} \quad (11.4)$$

$$\psi_{nk} := \begin{cases} \|X_0\| \sqrt{v_n}, & k = 0, \\ \sum_{j=1}^k \phi_{nj} \psi_{n-j, k-j}, & k \in \{1, \dots, n\}. \end{cases} \quad (11.5)$$

Współczynniki ψ_{nk} rozkładu X_{p+n} w bazie $(Z_n^p)_{n \in \mathbb{N}}$ są bliskie algebraicznie lepiej znanym współczynnikom θ_{nk} rozkładu X_{p+n} w bazie $(Y_n^{p+1:n-1})_{n \in \mathbb{N}}$ (Brockwell i Davis, 1987, stwierdzenie 5.2.2, algorytm innowacji). Ponieważ $\gamma(n-1) = \text{Cov}(X_n; X_1) = \psi_{n-1, n-1}^* \psi_{00}$, równanie (11.5) dla $k = n$ jest równoważne (8.15) dla $k = n$.

Współczynniki ψ_{nk} są silnie ograniczone. Z równości $\|X_{p+n+1}\|^2 = \|X_0\|^2$ wynika, że

$$\sum_{k=0}^n |\psi_{nk}|^2 = \|X_0\|^2. \quad (11.6)$$

Ponieważ $\gamma(n) = \psi_{nn}^* \psi_{00}$ oraz $\psi_{n0} = \|X_0\| \sqrt{v_n}$, z (11.6) wynika prosta nierówność

$$v_n + |\rho(n)|^2 \leq 1 \quad (11.7)$$

dla $n \geq 1$. Korzystając z równości (8.19) i (8.22), można zauważyć, że (11.7) jest szczególnym przypadkiem teoriainformacyjnej nierówności (2.21) dla procesów gaussowskich.

Naiwnie rozumując, jeżeli położymy $p = \text{const}$ i $n \rightarrow -\infty$ w (11.3) i (11.2), to otrzymamy przedstawienie odwracalne i kauzalne dla $Z_{\mathbb{Z}}^{-\infty}$ danego przez (10.2). Traktując rzecz formalnie, należy wykazać istnienie odpowiednich przejść granicznych. Nie jest oczywiste, czy istnieje granica $\lim_{n \rightarrow \infty} Z_p^{-n}$, czy $Z_p^{-\infty} = \lim_{n \rightarrow \infty} Z_p^{-n}$, czy istnieją granice współczynników π_{nk} i ψ_{nk} , oraz czy istnieje przedstawienie odwracalne i/lub kauzalne dla $Z_{\mathbb{Z}}^{-\infty}$.

Twierdzenie 11.3. *Dla dowolnego procesu niedeterministycznego $X_{\mathbb{Z}}$ zachodzi*

$$\lim_{n \rightarrow \infty} Z_p^{-n} = Z_p^{-\infty}, \quad (11.8)$$

$$\lim_{n \rightarrow \infty} \psi_{nk} = \psi_k, \quad (11.9)$$

$$\lim_{n \rightarrow \infty} \pi_{nk} = \pi_k, \quad (11.10)$$

gdzie $\psi_k := \text{Cov}(Z_{n-k}^{-\infty}, X_n)$ zaś współczynniki π_k dane są przez (10.8) (zob. Pourahmadi, 2001, twierdzenie 7.14, Dégerine, 1982).

(NB. Proces jest niedeterministyczny wtedy i tylko wtedy, gdy $\lim_{n \rightarrow \infty} \pi_{n0} < \infty$.)

Dowód: Mamy $Z_p^{-n} := (X_p - \Phi_p^{-n+1:p-1}) / \|X_p - \Phi_p^{-n+1:p-1}\|$ oraz $\|X_p - \Phi_p^{-\infty:p-1}\| > 0$. Stąd (11.8) wynika na mocy twierdzenia 9.2.

Ponieważ $\lim_{n \rightarrow \infty} Z_p^{-n} = Z_p^{-\infty}$, mamy $\lim_{p \rightarrow \infty} \text{Cov}(Z_{n-k}^{-p}, X_n) = \text{Cov}(Z_{n-k}^{-\infty}, X_n)$. Stąd wynika (11.9), gdyż $\text{Cov}(Z_{n-k}^{-p}, X_n) = \psi_{n+p-1, k}$. Mamy też $\psi_0 = \sigma > 0$.

Zauważmy, że $\sum_{j=0}^m \pi_{nj} \psi_{n-j, m-j} = \llbracket m = 0 \rrbracket$. Udowodnimy (11.10) stosując indukcję po k . Dla $k = 0$ mamy $\pi_{n0} = 1/\psi_{n0}$, więc teza jest oczywista. Jeżeli (11.10) zachodzi dla wszystkich $k < m$ i pewnego $m \geq 1$, to

$$\lim_{n \rightarrow \infty} \pi_{nm} = \lim_{n \rightarrow \infty} -\frac{1}{\psi_{n-m, 0}} \cdot \sum_{j=0}^{m-1} \pi_{nj} \psi_{n-j, m-j} = -\frac{1}{\psi_0} \cdot \sum_{j=0}^{m-1} \pi_j \psi_{m-j} = \pi_m,$$

a zatem (11.10) zachodzi dla $k = m$. \square

Twierdzenie 11.4. *Dla procesu niedeterministycznego szeregi $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$ i $\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k = 1/\psi(z)$ zbiegają bezwzględnie dla $|z| < 1$. Zachodzą związki*

$$|\pi(z)|^2 \cdot \sigma^2 \in \left[\exp\left(-A \cdot \frac{1+|z|}{1-|z|}\right), \exp\left(A \cdot \frac{1+|z|}{1-|z|}\right) \right], \quad |\pi(z)|^2 \geq \frac{1-|z|^2}{B}, \quad (11.11)$$

gdzie $A := \frac{1}{2\pi} \int_{-\pi}^{\pi} |\log(f(\omega)/\sigma^2)| d\omega \geq 0$, $f := dF_{\ll}/dm$, $B := \sum_{k=0}^{\infty} |\psi_k|^2 \in [\sigma^2, \|X_0\|^2]$.

Dowód: Z twierdzenia 9.9 $\psi(z)$ zbiega dla $|z| < 1$, gdyż $\sum_{k=0}^{\infty} |\psi_k|^2 < \infty$. Z nierówności Schwarza $|\psi(z)|^2 \leq \sum_{k=0}^{\infty} |\psi_k|^2 \cdot \sum_{m=0}^{\infty} |z|^{2m} = \sum_{k=0}^{\infty} |\psi_k|^2 / (1 - |z|^2)$. Ze wzoru (9.20) otrzymujemy oszacowanie

$$\begin{aligned} |h(z) - \log \sigma^2| &= \left| \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\omega} + z}{e^{i\omega} - z} \log(f(\omega)/\sigma^2) d\omega \right| \\ &\leq \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1+|z|}{1-|z|} \cdot |\log(f(\omega)/\sigma^2)| d\omega = C(z), \quad C(z) := A \cdot \frac{1+|z|}{1-|z|}. \end{aligned}$$

Ponieważ $|\psi(z)|^2 = \sigma^2 \exp[h(z) - \log \sigma^2]$, otrzymujemy $|\psi(z)|^2 / \sigma^2 \in [e^{-C(z)}, e^{C(z)}]$. Funkcja $|\psi(z)|^2$ nie ma zer dla $|z| < 1$, a zatem $\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k = 1/\psi(z)$ zbiega dla $|z| < 1$ z twierdzenia 10.10 pkt. (i). \square

Przypomnijmy, że proces jest czysto niedeterministyczny wtedy i tylko wtedy, gdy $\sum_{k=0}^{\infty} |\psi_k|^2 = \|\sum_{k=0}^{\infty} \psi_k Z_{-k}^{-\infty}\|^2 = \|X_0\|^2$. Dla procesów czysto niedeterministycznych o ustalonym $\|X_0\|^2$ zależność pomiędzy funkcjami $\alpha(\cdot)$, $\pi(z)$, $\psi(z)$ oraz $\rho(\cdot)$ jest wzajemnie jednoznaczna. Z postaci $\alpha(\cdot)$ wynika postać $\pi(z)$, z $\pi(z)$ wynika $\psi(z)$, z $\psi(z)$ wynika $\rho(\cdot)$ (twierdzenie 11.1), z którego z kolei można obliczyć $\alpha(\cdot)$ za pomocą algorytmu Durбина-Levinsona.

Twierdzenie 11.3 umożliwia pewną dyskusję warunków koniecznych i dostatecznych czystej niedeterministyczności procesu.

Twierdzenie 11.5. *Dla ψ_k danego przez (11.9) zachodzi*

$$\sum_{k=0}^{\infty} |\psi_k|^2 \leq \|X_0\|^2 \left[1 - \limsup_{n \rightarrow \infty} |\rho(n)|^2 \right], \quad (11.12)$$

czyli proces nie jest czysto niedeterministyczny, gdy $\limsup_{n \rightarrow \infty} |\rho(n)| > 0$.

Dowód: Dla ustalonego N i $\epsilon > 0$ istnieje M takie, że dla każdego $n > M$ zachodzi

$$\left| \sum_{k=1}^N |\psi_{nk}|^2 - \sum_{k=1}^N |\psi_k|^2 \right| < \epsilon.$$

Dla każdego $n \geq N$ zachodzi $\|X_0\|^2 \geq \sum_{k=1}^N |\psi_{nk}|^2 + |\psi_{nn}|^2 = \sum_{k=1}^N |\psi_{nk}|^2 + \|X_0\|^2 |\rho(n)|^2$. Stąd wynika, że $\sum_{k=0}^N |\psi_k|^2 \leq \|X_0\|^2 [1 - \limsup_{n \rightarrow \infty} |\rho(n)|^2]$. Kładąc $N \rightarrow \infty$ otrzymujemy tezę. \square

Przykładem procesów gaussowskich, dla których $\lim_{n \rightarrow \infty} |\rho(n)| > 0$, są procesy wymienne (przykład 13.8, $p \neq 1$). Dla procesów tych V_n z rozkładu Wolda nie zależy od n , $\pi_0 \in [1/\|X_0\|, 1/\|X_0\|\sqrt{1-p}]$, natomiast $\pi_j = 0$ i $\psi_j = 0$ dla $j \geq 1$.

Do następnych twierdzeń definiujemy współczynniki $\psi_{nk} = \pi_{nk} = 0$ dla $k > n$.

Twierdzenie 11.6. (Pourahmadi, 2001, lemat 7.15., twierdzenie 7.19) *Dla składowej deterministycznej V_Z określonej wzorem (10.3) zachodzi*

$$V_p = \lim_{n \rightarrow \infty} \sum_{k=0}^n [\psi_{nk} - \psi_k] Z_{p-k}^{p-n-1}. \quad (11.13)$$

Proces niedeterministyczny X_Z jest czysto niedeterministyczny wtedy i tylko wtedy, gdy

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\psi_k - \psi_{nk}|^2 = 0.$$

Podobnie, aby dowieść $Z_l^{-\infty} = \lim_{n \rightarrow \infty} \sum_{k=0}^n \pi_{nk} X_{l-k} = \sum_{k=0}^{\infty} \pi_k X_{l-k}$, zbieżność punktowa (11.10) jest za słaba. W pierwszej kolejności należy zapewnić zbieżność szeregu po prawej stronie.

Twierdzenie 11.7. (Pourahmadi, 2001, lemat 6.4) *Szereg $\sum_{k=0}^{\infty} \pi_k X_{l-k}$ jest zbieżny w L^2 wtedy i tylko wtedy, gdy $\lim_{n \rightarrow \infty} \sum_{k=0}^n \sum_{l=0}^n \pi_k \pi_l \rho(k-l) < \infty$.*

Przypomnijmy, że ciągi zbiegające w ℓ^1 są ciągami Cauchy'ego w ℓ^1 i vice versa.

Twierdzenie 11.8. *Niech $\sum_{k=0}^{\infty} |\pi_{nk}| < \infty$ dla każdego n . Zachodzi równoważność*

$$\lim_{n \rightarrow \infty} \sup_{m > n} \sum_{k=0}^{\infty} |\pi_{mk} - \pi_{nk}| = 0 \iff \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\pi_k - \pi_{nk}| = 0 \text{ dla } \pi_k := \lim_{n \rightarrow \infty} \pi_{nk}.$$

Jeżeli warunki te zachodzą, to $\sum_{k=0}^{\infty} \pi_k = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \pi_{nk}$ oraz $\sum_{k=0}^{\infty} |\pi_k| = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\pi_{nk}| < \infty$.

Analogiczne twierdzenia obowiązują dla ℓ^k , $k \in (0, \infty)$. Z twierdzenia 11.8 dla ℓ^1 mamy:

Twierdzenie 11.9. *Niech X_Z będzie procesem stacjonarnym. Jeżeli $\sum_{k=0}^{\infty} |\pi_{nk}| < \infty$ dla wszystkich n i $\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\pi_k - \pi_{nk}| = 0$, to $\sum_{k=0}^{\infty} \pi_k X_{l-k}$ zbiega w L^2 i zachodzi równość $\sum_{k=0}^{\infty} \pi_k X_{l-k} = \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \pi_{nk} X_{l-k}$.*

Dowód: Z twierdzenia 10.1 szereg $\sum_{k=0}^{\infty} \pi_k X_{l-k}$ zbiega w L^2 , gdyż $\sum_{k=0}^{\infty} |\pi_k| < \infty$ z twierdzenia 11.8. Mamy $\|\sum_{k=0}^{\infty} (\pi_{nk} - \pi_k) X_{l-k}\| \leq \|X_l\| \sum_{k=0}^{\infty} |\pi_{nk} - \pi_k|$, więc $\sum_{k=0}^{\infty} \pi_{nk} X_{l-k}$ zbiega do $\sum_{k=0}^{\infty} \pi_k X_{l-k}$ dla $n \rightarrow \infty$. \square

Zastanówmy się teraz, kiedy zachodzi $\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\pi_k - \pi_{nk}| = 0$ bądź równoważnie $\lim_{n \rightarrow \infty} \sup_{m > n} \sum_{k=0}^{\infty} |\pi_{mk} - \pi_{nk}| = 0$. Współczynniki π_{nk} są proporcjonalne do ϕ_{nk} . Rekursja Durбина-Levinsona definiująca te ostatnie przypomina definicję współczynników Newtona. Stąd otrzymujemy:

Twierdzenie 11.10. *Dla $j \in \{1, \dots, n\}$ zachodzi nierówność*

$$|\phi_{nj}| \leq \binom{n}{j} \max_{k \in \{1, \dots, n\}} |\alpha(k)|. \quad (11.14)$$

Dowód: Mamy $\binom{n-1}{j} = \binom{n}{j} \frac{n-j}{n}$ oraz $\binom{n-1}{n-j} = \binom{n}{j} \frac{j}{n}$, a zatem z równań (8.12)–(8.13) dowolna nierówność postaci $|\phi_{nj}| \leq \binom{n}{j} w_n$ zachodzi, jeżeli $w_n \geq |\alpha(n)|$ zachodzi dla $n \in \{1, 2, \dots\}$ zaś $w_n \geq w_{n-1} \left[\frac{n-j}{n} + \frac{j}{n} |\alpha(n)| \right]$ zachodzi dla $j \in \{1, \dots, n-1\}$ i $n \in \{2, 3, \dots\}$. Ponieważ $|\alpha(n)| \leq 1$ zachodzi dla wszystkich n , zatem mamy $w_n \geq w_{n-1} \left[\frac{n-j}{n} + \frac{j}{n} |\alpha(n)| \right]$, jeżeli mamy $w_n \geq w_{n-1}$. W rezultacie odpowiednim w_n aby otrzymać (11.14) jest $w_n = \max_{k \in \{1, \dots, n\}} |\alpha(k)|$. \square

Z twierdzenia 11.10 wynika, że $\sum_{j=1}^n |\phi_{nj}| \leq (2^n - 1) \max_{k \in \{1, \dots, n\}} |\alpha(k)|$, gdzie prawa strona rozbiega do nieskończoności dla $n \rightarrow \infty$. Następne twierdzenia to seria ograniczeń o coraz większej mocy.

Twierdzenie 11.11. *Oznaczmy $\phi_{n0} := -1$. Dla dowolnej funkcji częściowej autokorelacji α zdefiniujemy pomocniczą funkcję częściowej autokorelacji α^A , gdzie $\alpha^A(n) = -|\alpha(n)|$. Niech ϕ_{nj}^A będą współczynnikami najlepszych predyktorów liniowych odpowiadających α^A . Dla $n \geq 0$ i $j \in \{0, \dots, n\}$ zachodzą relacje*

$$\phi_{nj}^A \leq 0, \quad (11.15)$$

$$|\phi_{n+1,j}^A| \geq |\phi_{n,j}^A|, \quad (11.16)$$

$$\sum_{j=0}^n |\phi_{nj}^A| = \prod_{k=1}^n (1 + |\alpha(k)|), \quad (11.17)$$

$$|\phi_{nj}| \leq |\phi_{nj}^A|. \quad (11.18)$$

Dowód: Mamy $-\phi_{11}^A = |\alpha(1)|$, więc (11.15)–(11.18) zachodzą dla $n = 0$. Następnie stosujemy indukcję po n : Mamy $\phi_{mj}^A = \phi_{m-1,j}^A + |\alpha(m)| \phi_{m-1,m-j}^A$ dla $j < m$ i $\phi_{mm}^A = -|\alpha(m)|$. Zatem (11.15) dla $n = m-1$ implikuje (11.15) oraz (11.16) dla $n = m$. Jeśli (11.15) zachodzi dla $n = m$ i $n = m-1$, równość $\sum_{j=0}^m \phi_{mj}^A = (1 + |\alpha(m)|) \sum_{j=0}^{m-1} \phi_{m-1,j}^A$ implikuje (11.17) dla $n = m$, jeżeli (11.17) zachodzi dla $n = m-1$. W końcu założmy (11.18) dla $n = m-1$. Wtedy $|\phi_{mj}| \leq |\phi_{m-1,j}| + |\alpha(m)| |\phi_{m-1,m-j}| \leq -\phi_{m-1,j}^A - |\alpha(m)| \phi_{m-1,m-j}^A = -\phi_{mj}^A$ dla $j < m$ i $|\phi_{mm}| = |\alpha(m)| = -\phi_{mm}^A$. Tym samym udowodniliśmy (11.18) dla $n = m$. \square

Interesującym wnioskiem z powyższego twierdzenia jest to, że istnieje bardzo wiele procesów spełniających warunek $\lim_{n \rightarrow \infty} \Phi_p^{p-n:p-1} = \Phi_p^{-\infty:p-1}$, dla których sumy współczynników $|\phi_{nk}|$ rozbiegają wykładniczo do nieskończoności. Niekiedy jednak może pojawić się silna zbieżność.

Twierdzenie 11.12. *Niech $\phi_{n0} := -1$ oraz $\phi_{nj} := 0$ dla $j > n$. Dla $m > n$ zachodzi*

$$\sum_{j=0}^m |\phi_{mj} - \phi_{nj}| \leq \prod_{k=1}^m (1 + |\alpha(k)|) - \prod_{k=1}^n (1 + |\alpha(k)|). \quad (11.19)$$

Dowód: Z równań (8.12)–(8.13) wynika

$$\begin{aligned} \sum_{j=0}^n |\phi_{nj} - \phi_{n-1,j}| &\leq |\alpha(n)| \sum_{j=0}^{n-1} |\phi_{n-1,j}| \leq |\alpha(n)| \prod_{k=1}^{n-1} (1 + |\alpha(k)|) \\ &= \prod_{k=1}^n (1 + |\alpha(k)|) - \prod_{k=1}^{n-1} (1 + |\alpha(k)|). \end{aligned} \quad (11.20)$$

Sumując stronami otrzymujemy (11.19). \square

Obierając $C > \prod_{k=1}^m (1 + |\alpha(k)|)$ nierówność (11.20) można zapisać jako

$$\sum_{k=0}^n |\phi_{mk} - \phi_{nk}| \leq C \sum_{k=n+1}^m |\alpha(k)| - \sum_{k=n+1}^m |\phi_{mk}| \leq C \sum_{k=n+1}^m |\alpha(k)|. \quad (11.21)$$

Nierówność (11.21) przypomina nierówność Baxtera $\sum_{k=0}^n |\phi_{nk} - \phi_k| \leq c \sum_{k=n+1}^{\infty} |\phi_k|$ zachodzącą dla dodatniej i ciągłej gęstości spektralnej, pewnych stałych c i N oraz wszystkich $n \geq N$ (Pourahmadi, 2001, twierdzenie 7.22). Wyprowadzenie nierówności Baxtera nie odwoływało się do rekursji Durбина-Levinsona (Baxter, 1962, twierdzenie 2.2).

Zdefiniujmy wielomiany

$$\phi_n(z) := \sum_{k=0}^n \phi_{nk} z^k, \quad \bar{\phi}_n(z) := \sum_{k=0}^n \phi_{n,n-k}^* z^k = z^n [\phi_n(1/z^*)]^*, \quad (11.22)$$

$$\pi_n(z) := \sum_{k=0}^n \pi_{nk} z^k, \quad \bar{\pi}_n(z) := \sum_{k=0}^n \pi_{n,n-k}^* z^k = z^n [\pi_n(1/z^*)]^*. \quad (11.23)$$

Jeżeli $|z| = 1$, to $\bar{\phi}_n(z) = [z^{-n} \phi_n(z)]^*$ oraz $|\bar{\phi}_n(z)| = |\phi_n(z)|$ a także $\bar{\pi}_n(z) = [z^{-n} \pi_n(z)]^*$ oraz $|\bar{\pi}_n(z)| = |\pi_n(z)|$.

Dla $z \in \mathbb{C}$ rekursję Durбина-Levinsona (8.12)–(8.13) można zapisać równoważnie jako

$$\begin{bmatrix} \phi_n(z) \\ \bar{\phi}_n(z) \end{bmatrix} = T(\alpha(n)) \begin{bmatrix} \phi_{n-1}(z) \\ \bar{\phi}_{n-1}(z) \end{bmatrix}, \quad \begin{bmatrix} \phi_0(z) \\ \bar{\phi}_0(z) \end{bmatrix} = -\mathbf{1}, \quad (11.24)$$

$$\begin{bmatrix} \pi_n(z) \\ \bar{\pi}_n(z) \end{bmatrix} = \hat{T}(\alpha(n)) \begin{bmatrix} \pi_{n-1}(z) \\ z \bar{\pi}_{n-1}(z) \end{bmatrix}, \quad \begin{bmatrix} \pi_0(z) \\ \bar{\pi}_0(z) \end{bmatrix} = \frac{\mathbf{1}}{\|X_0\|}, \quad (11.25)$$

gdzie

$$T(u) = \begin{bmatrix} 1 & -u^* \\ -u & 1 \end{bmatrix}, \quad \hat{T}(u) = \frac{T(u)}{\sqrt{1 - |u|^2}}, \quad \mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Twierdzenie 11.13. *Zachodzą związki*

$$|\pi_n(z)|^2 - |\bar{\pi}_n(z)|^2 = |\pi_{n-1}(z)|^2 - |z|^2 |\bar{\pi}_{n-1}(z)|^2, \quad (11.26)$$

$$|\pi_n(z)|^2 = |\bar{\pi}_n(z)|^2 + (1 - |z|^2) \sum_{k=0}^{n-1} |\bar{\pi}_k(z)|^2 \geq \frac{1 - |z|^2}{\|X_0\|^2}, \quad (11.27)$$

$$\sum_{k=0}^n |\bar{\pi}_k(z)|^2 = \sum_{k=0}^n |z^{n-k}|^2 |\pi_k(z)|^2. \quad (11.28)$$

Dowód: Macierz $\hat{T}(u)$ jest macierzą transformacji Lorenza i zachowuje niezmiennik $|a|^2 - |b|^2 = |c|^2 - |d|^2$ dla $[a, b]^T = \hat{T}(u)[c, d]^T$. Stąd wynika (11.26). Równość tę można przepisać do postaci $|\pi_k(z)|^2 - |\pi_{k-1}(z)|^2 = |\bar{\pi}_k(z)|^2 - |z|^2 |\bar{\pi}_{k-1}(z)|^2$ dla dowolnego k . Sumując takie równości stronami od $k = 1$ do $k = n$ otrzymujemy

$$|\pi_n(z)|^2 - |\pi_0(z)|^2 = |\bar{\pi}_n(z)|^2 + (1 - |z|^2) \sum_{k=1}^{n-1} |\bar{\pi}_k(z)|^2 - |z|^2 |\bar{\pi}_0(z)|^2.$$

Ponieważ $\pi_0(z) = \bar{\pi}_0(z) = 1/\|X_0\|$, stąd wynika (11.27). Połóżmy $1/z^*$ zamiast z w (11.27) i wstawmy tam $|\pi_n(1/z^*)| = |z^{-n} \bar{\pi}_n(z)|$ oraz $|\bar{\pi}_n(1/z^*)| = |z^{-n} \pi_n(z)|$. Po przemnożeniu obu stron przez $|z^n|^2$ otrzymujemy $|\bar{\pi}_n(z)|^2 = |\pi_n(z)|^2 + (|z|^2 - 1) \sum_{k=0}^{n-1} |z^{n-1-k}|^2 |\pi_k(z)|^2$. Gdy od tej równości odejmiemy (11.27), dostajemy (11.28). \square

Z nierówności (11.27) wynika, że wielomiany $\phi_n(z)$ oraz $\pi_n(z)$ nie mają pierwiastków dla $|z| < 1$. Znacznie bardziej złożony dowód tego ostatniego faktu znaleźć można w pracy Grenandera i Szegö (1958, sekcja 2.3. (a)). Dowód ów odwołuje się do ortonormalności wielomianów $\bar{\pi}_n(z)$ względem miary spektralnej.

Twierdzenie 11.14. (zasada maksimum i minimum) *Niech $V \subset \mathbb{C}$ będzie zwartym podzbiorem płaszczyzny zespolonej o brzegu ∂V . Niech $w(z)$ będzie dowolną funkcją holomorficzną (analityczną) we wnętrzu $V \setminus \partial V$ i ciągłą na całym V . Prawdziwe są następujące zdania:*

- (i) $\sup_{z \in \partial V} |w(z)| = \sup_{z \in V} |w(z)|$;
- (ii) $\inf_{z \in \partial V} |w(z)| = \inf_{z \in V} |w(z)|$, jeżeli $w(z) \neq 0$ dla wszystkich $z \in V$.

Dowód: Pkt. (i) to znana zasada maksimum (Rudin, 1986, sekcja 12.1). Pkt. (ii) wynika z niej w sposób następujący: Jeżeli $w(z) \neq 0$ dla wszystkich $z \in V$, to $1/w(z)$ jest funkcją holomorficzną na V . Stosując zasadę maksimum do $1/w(z)$ otrzymujemy tezę. \square

Twierdzenie 11.15. *Zachodzą związki*

$$\phi_n(\pm 1) = \prod_{k=1}^n (1 - (\pm 1)^k \alpha(k)), \quad \text{gdy } \alpha(k) \in \mathbb{R} \text{ dla } k \leq n; \quad (11.29)$$

$$|\phi_n(z)| \in \left[\prod_{k=1}^n (1 - |\alpha(k)|), \prod_{k=1}^n (1 + |\alpha(k)|) \right], \quad \text{dla } |z| \leq 1. \quad (11.30)$$

Dowód: (11.24) ma dwa szczególne przypadki:

$$\phi_n(\pm 1) = (1 - (\pm 1)^n \alpha(n)) \phi_{n-1}(\pm 1), \quad \text{gdy } \alpha(k) \in \mathbb{R} \text{ dla } k \leq n; \quad (11.31)$$

$$\phi_n(z) = \phi_{n-1}(z) - [\alpha(n)z^{-n} \phi_{n-1}(z)]^*, \quad \text{gdy } |z| = 1. \quad (11.32)$$

Z (11.31) wynika (11.29).

Załóżmy teraz, że $|z| = 1$. Zauważmy, że dla $A \in \mathbb{C}$ zachodzi

$$(1 - |\alpha(n)|) \cdot |A| \leq |A - [\alpha(n)z^{-n}A]^*| \leq (1 + |\alpha(n)|) \cdot |A|.$$

Z nierówności tej i (11.32) otrzymujemy przez indukcję (11.30) dla $|z| = 1$. Z zasady maksimum wynika zatem, że $|\phi_n(z)| \leq \prod_{k=1}^n (1 + |\alpha(k)|)$ dla $|z| \leq 1$. Jeżeli $\prod_{k=1}^n (1 - |\alpha(k)|) = 0$, to druga nierówność w (11.30) jest trywialna. W przeciwnym razie należy zauważyć, że z twierdzenia 11.13 $\phi_n(z) \neq 0$ nie tylko dla $|z| = 1$, ale także dla $|z| < 1$. Tym samym otrzymujemy $|\phi_n(z)| \geq \prod_{k=1}^n (1 - |\alpha(k)|)$ dla $|z| \leq 1$ z zasady minimum. \square

Zdefiniujmy teraz dwie klasy procesów.

Definicja 11.16. (procesy autoregresyjne i kwazifinitarne) *Proces taki, że $\alpha(k) = 0$ dla $k > n$, nazywa się procesem autoregresyjnym $AR(n)$. Procesem kwazifinitarnym nazywamy proces, dla którego $\sum_{k=1}^{\infty} |\alpha(k)| < \infty$ oraz $|\alpha(m)| < 1$ dla wszystkich m .*

Podobnie jak procesy finitarne, procesy kwazifinitarne są uogólnieniem niedeterministycznych procesów autoregresyjnych.

Twierdzenie 11.17. *Każdy proces kwazifinitarny X_Z ma przedstawienie odwracalne $Z_l^{-\infty} = \sum_{k=0}^{\infty} \pi_k X_{l-k}$ i kauzalne $X_l = \sum_{k=0}^{\infty} \psi_k Z_{l-k}^{-\infty}$, gdzie współczynniki π_k i ψ_k dane są przez (11.9)–(11.10). Szeregi potęgowe $\pi(z) = \sum_{k=0}^{\infty} \pi_k z^k$ i $\psi(z) = \sum_{k=0}^{\infty} \psi_k z^k$ spełniają*

$$\sum_{k=0}^{\infty} |\pi_k| \cdot \|X_0\|, |\pi(z)| \cdot \|X_0\| \in \left[\prod_{k=1}^{\infty} \sqrt{\frac{1 - |\alpha(k)|}{1 + |\alpha(k)|}}, \prod_{k=1}^{\infty} \sqrt{\frac{1 + |\alpha(k)|}{1 - |\alpha(k)|}} \right] \quad (11.33)$$

oraz $\psi(z) = 1/\pi(z)$ dla $|z| \leq 1$. Zachodzi też

$$\sum_{k=-\infty}^{\infty} |\rho(k)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^4 d\omega / \|X_0\|^4 \leq \prod_{k=1}^{\infty} \left(\frac{1 + |\alpha(k)|}{1 - |\alpha(k)|} \right)^2 \quad (11.34)$$

oraz $\sum_{k=-\infty}^{\infty} \rho(k)e^{-ik\omega} = |\psi(e^{i\omega})|^2 / \|X_0\|^2$. Ponadto zachodzą implikacje:

- (i) Jeżeli $X_{\mathbb{Z}}$ jest procesem autoregresyjnym, to $\sum_{k=0}^{\infty} |\psi_k| < \infty$, $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$.
(ii) Jeżeli $\alpha(k) \in \mathbb{R}$ dla wszystkich k , to

$$\pi(\pm 1) \cdot \|X_0\| = \prod_{k=1}^{\infty} \sqrt{\frac{1 - (\pm 1)^k \alpha(k)}{1 + (\pm 1)^k \alpha(k)}}, \quad (11.35)$$

$$\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k) = \prod_{k=1}^{\infty} \frac{1 + (\pm 1)^k \alpha(k)}{1 - (\pm 1)^k \alpha(k)}. \quad (11.36)$$

Dowód: Ponieważ $\sum_{k=1}^{\infty} |\alpha(k)| < \infty$, to zachodzi $\prod_{k=1}^{\infty} (1 + |\alpha(k)|) < \infty$. W myśl twierdzenia 11.12 zachodzi $\lim_{n \rightarrow \infty} \sup_{m > n} \sum_{j=0}^m |\phi_{mj} - \phi_{nj}| = 0$. Ciąg $(\phi_n)_{n \in \mathbb{N}}$ jest ciągiem Cauchy'ego w przestrzeni ℓ^1 bezwzględnie sumowalnych ciągów. Istnieją zatem granice $\phi_k := \lim_{n \rightarrow \infty} \phi_{nk}$.

Ponieważ $\lim_{n \rightarrow \infty} v_n = \prod_{k=1}^{\infty} (1 - |\alpha(k)|^2) > 0$ oraz $\pi_n(z) \cdot \|X_0\| = -\phi_n(z)/\sqrt{v_n}$, to ciąg $(\pi_n)_{n \in \mathbb{N}}$ także jest ciągiem Cauchy'ego w przestrzeni ℓ^1 . Na podstawie twierdzenia 11.9 wnioskujemy, że $Z_l^{-\infty} = \lim_{n \rightarrow \infty} \sum_{k=0}^n \pi_{nk} X_{l-k} = \sum_{k=0}^{\infty} \pi_k X_{l-k}$, a zatem proces $X_{\mathbb{Z}}$ jest odwracalny względem $Z_{\mathbb{Z}}^{-\infty}$.

Dla $\phi(z) := \sum_{k=0}^{\infty} \phi_k z^k$ zachodzi $\lim_{n \rightarrow \infty} \sup_{|z| \leq 1} |\phi_n(z) - \phi(z)| \leq \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\phi_{nk} - \phi_k| = 0$. Dzięki temu własności (11.30) oraz (11.29) przenoszą się na $\phi(z)$, gdy dokonamy podstawień $\phi_n(z) \rightarrow \phi(z)$ oraz $\prod_{k=1}^n \rightarrow \prod_{k=1}^{\infty}$. Ponieważ z zależności $\pi_n(z) \cdot \|X_0\| = -\phi_n(z)/\sqrt{v_n}$ wynika $\pi(z) \cdot \|X_0\| = -\phi(z)/[\prod_{k=1}^{\infty} (1 - |\alpha(k)|^2)]^{1/2}$, z własności (11.30) oraz (11.29) wynikają wzory (11.33) i (11.35) dla $\pi(z)$. Z kolei ze wzorów (11.17)–(11.18), $\sum_{k=0}^{\infty} |\pi_k| \geq |\pi(0)|$ oraz (11.33) dla $\pi(z)$ wynika wzór (11.33) dla $\sum_{k=0}^{\infty} |\pi_k|$.

Ze wzoru (11.33) dla $\pi(z)$ wynika $\pi(z) \neq 0$ dla $|z| \leq 1$. Ze zbieżności jednostajnej $\lim_{n \rightarrow \infty} \sup_{|z| \leq 1} |\pi_n(z) - \pi(z)| \leq \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\pi_{nk} - \pi_k| = 0$ wynika, że $\pi(z)$ jest funkcją ciągłą dla $|z| \leq 1$. Z twierdzenia 10.10 pkt. (ii) wynika, że szereg $\psi(z)$ zbiega do $1/\pi(z)$ dla $|z| \leq 1$.

Jeżeli proces $X_{\mathbb{Z}}$ jest procesem autoregresyjnym $AR(n)$, to $\pi(z) = \pi_n(z)$ jest wielomianem i ma nieskończony promień zbieżności. W tej sytuacji twierdzenie 10.10 pkt. (i) orzeka, że $\sum_{k=0}^{\infty} |\psi_k| < \infty$. Z twierdzenia 10.8 wynika, że przedstawienie odwracalne $Z_n^{-\infty} = \pi(B)X_n$ można odwrócić do przedstawienia kauzalnego $X_n = \psi(B)Z_n^{-\infty}$. Z twierdzenia 11.2 pkt. (i) wynika $\sum_{k=-\infty}^{\infty} |\rho(k)| < \infty$.

Zrezygnujemy teraz z założenia, że $X_{\mathbb{Z}}$ jest procesem autoregresyjnym. Funkcja $1/\pi_n(z)$ to szereg potęgowy współczynników przedstawienia kauzalnego procesu $AR(n)$, którego autokorelacje i częściowe autokorelacje są identyczne z $\rho(k)$ i $\alpha(k)$ procesu $X_{\mathbb{Z}}$ dla $0 \leq k \leq n$. Na mocy twierdzenia 11.1 mamy zatem

$$\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |1/\pi_n(e^{i\omega})|^2 \exp(ik\omega) d\omega \quad \text{dla } 0 \leq k \leq n. \quad (11.37)$$

Ponieważ $1/|\pi(z)|$ jest ograniczone, to zachodzi $\lim_{n \rightarrow \infty} \sup_{|z| \leq 1} |1/\pi_n(z) - 1/\pi(z)| = 0$. Stąd (11.37) implikuje $\gamma(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\psi(e^{i\omega})|^2 \exp(ik\omega) d\omega$ dla wszystkich $k \geq 0$. Zatem na mocy twierdzenia 11.1 zachodzi $X_l = \sum_{k=0}^{\infty} \psi_k Z_{l-k}^{-\infty}$. Na mocy twierdzenia 11.2 pkt. (ii)–(iii) z (11.33) oraz (11.35) wynika (11.34) oraz (11.36). \square

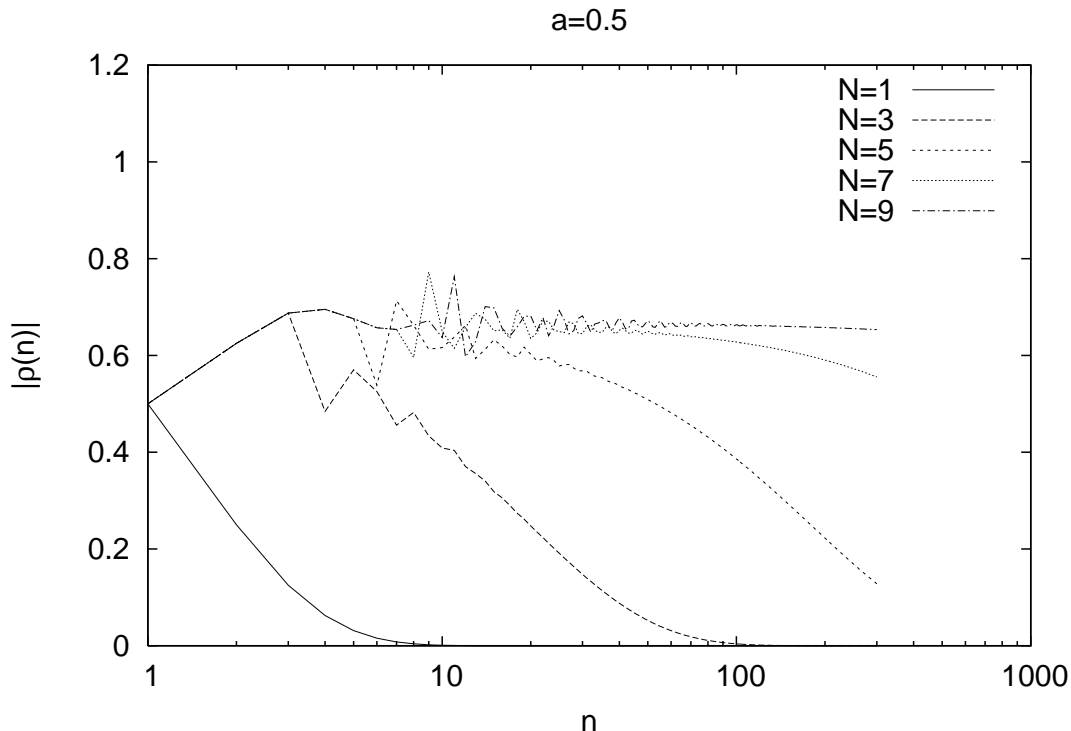


Diagram 11.1. Wykres $|\rho|$ dla $\alpha(n) = a \cdot \llbracket n \leq N \rrbracket$.

Oto kilka prostych wniosków z twierdzenia 11.17:

- (i) Jeżeli $(\pm 1)^k \alpha(k) > 0$, to zwiększenie $|\alpha(k)|$ powoduje zwiększenie $\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k)$.
 Jeżeli $(\pm 1)^k \alpha(k) < 0$, to zwiększenie $|\alpha(k)|$ powoduje zmniejszenie $\sum_{k=-\infty}^{\infty} (\pm 1)^k \rho(k)$.
- (ii) Jeżeli $(\pm 1)^k \alpha(k) > 0$ dla wszystkich k oraz $(\pm 1)^k \alpha(k) \geq a > 0$ dla N różnych k , to $(\pm 1)^m \rho(m) > 0$ zachodzi dla co najmniej $\left(\frac{1+a}{1-a}\right)^N$ różnych m , ponieważ $|\rho(k)| \leq 1$.

Diagram 11.1 ilustruje zjawisko opisane w pkt. (ii). Autokorelacje dla procesu kwazifinitarnego mogą zanikać bardzo powoli, choć suma szeregu kwadratów autokorelacji jest zawsze skończona.

Na koniec zauważmy, że oszacowania intensywności entropii i entropii nadwyżkowej można także wyrazić za pomocą wielomianów $\pi_n(z)$.

Twierdzenie 11.18. Niech $\hat{\pi}_n(z) := \pi_n(z) \cdot \|X_0\|$, gdzie $\pi_n(z) = \sum_{k=0}^n \pi_{nk} z^k$. Dla procesu takiego, że $|\alpha(k)| < 1$ dla $k \in \mathbb{N}$, zachodzą tożsamości:

$$-\sum_{k=1}^n \log [1 - |\alpha(k)|^2] = \log |\hat{\pi}_n(0)|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log |\hat{\pi}_n(e^{i\omega})|^2 d\omega, \quad (11.38)$$

$$-\sum_{k=1}^n k \log [1 - |\alpha(k)|^2] = \frac{1}{\pi} \int_{|z|<1} \left| \frac{d}{dz} \log \hat{\pi}_n(z) \right|^2 d(\operatorname{Re} z) d(\operatorname{Im} z). \quad (11.39)$$

Dowód: Na mocy twierdzeń 8.4 i 8.8 bez zmniejszania ogólności możemy założyć, że proces X_Z jest procesem autoregresyjnym $\operatorname{AR}(n)$. Mamy dla $\alpha(k) = 0$ dla $k > n$, czyli $-\sum_{k=1}^n \log [1 - |\alpha(k)|^2] = 2[H_X(1) - h_X]/d = -\log[\sigma^2/\|X_0\|^2]$ a także $-\sum_{k=1}^n k \log [1 - |\alpha(k)|^2] = 2E_X(n+1)/d = 2E_X/d = \log g$. Wariancja innowacji i gęstość spektralna mają postać $\sigma^2/\|X_0\|^2 = |\psi(0)|^2/\|X_0\|^2$ i $f(\omega) = |\psi(e^{i\omega})|^2$, gdzie $\psi(z) = 1/\pi_n(z)$. Stąd wynika (11.38). Ponieważ $\psi(z)$ jest odwrotnością wielomianu, który

nie zeruje się dla $|z| = 1$, to f spełnia $f \in [a, b] \subset (0, \infty)$ oraz istnieje druga pochodna f'' . Spełniony jest zatem warunek $|f'(\omega_1) - f'(\omega_2)| < \text{const} |\omega_1 - \omega_2|^\alpha$ dla $0 < \alpha < 1$. Korzystając z twierdzeń 9.9 i 9.12 otrzymujemy (11.39). \square

W istocie twierdzenie 9.9 było wyprowadzane przez Grenandera i Szegö (1958) w oparciu o algebraiczne własności wielomianów $\pi_n(z)$. Odpowiedników wzorów (11.38)–(11.39) nie udało nam się jednak w ich pracy zidentyfikować.

Rozdział 12

Wzajemne ograniczenia zaniku ACF i PACF

Ponieważ $\rho(n)$ jest funkcją $\alpha(k)$, gdzie $k \leq n$, i możemy wyzerować $\alpha(k)$ dla $k > n$, to ze wzoru (11.34) dla procesu kwazifinitarnego wynika ograniczenie

$$\sum_{k=-n}^n |\rho(k)|^2 \leq \prod_{k=1}^n \left(\frac{1 + |\alpha(k)|}{1 - |\alpha(k)|} \right)^2, \quad (12.1)$$

które jest prawdziwe dla jakiegokolwiek procesu.

W pewnych przypadkach istnieje podobne proste ograniczenie $\sum_{k=0}^n |\rho(k)|$.

Twierdzenie 12.1. *Dla każdego procesu spełniającego $\prod_{k=1}^n (1 + |\alpha(k)|) < 2$ zachodzi*

$$\sum_{k=0}^n |\rho(k)| \leq \left[2 - \prod_{k=1}^n (1 + |\alpha(k)|) \right]^{-1}. \quad (12.2)$$

Dowód: Po pierwsze, dla $k \geq 1$ mamy

$$\sum_{k=1}^n |\phi_{i+k,k}| \leq \sum_{k=1}^n |\phi_{i-1+k,k}| + \sum_{k=1}^n |\alpha(i+k)| |\phi_{i-1+k,i}|.$$

Oznaczmy $\phi_{n0} := -1$. Używając wzorów (11.16)–(11.18), uzyskujemy

$$\begin{aligned} \sum_{k=1}^n |\phi_{i+k,k}| &\leq \sum_{m=0}^i \sum_{k=1}^n |\alpha(m+k)| |\phi_{m-1+k,m}| \\ &= \sum_{k=1}^{\infty} |\alpha(k)| \sum_{m=0}^i |\phi_{k-1,m}| \mathbb{I}[k-m \geq 1 \wedge k-m \leq n] \\ &\leq \sum_{k=1}^{n+i} |\alpha(k)| \sum_{m=0}^{k-1} |\phi_{k-1,m}^A| \\ &= \sum_{k=1}^{n+i} |\alpha(k)| \cdot \prod_{m=1}^{k-1} (1 + |\alpha(m)|) = \prod_{k=1}^{n+i} (1 + |\alpha(k)|) - 1, \end{aligned}$$

gdzie ostatnie przekształcenie wynika z równości w (11.20).

Dla $k = n$ z (8.15) mamy $|\rho(k)| \leq \sum_{i=1}^k |\phi_{ki}| |\rho(k-i)| = \sum_{i=0}^{k-1} |\rho(i)| |\phi_{k,k-i}|$. Zatem

$$\begin{aligned} \sum_{k=0}^n |\rho(k)| - 1 &= \sum_{k=1}^n |\rho(k)| \leq \sum_{k=1}^n \sum_{i=0}^{k-1} |\rho(i)| |\phi_{k,k-i}| \\ &= \sum_{i=0}^{n-1} \sum_{k=i+1}^n |\rho(i)| |\phi_{k,k-i}| = \sum_{i=0}^{n-1} |\rho(i)| \sum_{k=1}^{n-i} |\phi_{i+k,k}| \\ &\leq \sum_{i=0}^{n-1} |\rho(i)| \cdot \left[\prod_{k=1}^n (1 + |\alpha(k)|) - 1 \right]. \end{aligned}$$

Przegrupowanie $\sum_{k=0}^n |\rho(k)| - 1 \leq \sum_{k=0}^n |\rho(k)| \cdot [\prod_{m=1}^n (1 + |\alpha(m)|) - 1]$ daje (12.2). \square

Warunek $\prod_{k=1}^n (1 + |\alpha(k)|) < 2$ nie jest trudny do spełnienia. Przypomnijmy, że $\log(1+x) \leq x$ dla $x \geq 0$, więc $\prod_{k=1}^n (1 + |\alpha(k)|) \leq \exp \sum_{k=1}^n |\alpha(k)|$. Zatem

$$\sum_{k=1}^n |\alpha(k)| < \ln 2 \implies \prod_{k=1}^n (1 + |\alpha(k)|) < 2. \quad (12.3)$$

Nie tylko ograniczoność skończonej sumy częściowych autokorelacji implikuje ograniczoność skończonej sumy autokorelacji. Zachodzą także pewne relacje odwrotne.

Twierdzenie 12.2. *Niech funkcja $G : \mathbb{N} \rightarrow \mathbb{R}$ będzie dowolną funkcją spełniającą następujące warunki:*

- (i) G jest nieujemna i nierosnąca,
- (ii) $G(0) := G(1)$,
- (iii) istnieje stała $B \geq 1$ taka, że $G(\lceil n/2 \rceil) \leq BG(n)$ dla każdego n ,
- (iv) $4B \sum_{k=0}^{\infty} G(k) < 1$.

Niech ρ będzie funkcją autokorelacji pewnego procesu taką, że $|\rho(n)| \leq G(n)$ dla $n \in \mathbb{N}$. (Nie żądamy $G(0) \geq \rho(0) = 1$.) Wówczas istnieje stała A taka, że współczynniki ϕ_{ni} najlepszych liniowych predyktorów spełniają

$$|\phi_{ni}| \leq AG(i) \quad (12.4)$$

dla $n \in \mathbb{N}$ oraz $i \in \{1, \dots, n\}$.

Dowód: Dla procesu stacjonarnego równania Yule'a-Walkera (8.5) przybierają postać

$$\phi_{ij}^K = \rho(j-i) - \sum_{k \in K \setminus \{j\}} \phi_{ik}^K \rho(j-k), \quad j \in K. \quad (12.5)$$

Równania te mają dokładnie jedno rozwiązanie ze względu na $(\phi_{ij}^K)_{j \in K}$. Oznaczmy

$$\rho_n^K(i; j) := \begin{cases} \rho(j-i), & n = 1, \\ -\sum_{k \in K \setminus \{j\}} \rho_{n-1}^K(i; k) \rho(j-k), & n \geq 2. \end{cases} \quad (12.6)$$

Jeżeli szeregi $\sum_{n=1}^{\infty} \rho_n^K(i; j)$ są zbieżne dla wszystkich $j \in K$, to $\phi_{ij}^K = \sum_{n=1}^{\infty} \rho_n^K(i; j)$.

Ograniczenie funkcji autokorelacji przyjęte w założeniach wystarcza, by dowieść zbieżności szeregów $\sum_{n=1}^{\infty} \rho_n^K(i; j)$. Oznaczmy

$$G_n(i; j) := \begin{cases} G(|i-j|), & n = 1, \\ \sum_{k \in \mathbb{Z}} G_{n-1}(i; k) G(|k-j|), & n \geq 2. \end{cases} \quad (12.7)$$

Z nierówności $|\rho(k)| \leq G(k)$ dla $k \in \mathbb{N}$ wynika

$$\begin{aligned} |\rho_n^K(k_0; k_n)| &\leq \sum_{k_1 \in K \setminus \{k_2\}} \dots \sum_{k_{n-1} \in K \setminus \{k_n\}} \prod_{j=1}^n |\rho(k_j - k_{j-1})| \\ &\leq \sum_{k_1 \in \mathbb{Z}} \dots \sum_{k_{n-1} \in \mathbb{Z}} [\llbracket k_1 = k_0 \rrbracket + G(|k_1 - k_0|)] \prod_{j=2}^n G(|k_j - k_{j-1}|) \\ &= G_{n-1}(k_0; k_n) + G_n(k_0; k_n) \quad \text{dla} \quad G_0(k_0; k_n) := \llbracket k_1 = k_0 \rrbracket. \end{aligned} \quad (12.8)$$

Ponadto dla $i \leq j$ otrzymujemy

$$\begin{aligned}
& \sum_{k \in \mathbb{Z}} G(|i - k|)G(|k - j|) \\
& \leq \sum_{k \leq \frac{i+j}{2}} G(|i - k|)G\left(\left\lceil \frac{j-i}{2} \right\rceil\right) + \sum_{k \geq \frac{i+j}{2}} G\left(\left\lceil \frac{j-i}{2} \right\rceil\right) G(|k - j|) \\
& \leq \left[4 \sum_{k=0}^{\infty} G(k)\right] G\left(\left\lceil \frac{j-i}{2} \right\rceil\right) \leq CG(|i - j|), \tag{12.9}
\end{aligned}$$

gdzie $C := 4B \sum_{k=0}^{\infty} G(k)$. Tę samą nierówność dowodzimy analogicznie dla $i \geq j$.

Z (12.9) mamy $G_n(i; j) \leq C^{n-1}G(|i - j|)$ dla $n \geq 1$. Zatem z (12.8) wynika

$$|\phi_{ij}^K| \leq \sum_{n=1}^{\infty} |\rho_n^K(i; j)| \leq \llbracket i = j \rrbracket + \frac{2}{1-C} \cdot G(|i - j|). \tag{12.10}$$

Ponieważ $\phi_{ni} := \phi_{k+n+1, k+n+1-i}^{k+1: k+n}$, to z (12.10) mamy (12.4) dla $A = 2(1-C)^{-1}$. \square

Jeżeli funkcja autokorelacji pewnego procesu spełnia założenia twierdzenia 12.2, to

$$|\alpha(n)| = |\phi_{nn}| \leq AG(n), \tag{12.11}$$

czyli $\sum_{k=0}^{\infty} |\alpha(k)| < \infty$. W szczególności dla $AG(1) < 1$ proces jest kwazifinitarny.

Warto zauważyć, że twierdzenie 12.2 odnosi się do wszystkich procesów, których funkcja autokorelacji spełnia $|\rho(n)| \leq G(n) = \delta|n|^{-\beta}$ dla

$$\delta < \delta_G(\beta) := \frac{2^{-2-\beta}(\beta-1)}{2\beta-1}, \quad \beta > 1. \tag{12.12}$$

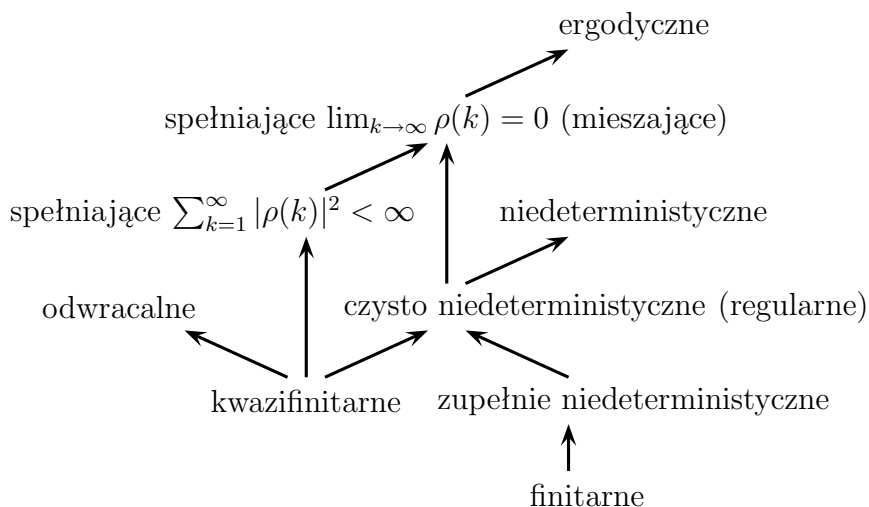
W istocie mamy $G(n/2) \leq BG(n)$ dla $B = 2^\beta$ oraz

$$\sum_{k=0}^{\infty} G(k) \leq 2G(1) + \int_1^{\infty} G(k)dk \leq \frac{(2\beta-1)\delta}{\beta-1} < \frac{(2\beta-1)\delta_G(\beta)}{\beta-1} = (4B)^{-1}.$$

Rozdział 13

Przykłady procesów gaussowskich

W świetle poprzednich rozdziałów pewna klasyfikacja stacjonarnych procesów gaussowskich przedstawia się następująco:



Strzałki na diagramie należy interpretować jako wynikania. Na przykład, każdy proces kwazifinitarny jest odwracalny, czysto niedeterministyczny i spełnia $\sum_{k=1}^{\infty} |\rho(k)|^2 < \infty$. Powyższy diagram warto zilustrować pewnymi przykładami procesów. Osobno rozpatrzemy procesy finitarne i niefinitarne.

Klasycznym przykładem procesów finitarnych są nieosobliwe procesy ARMA.

Przykład 13.1. (procesy ARMA) Mówimy, że proces $X_{\mathbb{Z}}$ jest procesem ARMA(p, q) (autoregressive moving-average), jeżeli istnieje standardowy biały szum $Z_{\mathbb{Z}}$, stała $\tilde{\sigma} > 0$ i wielomiany $\eta(z) = 1 - \sum_{k=1}^p \eta_k z^k$, $\theta(z) = 1 - \sum_{k=1}^q \theta_k z^k$ takie, że

$$\eta(B)X_n = \tilde{\sigma}\theta(B)Z_n, \quad (13.1)$$

zachodzi dla wszystkich $n \in \mathbb{Z}$. Nieosobliwym procesem ARMA(p, q) nazywamy proces taki, że $\theta(z)\eta(z) \neq 0$ dla wszystkich $|z| = 1$.

Pomimo tego, że gaussowskie procesy ARMA są ukrytymi łańcuchami Markowa w sensie definicji A.46, finitarności procesów ARMA nie można dowieść w oparciu o twierdzenie 2.11 ani przez podobnie proste zastosowanie twierdzenia 1.17. Klasyczny dowód finitarności procesów ARMA wykorzystuje wzór (9.22). Przypomnijmy pokrótce to rozumowanie, by skontrastować je z dowodem twierdzenia 2.11.

Wiadomo, że proces ARMA(p, q) jest kauzalny względem białego szumu $Z_{\mathbb{Z}}$ z równości (13.1), jeżeli $\eta(z) \neq 0$ dla wszystkich $|z| \leq 1$, zaś jest on odwracalny względem tego szumu, jeżeli $\theta(z) \neq 0$ dla wszystkich $|z| \leq 1$ (Brockwell i Davis, 1987, twierdzenia 3.1.1,

3.1.2, i sekcja 3.3). W szczególności każdy proces nieosobliwy można przedstawić jako ARMA(p, q) postaci

$$\tilde{\eta}(B)X_n = \sigma\tilde{\theta}(B)\tilde{Z}_n \quad (13.2)$$

takiej, że X_Z jest odwracalny i kauzalny względem standardowego białego szumu \tilde{Z}_Z (Brockwell i Davis, 1987, sekcja 4.4). Dla $\eta(z) = \prod_{i=1}^p(1 - a_i^{-1}z)$, $\theta(z) = \prod_{j=1}^q(1 - b_j^{-1}z)$, gdzie $|a_1|, \dots, |a_l| < 1 < |a_{l+1}|, \dots, |a_p|$ i $|b_1|, \dots, |b_k| < 1 < |b_{k+1}|, \dots, |b_q|$, mamy $\sigma = \tilde{\sigma} \prod_{j=1}^q |b_j| / \prod_{i=1}^p |a_i|$, $\tilde{\eta}(z) = \prod_{i=1}^p(1 - \tilde{a}_i^{-1}z)$ i $\tilde{\theta}(z) = \prod_{j=1}^q(1 - \tilde{b}_j^{-1}z)$ ze współczynnikami

$$\begin{aligned} (\tilde{a}_1, \dots, \tilde{a}_l, \tilde{a}_{l+1}, \dots, \tilde{a}_p) &= ([a_1^*]^{-1}, \dots, [a_l^*]^{-1}, a_{l+1}, \dots, a_p), \\ (\tilde{b}_1, \dots, \tilde{b}_k, \tilde{b}_{k+1}, \dots, \tilde{b}_q) &= ([b_1^*]^{-1}, \dots, [b_k^*]^{-1}, b_{k+1}, \dots, b_q). \end{aligned}$$

Funkcja $\psi(z) := \sigma\tilde{\theta}(z)/\tilde{\eta}(z)$ spełnia (9.20). Zatem σ jest wariancją innowacji procesu X_Z , zaś wariancja uogólniona wynosi

$$g = \frac{\prod_{j=1}^q \prod_{i=1}^p |\tilde{b}_j \tilde{a}_i^* - 1|^2}{\prod_{j=1}^q \prod_{j'=1}^q |\tilde{b}_j \tilde{b}_{j'}^* - 1| \prod_{i=1}^p \prod_{i'=1}^p |\tilde{a}_i \tilde{a}_{i'}^* - 1|} \quad (13.3)$$

(Finch, 1960).¹ Dla nieosobliwych procesów ARMA zachodzi $g < \infty$.

Nieosobliwe procesy ARMA są nie tylko finitarne, ale także kwazifinitarne. Wynika to w sposób następujący. Brockwell i Davis (1987, sekcje 3.3 i 3.6) dowodzą, że funkcja autokorelacji nieosobliwych procesów ARMA jest ograniczona co do modułu przez pewien zanik wykładniczy. Oznacza to, że $\sum_{k=1}^{\infty} k^l |\rho(k)| < \infty$ dla każdej liczby całkowitej $l \geq 0$. Z kolei gęstość spektralna nieosobliwego ARMA spełnia $f(\omega) = |\psi(e^{i\omega})|^2 = |\tilde{\sigma}\tilde{\theta}(e^{i\omega})/\tilde{\eta}(e^{i\omega})|^2$, a zatem jest ciągła i ściśle dodatnia. Korzystając z twierdzenia 9.11 otrzymujemy, że $\sum_{k=1}^{\infty} k^l |\alpha(k)| < \infty$ dla każdej liczby całkowitej $l \geq 0$. Kładąc $l = 0$, wnioskujemy, że proces jest kwazifinitarny.

Przykład 13.2. (procesy o potęgowym zaniku PACF) Niech proces X_Z będzie procesem o funkcji częściowej autokorelacji

$$\alpha(n) = \delta_\alpha n^{-\beta_\alpha}, \quad (13.4)$$

gdzie $|\delta_\alpha| < 1$, $\beta_\alpha > 0$.

Procesy o potęgowym zaniku PACF są finitarne i kwazifinitarne dla $\beta_\alpha > 1$.

Przykład 13.3. (procesy o potęgowym zaniku ACF) Niech proces X_Z będzie procesem o funkcji autokorelacji

$$\rho(n) = a_\rho \delta_M(\beta_\rho) |n|^{-\beta_\rho}, \quad (13.5)$$

gdzie $\delta_M(\beta) := (\beta-1)/(2\beta)$, $|a_\rho| \leq 1$, $\beta_\rho > 1$, $n \geq 1$. Tak określona funkcja ρ jest nieujemnie określona, gdyż $\sum_{k=1}^{\infty} |\rho(k)| \leq \sum_{k=1}^{\infty} \delta_M(\beta_\rho) |k|^{-\beta_\rho} \leq \delta_M(\beta_\rho) [1 + \int_1^{\infty} |k|^{-\beta_\rho} dk] = 1/2$, a zatem jej transformata Fouriera $f/\|X_0\|^2$ jest nieujemna.

Twierdzenie 12.2 dowodzi finitarności i kwazifinitarności pewnego podzbioru procesów o potęgowym zaniku ACF. Niestety, jak widać z diagramu 13.6, jest to bardzo mały podzbiór.

¹ Prawa strona (13.3) w pracy Fincha jest błędnie pomnożona przez $[(\prod_{j=1}^q \tilde{b}_j)/(\prod_{i=1}^p \tilde{a}_i)]^{2(p-q)}$.

W świetle obecnego stanu badań wydaje się, że istnieje szeroka klasa procesów, dla których zachodzi jednoczesny przybliżony potęgowy zanik autokorelacji i częściowej autokorelacji ze zbliżonym wykładnikiem (Baxter, 1962; McLeod, 1998; Inoue, 2005). Częstokowe ścisłe wyniki są trudne do osiągnięcia i wymagają wielu dodatkowych założeń. Aby zorientować się, jakiego typu regularności mogą zachodzić dla procesów o potęgowym zaniku ACF lub PACF, przeprowadziliśmy dwa eksperymenty numeryczne. W pierwszym eksperymencie obliczaliśmy α dla danego ρ , w drugim obliczaliśmy ρ dla danego α . Wyniki zostały przedstawione na diagramach 13.1, 13.2, 13.3, 13.4 i 13.5.

Dla $\beta_\rho > 1$, ściśle potęgowy zanik $\rho(n)$ wydaje się odpowiadać asymptotycznie potęgowemu zanikowi $|\alpha(n)|$ (diagram 13.4). Regularność ta nie zależy od amplitudy a_ρ , a oszacowanie wykładnika β_α zaniku $|\alpha(n)|$ jest w dobrym przybliżeniu równe wykładnikowi β_ρ zaniku funkcji autokorelacji. Analogicznie, dla szerokiej klasy ściśle potęgowych zaników $\alpha(n)$, wartość bezwzględna $|\rho(n)|$ zanika asymptotycznie potęgowo z wykładnikiem β_ρ równym wykładnikowi β_α zaniku funkcji częściowej autokorelacji (diagram 13.5).

Określenie z diagramu 13.5 hipotetycznego zbioru parametrów, dla których $\alpha(n) = \delta_\alpha n^{-\beta_\alpha}$ przekłada się na $|\rho(n)| \approx \delta_\rho n^{-\beta_\rho}$, gdzie $\beta_\rho = \beta_\alpha$, nie wydaje się w pełni wiarygodne. Diagram 13.5 sugeruje, że dla $\beta_\alpha < 1.5$ asymptotycznie mamy $\beta_\rho < \beta_\alpha$ dla $\delta_\alpha > 0$ i oraz $\beta_\rho > \beta_\alpha$ dla $\delta_\alpha < 0$. Wyniki te uzyskaliśmy obserwując $\rho(n)$ dla $n \leq 300$. Zbiór procesów o potęgowym zaniku częściowej autokorelacji, dla których $\beta_\rho = \beta_\alpha$, może jednak obejmować część procesów o $\beta_\alpha < 1.5$.

W świetle wniosków (ii)–(iii) z twierdzenia 11.17, dla bezwzględnie sumowalnej funkcji częściowej autokorelacji α , sumowalna z kwadratem funkcja autokorelacji ρ przybiera wysokie wartości dodatnie dla o wiele większej liczby argumentów niż α — dopiero po tej fazie przebiegu funkcja ρ może “naśladować zanikiem” funkcję α . Zjawisko to było ilustrowane na diagramie 11.1 dla pewnej klasy prostych procesów autoregresyjnych. Podobnie diagram 13.3 dla procesów o potęgowym zaniku częściowej autokorelacji pokazuje, jak dla ustalonego $\beta_\alpha > 1$ i rosnącego $|\delta_\alpha|$ gwałtownie zwiększa się zakres argumentów, dla których funkcja autokorelacji ρ przybiera wysokie wartości. Pomimo tego, niezależnie od amplitudy $|\delta_\alpha|$, dla dostatecznie dużych argumentów funkcja autokorelacji ρ staje się proporcjonalna do funkcji częściowej autokorelacji α .

Jak przekonamy się z dalszych przykładów, w tym ścisłego rozwiązania rekursji Durбина-Levinsona dla procesu ARIMA(0, d , 0), dla ujemnej i bezwzględnie niesumowalnej funkcji częściowej autokorelacji α , sytuacja zwykle kształtuje się odmiennie: Funkcja ρ zanika szybciej niż α i wcale nie musi naśladować zanikiem funkcji α . Dla procesów o $\alpha(n) = -0.5n^{-\beta_\alpha}$, gdzie $\beta_\alpha < 1$, pojawiają się wyraźne oscylacje $|\rho(n)|$, które wydają się sięgać poza przedział $n \leq 300$ i zaburzają oszacowania β_ρ na diagramie 13.5.

W obrębie klasy procesów gaussowskich o potęgowym zaniku PACF podzbiór procesów finitarnych równy jest podzbiorowi procesów kwazifinitarnych. W obrębie całej klasy procesów gaussowskich równość ta nie zachodzi.

Przykład 13.4. (niekwazifinitarne procesy finitarne) Niech proces $X_{\mathbb{Z}}$ będzie procesem o funkcji częściowej autokorelacji

$$\alpha(n) = an^{-1}(\log n)^{-r}, \quad n \geq 2, \quad (13.6)$$

gdzie $r \in (1/2, 1]$, $|a| < 2(\log 2)^r$.

Dla tak zdefiniowanego procesu zachodzi $\sum_{k=2}^{\infty} |\alpha(k)| = |a| \sum_{k=2}^{\infty} k^{-1}(\log k)^{-r} = \infty$ natomiast $\sum_{k=2}^{\infty} k|\alpha(k)|^2 = |a|^2 \sum_{k=2}^{\infty} k^{-1}(\log k)^{-2r} < \infty$.

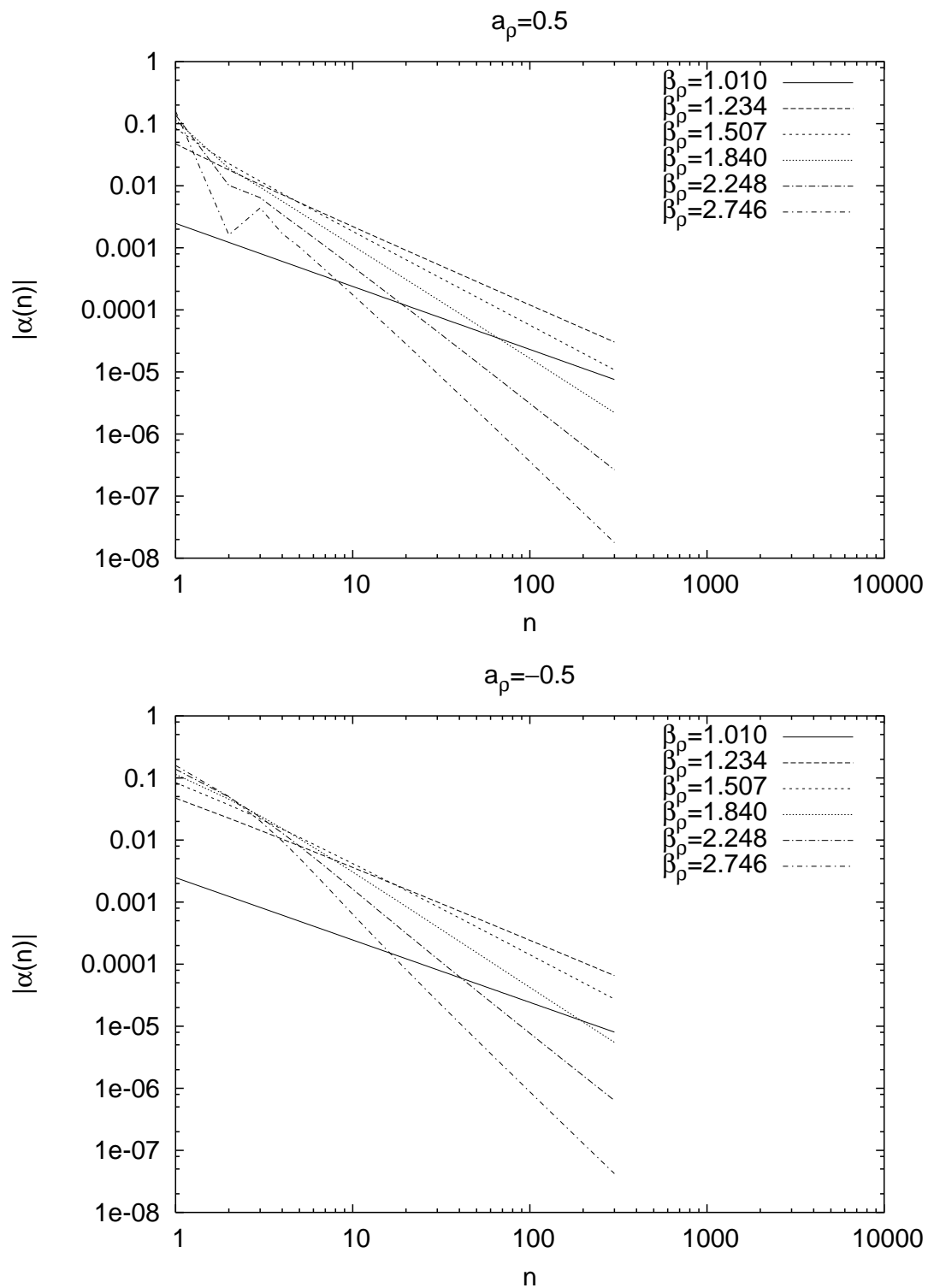


Diagram 13.1. Wykres $|\alpha(n)|$ dla $\rho(n) = a_p \delta_M(\beta) |n|^{-\beta_p}$ gdzie $a_p \in \{0.5, -0.5\}$.

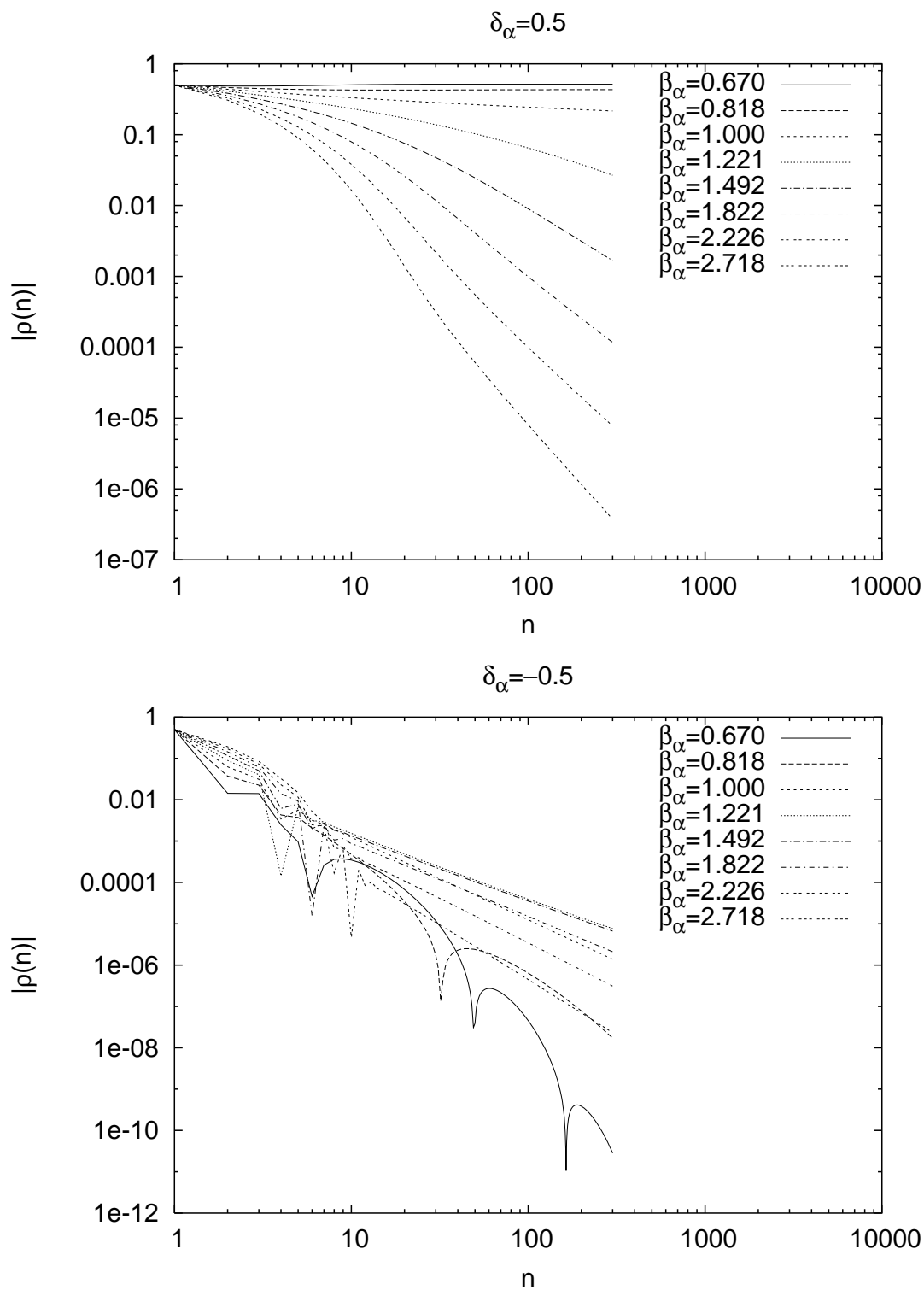


Diagram 13.2. Wykres $|\rho(n)|$ dla $\alpha(n) = \delta_\alpha n^{-\beta_\alpha}$ gdzie $\delta_\alpha \in \{0.5, -0.5\}$.

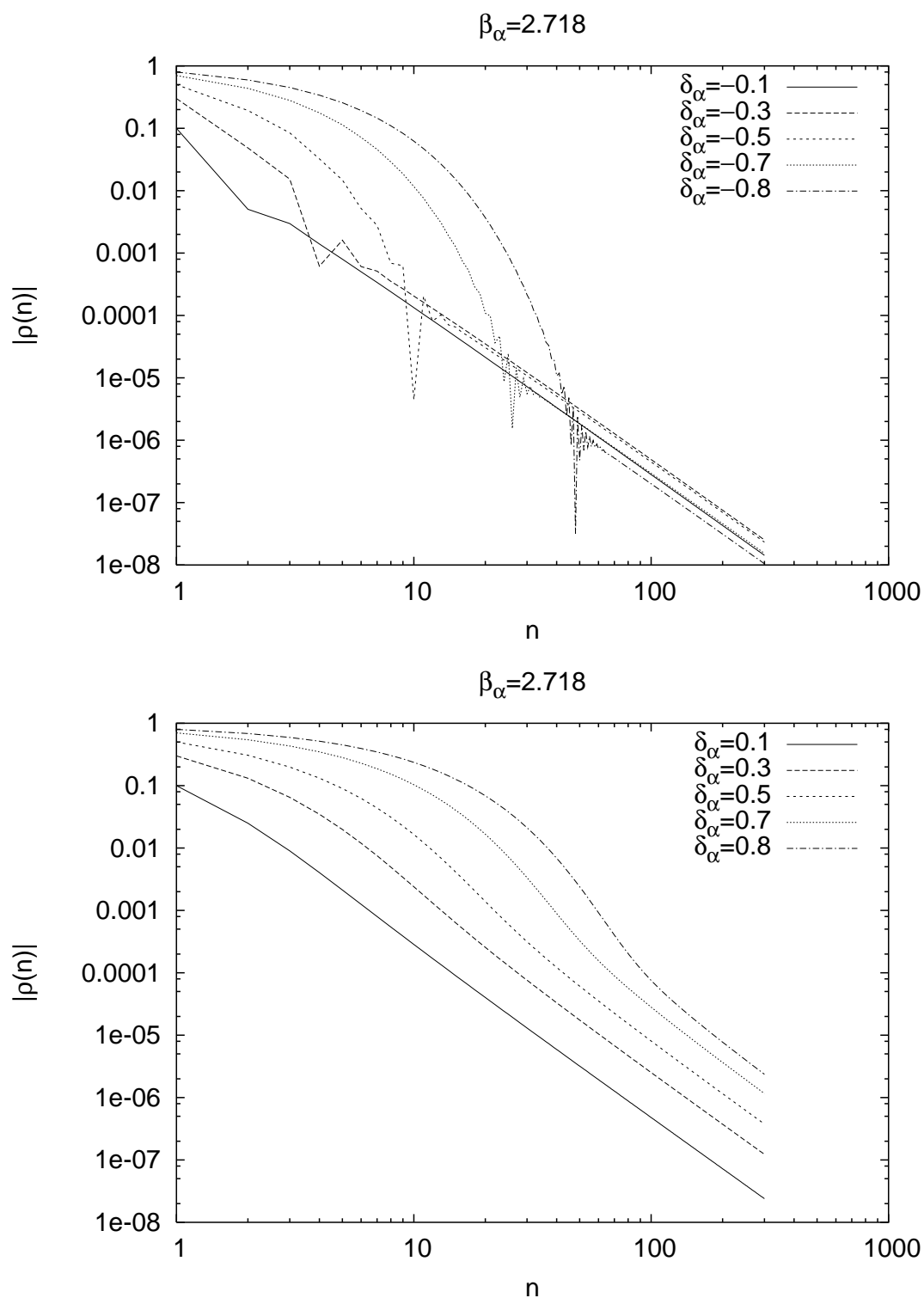


Diagram 13.3. Wykres $|\rho(n)|$ dla $\alpha(n) = \delta_\alpha n^{-\beta_\alpha}$ i ustalonego β_α .

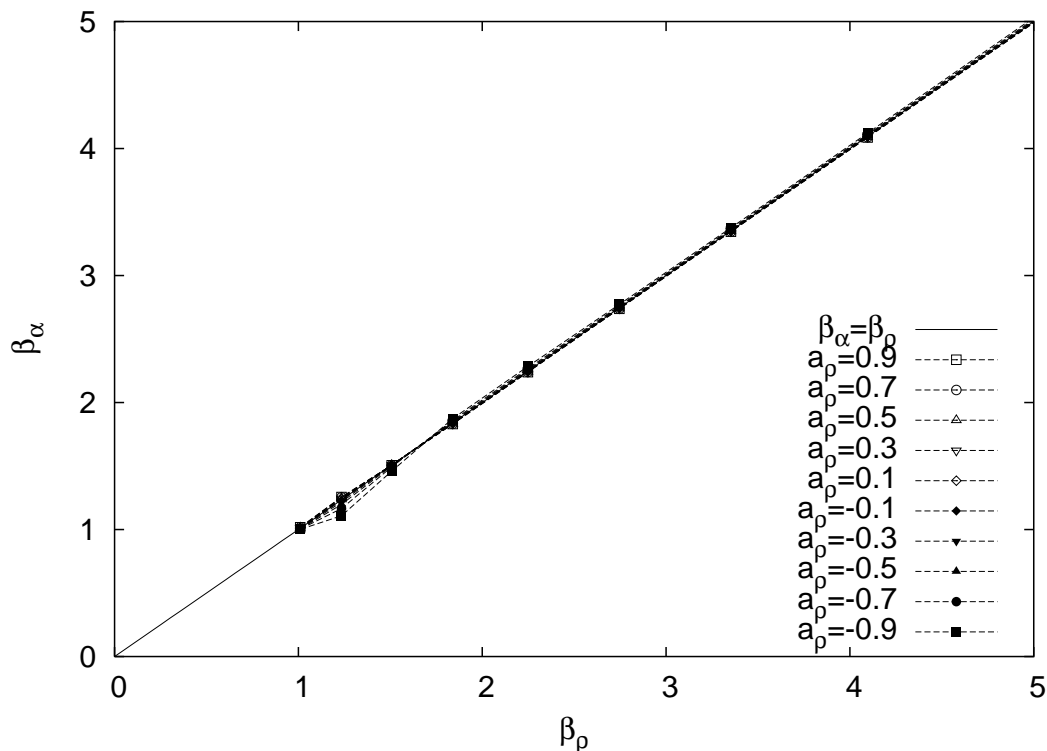


Diagram 13.4. Oszacowanie wykładnika β_α dla $|\alpha(n)| \approx \delta_\alpha n^{-\beta_\alpha}$ dane jako $\beta_\alpha := [\log \alpha(Q) - \log \alpha(P)] / [\log P - \log Q]$ w funkcji β_ρ , gdzie $\rho(n) = a_\rho \delta_M(\beta) \cdot |n|^{-\beta_\rho}$, $P = 295$, $Q = 300$.

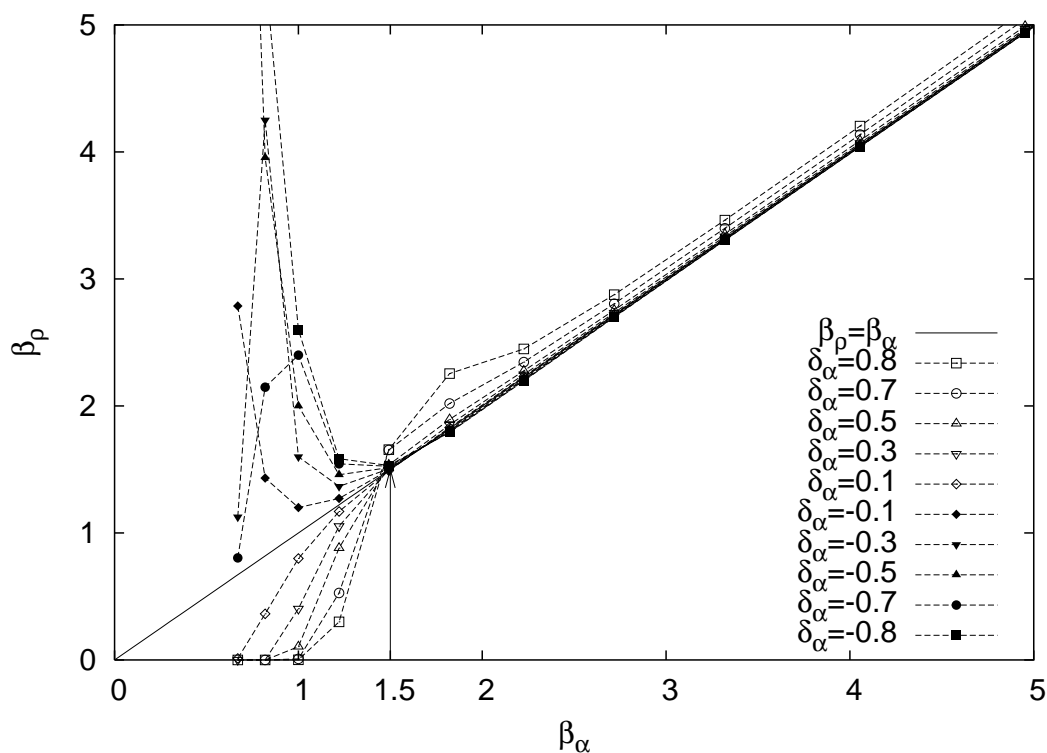


Diagram 13.5. Oszacowanie wykładnika β_ρ dla $|\rho(n)| \approx \delta_\rho n^{-\beta_\rho}$ dane jako $\beta_\rho := [\log \rho(Q) - \log \rho(P)] / [\log P - \log Q]$ w funkcji β_α , gdzie $\alpha(n) = \delta_\alpha n^{-\beta_\alpha}$, $P = 295$, $Q = 300$.

Przejdźmy teraz do przykładów procesów niefinitarnych.

Przykład 13.5. (procesy ARIMA) Proces $X_{\mathbb{Z}}$ nazywamy procesem ARIMA(p, d, q) (autoregressive integrated moving-average), jeżeli istnieje standardowy biały szum $Z_{\mathbb{Z}}$ taki, że $\eta(B)(1-B)^d X_n = \tilde{\sigma}\theta(B)Z_n$ dla $\eta(z) := 1 - \sum_{k=1}^p \eta_k z^k$, $\theta(z) := 1 - \sum_{k=1}^q \theta_k z^k$ i $(1-B)^d := \sum_{k=1}^{\infty} \binom{d}{k} (-B)^k$. Hosking (1981) wykazał, że dla $p = q = 0$ i $d \in (-1/2, 0) \cup (0, 1/2)$ zachodzą związki

$$\|X_0\|^2 = (-2d)!/(-d)!, \quad (13.7)$$

$$\rho(n) = \frac{(-d)!(n+d-1)!}{(n-d)!(d-1)!}, \quad (13.8)$$

$$\phi_{nk} = -\binom{n}{k} \frac{(k-d-1)!(n-d-k)!}{(-d-1)!(n-d)!}, \quad (13.9)$$

$$\alpha(n) = \frac{d}{n-d}, \quad (13.10)$$

gdzie $z! := \Gamma(z+1)$.

Dla $d \notin \mathbb{Z}$ procesy ARIMA nazywane są również procesami FARIMA.

W istocie proces $X_{\mathbb{Z}}$ o częściowej autokorelacji danej przez równość (13.10) istnieje dla każdego $d \in (-\infty, 0) \cup (0, 1/2)$. Dla wszystkich tych procesów wzory (13.9) i (13.8) są rozwiązaniem równań Durбина-Levinsona dla ϕ_{nk} i Yule'a-Walkera $\sum_{i=1}^n \phi_{ni}^* \rho(n-i) = \rho(n)$ dla $\rho(n)$. Dowód wynika z zastosowania reguły górnej negacji $\binom{r}{k} = (-1)^k \binom{k-r-1}{k}$ i wzoru Cauchy'ego $\sum_{k=0}^n \binom{r}{k} \binom{s}{n-k} = \binom{r+s}{n}$ dla $n, k \in \mathbb{Z}$, $r, s \in \mathbb{R}$ (Graham et al., 1994, rozdział 5, tabela 202).

Procesy ARIMA($0, d, 0$) są niefinitarne dla wszystkich $d \in (-\infty, 0) \cup (0, 1/2)$. Z drugiej strony z równości $\lim_{n \rightarrow \infty} \rho(n)/n^{-1+2d} = (-d)!/(d-1)!$ dla (13.8) (Hosking, 1981) wynika, że procesy te wykazują zależność dalekiego zasięgu dla $0 < d < 1/2$, zaś nie wykazują jej dla $d < 0$. Równoważnie procesy ARIMA($0, d, 0$) nie wykazują zależności dalekiego zasięgu wtedy i tylko wtedy, gdy funkcja częściowej autokorelacji jest asymptotycznie ujemna. Prawdliwość ta była wzmiankowana wcześniej i powtórzy się w następnych przykładach. Warto zauważyć, że dla $d < 0$ mamy także $\sum_{k=-\infty}^{\infty} \rho(k) = 0$, co sugeruje hipotetyczne uogólnienie wzoru (11.36).

Diagram 13.6 przedstawia amplitudę $(-d)!/(d-1)!$ asymptotycznego zaniku $\rho(n) \sim n^{2d-1}$ dla procesu ARIMA($0, d, 0$). Zwróćmy uwagę, że proces ARIMA($0, d, 0$) wykazujący asymptotyczny potęgowy zanik funkcji autokorelacji jest niefinitarny również dla tych d , dla których autokorelacje ściśle równe $\rho(n) = [(-d)!/(d-1)!] n^{-1+2d}$ odpowiadają procesom finitarnym na mocy twierdzenia 12.2.

Asymptotykę $|\alpha(n)| \sim |d|/n$ dla $d \in (-1/2, 0) \cup (0, 1/2)$ wykazują także procesy ARIMA(p, d, q) dla $p, q \neq 0$ (Inoue i Kasahara, 2004).

Przykład 13.6. (ułankowy szum gaussowski) Ułankowy ruch Browna o parametrze $H \in (0, 1)$ to proces $Y_{\mathbb{N}}$ o kowariancji $\text{Cov}(Y_m; Y_n) = \frac{\sigma^2}{2} (m^{2H} - |m-n|^{2H} + n^{2H})$. Podstawienie $X_n := Y_n - Y_{n-1}$ określa pewien proces stacjonarny $X_{\mathbb{Z}}$. Proces $X_{\mathbb{Z}}$ nazywa się ułankowym szumem gaussowskim (fractional Gaussian noise). Jeżeli $H = \frac{1}{2}$, to zmienne X_i są niezależne, a $Y_{\mathbb{N}}$ redukuje się do zwykłego ruchu Browna. W ogólności funkcja autokorelacji dla $X_{\mathbb{Z}}$ wynosi

$$\rho(k) = \text{Corr}(X_{n+k}; X_n) = \frac{1}{2} (|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H}). \quad (13.11)$$

Mamy $\lim_{k \rightarrow \infty} \rho(k)/k^{-2+2H} = 2H(H - \frac{1}{2})$ (Beran, 1994). Dla $H \in (\frac{1}{2}, 1)$ ułankowy szum gaussowski wykazuje zależność dalekiego zasięgu, zaś nie wykazuje jej dla $H \in (0, \frac{1}{2})$.

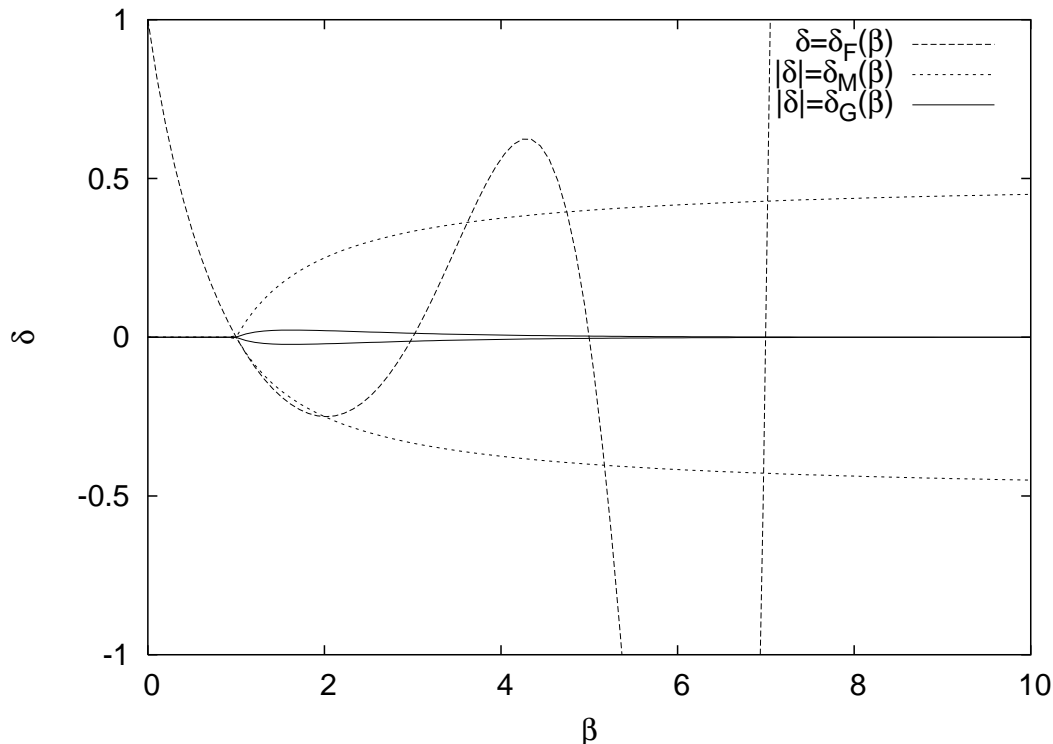


Diagram 13.6. Wykres parametrów funkcji $\rho(n) = \delta|n|^{-\beta}$. Dla $|\delta| < \delta_M(\beta) := (\beta - 1)/(2\beta)$ funkcja ρ jest nieujemnie określona (czyli jest funkcją autokorelacji pewnego procesu). Dla $|\delta| < \delta_G(\beta) := 2^{-2-\beta}(\beta - 1)/(2\beta - 1)$, proces o autokorelacji ρ jest finitarny (twierdzenie 12.2). Wartość $\delta_F(\beta) := (-d)!/(d - 1)!$ dla $2d - 1 = -\beta$ odpowiada asymptotycznej równości $\rho(n) \approx [(-d)!/(d - 1)!] n^{-1+2d}$ dla procesu ARIMA(0, d , 0).

W zasadzie we wzorze $\text{Cov}(Y_m; Y_n) = \frac{\sigma^2}{2} (m^{2H} - |m - n|^{2H} + n^{2H})$ możemy też położyć $H = 0$ lub $H = 1$. Dla $H = 0$ otrzymujemy $X_n = 0$, zaś dla $H = 1$ mamy $X_{n+1} = X_n$.

Zwartego przedstawienia funkcji częściowej autokorelacji dla ułamkowego szumu gaussowskiego nie znamy. Rozwiązanie numeryczne rekursji Durбина-Levinsona (diagramy 13.7 i 13.8) sugeruje, że zachodzi

$$\lim_{k \rightarrow \infty} \alpha(k)/k^{-1} = H - \frac{1}{2}. \quad (13.12)$$

Oznaczałoby to, że ułamkowe szumy gaussowskie dla $H \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$ są procesami niefinitarnymi. Podobnie jak procesy ARIMA(0, d , 0), ułamkowe szumy gaussowskie nie wykazują zależności dalekiego zasięgu wtedy i tylko wtedy, gdy funkcja częściowej autokorelacji jest asymptotycznie ujemna.

Własności asymptotyczne ułamkowego szumu gaussowskiego są dość podobne do własności procesu ARIMA(0, d , 0) o $d = H - \frac{1}{2} \in (-\frac{1}{2}, 0) \cup (0, \frac{1}{2})$. Dla obydwu procesów

$$\text{Power } \rho = \lim_{k \rightarrow \infty} \frac{\log \rho(k)}{\log k} = -1 + 2 \lim_{k \rightarrow \infty} \alpha(k)/k^{-1}. \quad (13.13)$$

Wzór ten nie jest prawdziwy w przypadku ogólnym. Zauważmy, że $\text{Power } \rho \leq 0$ zachodzi dla wszystkich procesów, dla których $\text{Power } \rho$ jest określone. Zatem wzór (13.13) nie może być spełniony, gdy $\lim_{k \rightarrow \infty} \alpha(k)/k^{-1} > 1/2$. W szczególności dla procesów wymiernych (przykład 13.8) różnych od białego szumu zachodzi $\lim_{k \rightarrow \infty} \alpha(k)/k^{-1} = 1$, natomiast $\text{Power } \rho = 0 \neq -1 + 2 = 1$.

Procesy ARIMA są przykładem procesów niekwazifinitarnych i niefinitarnych, które są czysto niedeterministyczne, a zatem ergodyczne. Dla procesów ARIMA(p, d, q) obserwujemy asymptotykę $|\alpha(n)| \sim |d|/n$, a zatem entropie nadwyżkowe rzędu n wynoszą $\bar{E}_X(n) = -\sum_{k=2}^n (k-1)\Delta^2 H(k) \approx \text{const}_1 + |d|^2 \log n$.

Rozbieżność entropii nadwyżkowych skończonego rzędu dla procesów ARIMA jest dość powolna. Istnieją jednakże czysto niedeterministyczne, ergodyczne i kwazifinitarne procesy niefinitarne, dla których podsumy $\bar{E}_X(n)$ rosną w przybliżeniu jak funkcje potęgowe.

Przykład 13.7. (niefinitarne procesy kwazifinitarne) Niech proces X_Z będzie procesem o funkcji częściowej autokorelacji

$$\alpha(n) = \begin{cases} 0, & n^{1/q} \notin \mathbb{N}, \\ an^{-r/q}, & n^{1/q} \in \mathbb{N}, \end{cases} \quad (13.14)$$

gdzie $q \in \mathbb{N}$, $r \in (1, (q+1)/2]$, $|a| < 1$.

Dla tak zdefiniowanego procesu zachodzi $\sum_{k=1}^{\infty} |\alpha(k)| = |a| \sum_{m=1}^{\infty} m^{-r} < \infty$ natomiast $\sum_{k=1}^{\infty} k|\alpha(k)|^2 = |a|^2 \sum_{m=1}^{\infty} m^{q-2r} = \infty$. Entropie nadwyżkowe rzędu n wynoszą $\bar{E}_X(n) = -\sum_{k=2}^n (k-1)\Delta^2 H(k) \approx \text{const}_1 + \text{const}_2 [n^{1/q}]^{q+1-2r}$, gdzie $[n^{1/q}]^{q+1-2r} < n^{1-1/q}$. Pominąwszy skokowość wzrostu $\bar{E}_X(n)$, możemy tak dobrać parametry q i r , aby dla skończonego podzbioru argumentów funkcja $\bar{E}_X(n)$ rosła prawie liniowo.

Na koniec przedstawmy dwa przykłady procesów niefinitarnych, które nie są czysto niedeterministyczne.

Przykład 13.8. (procesy wymienialne) Proces X_Z nazywamy procesem wymienialnym (exchangeable/interchangeable process), gdy zmienne $X_{i_1} \times \dots \times X_{i_k}$ i $X_{j_1} \times \dots \times X_{j_k}$ mają identyczny rozkład prawdopodobieństwa dla dowolnych i_m, j_m i k (Chow i Teicher, 1978; Haag, 1928). Z definicji wszystkie procesy wymienialne są stacjonarne. Gaussowskie procesy wymienialne mają funkcję autokorelacji postaci

$$\rho(n) = \begin{cases} 1, & n = 0, \\ p, & n \neq 0, \end{cases} \quad (13.15)$$

która jest nieujemnie określona dla $0 \leq p \leq 1$.

Proces wymienialny ma rozkład Wolda $X_n = \|X_0\| \sqrt{1-p} Z_n^{-\infty} + \|X_0\| \sqrt{p} V$, gdzie V jest standardową zmienną niezależną od standardowego białego szumu $Z_Z^{-\infty}$.

Łatwo sprawdzić, że dla procesu wymienialnego jedynym rozwiązaniem układu równań Yule'a-Walkera $\sum_{i=1}^n \phi_{ni}^* \rho(k-i) = \rho(k)$ dla $k \in \{1, \dots, n\}$ jest wektor stałych współczynników $\phi_{ni} = p/[(n-1)p+1]$, $i \in \{1, \dots, n\}$. Stąd wynika wzór na funkcję częściowej autokorelacji,

$$\alpha(n) = \phi_{nn} = \frac{p}{(n-1)p+1}. \quad (13.16)$$

Dla $p \neq 0$ procesy wymienialne są niedeterministyczne, lecz nie są czysto deterministyczne. Są zatem niefinitarne na mocy twierdzenia 10.4. Niefinitarność procesów wymienialnych można wywnioskować także z jawnej postaci autokorelacji częściowej (13.16).

Przykład 13.9. (procesy o stałej częściowej autokorelacji) Niech proces X_Z będzie procesem o częściowej autokorelacji $\alpha(k) = a$ dla wszystkich $k \in \mathbb{N}$ i $a \in (-1, 1)$.

Tak zdefiniowany proces jest deterministyczny. Automatyczne obliczenia symboliczne dla $k \in \{0, \dots, 12\}$ zwracają $\rho(k+1) = a - (1-a)w_k(-a^2)$, gdzie $w_k(x)$ jest wielomianem k -tego stopnia, $w_0(x) = 0$ i $w_k(x) = \sum_{i=1}^k w_{ki}x^i$ dla $k > 0$. Dla kilku pierwszych k mamy

w_{ki}	$i :$					
	1	2	3	4	5	6
$k : 1$	1					
2	2	2				
3	3	7	5			
4	4	16	26	14		
5	5	30	83	98	42	
6	6	50	202	398	372	132 .

Zakładając, że $\rho(k+1) = a - (1-a)w_k(-a^2)$ zachodzi dla wszystkich $k \in \mathbb{N}$, i spoglądając na diagramy 13.10 i 13.11, skłaniamy się do przypuszczenia, że (i) $0 \leq -w_k(-a^2) < |a|$ zachodzi dla wszystkich $k \in \mathbb{N}$ i $|a| < 1$ oraz (ii) $\lim_{k \rightarrow \infty} \rho(k) = 0$ zachodzi dla $-1 < a \leq 0$. Oba te przypuszczenia implikują $\lim_{k \rightarrow \infty} w_k(-a^2) = -|a|/(1+|a|)$ dla $|a| < 1$. Stąd $\lim_{k \rightarrow \infty} \rho(k) = 2a/(1+a)$ dla $0 \leq a < 1$. Oprócz tego mielibyśmy $\rho(k) \geq a = \alpha(k)$ dla $|a| \leq 1$ i $\rho(k) \leq -a$ dla $-1 \leq a \leq 0$. W rezultacie potrójna informacja wzajemna $I(X_1; X_k; X_{2:k-1}) = I(X_1; X_k) - I(X_1; X_k | X_{2:k-1})$ dla procesu gaussowskiego byłaby dodatnia dla $\alpha(k) > 0$ zaś ujemna dla $\alpha(k) < 0$. Co więcej wydaje się, że dla ujemnej częściowej autokorelacji moduł funkcji autokorelacji $|\rho(k)|$ jest ograniczony przez $k^{-1.5}$, a zatem sumowalny (diagram 13.9).

Podsumowując rozdział, poczujemy pewną ogólniejszą uwagę. Rozwijana dotychczas teoria silnej zależności koncentruje się procesach o niesumowalnej autokorelacji i dodatniej niesumowalnej częściowej autokorelacji. Niektóre przykłady wskazane w niniejszym rozdziale sugerują, że dla ujemnej i bezwzględnie niesumowalnej funkcji częściowej autokorelacji α funkcja autokorelacji ρ prawdopodobnie zanika szybciej niż α . Jak wskazuje przykład procesów ARIMA(0, d , 0), $d < 0$, funkcja autokorelacji w tym przypadku nie musi być niesumowalna. Tego typu procesy gaussowskie na korelogramie wyglądają na słabo zależne, natomiast w rzeczywistości mogą być bliskie determinizmowi. Sądzymy, że takie procesy warto byłoby zbadać dokładniej.

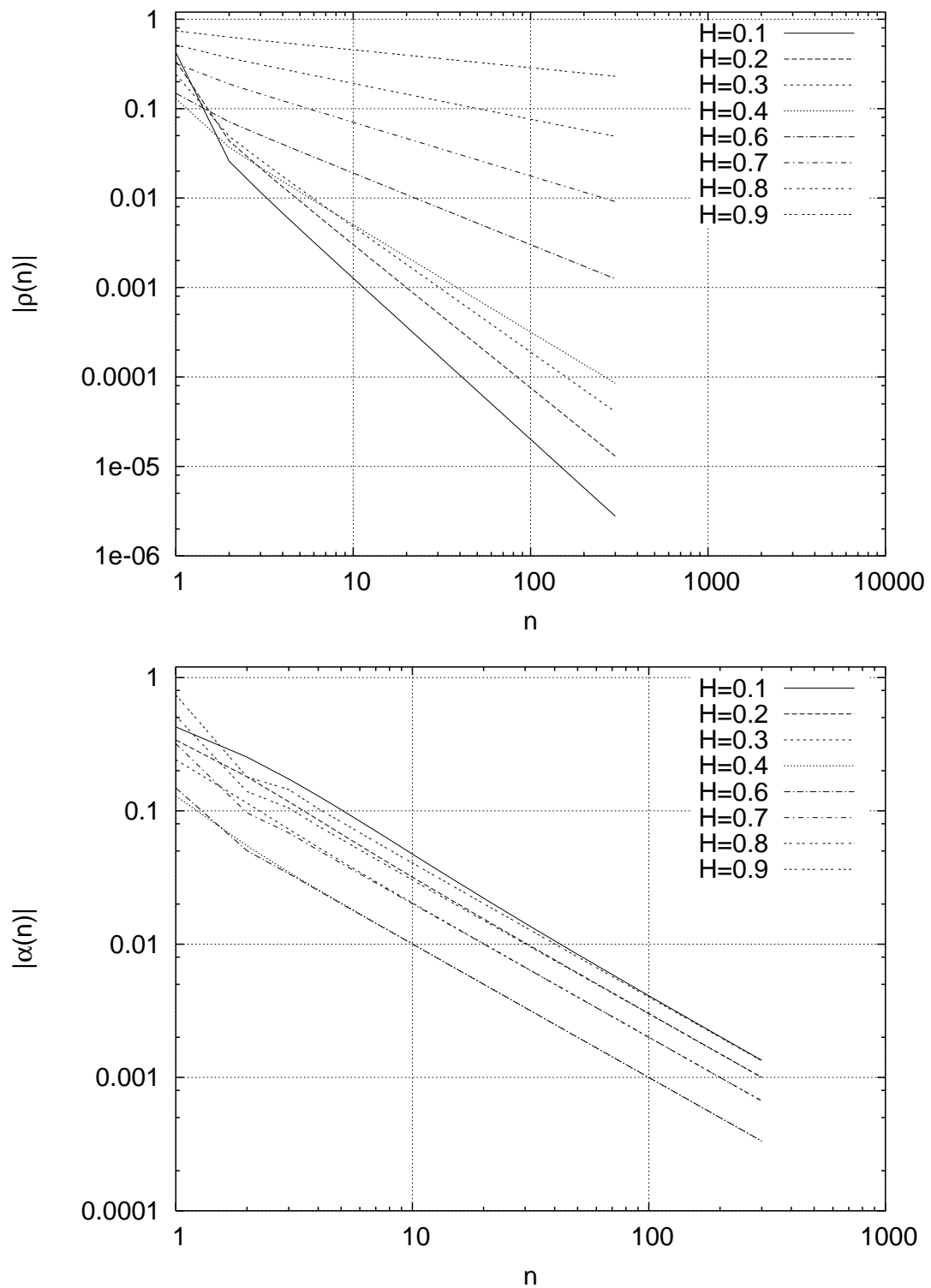


Diagram 13.7. Wykres dwóch funkcji autokorelacji dla ułamkowych szumów gaussowskich.

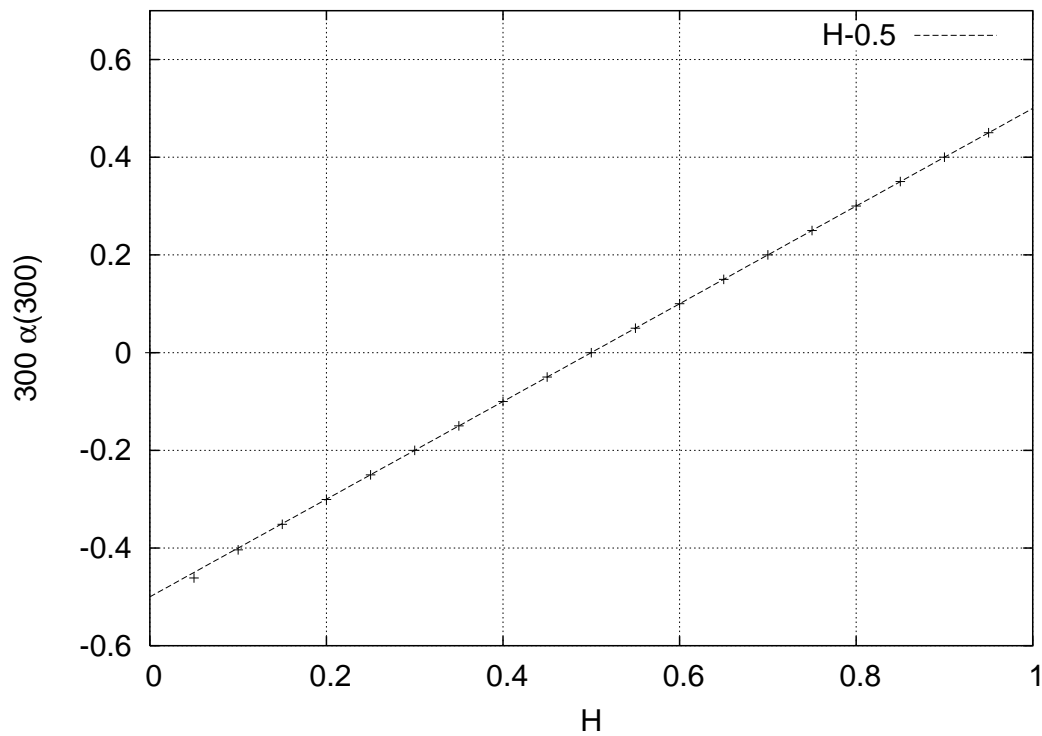


Diagram 13.8. Wartość $300\alpha(300)$ jako oszacowanie granicy $\lim_{k \rightarrow \infty} \alpha(k)/k^{-1}$ dla ułamkowego szumu gaussowskiego.

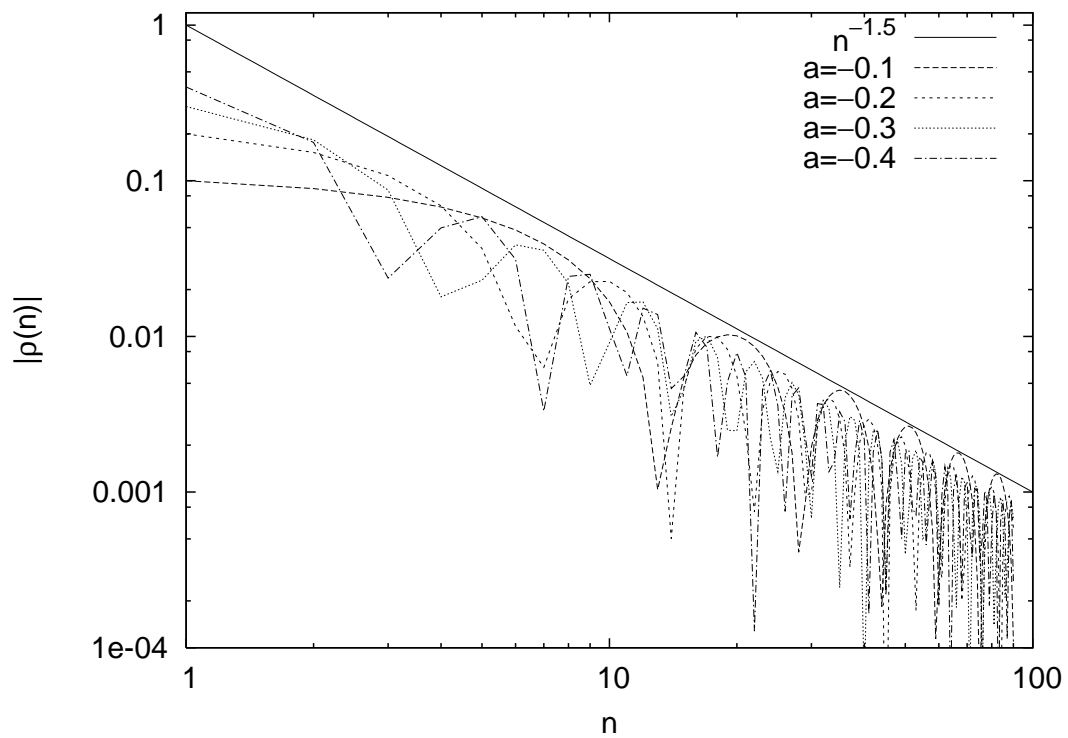


Diagram 13.9. Wykresy autokorelacji dla procesu o częściowej autokorelacji $\alpha(k) = a$.

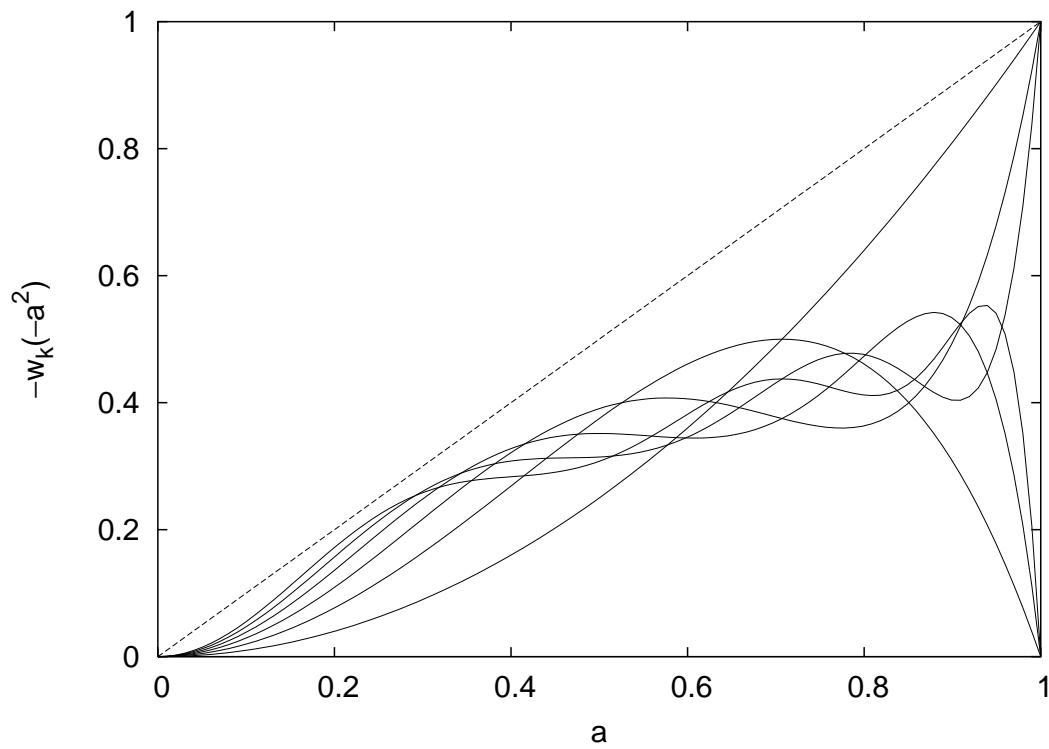


Diagram 13.10. Wykresy $-w_k(-a^2)$ dla $k \in \{0, \dots, 6\}$.

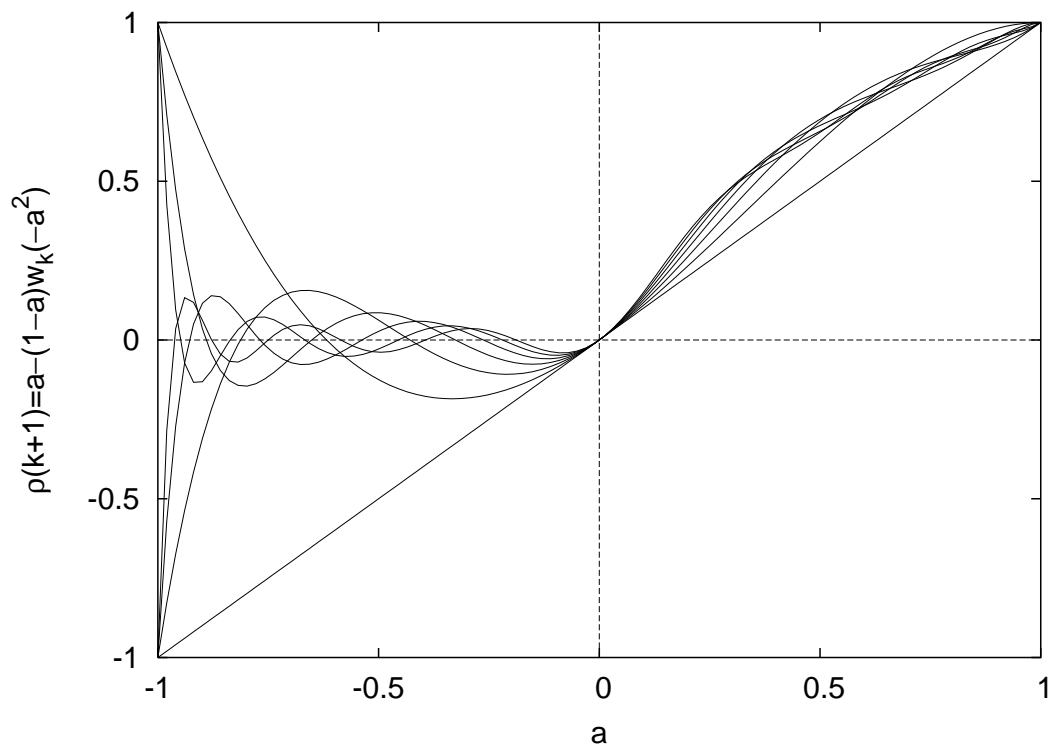


Diagram 13.11. Wykresy $\rho(k+1) = a - (1-a)w_k(-a^2)$ dla $k \in \{0, \dots, 6\}$.

Zakończenie

W niniejszej pracy omówiliśmy kilka jakościowo różnych związków pomiędzy asymptotyką entropii nadwyżkowej skończonego rzędu a niektórymi innymi własnościami procesów stacjonarnych. Kilkakrotnie wspominaliśmy o problemach wartych dalszych badań. Tytułem zakończenia spiszmy zbiorczą listę naszych pytań:

1. Jakie są warunki konieczne istnienia warunkowej miary produktowej oraz addytywności (1.17) dla teoriomiarowej warunkowej informacji wzajemnej? Czy nierówność (1.41) można uogólnić na przypadek dowolnych σ -ciał?
2. Jaka jest przestrzeń osiągalnych wartości entropii blokowej dla stacjonarnych procesów dyskretnych o nieskończonym i skończonym zbiorze wartości?
3. Jak definiować procesy nieprzeliczalnego opisu o wartościach binarnych?
4. Czy istnieją ergodyczne procesy niefinitarne o wartościach binarnych?
5. Czy istnieją binarne procesy niefinitarne o potęgowo rosnących entropiach nadwyżkowych skończonego rzędu? (Istnienie takich procesów sugerują pośrednio wyniki eksperymentów dotyczących kodowania tekstu w języku naturalnym, zob. rozdział 6.)
6. Jakie są związki pomiędzy nadwyżkową długością kodu a innymi parametrami kodów uniwersalnych? (Dębowski, 2005)
7. Czy istnieje kod uniwersalny, którego nadwyżkowa długość lub inne parametry mogą służyć do szacowania wartości entropii nadwyżkowej skończonego rzędu nie tylko od góry, ale i od dołu?
8. W jakich przypadkach przeliterowanie procesu o nieskończonej liczbie wartości na proces o skończonej liczbie wartości zachowuje wartość entropii nadwyżkowej? Co dzieje się wówczas z entropiami nadwyżkowymi skończonego rzędu?
9. W jakich przypadkach można zdefiniować segmentację procesu o skończonej liczbie wartości na proces o nieskończonej liczbie wartości?
10. W jakich przypadkach przeliterowanie jest operacją odwrotną do segmentacji?
11. Czy w terminach asymptotyki funkcji autokorelacji częściowej można łatwo wyrazić warunki konieczne istnienia przedstawień $AR(\infty)$ oraz $MA(\infty)$?
12. Jak szeroka jest klasa procesów gaussowskich o wolno zanikającej funkcji częściowej autokorelacji i szybko zanikającej funkcji autokorelacji? Jakie inne własności łączą takie procesy?
13. Czy istnieje analogon funkcji autokorelacji częściowej i wyrażenia przez nią entropii nadwyżkowej dla procesów $X_{\mathbb{R}}$ z czasem rzeczywistym? Czy istnieje związek pomiędzy hipotetycznym analogonem PACF a pojęciem bezpośredniej funkcji korelacji (*direct correlation function*) dyskutowanym w mechanice statystycznej? (Dębowski, 1999)

Dodatek A

Podstawy probabilistyczne

W rozdziale 1 dyskutujemy teoriomiarowe definicje pojęć warunkowej entropii i warunkowej informacji wzajemnej. Z tego względu w niniejszym dodatku przedstawiamy pewne podstawowe idee dotyczące teoriomiarowego rachunku prawdopodobieństwa (zob. Billingsley, 1979, wydanie polskie — Billingsley, 1987). Do wielu wyników odwołujemy się także w innych częściach pracy.

A.1. Miary prawdopodobieństwa i procesy stochastyczne

Ustalmy pewien zbiór Ω , który nazwiemy przestrzenią zdarzeń. Elementy $\omega \in \Omega$ nazywać będziemy zdarzeniami elementarnymi. Zbiory $A \subset \Omega$ nazwiemy zdarzeniami. Zbiory $\mathcal{J} \subset 2^\Omega$ nazywać będziemy klasami (zdarzeń lub zbiorów).

Definicja A.1. (π -układ, ciało, σ -ciało) Dla klasy $\mathcal{J} \subset 2^\Omega$ określmy warunki:

- (i) $A, B \in \mathcal{J}$ implikuje $A \cap B \in \mathcal{J}$,
- (ii) $\Omega \in \mathcal{J}$,
- (iii) $A \in \mathcal{J}$ implikuje $A^c \in \mathcal{J}$, gdzie $A^c := \Omega \setminus A$,
- (iv) $A_1, A_2, A_3, \dots \in \mathcal{J}$ implikuje $\bigcap_{n \in \mathbb{N}} A_n \in \mathcal{J}$.

Jeżeli spełniony jest warunek (i), to \mathcal{J} nazywa się π -układem. Jeżeli spełnione są warunki (i)–(iii), to \mathcal{J} nazywa się ciałem. Jeżeli spełnione są warunki (i)–(iv), to \mathcal{J} nazywa się σ -ciałem.

Jeżeli \mathcal{J} jest σ -ciałem, to parę (Ω, \mathcal{J}) nazywa się przestrzenią mierzalną. Symbol $\sigma(\mathcal{A})$ oznacza σ -ciało generowane przez zbiór \mathcal{A} , czyli przecięcie wszystkich σ -ciał $\mathcal{J} \subset 2^\Omega$ takich, że $\mathcal{A} \subset \mathcal{J}$. Klasa $\sigma(\mathcal{A})$ zależy od wyboru Ω . Jeżeli Ω nie jest jednoznaczne, to zamiast $\sigma(\mathcal{A})$ piszemy $\sigma_\Omega(\mathcal{A})$. Najmniejszym σ -ciałem jest $\{\emptyset, \Omega\}$.

Definicja A.2. (miara) Dla ciała \mathcal{J} i funkcji $\mu : \mathcal{J} \rightarrow \mathbb{R} \cup \{-\infty, \infty\}$ określmy warunki:

- (i) $\mu(A \cup B) = \mu(A) + \mu(B)$ dla $A \cup B \in \mathcal{J}$ i rozłącznych $A, B \in \mathcal{J}$,
- (ii) $\mu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mu(A_n)$ dla $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{J}$ i rozłącznych $A_1, A_2, A_3, \dots \in \mathcal{J}$,
- (iii) $|\mu(A)| < \infty$ dla $A \in \mathcal{J}$,
- (iv) $\mu(A) \geq 0$ dla $A \in \mathcal{J}$,
- (v) $\Omega = \bigcup_{n \in \mathbb{N}} A_n$, gdzie $\mu(A_n) < \infty$,
- (vi) $\mu(\Omega) = 1$.

Funkcję μ spełniającą warunek (i) nazywać będziemy addytywną miarą znakowaną, przy czym stosuje się dodatkowe określenia w zależności od tego, które z warunków (ii)–(vi) są spełnione:

- Określenie “addytywna” zastępuje się przez “ σ -addytywna”, jeżeli zachodzi (ii).
- Określenie “znakowana” zastępuje się przez: “rzeczywista”, jeżeli zachodzi (iii); “nieujemna”, jeżeli zachodzi (iv); “ σ -skończona”, jeżeli zachodzi (iv)–(v); “skończona”, jeżeli zachodzi (iii)–(iv); “prawdopodobieństwa”, jeżeli zachodzi (iii)–(vi).

Po dokonaniu wyżej wymienionych zastąpień, określenia “ σ -addytywna” i “nieujemna” zwyczajowo się opuszcza. (Gdy mówimy po prostu, że μ jest miarą prawdopodobieństwa, rozumiemy, że μ spełnia wszystkie warunki (i)–(vi).)

Przestrzenią probabilistyczną nazywamy trójkę (Ω, \mathcal{J}, P) , gdzie (Ω, \mathcal{J}) jest przestrzenią mierzalną zaś P jest miarą prawdopodobieństwa na \mathcal{J} .

Elementarny rachunek prawdopodobieństwa koncentruje się na własnościach miary prawdopodobieństwa P określonej na pewnym skończonym ciele \mathcal{J} . W tym przypadku σ -addytywność P sprowadza się do addytywności. Rozróżnienie addytywności i σ -addytywności pojawia się dla ciał nieskończonych. Niech $\mathcal{J} \subset 2^{\mathbb{N}}$ będzie klasą zbiorów A takich, że określone jest $P(A) := \lim_{n \rightarrow \infty} n^{-1} \sum_{k=1}^n \mathbb{I}[k \in A]$. ($P(A)$ jest nieokreślone np. dla $A = \{n \in \mathbb{N} : 3^{2k} < n \leq 3^{2k+1}, k \in \mathbb{N}\}$.) Funkcja P jest miarą addytywną na \mathcal{J} , ale nie spełnia intuicyjnej równości dla zdarzenia pewnego $P(\mathbb{N}) = 1 = \sum_{n \in \mathbb{N}} P(\{n\})$. Równość ta jest spełniona wszakże przy założeniu, że P jest miarą σ -addytywną, więc i tego ostatniego oczekujemy od rozsądnej miary prawdopodobieństwa.

Z definicji nie wymagamy od miar prawdopodobieństwa nieprzeliczalnej addytywności. Dla nieprzeliczalnego Ω istnieją miary P takie, że $P(\Omega) = 1$ natomiast $P(\{\omega\}) = 0$ dla każdego $\omega \in \Omega$ (miary takie nazywa się bezatomowymi, a ważnym przykładem miary bezatomowej jest miara Lebesgue’a). Zwróćmy uwagę, że warunek $P(\Omega) = \sum_{\omega \in \Omega} P(\{\omega\})$ implikuje, że $P(\{\omega\}) > 0$ zachodzi tylko dla przeliczalnie wielu ω . Dowolną nieprzeliczalną addytywną miarą prawdopodobieństwa możemy zatem obciąć do przestrzeni zdarzeń, która jest przeliczalna.

Własność σ -addytywności przydaje się do rozwiązania jeszcze jednej kwestii. Jeżeli \mathcal{J} nie jest σ -ciałem, to istnieją takie zdarzenia U nie należące do \mathcal{J} , że $U = \bigcup_{n \in \mathbb{N}} A_n$, gdzie $A_j \in \mathcal{J}$ są rozłączne. Ad hoc możemy próbować dedefiniować $P(U) := \sum_{n \in \mathbb{N}} P(A_n)$. Mocnym rezultatem jest poniższe twierdzenie, według którego w niektórych przypadkach funkcję P można rozszerzyć do miary na pewnym σ -ciele zawierającym \mathcal{J} jednoznacznie i bez złamania σ -addytywności.

Niech $P|_{\mathcal{P}}$ oznacza funkcję P obciętą do klasy \mathcal{P} .

Twierdzenie A.3. (o rozszerzeniu)

- (i) Niech \mathcal{P} będzie pewnym π -układem. Jeżeli P_1 i P_2 są miarami prawdopodobieństwa na $\sigma(\mathcal{P})$ i spełniają $P_1|_{\mathcal{P}} = P_2|_{\mathcal{P}}$, to $P_1 = P_2$.
- (ii) Niech P będzie miarą prawdopodobieństwa na ciele \mathcal{J}_0 . Dla każdego $A \in 2^{\Omega}$ określmy $P^*(A) := \inf \sum_{n \in \mathbb{N}} P(A_n)$, gdzie kres dolny liczony jest względem wszystkich ciągów $A_1, A_2, A_3, \dots \in \mathcal{J}_0$ spełniających $A \subset \bigcup_{n \in \mathbb{N}} A_n$. Niech \mathcal{M}_P będzie klasą wszystkich zbiorów A takich, że $P^*(A \cap E) + P^*(A^c \cap E) = P^*(E)$ dla każdego $E \in 2^{\Omega}$. Zachodzą następujące fakty:
 - a) Klasa \mathcal{M}_P jest σ -ciałem, przy czym $\mathcal{M}_P \supset \sigma(\mathcal{J}_0)$.
 - b) Funkcja $P^*|_{\mathcal{M}_P}$ jest miarą prawdopodobieństwa, przy czym $P^*|_{\mathcal{J}_0} = P$.
 - c) Przestrzeń $(\Omega, \mathcal{M}_P, P^*|_{\mathcal{M}_P})$ jest zupełna, tzn. warunki $B \in \mathcal{M}_P$, $A \subset B$ oraz $P^*(B) = 0$ implikują $A \in \mathcal{M}_P$.

NB. Powyższe twierdzenie występuje w literaturze w kilku wariantach (zob. Billingsley, 1979, sekcje 2 i 3). Jednym z wariantów punktu (i) jest twierdzenie o π - λ -układach (Billingsley, 1979, sekcja 2).

W ogólności nie musi zachodzić równość $\mathcal{M}_P = 2^{\Omega}$. Dla $\Omega = (0, 1]$ i bezatomowej miary prawdopodobieństwa P takiej jak np. miara Lebesgue’a istnieją takie $A \in 2^{\Omega}$, że wartość rozszerzenia $P(A)$ nie daje się jednoznacznie określić (Billingsley, 1979, sekcja 3). Z tego powodu interesująca nas w następnej kolejności definicja zmiennej losowej wymaga pewnej ostrożności sformułowania.

Dla funkcji $f : U \rightarrow U'$, zbioru $A \subset U$ i dowolnego zbioru $A' \subset U'$ określmy $f(A) := \{f(x) : x \in A\}$, $f^{-1}(A') := \{x : f(x) \in A'\}$. Dla klas zbiorów $\mathcal{J}, \mathcal{J}'$ określmy też $f(\mathcal{J}) := \{f(A) : A \in \mathcal{J}\}$, $f^{-1}(\mathcal{J}') := \{f^{-1}(A') : A' \in \mathcal{J}'\}$.

Definicja A.4. (funkcja mierzalna) Dla pary przestrzeni mierzalnych (Ω, \mathcal{J}) i (Ω', \mathcal{J}') mówimy, że funkcja $f : U \rightarrow U'$ jest mierzalna \mathcal{J}/\mathcal{J}' , jeżeli $f^{-1}(\mathcal{J}') \subset \mathcal{J}$. Funkcję $g : \Omega \rightarrow \Omega'$ nazywamy będziemy bijektywnie mierzalną \mathcal{J}/\mathcal{J}' , jeżeli g jest bijekcją mierzalną \mathcal{J}/\mathcal{J}' , zaś $g^{-1} : \Omega' \rightarrow \Omega$ jest mierzalna \mathcal{J}'/\mathcal{J} .

Istnienie choć jednej funkcji g bijektywnie mierzalnej \mathcal{J}/\mathcal{J}' implikuje izomorficzność przestrzeni mierzalnych (Ω, \mathcal{J}) i (Ω', \mathcal{J}') względem algebry Boole'a podzbiorów, tzn. zachodzi równoważność $A' \in \mathcal{J}' \wedge A = g^{-1}(A') \iff A \in \mathcal{J} \wedge A' = g(A)$, zaś dla $A_i \in \mathcal{J}$ mamy $g(\bigcap_i A_i) = \bigcap_i g(A_i)$ i $\Omega \setminus A_i = \Omega' \setminus g(A_i)$. Izomorfizm ten będziemy nazywać po prostu izomorfizmem przestrzeni mierzalnych.

Obcięcie σ -ciała \mathcal{G} do zbioru U określamy jako

$$\mathcal{G}|_U := \{A \cap U : A \in \mathcal{G}\}. \quad (\text{A.1})$$

Zauważmy, że funkcja $f : U \rightarrow U'$ jest mierzalna \mathcal{J}/\mathcal{J}' wtedy i tylko wtedy, gdy jest mierzalna $\mathcal{J}/\mathcal{J}''$ dla $\mathcal{J}'' := \mathcal{J}'|_{f(U)}$. W przypadku tak mierzalnego f operowanie klasą \mathcal{J}'' jest niekiedy wygodniejsze niż operowanie \mathcal{J}' , gdyż $(f(U), \mathcal{J}'')$ jest przestrzenią mierzalną izomorficzną z $(U, f^{-1}(\mathcal{J}'))$. W izomorfizmie tym rolę odwzorowania g pełni odwzorowanie f , które aplikowane do zbiorów z klasy $f^{-1}(\mathcal{J}'')$ jest bijekcją.

Definicja A.5. (zmienna losowa) Zmienną losową na (Ω, \mathcal{J}, P) nazywa się dowolną funkcję $X : \Omega \rightarrow \mathcal{A}(X)$ mierzalną $\mathcal{J}/\tilde{\sigma}(X)$, gdzie $(X(\Omega), \tilde{\sigma}(X))$ jest pewną przestrzenią mierzalną. Zbiór $\mathcal{A}(X)$ nazywany jest alfabetem zmiennej X , $X(\Omega) \subset \mathcal{A}(X)$ nazywane jest zbiorem wartości zmiennej X , $\tilde{\sigma}(X)$ nazywane jest σ -ciałem obrazów zmiennej X , zaś $\sigma(X) := X^{-1}(\tilde{\sigma}(X)) \subset \mathcal{J}$ nazywane jest σ -ciałem generowanym przez zmienną X .

Uwaga 1. Postać $\tilde{\sigma}(X)$ nie wynika jednoznacznie z postaci funkcji X . Aby nie operować jawnie parami $(X, \tilde{\sigma}(X))$, przyjmujemy, że $\tilde{\sigma}(X)$ jest funkcją obiektu X rozumianego jako funktor. Nie zakładamy też, że dla każdego $\omega \in \Omega$ musi zachodzić $\{X(\omega)\} \in \tilde{\sigma}(X)$.

Uwaga 2. Ogólny obiekt, który nazywamy zmienną losową, bywa niekiedy dla większej wyrazistości nazywany obiektem losowym (Kingman, 2002), gdyż pojęcie zmiennej losowej X często implícite rezerwuje się na przypadek $X(\Omega)$ będącego podzbiorem liczb.

Za powyższą definicją kryje się prosta idea, że dla każdego podzbioru $\tilde{A} \in \tilde{\sigma}(X)$ prawdopodobieństwo $P(X \in \tilde{A}) := P(X^{-1}(\tilde{A}))$ jest dobrze określone. Klasa $\sigma(X) = X^{-1}(\tilde{\sigma}(X))$ jest z kolei izomorficzna z σ -ciałem $\tilde{\sigma}(X)$.

Uwaga A.6. Według konwencji probabilistycznej (ϕ) to zdarzenie $\{\omega \in \Omega : \phi(\omega)\}$ dla formy zdaniowej $\phi : \Omega \rightarrow \{T, F\}$ (nazywanej zdaniem losowym). Przecięcia zdarzeń tego typu zapisujemy zastępując znak \cap przecinkiem, $(\phi_1) \cap (\phi_2) = (\phi_1, \phi_2)$. Ciąg zmiennych (słowo) X_1, \dots, X_k zapisujemy jako (X_1, \dots, X_k) lub $X_{1:k}$ — zgodnie z konwencją informatyczną. Oznaczenie operacji konkatencji utożsamiamy z oznaczeniem iloczynu kartezjańskiego: $(X_1, \dots, X_k) \times (X_{k+1}, \dots, X_l) = (X_1, \dots, X_l)$. Długość słowa w oznaczamy $\text{len } w$, gdzie $\text{len } w = k$ dla $w = a_{1:k} = (a_1, \dots, a_k)$. Zatem napisy $(X_{1:\text{len } w} = w)$, $((X_1, \dots, X_k) = w)$ oraz $(X_1 = a_1, \dots, X_k = a_k)$ oznaczają to samo zdarzenie.

W szczególności wyróżniamy następujące typy zmiennych:

- (i) Dyskretną zmienną losową X nazywamy zmienną taką, że $\mathcal{A}(X)$ jest przeliczalny ($\text{card } \mathcal{A}(X) \leq \aleph_0$), zaś $\tilde{\sigma}(X) = \sigma(\{\{x\} : x \in X(\Omega)\})$. Alfabet $\mathcal{A}(X)$ nie musi być podzbiorem liczb. Zmienną dyskretną o skończonej liczbie wartości nazywa się zmienną prostą.

- (ii) Rzeczywistą zmienną losową Y nazywamy zmienną taką, że $\mathcal{A}(Y) = \mathbb{R}$, $Y(\Omega) \in \mathcal{R}$, zaś $\tilde{\sigma}(Y) = \mathcal{R}|_{Y(\Omega)} \subset \mathcal{R}$, gdzie

$$\mathcal{R} := \sigma(\{\{x : x \leq r, x \in \mathbb{R}\} : r \in \mathbb{R}\}) \quad (\text{A.2})$$

jest σ -ciałem borelowskim na \mathbb{R} .

- (iii) Zmienną wielowymiarową $X_{1:k} = (X_1, \dots, X_k)$ definiujemy jako ciąg dyskretnych lub rzeczywistych zmiennych X_i na tej samej przestrzeni probabilistycznej, zaś $\tilde{\sigma}(X_{1:k}) = \sigma(\tilde{\sigma}(X_1) \times \dots \times \tilde{\sigma}(X_k))$.

Zauważmy, że w myśl powyższych definicji dowolna zmienna rzeczywista Y o przeliczalnym zbiorze wartości $Y(\Omega)$ jest zmienną dyskretną. Przy obydwu interpretacjach otrzymujemy dokładnie to samo σ -ciało obrazów i σ -ciało generowane.

Dla zmiennych dyskretnych X_1, \dots, X_k dowolna funkcja $f(X_1, \dots, X_k)$ jest także dyskretną zmienną losową na tej samej przestrzeni (Ω, \mathcal{J}, P) . Dowolna funkcja rzeczywista $g(Y_1, \dots, Y_k)$ rzeczywistych zmiennych losowych Y_1, \dots, Y_k nie musi być rzeczywistą zmienną losową na tej samej przestrzeni, gdyż zdarzenia $g(Y_1, \dots, Y_k) \leq r$, $r \in \mathbb{R}$, mogą nie należeć do ciała \mathcal{J} . Jednakże zdarzenia $g(Y_1, \dots, Y_k) \leq r$ są elementami \mathcal{J} , jeżeli istnieje takie pokrycie dziedziny funkcji g rodziną zbiorów $\{A_n\}_{n \in \mathbb{N}}$, $A_i \in \mathcal{J}$, że $g|_{A_n}$ (g obcięta do dziedziny A_n) jest ciągła dla każdego $n \in \mathbb{N}$ (Billingsley, 1979, sekcja 13).

Procesem stochastycznym nazywa się dowolną kolekcję zmiennych losowych, którą można uznać za pojedynczą zmienną losową w myśl definicji A.5. W tym celu należy zdefiniować odpowiednie σ -ciało obrazów. Tak jak dla zmiennej wielowymiarowej (skończonej kolekcji) chcemy by σ -ciało generowane przez proces było równe najmnieszemu σ -ciału zawierającemu σ -ciała generowane przez poszczególne zmienne procesu. Niech \mathbb{T} będzie dowolnym zbiorem. Przez $(x_t)_{t \in \mathbb{T}}$ będziemy rozumieć funkcję taką, że $x_{\mathbb{T}}(s) = x_s$ dla $s \in \mathbb{T}$ i $x_{\mathbb{T}} := (x_t)_{t \in \mathbb{T}}$. Dla zmiennych losowych X_t , funkcji $X_{\mathbb{T}} = (X_t)_{t \in \mathbb{T}}$ oraz zbiorów $A \in \tilde{\sigma}(X_s)$, $s \in \mathbb{T}$, definiujemy cylindry oraz klasę cylindrów

$$\text{cyl}_s^{\mathbb{T}}(A) := \{x_{\mathbb{T}} : x_r \in X_r(\Omega), r \in \mathbb{T}, x_s \in A\}, \quad (\text{A.3})$$

$$\text{Cyl}_s^{\mathbb{T}}(X_s) := \{\text{cyl}_s^{\mathbb{T}}(A) : A \in \tilde{\sigma}(X_s)\}. \quad (\text{A.4})$$

Notacja $\text{cyl}_s^{\mathbb{T}}$ jest trochę niejednoznaczna, gdyż oznaczony nią obiekt zależy implícite od $X_r(\Omega)$. Sądzimy jednak, że lepiej nie komplikować notacji i zdać się na czytelnika, który domyśli się właściwych $X_r(\Omega)$ z kontekstu.

Korzystając z definicji klasy cylindrów $\text{Cyl}_s^{\mathbb{T}}$, określmy σ -ciało obrazów dla procesu.

Definicja A.7. (proces stochastyczny) *Procesem stochastycznym $X_{\mathbb{T}} = (X_t)_{t \in \mathbb{T}}$ nazywamy kolekcję zmiennych losowych X_t indeksowanych pewnymi wartościami $t \in \mathbb{T}$ i określonych na tej samej przestrzeni probabilistycznej. Definiujemy σ -ciało obrazów jako $\tilde{\sigma}(X_{\mathbb{T}}) = \sigma(\bigcup_{t \in \mathbb{T}} \text{Cyl}_t^{\mathbb{T}}(X_t))$.*

Zbiór \mathbb{T} w powyższej definicji może być dowolnym zbiorem, np. nieprzeliczalnym. W myśl definicji A.5 proces stochastyczny jest także zmienną losową, przy czym $X_{(1,2,\dots,n)} = (X_1, X_2, \dots, X_n)$. Mamy też $\sigma(X_{\mathbb{T}}) = \sigma(\bigcup_{t \in \mathbb{T}} \sigma(X_t))$. Wygodnie też jest uogólnić pojęcie cylindra i klasy cylindrów jako

$$\text{cyl}_{\mathbb{T}'}^{\mathbb{T}}(A) := \{x_{\mathbb{T}} : x_r \in X_r(\Omega), r \in \mathbb{T}, x_{\mathbb{T}'} \in A\}, \quad (\text{A.5})$$

$$\text{Cyl}_{\mathbb{T}'}^{\mathbb{T}}(X_{\mathbb{T}'}) := \{\text{cyl}_{\mathbb{T}'}^{\mathbb{T}}(A) : A \in \tilde{\sigma}(X_{\mathbb{T}'})\}, \quad (\text{A.6})$$

gdzie \mathbb{T}' jest dowolnym zbiorem indeksów.

Uwaga A.8. Dla zbioru liczb całkowitych \mathbb{Z} , oprócz $X_{\mathbb{Z}} = (X_i)_{i \in \mathbb{Z}}$ oraz $X_{m:n} = (X_i)_{m \leq i \leq n} = (X_m, X_{m+1}, \dots, X_n)$ wprowadzamy oznaczenia $X_{-\infty:k} = (X_i)_{i \leq k, i \in \mathbb{Z}} = (\dots, X_{k-1}, X_k)$ oraz $X_{k:\infty} = (X_i)_{i \geq k, i \in \mathbb{Z}} = (X_k, X_{k+1}, \dots)$.

Dla procesu $X_{\mathbb{Z}}$ definiujemy także σ -ciała ogonowe

$$\sigma_X^{-\infty} := \bigcap_{k \in \mathbb{Z}} \sigma(X_{-\infty:k}), \quad \tilde{\sigma}_X^{-\infty} := X_{\mathbb{Z}}(\sigma_X^{-\infty}), \quad (\text{A.7})$$

$$\sigma_X^{\infty} := \bigcap_{k \in \mathbb{Z}} \sigma(X_{k:\infty}), \quad \tilde{\sigma}_X^{\infty} := X_{\mathbb{Z}}(\sigma_X^{\infty}). \quad (\text{A.8})$$

Klasy te są niepuste. Przykładowo $(\liminf_{n \rightarrow \infty} [X_n = f_n]) \in \sigma_X^{\infty}$.

Szczególnym przypadkiem procesów stochastycznych są miary losowe.

Definicja A.9. (miara losowa) *Miarą losową M na σ -ciele \mathcal{F} nazywa się funkcję taką, że $M(A)$ dla $A \in \mathcal{F}$ są rzeczywistymi zmiennymi losowymi, zaś dla każdego $\omega \in \Omega$ funkcja $M(\cdot)(\omega) : \mathcal{F} \rightarrow \mathbb{R}$ jest miarą. Pominąwszy specjalną konwencję notacyjną $M(\cdot)(\cdot)$, miarę losową M traktujemy jako proces stochastyczny $M = (M(A))_{A \in \mathcal{F}}$, któremu przypisujemy standardowe σ -ciało obrazów $\tilde{\sigma}(M) = \tilde{\sigma}((M(A))_{A \in \mathcal{F}}) = \sigma(\bigcup_{A \in \mathcal{F}} \text{Cyl}_A^{\mathcal{F}}(M(A)))$ (*Kallenberg, 1997, strona 83*).*

A.2. Definiowanie miar przez rozkłady

Zdarzenia elementarne $\omega \in \Omega$ można interpretować jako kolekcje $\omega = (x_t)_{t \in \mathbb{T}}$ wartości pewnych wyróżnionych zmiennych losowych X_t określonych na (Ω, \mathcal{J}, P) i spełniających $X_t(\omega) := x_t$. Interesujące jest pytanie, czy miarę P można określić jednoznacznie za pomocą prawdopodobieństw tylko niektórych zdarzeń postaci $P(X_{t_1} \in A_{t_1}, \dots, X_{t_k} \in A_{t_k})$. Dovolna zmienna dyskretna może być wzajemnie jednoznacznie reprezentowana jako pewna zmienna rzeczywista (wartości dyskretne reprezentujemy jako liczby naturalne), a zatem w stosunkowo ogólnym przypadku możemy założyć, że dla $\Omega \ni \omega = (x_t)_{t \in \mathbb{T}}$ współrzędne x_t są liczbami rzeczywistymi, co oznaczamy jako $\Omega = \mathbb{R}^{\mathbb{T}}$. Najpierw rozpatrzmy przypadek $\mathbb{T} = \{1, \dots, k\}$, czyli $\Omega = \mathbb{R}^{\{1, \dots, k\}} = \mathbb{R}^k$. Niech \mathcal{R}^k oznacza iloczyn kartezjański σ -ciał zgodnie ze wzorem (A.24).

Prostokątem w \mathbb{R}^k nazywamy zbiór $A_{a,b} = \{x = (x_1, \dots, x_k) : a_i < x_i \leq b_i\}$, gdzie $a = (a_1, \dots, a_k)$, $b = (b_1, \dots, b_k)$, $b_i \geq a_i$. Zbiór jego wierzchołków to

$$V_{a,b} = \{x = (x_1, \dots, x_k) : x_i = a_i \text{ lub } x_i = b_i\}.$$

Każdemu wierzchołkowi x przypisujemy znak $\text{sgn}^{a,b} x = \prod_{i=1}^k \text{sgn}_i^{a,b} x_i$, gdzie $\text{sgn}_i^{a,b} b_i = 1$ zaś $\text{sgn}_i^{a,b} a_i = -1$. Niech $\Delta_{a,b} F := \sum_{x \in V_{a,b}} \text{sgn}^{a,b} x \cdot F(x)$.

Twierdzenie A.10. (Billingsley, 1979, sekcja 12) *Jeżeli funkcja $F : \mathbb{R}^k \rightarrow \mathbb{R}$ spełnia*

(i) $\Delta_{a,b} F \geq 0$ dla każdego prostokąta $A_{a,b}$,

(ii) $\lim_{n \rightarrow \infty} F(x^{(n)}) = F(x)$ dla każdego $x = (x_1, \dots, x_k)$ oraz $x^{(n)} = (x_1 + 1/n, \dots, x_k + 1/n)$,

to istnieje dokładnie jedna σ -skończona miara μ na $\sigma(\mathcal{R}^k)$ spełniająca $\mu(A_{a,b}) = \Delta_{a,b} F$.

Odwrotnie, jeżeli μ jest miarą skończoną na $\sigma(\mathcal{R}^k)$, to

$$F(x) := \mu(\{y = (y_1, \dots, y_k) : y_i \leq x_i\})$$

spełnia $\mu(A_{a,b}) = \Delta_{a,b} F$ oraz warunki (i) i (ii).

Warunek (i) zapewnia nieujemność i addytywność μ . Dla $k > 1$ aby spełnić (i) nie wystarczy, by $F(x)$ była niemalejąca względem dowolnej współrzędnej x_i . Warunek (ii) zapewnia σ -addytywność. Korzystając z drugiej części twierdzenia, można sprawdzić, czy pewne μ jest miarą.

W istocie miarę μ na $\sigma(\mathcal{R}^k)$ można zdefiniować jednoznacznie podając jej wartości tylko dla przeliczalnego zbioru zdarzeń. Funkcja $F : \mathbb{R}^k \rightarrow \mathbb{R}$ spełnia warunki (i) i (ii) wtedy i tylko wtedy, gdy $F(x) = \inf_{q_i \geq x_i} \bar{F}(q)$ dla $\bar{F} := F|_{\mathbb{Q}^k} : \mathbb{Q}^k \rightarrow \mathbb{R}$ spełniającego (i) i (ii), gdzie $a, b, x \in \mathbb{Q}^k$. Funkcja \bar{F} ma nie tylko przeliczalną dziedzinę, ale także zbiór narzuconych na nią warunków postaci (i) i (ii) jest przeliczalny.

Teraz rozpatrzmy dowolne \mathbb{T} . Oznaczmy klasę uogólnionych cylindrów borelowskich

$$\mathcal{R}_0^{\mathbb{T}} := \{ \text{cyl}_{(t_1, \dots, t_k)}^{\mathbb{T}}(A) : A \in \sigma(\mathcal{R}^k), t_1, \dots, t_k \in \mathbb{T}, k \in \mathbb{N} \}. \quad (\text{A.9})$$

Klasa $\mathcal{R}_0^{\mathbb{T}}$ jest ciałem.

Twierdzenie A.11. (o procesie, Billingsley, 1979, sekcja 36) *Jeżeli dla dowolnych $k \in \mathbb{N}$ i $t_i \in \mathbb{T}$ istnieją skończone miary μ_{t_1, \dots, t_k} na $\sigma(\mathcal{R}^k)$ spełniające warunki*

(i) $\mu_{t_1, \dots, t_k} = \mu_{t_{\pi_1}, \dots, t_{\pi_k}} \circ \phi_{\pi}^{-1}$, gdzie π jest dowolną permutacją ciągu $(1, \dots, k)$, zaś $\phi_{\pi}(x_{\pi_1}, \dots, x_{\pi_k}) := (x_1, \dots, x_k)$,

(ii) $\mu_{t_1, \dots, t_k}(A) = \mu_{t_1, \dots, t_k, t_{k+1}}(A \times \mathbb{R})$ dla każdego $A \in \sigma(\mathcal{R}^k)$,

to istnieje dokładnie jedna skończona miara μ na $\sigma(\mathcal{R}_0^{\mathbb{T}})$ taka, że $\mu(\text{cyl}_{(t_1, \dots, t_k)}^{\mathbb{T}}(A)) = \mu_{t_1, \dots, t_k}(A)$ dla każdego $A \in \sigma(\mathcal{R}^k)$.

Odwrotnie, jeżeli μ jest skończoną miarą na $\sigma(\mathcal{R}_0^{\mathbb{T}})$, to $\mu_{t_1, \dots, t_k}(A) := \mu(\text{cyl}_{(t_1, \dots, t_k)}^{\mathbb{T}}(A))$ są miarami skończonymi na $\sigma(\mathcal{R}^k)$ i spełniają warunki (i) i (ii).

Dowód twierdzenia A.11 przy danych μ_{t_1, \dots, t_k} składa z dwóch części. Po pierwsze dowodzi się, że miara μ jest jednoznacznie określona na ciele $\mathcal{R}_0^{\mathbb{T}}$ i addytywna. W drugim, trudniejszym, kroku dowodzi się, że miara μ jest σ -addytywna na $\mathcal{R}_0^{\mathbb{T}}$. (Jest to własność mocniejsza niż σ -addytywność na zbiorze uogólnionych cylindrów $\text{cyl}_{(t_1, \dots, t_k)}^{\mathbb{T}}(A)$ o ustalonych t_1, \dots, t_k .) Dopiero wówczas można skorzystać z twierdzenia A.3, z którego wynika, że μ jest miarą na $\sigma(\mathcal{R}_0^{\mathbb{T}})$.

A.3. Kodowanie i mierzalność funkcji rzeczywistych

Rozważania teorii informacji wiążą się dość ściśle z tematyką transformacji bijektywnie mierzalnych. W niniejszej pracy interesują nas szczególnie bijektywnie mierzalne transformacje przeliczalnych procesów o wartościach rzeczywistych lub dyskretnych. Przykładowo twierdzenie 5.1 orzeka, że niefinitarność procesu może być skutkiem istnienia pewnej ciągłej rzeczywistej zmiennej losowej, która jest funkcją zarówno przeszłości jak i przyszłości procesu. Jeżeli rozważany proces jest dyskretny, to wygodnie jest móc reprezentować ową zmienną rzeczywistą jako ciąg zmiennych o wartościach dyskretnych i vice versa. Najpierw jednak należy wykazać bijektywną mierzalność tejże reprezentacji.

Twierdzenie A.12. *Niech (Ω, \mathcal{J}, P) będzie przestrzenią probabilistyczną, zaś $(\Omega_1, \mathcal{G}_1)$ i $(\Omega_2, \mathcal{G}_2)$ niech będą przestrzeniami mierzalnymi. Niech $f : \Omega_1 \rightarrow \Omega_2$, $X : \Omega \rightarrow \Omega_1$ oraz $Y = f(X) : \Omega \rightarrow \Omega_2$ będą dowolnymi funkcjami (niekoniecznie mierzalnymi). Oznaczmy $\mathcal{F}_1 := \mathcal{G}_1|_{X(\Omega)}$, $\mathcal{F}_2 := \mathcal{G}_2|_{Y(\Omega)}$.*

(i) *Jeżeli f jest mierzalna $\mathcal{G}_1/\mathcal{G}_2$, to $f|_{X(\Omega)}$ jest mierzalna $\mathcal{F}_1/\mathcal{F}_2$. Jeżeli X jest zmienną losową o $\tilde{\sigma}(X) := \mathcal{F}_1$ i $f|_{X(\Omega)}$ jest mierzalna $\mathcal{F}_1/\mathcal{F}_2$, to dla $\tilde{\sigma}(Y) := \mathcal{F}_2$ funkcja Y jest zmienną losową o $\sigma(Y) \subset \sigma(X)$.*

(ii) Jeżeli f jest bijektywnie mierzalna $\mathcal{G}_1/\mathcal{G}_2$, to $f|_{X(\Omega)}$ jest bijektywnie mierzalna $\mathcal{F}_1/\mathcal{F}_2$. Jeżeli X jest zmienną losową o $\tilde{\sigma}(X) := \mathcal{F}_1$ i $f|_{X(\Omega)}$ jest bijektywnie mierzalna $\mathcal{F}_1/\mathcal{F}_2$, to dla $\tilde{\sigma}(Y) := \mathcal{F}_2$ funkcja Y jest zmienną losową o $\sigma(Y) = \sigma(X)$.

NB. W pkt (i) i (ii) nie wymagamy, aby zachodziło $X(\Omega) \in \mathcal{G}_1$ i $Y(\Omega) \in \mathcal{G}_2$.

Dowód:

(i) Dla każdego $A_2 \in \mathcal{F}_2$ istnieje $A'_2 \in \mathcal{G}_2$ takie, że $A_2 = A'_2 \cap f(X(\Omega))$, $f^{-1}(A'_2) \in \mathcal{G}_1$ i $f^{-1}(A'_2) \cap X(\Omega) \in \mathcal{F}_1$. Stąd $(f|_{X(\Omega)})^{-1}(A_2) = (f|_{X(\Omega)})^{-1}(A'_2 \cap f(X(\Omega))) = f^{-1}(A'_2 \cap f(X(\Omega))) \cap X(\Omega) = f^{-1}(A'_2) \cap f^{-1}(f(X(\Omega))) \cap X(\Omega) = f^{-1}(A'_2) \cap X(\Omega) \in \mathcal{F}_1$, czyli $f|_{X(\Omega)}$ jest mierzalna $\mathcal{F}_1/\mathcal{F}_2$. Jeżeli X jest zmienną losową o $\tilde{\sigma}(X) := \mathcal{F}_1$, czyli mierzalną $\mathcal{J}/\mathcal{F}_1$, to Y jest mierzalna $\mathcal{J}/\mathcal{F}_2$, czyli dla $\tilde{\sigma}(Y) := \mathcal{F}_2$ także jest zmienną losową. Mamy $(f|_{X(\Omega)})^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1$, czyli $\sigma(Y) = Y^{-1}(\mathcal{F}_2) = X^{-1}(f|_{X(\Omega)})^{-1}(\mathcal{F}_2) \subset X^{-1}(\mathcal{F}_1) = \sigma(X)$.

(ii) Teza wynika z dwukrotnego zastosowania pkt. (i). □

Dzięki powyższemu twierdzeniu można się nie przejmować tym, że zbiory wartości zmiennych powiązanych bijekcją różnią się od maksymalnych dziedzin i przeciwdziedzin dla danej bijekcji mierzalnej. Równość σ -ciał generowanych przez zmienne będzie zachodzić nawet i w takim przypadku (por. twierdzenie 1.17).

Twierdzenie A.13. Niech (Ω, \mathcal{J}) i (Ω', \mathcal{J}') będą przestrzeniami mierzalnymi, a \mathcal{P}' — π -układem takim, że $\mathcal{J}' = \sigma_{\Omega'}(\mathcal{P}')$. Funkcja $f : \Omega \rightarrow \Omega'$ jest mierzalna \mathcal{J}/\mathcal{J}' , jeżeli $f^{-1}(\mathcal{P}') \subset \mathcal{J}$.

Dowód: Para $(\Omega'', \mathcal{J}'') := (f^{-1}(\Omega'), f^{-1}(\mathcal{J}'))$ jest przestrzenią mierzalną, gdzie $\Omega'' = \Omega$. Zbiór $\mathcal{P}'' := f^{-1}(\mathcal{P}')$ jest π -układem spełniającym $\mathcal{J}'' = \sigma_{\Omega''}(\mathcal{P}'') = \sigma_{\Omega}(\mathcal{P}'')$. Z twierdzenia o π - λ -układach (Billingsley, 1979, sekcja 2) dla każdego π -układu \mathcal{P}'' spełniającego $\mathcal{J}'' = \sigma_{\Omega}(\mathcal{P}'')$ relacja $\mathcal{P}'' \subset \mathcal{J}$ implikuje $\mathcal{J}'' \subset \mathcal{J}$. Stąd f jest mierzalna \mathcal{J}/\mathcal{J}' . □

Zmienne losowe, których σ -ciała są izomorficzne z σ -ciałami zmiennych rzeczywistych, warto jest móc nazywać jednym słowem.

Definicja A.14. (zmienna borelowska) Zmienną U nazywamy borelowską zmienną losową, jeżeli istnieje zmienna rzeczywista $Y = f(U)$, gdzie f jest bijektywnie mierzalna $\tilde{\sigma}(U)/\tilde{\sigma}(Y)$ zaś $\tilde{\sigma}(Y) = f(\tilde{\sigma}(U))$.

Pokażemy teraz, że szeroka klasa procesów stochastycznych to w istocie borelowskie zmienne losowe.

Twierdzenie A.15. Niech $D \geq 2$ będzie ustaloną liczbą naturalną. Korzystając z oznaczenia (A.9), określmy $\mathbb{B} := \{0, 1, \dots, D-1\}^{\mathbb{N}}$, $\mathcal{B} := \sigma(\mathcal{R}_0^{\mathbb{N}})|_{\mathbb{B}}$,

$$U := \left\{ x_{\mathbb{N}} \in \mathbb{B} : \limsup_{n \rightarrow \infty} x_n = 0 \right\} \quad (\text{A.10})$$

oraz $U^c := \mathbb{B} \setminus U$. Funkcja

$$g_D : U^c \ni x_{\mathbb{N}} \mapsto \sum_{i=1}^{\infty} D^{-i} x_i \in (0, 1] \quad (\text{A.11})$$

jest bijektywnie mierzalna $\mathcal{B}|_{U^c}/\mathcal{R}|_{(0,1]}$.

Dowód: Dla dowolnego $x_{\mathbb{N}} \in \mathbb{B}$ mamy $g_D(x_{\mathbb{N}}) = \sum_{i=1}^{\infty} D^{-i} x_i \in [0, 1]$. Zbiór $g_D(U) \setminus \{0\}$ to zbiór liczb y z przedziału $(0, 1]$, dla których zbiór rozwinięć D -arnych $g_D^{-1}(\{y\})$ ma

więcej niż jeden element. Dla każdego $y \in g_D(U^c) = (0, 1]$ zbiór rozwinięć $g_D^{-1}(\{y\}) \setminus U$ jest jednak jednoelementowy. Funkcja $g_D : U^c \rightarrow (0, 1]$ jest zatem bijekcją.

Przejdźmy do dowodu mierzalności. Ustalmy $g_D(x_{\mathbb{N}}) = y$, $g_D(x'_{\mathbb{N}}) = y'$, gdzie $x_{\mathbb{N}}, x'_{\mathbb{N}} \in U^c$. Nierówność $y \leq y'$ zachodzi wtedy i tylko wtedy, gdy $x_{1:m-1} = x'_{1:m-1}$ oraz $x_m < x'_m$ dla pewnego $m \geq 1$.

Klasa zbiorów $\{y \in (0, 1] : y \leq y'\}$, gdzie $y' \in (0, 1]$, jest π -układem generującym σ -ciało $\mathcal{R}|_{(0,1]}$. Mamy

$$g_D^{-1}(\{y \in (0, 1] : y \leq y'\}) = \{x'_{\mathbb{N}}\} \cup \bigcup_{m \in \mathbb{N}} \{x_{\mathbb{N}} \in U^c : x_{1:m-1} = x'_{1:m-1}, x_m < x'_m\} \in \mathcal{B}|_{U^c},$$

czyli g_D jest mierzalna $\mathcal{B}|_{U^c}/\mathcal{R}|_{(0,1]}$ na mocy twierdzenia A.13. Z kolei klasa zbiorów $\{x_{\mathbb{N}} \in U^c : x_{1:n} = b_{1:n}\}$, gdzie $b_i \in \{0, 1, \dots, D-1\}$, $n \in \mathbb{N}$, jest π -układem generującym σ -ciało $\mathcal{B}|_{U^c}$. Mamy $g_D(\{x_{\mathbb{N}} : x_{1:n} = b_{1:n}\}) \in \mathcal{R}|_{(0,1]}$, czyli g_D^{-1} jest mierzalna $\mathcal{R}|_{(0,1]}/\mathcal{B}|_{U^c}$ także na mocy twierdzenia A.13. \square

Twierdzenie A.16. *Dla dowolnej zmiennej rzeczywistej Y istnieje proces dyskretny $X_{\mathbb{N}}$ o wartościach ze zbioru $\{0, 1, \dots, D-1\}$ spełniający $\sigma(Y) = \sigma(X_{\mathbb{N}})$ oraz funkcja f bijektywnie mierzalna $\tilde{\sigma}(X_{\mathbb{N}})/\tilde{\sigma}(Y)$ taka, że $Y = f(X_{\mathbb{N}})$. Odwrotnie, dla dowolnego procesu dyskretnego $X_{\mathbb{N}}$ o wartościach ze zbioru $\{0, 1, \dots, D-1\}$ istnieje zmienna rzeczywista Y spełniająca $\sigma(Y) = \sigma(X_{\mathbb{N}})$ oraz funkcja f bijektywnie mierzalna $\tilde{\sigma}(X_{\mathbb{N}})/\tilde{\sigma}(Y)$ taka, że $Y = f(X_{\mathbb{N}})$.*

Dowód: Niech Y będzie dowolną zmienną rzeczywistą. Na mocy twierdzeń A.12 i A.15 teza pierwszej części twierdzenia jest spełniona dla $f = h^{-1} \circ g_D$, gdzie $h(x) = (\operatorname{tgh}(x) + 1)/2$ oraz $X_{\mathbb{N}}$, gdzie $X_{\mathbb{N}}(\Omega) \subset U^c$ i $X_{\mathbb{N}} = f^{-1}(Y)$. Odwrotnie, niech $X_{\mathbb{N}}$ będzie procesem o wartościach $\{0, 1, \dots, D-1\}$. Na mocy twierdzeń A.12 i A.15, teza drugiej części twierdzenia jest spełniona dla $f = g_{D+1}$. (Bierzemy g_{D+1} zamiast g_D aby zapewnić iniektowność odwzorowania.) \square

W istocie możemy pójść dalej i reprezentować na pojedynczej zmiennej rzeczywistej Y lub ciągu losowych bitów $X_{\mathbb{N}}$ dowolny proces $U_{\mathbb{T}}$, gdzie zmienne U_i są dowolnymi zmiennymi dyskretnymi lub rzeczywistymi, a zbiór \mathbb{T} jest przeliczalny.

Twierdzenie A.17. *Dla dowolnego procesu $U_{\mathbb{T}}$, gdzie zmienne U_i są zmiennymi dyskretnymi lub rzeczywistymi, a zbiór \mathbb{T} jest przeliczalny, istnieje zmienna rzeczywista Y spełniająca $\sigma(Y) = \sigma(U_{\mathbb{T}})$ oraz funkcja f bijektywnie mierzalna $\tilde{\sigma}(U_{\mathbb{T}})/\tilde{\sigma}(Y)$ taka, że $Y = f(U_{\mathbb{T}})$.*

Dowód: Bijekcja $U_{\mathbb{T}}$ na Y jest następująca. Dla każdej zmiennej dyskretniej U_i jej wartości numerujemy liczbami naturalnymi za pomocą bijekcji $\phi : U_i(\Omega) \rightarrow \mathbb{N}$. W ten sposób otrzymujemy proces $U'_{\mathbb{T}}$, gdzie $U'_i := \phi(U_i)$, gdy U_i jest dyskretne, zaś $U'_i := U_i$, gdy U_i jest rzeczywiste. Tak uzyskane zmienne U'_i traktujemy jako zmienne rzeczywiste. Korzystając z twierdzenia A.16, konstruujemy zmienne X'_{ij} o wartościach ze zbioru $\{0, 1\}$ takie, że $U'_i = h^{-1} \circ g_2((X'_{ij})_{j \in \mathbb{N}})$. Obierając dowolną bijekcję $\psi : \mathbb{T} \times \mathbb{N} \rightarrow \mathbb{N}$, otrzymujemy proces $X_{\mathbb{N}}$, gdzie $X_t := X'_{\psi^{-1}(t)}$. Korzystając ponownie z twierdzenia A.16, konstruujemy zmienną rzeczywistą $Y = g_3(X_{\mathbb{N}})$. \square

Zauważmy, że twierdzenie A.17 przydaje się nie tylko w teorii informacji, ale także w dyskusji miar losowych (w tym regularnego prawdopodobieństwa warunkowego) i ukrytych łańcuchów Markowa, o czym wspominamy w dalszych dodatkach. Reprezentując przeliczalne zbiory zmiennych rzeczywistych jako przeliczalne zbiory zmiennych binarnych, można też otrzymać dość podstawowe rezultaty dotyczące kryteriów zbieżności sto-

chastycznej procesów oraz kryteriów wyrażalności procesów jako funkcji innych procesów, por. twierdzenia 3.5 oraz A.20.

Warianty twierdzeń A.16, A.17 oraz 1.28 są znane (Kallenberg, 1997, twierdzenie A1.6; Breiman, 1992, twierdzenie A.47). Mimo to nie natrafiliśmy na dostatecznie szczegółową dyskusję równości σ -ciał generowanych przez powiązane bijekcją zmienne losowe (twierdzenie A.12). Ścisła równość takich σ -ciał jest tymczasem istotna w rozważaniach teorii informacyjnych, gdyż miary informacji są funkcjami σ -ciał generowanych przez zmienne, a nie samych wartości tych zmiennych.

A.4. Równość prawie na pewno

Ścisłą równość σ -ciał nie zawsze można zapewnić. Do wielu zastosowań wystarcza, aby rozpatrywane klasy zdarzeń równały się sobie z dokładnością do zbiorów miary 0. Niech $P^* : \mathcal{M}_P \rightarrow \mathbb{R}$ będzie rozszerzeniem miary P określonym w twierdzeniu A.3 pkt. (ii).

Definicja A.18. (uzupełnienie) *Uzupełnieniem σ -ciała \mathcal{G} nazywać będziemy najmniejsze σ -ciało \mathcal{G}^P zawierające \mathcal{G} oraz wszystkie zbiory P^* -miary 0.*

Zauważmy, że $\mathcal{G}^P = \{A' \in \mathcal{M}_P : \exists_{A \in \mathcal{G}} P^*(A \ominus A') = 0\}$. Dla uproszczenia notacji dalej będziemy zakładać, że przestrzeń probabilistyczna jest zupełna, tzn. $\mathcal{M}_P = \mathcal{J}$, $P^* = P$.

Dla odmiany nie zakładamy, że σ -ciała zmiennych losowych są zupełne. Nie chcemy bowiem tracić ścisłego izomorfizmu klas $\sigma(X)$ i $\tilde{\sigma}(X)$ dla żadnej zmiennej X . W ogólnym przypadku zachodzi zawieranie $\sigma(X)^P \supset X^{-1}(\tilde{\sigma}(X)^{P \circ X}) \supset \sigma(X) := X^{-1}(\tilde{\sigma}(X))$ i te trzy klasy mogą się różnić.

Dla miary prawdopodobieństwa P o zdaniu losowym ϕ mówimy, że ϕ prawie na pewno, jeżeli $P^*(\neg\phi) = 0$, gdzie P^* jest rozszerzeniem miary P . Przypomnijmy, że jeżeli $X = Y$ i $\tilde{\sigma}(X) = \tilde{\sigma}(Y)$ dla zmiennych losowych X i Y , to $\sigma(X) = \sigma(Y)$. Jeżeli $X = f(Y)$, gdzie f jest mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(X)$, to $\sigma(X) \subset \sigma(Y)$. Jeżeli $X = f(Y)$, gdzie f jest bijektywnie mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(X)$, to $\sigma(X) = \sigma(Y)$. Wszystkie trzy własności można uogólnić na równość prawie na pewno.

Twierdzenie A.19. *Niech X i Y będą dowolnymi zmiennymi losowymi.*

- (i) *Mamy $X(X = Y) = Y(X = Y)$.*
- (ii) *Jeżeli $X = Y$ prawie na pewno i $\tilde{\sigma}(X)|_{X(X=Y)} = \tilde{\sigma}(Y)|_{Y(X=Y)}$, to $\sigma(X)^P = \sigma(Y)^P$.*
- (iii) *Niech $Z = f(Y)$, gdzie f jest mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(Z)$. Jeżeli $X = f(Y)$ prawie na pewno i $\tilde{\sigma}(X)|_{X(X=Z)} = \tilde{\sigma}(Z)|_{Z(X=Z)}$, to $\sigma(X)^P \subset \sigma(Y)^P$.*
- (iv) *Niech $Z = f(Y)$, gdzie f jest bijektywnie mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(Z)$. Jeżeli $X = f(Y)$ prawie na pewno i $\tilde{\sigma}(X)|_{X(X=Z)} = \tilde{\sigma}(Z)|_{Z(X=Z)}$, to $\sigma(X)^P = \sigma(Y)^P$.*

Dowód:

- (i) $X(X = Y) = \{X(\omega) : \omega \in (X = Y)\} = \{Y(\omega) : \omega \in (X = Y)\} = Y(X = Y)$.
- (ii) Zauważmy, że $X^{-1}(X(X = Y)) \supset (X = Y)$. Ponieważ dla każdego $A \in \sigma(X)$ mamy $A \cap X^{-1}(X(X = Y)) = X^{-1}[X(A) \cap X(X = Y)]$, to $A \cap (X = Y) = X^{-1}(X(A) \cap X(X = Y)) \cap (X = Y)$. Inaczej mówiąc, $\sigma(X)|_{(X=Y)} = X^{-1}(\tilde{\sigma}(X)|_{X(X=Y)})|_{(X=Y)} = (X|_{(X=Y)})^{-1}(\tilde{\sigma}(X)|_{X(X=Y)})$. Wykorzystując analogiczną równość dla zmiennej Y otrzymujemy

$$\begin{aligned} \sigma(X)^P &= [\sigma(X)|_{(X=Y)}]^P = [(X|_{(X=Y)})^{-1}(\tilde{\sigma}(X)|_{X(X=Y)})]^P \\ &= [(Y|_{(X=Y)})^{-1}(\tilde{\sigma}(Y)|_{Y(X=Y)})]^P = [\sigma(Y)|_{(X=Y)}]^P = \sigma(Y)^P. \end{aligned}$$

- (iii) Teza wynika z pkt. (ii) i twierdzenia A.12 pkt. (i).

(iv) Teza wynika z pkt. (ii) i twierdzenia A.12 pkt. (ii). □

Dla zmiennych dyskretnych prawdziwe jest twierdzenie rekonstruujące funkcje mierzalne z zawierających się prawie na pewno σ -ciała.

Twierdzenie A.20. *Niech X i Y będą zmiennymi losowymi, gdzie X jest dyskretną zmienną losową (równoważnie $\sigma(X)$ jest σ -ciałem dyskretnym, por. definicja 1.15). Zachodzą równoważności*

(i) $\sigma(X) \subset \sigma(Y) \iff$ istnieje f taka, że $X = f(Y)$,

(ii) $\sigma(X)^P \subset \sigma(Y)^P \iff$ istnieje f taka, że $X = f(Y)$ prawie na pewno, gdzie zakładamy, że f jest mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(X)$.

Dowód:

(i) Jeżeli $X = f(Y)$, gdzie f jest mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(X)$, to oczywiście $\sigma(X) \subset \sigma(Y)$.

Odwrotnie, jeżeli $\sigma(X) \subset \sigma(Y)$, to dla każdego $x \in X(\Omega)$ mamy $(X = x) \in \sigma(Y)$.

Zbiory postaci $(X = x)$ są rozłączne, a zatem zbiory $Y(X = x)$ są także rozłączne.

Funkcję f możemy więc określić jako $f(y) := x$ dla każdego $y \in Y(X = x)$.

(ii) Jeżeli $X = f(Y)$ prawie na pewno, gdzie f jest mierzalna $\tilde{\sigma}(Y)/\tilde{\sigma}(X)$, to dla każdego $x \in X(\Omega)$ istnieją zbiory $(f(Y) = x) \in \sigma(Y)$ takie, że $P((X = x) \ominus (f(Y) = x)) = 0$, czyli $\sigma(X)^P \subset \sigma(Y)^P$.

Odwrotnie, jeżeli $\sigma(X)^P \subset \sigma(Y)^P$, to dla każdego $x \in X(\Omega)$ istnieją zbiory $B_x \in \sigma(Y)$ takie, że $P((X = x) \ominus B_x) = 0$. Ponieważ $(X = x) \cap (X = x') = \emptyset$ dla $x \neq x'$, to zachodzi także $P(B_x \cap B_{x'}) = 0$ dla $x \neq x'$.

Dowolny zbiór $B_x \in \sigma(Y)$ spełnia $B_x = Y^{-1}(\hat{B}_x) = (Y \in \hat{B}_x)$ dla pewnego $\hat{B}_x \in \tilde{\sigma}(Y)$.

Bez zmniejszania ogólności położmy $X(\Omega) \subset \mathbb{N}$. Definiujemy rozłączne zbiory $\tilde{B}_x := \hat{B}_x \setminus \bigcup_{x' < x} \hat{B}_{x'} \in \tilde{\sigma}(Y)$ oraz $f(y) := x$ dla $y \in \tilde{B}_x$. Ponieważ $(Y \in \tilde{B}_x) = B_x \setminus \bigcup_{x' < x} B_{x'}$, to $P(B_x \ominus (Y \in \tilde{B}_x)) = 0$. Stąd $P((X = x) \ominus (Y \in \tilde{B}_x)) = 0$, czyli $X = f(Y)$ prawie na pewno. □

Korzystając z twierdzenia A.16, twierdzenie A.20 można też uogólnić na przypadek, gdy X jest zmienną borelowską.

A.5. Prawdopodobieństwo warunkowe jako miara i martyngał

Fundamentalną rolę w teoriomiarowej definicji prawdopodobieństwa warunkowego odgrywa twierdzenie Lebesgue'a-Radona-Nikodyma o rozkładzie miary.

Definicja A.21. (miary absolutnie ciągłe i wzajemnie osobliwe) *Niech ν i μ będą miarami na przestrzeni mierzalnej (Ω, \mathcal{G}) . Oznaczamy $\nu \perp \mu$ i mówimy, że ν i μ są wzajemnie osobliwe, gdy istnieje $A \in \mathcal{G}$ takie, że $\nu(A) = \mu(A^c) = 0$. Oznaczamy $\nu \ll \mu$ i mówimy, że ν jest absolutnie ciągła względem μ , gdy z $\mu(A) = 0$ wynika $\nu(A) = 0$ dla każdego $A \in \mathcal{G}$.*

Dla dowolnej σ -skończonej miary μ będziemy mówić o zdaniu losowym ϕ , że ϕ μ -prawie na pewno, jeżeli $\mu^*(-\phi) = 0$, gdzie μ^* jest rozszerzeniem miary μ na przestrzeń zupełną.

Twierdzenie A.22. (o rozkładzie miary) *Niech ν będzie σ -skończoną miarą na przestrzeni mierzalnej (Ω, \mathcal{G}_1) , zaś μ — σ -skończoną miarą na (Ω, \mathcal{G}_2) , gdzie $\mathcal{G}_1 \subset \mathcal{G}_2$. Istnieje dokładnie jedna para miar ν_a, ν_s na (Ω, \mathcal{G}_1) takich, że $\nu = \nu_a + \nu_s$, gdzie $\nu_a \perp \mu|_{\mathcal{G}_1}$ oraz $\nu_s \ll \mu|_{\mathcal{G}_1}$. Dla pary miar $\nu_s \ll \mu|_{\mathcal{G}_1}$ istnieje także nieujemna pochodna Radona-Nikodyma $dv_s/d\mu : \Omega \rightarrow \mathbb{R}$ taka, że*

(i) $dv_s/d\mu$ jest mierzalna $\mathcal{G}_1/\mathcal{R}$,

(ii) $\nu_s(A) = \int_A (dv_s/d\mu)d\mu$ dla każdego $A \in \mathcal{G}_1$.

Pochodna Radona-Nikodyma jest jednoznaczna z dokładnością do zbiorów μ -miary 0, tzn. jeżeli g spełnia te same warunki co $dv_s/d\mu$, to $dv_s/d\mu = g$ μ -prawie na pewno.

NB. Rudin (1986, sekcja 6.9) dowodzi powyższego twierdzenia dla $\mathcal{G}_1 = \mathcal{G}_2$. Uogólnienie na $\mathcal{G}_1 \neq \mathcal{G}_2$ wynika z własności całki Lebesgue'a.

Istnieją dwa dość odmienne dowody twierdzenia A.22. Jeden wykorzystuje rozkład Hahna-Jordana (Billingsley, 1979, sekcje 32 i 33), a drugi zupełność przestrzeni L^2 (Rudin, 1986, sekcja 6.9; Kallenberg, 1997, rozdział 5).

Poczyńmy kilka uwag.

- (i) Warunki $\nu \ll \mu$ i $\mu \ll \rho$ implikują $\nu \ll \rho$ oraz $dv/d\rho = dv/d\mu \cdot d\mu/d\rho$ ρ -prawie na pewno. Stąd warunki $\nu \ll \mu$ i $\mu \ll \nu$ implikują $dv/d\mu \cdot d\mu/d\nu = 1$ ν -prawie na pewno.
- (ii) Dla miary obciętej $\mu|_{\mathcal{G}}$ zachodzi $d\mu|_{\mathcal{G}}/d\mu = 1$ μ -prawie na pewno. Stąd $dv/d\mu|_{\mathcal{G}} = dv/d\mu$ μ -prawie na pewno, jeżeli $\nu \ll \mu$, ν jest miarą na (Ω, \mathcal{G}_1) , μ jest miarą na (Ω, \mathcal{G}_2) , zaś $\mathcal{G}_1 \subset \mathcal{G} \subset \mathcal{G}_2$.
- (iii) Zwykle $\mu(dv/d\mu \neq dv|_{\mathcal{G}}/d\mu) \neq 0$.

Z rozkładu Hahna-Jordana dowolną miarę rzeczywistą ν na \mathcal{G} można przedstawić jako sumę $\nu = \nu_+ - \nu_-$, gdzie ν_+ i ν_- są miarami skończonymi spełniającymi $\nu_+ \perp \nu_-$. Dla miary rzeczywistej ν piszmy $\nu \ll \mu$, gdy $\nu_+, \nu_- \ll \mu$. Dowolną miarę rzeczywistą ν taką, że $\nu \ll \mu$ dla pewnej miary σ -skończonej μ , można przedstawić jako $\nu(A) = \int_A (dv/d\mu)d\mu$, gdzie pochodna $dv/d\mu = dv_+/d\mu - dv_-/d\mu$ jest mierzalna \mathcal{G}/\mathcal{R} oraz $\int |dv/d\mu|d\mu < \infty$.

Dowolną miarę rzeczywistą ν możemy zadać jednoznacznie także przez pewien ciąg funkcji przybliżających pochodną $dv/d\mu$.

Definicja A.23. (martyngał) Ciąg funkcji i σ -ciał $(f_i, \mathcal{G}_i)_{i \in \mathbb{Z}}$ nazywa się martyngałem względem σ -skończonej miary μ na przestrzeni mierzalnej (Ω, \mathcal{G}) , jeżeli każda funkcja f_i jest mierzalna $\mathcal{G}_i/\mathcal{R}$ i zachodzą warunki $\int_{\Omega} |f_i|d\mu < \infty$, $\mathcal{G}_i \subset \mathcal{G}_{i+1} \subset \mathcal{G}$ oraz $\int_A f_i d\mu = \int_A f_{i+1} d\mu$ dla każdego $A \in \mathcal{G}_i$.

Prostym przykładem martyngału jest $(f_i, \mathcal{G}_i)_{i \in \mathbb{Z}}$, gdzie $f_i = dv|_{\mathcal{G}_i}/d\mu$ dla pewnej miary rzeczywistej $\nu \ll \mu$. W istocie taka miara ν istnieje dla dowolnego martyngału.

Twierdzenie A.24. (o zbieżności martyngałów) Niech $(f_i, \mathcal{G}_i)_{i \in \mathbb{Z}}$ będzie martyngałem względem miary μ na (Ω, \mathcal{G}) . Oznaczmy $\mathcal{G}_{-\infty} = \bigcap_{i \in \mathbb{Z}} \mathcal{G}_i$ oraz $\mathcal{G}_{\infty} = \sigma(\bigcup_{i \in \mathbb{Z}} \mathcal{G}_i)$. Funkcję określoną jako $\nu(A) := \int_A f_i d\mu$ dla każdego $i \in \mathbb{Z}$ i $A \in \mathcal{G}_i$ można jednoznacznie rozszerzyć do miary rzeczywistej ν na \mathcal{G}_{∞} . Zachodzą równości

$$\lim_{n \rightarrow \infty} f_n = f_{\infty} \quad \mu\text{-prawie na pewno,} \quad (\text{A.12})$$

$$\lim_{n \rightarrow -\infty} f_n = f_{-\infty} \quad \mu\text{-prawie na pewno,} \quad (\text{A.13})$$

gdzie $f_i = dv|_{\mathcal{G}_i}/d\mu$ dla $i \in \{-\infty, \infty\}$.

NB. Istnienie mierzalnych granic po lewej stronie (A.12)–(A.13) wynika ze standardowego twierdzenia Dooba o zbieżności martyngałów (Doob, 1953, rozdział 8, twierdzenie 4.3). Wówczas określa się $\nu(A) := \int_A (\lim_{n \rightarrow \infty} f_n)d\mu$, skąd wynikają pozostałe tezy.

Definicja A.25. (warunkowa wartość oczekiwana) Dla rzeczywistej zmiennej losowej X spełniającej $\int |X|dP < \infty$ i σ -ciała \mathcal{G} określamy warunkową wartość oczekiwaną $\langle X|\mathcal{G} \rangle$ jako rzeczywistą zmienną losową

$$\langle X|\mathcal{G} \rangle := \frac{dX_P|_{\mathcal{G}}}{dP}, \quad X_P(A) := \int_A X dP. \quad (\text{A.14})$$

Podobnie, jak dla innych zmiennych rzeczywistych, przyjmujemy $\tilde{\sigma}(\langle X|\mathcal{G} \rangle) := \mathcal{R}|_{\langle X|\mathcal{G} \rangle(\Omega)}$.

Mamy $\langle X|\mathcal{J} \rangle = X$ oraz $\langle X|\{\emptyset, \Omega\} \rangle = \langle X \rangle$, gdzie $\langle X \rangle = \int X dP$. Ciąg $(X_i, \mathcal{G}_i)_{i \in \mathbb{Z}}$ jest martyngałem dla $X_i = \langle X|\mathcal{G}_i \rangle$ oraz $\mathcal{G}_i \subset \mathcal{G}_{i+1} \subset \mathcal{J}$, przy czym z definicji martyngału $X_i = \langle X_{i+1}|\mathcal{G}_i \rangle$.

Zauważmy jeszcze jeden fakt. Przez $L^q(\mathcal{G}, \mu)$ oznaczmy przestrzeń funkcji f mierzalnych \mathcal{G}/\mathcal{R} takich, że $\int |f|^q d\mu < \infty$. Jeżeli $X \in L^2(\mathcal{J}, P)$, to zmienna $\langle X|\mathcal{G} \rangle$ może być zdefiniowana jako rzut ortogonalny zmiennej X na podprzestrzeń $L^2(\mathcal{G}, P)$, tzn. $\langle X|\mathcal{G} \rangle \in L^2(\mathcal{G}, P)$ oraz

$$\int [\langle X|\mathcal{G} \rangle - X] f dP = 0 \quad (\text{A.15})$$

dla każdej funkcji $f \in L^2(\mathcal{G}, P)$. Dla $X \in L^2(\mathcal{J}, P)$ definicje (A.14) i (A.15) są sobie równoważne, gdyż obie są równoważne stwierdzeniu, że $\langle X|\mathcal{G} \rangle \in L^2(\mathcal{G}, P)$ oraz $\int_G [\langle X|\mathcal{G} \rangle - X] dP = 0$ dla każdego $G \in \mathcal{G}$.

Definicja A.26. (prawdopodobieństwo warunkowe) *Prawdopodobieństwo warunkowe zdarzenia $A \in \mathcal{J}$ przy danym zdarzeniu $B \in \mathcal{J}$, gdzie $P(B) > 0$, to liczba $P(A|B) := P(A \cap B)/P(B)$. Prawdopodobieństwo warunkowe zdarzenia $A \in \mathcal{J}$ przy danym σ -ciele $\mathcal{G} \subset \mathcal{J}$ to rzeczywista zmienna losowa $P(A|\mathcal{G}) := \langle I_A|\mathcal{G} \rangle$, gdzie $I_A(\omega) := \llbracket \omega \in A \rrbracket$. Podobnie, jak dla innych zmiennych rzeczywistych, przyjmujemy $\tilde{\sigma}(P(A|\mathcal{G})) := \mathcal{R}|_{P(A|\mathcal{G})(\Omega)}$ (Billingsley, 1979, sekcja 33).*

Definicja $P(A|\mathcal{G})$ może być mało przejrzysta dla czytelników, którzy nie mieli dotychczas styczności z teorią miary, więc poczynimy kilka uwag.

- (i) Oznaczmy miarę skończoną $P_A : \mathcal{J} \ni G \mapsto P(A \cap G)$. Zachodzi $P_A|_{\mathcal{G}} \ll P$ oraz $P(A|\mathcal{G}) := dP_A|_{\mathcal{G}}/dP$.
- (ii) Jeżeli $\mathcal{G} = \sigma(\{B_k\}_{k \in \mathbb{N}})$, gdzie zbiory B_k są rozłączne oraz $P(B_k) > 0$, to dowolne $G \in \mathcal{G}$ ma postać $G = \bigcup_{k \in I_G} B_k$, $I_G \subset \mathbb{N}$. Wtedy $P(A \cap G) = \sum_{k \in I_G} P(A|B_k)P(B_k)$ i mamy

$$P(A|\mathcal{G})(\omega) = \sum_{k \in \mathbb{N}} P(A|B_k) \llbracket \omega \in B_k \rrbracket.$$

- (iii) Jeżeli \mathcal{G} zawiera zbiory o zerowym prawdopodobieństwie, to wartości prawdopodobieństwa warunkowego $P(A|\mathcal{G})(\omega)$ nie są jednoznacznie określone. Prawdopodobieństwa $P(P(A|\mathcal{G}) \leq r)$ są jednak określone jednoznacznie dla każdego $r \in \mathbb{R}$.
- (iv) Jeżeli $P(B) \in \{0, 1\}$ dla każdego $B \in \mathcal{G}$, to prawie na pewno $P(A|\mathcal{G}) = P(A)$. W szczególności $P(A|\{\emptyset, \Omega\})$ jest prawie na pewno równe $P(A)$.
- (v) Jeżeli $P(A \ominus \tilde{A}) = 0$, to $P(A|\mathcal{G}) = P(\tilde{A}|\mathcal{G})$ prawie na pewno.
- (vi) Dla $\mathcal{G}_i \subset \mathcal{G}_{i+1}$ ciąg $(P(A|\mathcal{G}_i), \mathcal{G}_i)_{i \in \mathbb{Z}}$ jest martyngałem względem miary P . Z twierdzenia A.24 dla $\mathcal{G}_\infty = \sigma(\bigcup_{i \in \mathbb{Z}} \mathcal{G}_i)$ zachodzi zbieżność

$$\lim_{n \rightarrow \infty} P(A|\mathcal{G}_n) = P(A|\mathcal{G}_\infty) \quad \text{prawie na pewno.} \quad (\text{A.16})$$

Twierdzenie A.27. (Billingsley, 1979, twierdzenie 33.2)

- (i) Dla $A \in \mathcal{J}$ prawie na pewno $P(A|\mathcal{G}) \geq 0$.
- (ii) Prawie na pewno $P(\Omega|\mathcal{G}) = 1$.

$$(iii) \text{ Dla rozłącznych } A_1, A_2, \dots \in \mathcal{J} \text{ prawie na pewno } P\left(\bigcup_{n \in \mathbb{N}} A_n|\mathcal{G}\right) = \sum_{n \in \mathbb{N}} P(A_n|\mathcal{G}).$$

Jak wspomnieliśmy, wartości $P(A|\mathcal{G})(\omega)$ nie są ściśle jednoznaczne. Dla każdego $A \in \mathcal{J}$ może istnieć więcej niż jedna wersja $P(A|\mathcal{G})$.

Definicja A.28. (miara warunkowa) Dla procesu $(P(A|\mathcal{G}))_{A \in \mathcal{F}}$ określmy zbiór

$$K = \{\omega \in \Omega : (P(A|\mathcal{G}))_{A \in \mathcal{F}} : \mathcal{F} \ni A \mapsto P(A|\mathcal{G})(\omega) \in \mathbb{R} \text{ jest miarą prawdopodobieństwa}\}. \quad (\text{A.17})$$

Proces stochastyczny $(P(A|\mathcal{G}))_{A \in \mathcal{F}}$ nazywa się *regularnym prawdopodobieństwem warunkowym (miarą warunkową)*, jeżeli $P(K)$ jest określone i $P(K) = 1$.

NB. Jeżeli istnieje regularne prawdopodobieństwo warunkowe $(P(A|\mathcal{G}))_{A \in \mathcal{F}}$, to można tak dobrać wersję $P(A|\mathcal{G})$, aby było ono miarą losową na \mathcal{F} .

Dla dowolnego procesu $(P(A|\mathcal{G}))_{A \in \mathcal{F}}$ oznaczmy klasę zdarzeń

$$W = \{(0 \leq P(A|\mathcal{G})) : A \in \mathcal{F}\} \cup \{(P(\Omega|\mathcal{G}) = 1)\} \\ \cup \left\{ \left(P \left(\bigcup_{n \in \mathbb{N}} A_n | \mathcal{G} \right) = \sum_{n \in \mathbb{N}} P(A_n | \mathcal{G}) \right) : A_1, A_2, \dots \in \mathcal{F} \text{ rozłączne} \right\}.$$

Mamy $K = \bigcap_{B \in W} B = (\bigcup_{B \in W} B^c)^c$, gdzie $P(B^c) = 0$ dla każdego $B \in W$. Jeżeli \mathcal{F} jest skończone, to W jest skończone, a więc $P(K) = 1$. Dla nieskończonego \mathcal{F} zbiór W jest nieprzeliczalny, a więc może zachodzić $P(K) < 1$. Przykłady przestrzeni probabilistycznych, dla których nierówność $P(K) < 1$ zachodzi, są znane (Swart, 1996).

Kwestia istnienia regularnych prawdopodobieństw jest istotna w twierdzeniu A.41 i jego zastosowaniach w rozdziale 1, więc przypomnijmy poniższe twierdzenie.

Twierdzenie A.29. (por. Chow i Teicher, 1978, twierdzenie 7.2.2) Dla dowolnego σ -ciała \mathcal{G} i zmiennej borelowskiej U istnieje miara warunkowa $(P(A|\mathcal{G}))_{A \in \sigma(U)}$.

NB. W oryginale twierdzenie było dowodzone dla zmiennej rzeczywistej. Dowód przenosi się na zmienną borelowską automatycznie.

Dowód twierdzenia A.29 wykorzystuje twierdzenie A.10 dla $k = 1$. Niech Y będzie zmienną rzeczywistą z definicji A.14. Definiuje się $\bar{F} : \mathbb{Q} \times \Omega \rightarrow \mathbb{R}$ jako $\bar{F}(q, \omega) := P(Y \leq q | \mathcal{G})(\omega)$. Funkcje $\bar{F}(\cdot, \omega)$ nie spełniają przeliczalnej liczby warunków postaci (i) i (ii) z twierdzenia A.10 tylko dla zbioru ω P -miary 0. Można zatem z \bar{F} odtworzyć regularne prawdopodobieństwo warunkowe $(P(A|\mathcal{G}))_{A \in \sigma(Y)} = (P(A|\mathcal{G}))_{A \in \sigma(U)}$.

Przypomnijmy także, że z twierdzenia A.17 zmienną borelowską jest dowolny ciąg dyskretnych lub rzeczywistych zmiennych losowych. Zatem miara warunkowa $(P(A|\mathcal{G}))_{A \in \sigma(U_{\mathbb{T}})}$ istnieje dla dowolnego procesu $U_{\mathbb{T}}$, gdzie U_i są zmiennymi rzeczywistymi lub dyskretnymi, zaś \mathbb{T} jest zbiorem przeliczalnym.

Wróćmy do elementarnych własności $P(A|\mathcal{G})$. Udowodnijmy kilka twierdzeń, które wykorzystujemy w pracy.

Twierdzenie A.30. Niech $A \in \mathcal{F}$. Prawie na pewno zachodzi równość

$$P(A | \sigma(P(A|\mathcal{G}))) = P(A | \sigma((P(B|\mathcal{G}))_{B \in \mathcal{F}})) = P(A|\mathcal{G}). \quad (\text{A.18})$$

Dowód: Funkcja $P(A|\mathcal{G})$ mierzalna $\mathcal{G}/\mathcal{R}|_{P(A|\mathcal{G})(\Omega)}$ jest rzeczywistą zmienną losową, gdzie $\tilde{\sigma}(P(A|\mathcal{G})) = \mathcal{R}|_{P(A|\mathcal{G})(\Omega)}$. Ponieważ $\sigma(P(A|\mathcal{G})) = P(A|\mathcal{G})^{-1}(\tilde{\sigma}(P(A|\mathcal{G})))$, to $\sigma(P(A|\mathcal{G})) \subset \sigma((P(B|\mathcal{G}))_{B \in \mathcal{F}}) \subset \mathcal{G}$.

Zmienna $P(A|\mathcal{G})$ jest równa prawie na pewno $P(A | \sigma(P(A|\mathcal{G})))$ wtedy i tylko wtedy, gdy dla każdego $G \in \sigma(P(A|\mathcal{G}))$ zachodzi $P(A \cap G) = \int_G P(A|\mathcal{G}) dP$ oraz $P(A|\mathcal{G})$ jest mierzalne $\sigma(P(A|\mathcal{G}))/\mathcal{R}$. W istocie pierwsza własność wynika z relacji $\sigma(P(A|\mathcal{G})) \subset \mathcal{G}$ zaś druga z faktu, że $P(A|\mathcal{G})$ jest mierzalna $\sigma(P(A|\mathcal{G}))/\mathcal{R}|_{P(A|\mathcal{G})(\Omega)}$.

Równość (A.18) wynika z zawierania $\sigma(P(A|\mathcal{G})) \subset \sigma((P(B|\mathcal{G}))_{B \in \mathcal{F}}) \subset \mathcal{G}$ oraz z ogólnego faktu, że dla $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3$ równość $Y = P(A|\mathcal{F}_1) = P(A|\mathcal{F}_3)$ prawie na pewno

implikuje $P(A|\mathcal{F}_2) = Y$ prawie na pewno. Jest tak, gdyż dla każdego $G \in \mathcal{F}_2 \subset \mathcal{F}_3$ zachodzi $P(A \cap G) = \int_G Y dP$ zaś Y jest mierzalna $\mathcal{F}_2/\mathcal{R}$, gdyż jest mierzalna $\mathcal{F}_1/\mathcal{R}$. \square

Twierdzenie A.31.

(i) Jeżeli $P(A \ominus \tilde{A}) = 0$ i $\tilde{A} \in \mathcal{G}$, to

$$P(A|\mathcal{G})(\omega) = \llbracket \omega \in A \rrbracket \quad \text{prawie na pewno.} \quad (\text{A.19})$$

(ii) Jeżeli $P(A|\mathcal{G}) \in \{0, 1\}$ prawie na pewno, to istnieje $\tilde{A} \in \mathcal{G}$ takie, że $P(A \ominus \tilde{A}) = 0$.

Dowód:

(i) Funkcja $I_{\tilde{A}}$, gdzie $I_{\tilde{A}} = \llbracket \omega \in \tilde{A} \rrbracket$ jest mierzalna \mathcal{G}/\mathcal{R} i spełnia $\int_G I_{\tilde{A}} dP = P(\tilde{A} \cap G)$, czyli $P(\tilde{A}|\mathcal{G}) = I_{\tilde{A}}$ prawie na pewno. Ponieważ $P(A \ominus \tilde{A}) = 0$, to mamy $P(A|\mathcal{G}) = P(\tilde{A}|\mathcal{G})$ prawie na pewno, a także (A.19).

(ii) $P(A|\mathcal{G})$ jest mierzalne \mathcal{G}/\mathcal{R} , a zatem $\tilde{A} := (P(A|\mathcal{G}) = 1) \in \mathcal{G}$. Mamy $P(A \cap \tilde{A}) = \int_{\tilde{A}} P(A|\mathcal{G}) dP$, gdzie $\int_{\tilde{A}} P(A|\mathcal{G}) dP = P(\tilde{A})$ z definicji \tilde{A} , zaś $\int_{\tilde{A}} P(A|\mathcal{G}) dP = \int P(A|\mathcal{G}) dP = P(A)$ z relacji $P(A|\mathcal{G}) \in \{0, 1\}$ prawie na pewno. Z właśnie wykazanej równości $P(A \cap \tilde{A}) = P(\tilde{A}) = P(A)$ wynika $P(A \setminus \tilde{A}) = P(\tilde{A} \setminus A) = P(A \ominus \tilde{A}) = 0$. \square

Twierdzenie A.32. Dla nieujemnej funkcji f mierzalnej \mathcal{G}/\mathcal{R} zachodzi

$$\int_A f dP = \int P(A|\mathcal{G}) f dP. \quad (\text{A.20})$$

Dowód: Dla miary $\mu : \mathcal{J} \rightarrow [0, \infty]$ i funkcji $h : \Omega \rightarrow [0, \infty]$ mierzalnej \mathcal{J}/\mathcal{R} całkę Lebesgue'a definiuje się jako

$$\int h d\mu := \sup_{h': 0 \leq h' \leq h, \text{card } h'(\Omega) < \infty} \sum_{r \in h'(\Omega)} r \mu(h' = r),$$

gdzie h' są funkcjami prostymi mierzalnymi \mathcal{J}/\mathcal{R} .

Dla dowolnej funkcji prostej $f' \leq f$ mierzalnej \mathcal{J}/\mathcal{R} istnieje funkcja prosta $f'' : \omega \mapsto \sup \{f'(\omega') : f(\omega') \leq f(\omega), \omega' \in \Omega\}$ mierzalna \mathcal{G}/\mathcal{R} i spełniająca $f' \leq f'' \leq f$. Łatwo zauważyć, że własność $\int_{A \cap G} dP = \int_G dP_A|_{\mathcal{G}}$ dla dowolnego $G \in \mathcal{G}$ implikuje $\int_A f'' dP = \int f'' dP_A|_{\mathcal{G}}$. A zatem mamy także $\int_A f dP = \int f dP_A|_{\mathcal{G}}$. Z tegoż wynika teza na mocy ogólnej własności $\int f d\mu = \int (d\mu/d\nu) f d\nu$. \square

Twierdzenie A.32 jest trywialne w dwóch przypadkach: (i) dla $\mathcal{G} = \sigma(\{B_k\}_{k \in \mathbb{N}})$ i rozłącznych B_k mamy $\sum_k f(\omega_k) P(A \cap B_k) = \sum_k P(A|B_k) f(\omega_k) P(B_k)$, $\omega_k \in B_k$; (ii) dla $A \in \mathcal{G}$ mamy $\int P(A|\mathcal{G}) f dP = \int \llbracket \omega \in A \rrbracket f(\omega) dP(\omega)$.

A.6. Warunkowa niezależność i warunkowa miara produktowa

W oparciu o twierdzenia A.10 i A.11 można zdefiniować proces stochastyczny zadając pewien nieskończony zbiór zgodnych ze sobą miar μ_{t_1, \dots, t_k} . Jeżeli interesuje nas możliwość efektywnego obliczania wartości miar μ_{t_1, \dots, t_k} , musimy zakładać pewną ich regularność. Pewnym typem założeń upraszczających obliczenia są założenia o niezależności bądź warunkowej niezależności. Przy okazji założenia te prowadzą do innych ciekawych rezultatów.

Definicja A.33. (zdarzenia niezależne) Zdarzenia $(A_t)_{t \in \mathbb{T}}$ nazywane są niezależnymi, jeżeli

$$P(A_{t_1} \cap \dots \cap A_{t_n}) = P(A_{t_1}) \cdot \dots \cdot P(A_{t_n}) \quad (\text{A.21})$$

dla każdego $n \in \mathbb{N}$ oraz $t_i \in \mathbb{T}$, gdzie $t_i \neq t_j$ dla $i \neq j$. Niezależność zdarzeń A_1, \dots, A_k oznaczamy $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_k$.

Mówi się, że σ -ciała $(\mathcal{G}_t)_{t \in \mathbb{T}}$ są niezależne, gdy niezależne są dowolne zbiory zdarzeń $(A_t)_{t \in \mathbb{T}}$, gdzie $A_t \in \mathcal{G}_t$. Zmienne $(X_t)_{t \in \mathbb{T}}$ są niezależne, gdy niezależne są σ -ciała $(\sigma(X_t))_{t \in \mathbb{T}}$. Niezależność $\mathcal{G}_1, \dots, \mathcal{G}_n$ i X_1, \dots, X_n oznaczamy $\mathcal{G}_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp \mathcal{G}_n$ i $X_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp X_n$. Dopuszczamy także oznaczenia mieszane $\dots \perp\!\!\!\perp A_i \perp\!\!\!\perp \dots \perp\!\!\!\perp \mathcal{G}_j \perp\!\!\!\perp \dots \perp\!\!\!\perp X_k \perp\!\!\!\perp \dots \iff \dots \perp\!\!\!\perp \sigma(A_i) \perp\!\!\!\perp \dots \perp\!\!\!\perp \mathcal{G}_j \perp\!\!\!\perp \dots \perp\!\!\!\perp \sigma(X_k) \perp\!\!\!\perp \dots$.

Definicja A.34. (zdarzenia niezależne warunkowo) Zdarzenia $(A_t)_{t \in \mathbb{T}}$ nazywane są niezależnymi przy danym zdarzeniu A , jeżeli

$$P(A_{t_1} \cap \dots \cap A_{t_n} | A) = P(A_{t_1} | A) \cdot \dots \cdot P(A_{t_n} | A) \quad (\text{A.22})$$

dla każdego $n \in \mathbb{N}$ oraz $t_i \in \mathbb{T}$, gdzie $t_i \neq t_j$ dla $i \neq j$. Niezależność zdarzeń A_1, \dots, A_k przy danym A oznaczamy $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_k | A$.

Zdarzenia $(A_t)_{t \in \mathbb{T}}$ nazywane są niezależnymi względem σ -ciała \mathcal{G} , jeżeli

$$P(P(A_{t_1} \cap \dots \cap A_{t_n} | \mathcal{G}) \neq P(A_{t_1} | \mathcal{G}) \cdot \dots \cdot P(A_{t_n} | \mathcal{G})) = 0 \quad (\text{A.23})$$

dla każdego $n \in \mathbb{N}$ oraz $t_i \in \mathbb{T}$, gdzie $t_i \neq t_j$ dla $i \neq j$. Niezależność zdarzeń A_1, \dots, A_k względem \mathcal{G} oznaczamy $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_k | \mathcal{G}$.

Mamy $A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_n | \{\emptyset, \Omega\} \iff A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_n | \Omega \iff A_1 \perp\!\!\!\perp \dots \perp\!\!\!\perp A_n$. Analogicznie definiuje się i oznacza σ -ciała i zmienne niezależne warunkowo.

Pojęcie warunkowej niezależności często bywa rozpatrywane w kontekście zależności przyczynowych lub funkcyjnych między zmiennymi losowymi (Pearl, 2000). Mamy w istocie następujące twierdzenie.

Twierdzenie A.35. *Prawdziwe są następujące stwierdzenia:*

- (i) Dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3$ zachodzi $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2 | \mathcal{G}_3$, jeżeli $\mathcal{G}_2 \subset \mathcal{G}_3$.
- (ii) Dla zmiennych X, Y, Z zachodzi $X \perp\!\!\!\perp Y | Z$, jeżeli $X = f(Z)$, gdzie f jest mierzalna $\tilde{\sigma}(Z)/\tilde{\sigma}(X)$.

Dowód:

- (i) Weźmy dowolne $A \in \mathcal{G}_1$ oraz $B \in \mathcal{G}_2$. Ponieważ $B \in \mathcal{G}_3$, to korzystając z (A.19) dla dowolnego $G \in \mathcal{G}_3$ otrzymujemy

$$\begin{aligned} \int_G P(A \cap B | \mathcal{G}_3) dP &= P(A \cap B \cap G) \\ &= \int_{B \cap G} P(A | \mathcal{G}_3) dP = \int_G P(A | \mathcal{G}_3) P(B | \mathcal{G}_3) dP, \end{aligned}$$

czyli na mocy jednoznaczności pochodnej Radona-Nikodyma zdarzenia $A \in \mathcal{G}_1$ oraz $B \in \mathcal{G}_2$ są niezależne względem \mathcal{G}_3 .

- (ii) Ponieważ f jest mierzalna $\tilde{\sigma}(Z)/\tilde{\sigma}(X)$, to $\sigma(X) \subset \sigma(Z)$ z twierdzenia A.12. Stosując pkt. (i) otrzymujemy tezę. □

Pojęcia niezależności i niezależności warunkowej wiążą się z pojęciami miary produktowej i warunkowej miary produktowej (Swart, 1996). Omówmy ten temat dokładniej.

Definicja A.36. (sumy i produkty σ -ciał) Dla dowolnych σ -ciał $\mathcal{G}_1, \mathcal{G}_2$ definiujemy

- (i) π -układ sumowy $\mathcal{G}_1 + \mathcal{G}_2 := \{A_1 \cap A_2 : A_i \in \mathcal{G}_i\}$,
- (ii) π -układ produktowy $\mathcal{G}_1 \times \mathcal{G}_2 := \{A_1 \times A_2 : A_i \in \mathcal{G}_i\}$,
- (iii) σ -ciało sumowe $\mathcal{G}_1 \oplus \mathcal{G}_2 := \sigma(\mathcal{G}_1 \cup \mathcal{G}_2) = \sigma(\mathcal{G}_1 + \mathcal{G}_2)$.
- (iv) σ -ciało produktowe $\mathcal{G}_1 \otimes \mathcal{G}_2 := \sigma(\mathcal{G}_1 \times \mathcal{G}_2)$.

Operacja \oplus jest przemienne i łączna, zaś operacja \otimes jest łączna. Mamy też $\{\emptyset, \Omega\} \oplus \mathcal{G} = \mathcal{G}$, $\{\emptyset, \Omega\} \otimes \mathcal{G} = \{\Omega \times A : A \in \mathcal{G}\}$. Zgodnie ze standardowym oznaczeniem potęgi kartezyjskiej oznaczamy

$$\mathcal{G}^n := \underbrace{\mathcal{G} \times \dots \times \mathcal{G}}_{n \text{ elementów}}. \quad (\text{A.24})$$

W literaturze spotyka się niekiedy symbol \mathcal{G}^n o znaczeniu $\mathcal{G} \otimes \dots \otimes \mathcal{G}$, co z kolei oznaczamy jako $\sigma(\mathcal{G}^n)$.

Ponieważ $\mathcal{G}_1 \oplus \mathcal{G}_2 = \sigma(\mathcal{G}_1 + \mathcal{G}_2)$, $\mathcal{G}_1 \otimes \mathcal{G}_2 = \sigma(\mathcal{G}_1 \times \mathcal{G}_2)$, więc z twierdzenia A.3 pkt. (i) miary μ_1 na $\mathcal{G}_1 \oplus \mathcal{G}_2$ i μ_2 na $\mathcal{G}_1 \otimes \mathcal{G}_2$ są jednoznacznie określone przez obciążenia $\mu_1|_{\mathcal{G}_1 + \mathcal{G}_2}$ oraz $\mu_2|_{\mathcal{G}_1 \times \mathcal{G}_2}$. Obciążenia $\mu_1|_{\mathcal{G}_1 + \mathcal{G}_2}$ oraz $\mu_2|_{\mathcal{G}_1 \times \mathcal{G}_2}$ muszą być σ -addytywne.

Definicja A.37. (miara diagonalna) Dla $n \in \mathbb{N}$ i miary $P : \mathcal{J} \rightarrow \mathbb{R}$ miarę diagonalną $P^{(n)} : \sigma(\mathcal{J}^n) \rightarrow \mathbb{R}$ określamy jako

$$P^{(n)}(A) := P \left(\left\{ \omega \in \Omega : \left(\underbrace{\omega, \dots, \omega}_{n \text{ elementów}} \right) \in A \right\} \right) \quad (\text{A.25})$$

dla każdego $A \in \sigma(\mathcal{J}^n)$.

Łatwo sprawdzić, że $P^{(n)}$ jest miarą. Co więcej, $P^{(n)}$ jest jedyną miarą na $\sigma(\mathcal{J}^n)$ spełniającą warunek $P^{(n)}(A_1 \times A_2 \times \dots \times A_n) = P(\bigcap_{i=1}^n A_i)$ dla $A_i \in \mathcal{J}$.

Twierdzenie A.38. Dla σ -skończonej miary μ na \mathcal{G}_1 i σ -skończonej miary ν na \mathcal{G}_2 funkcję

$$\mu \times \nu : \mathcal{G}_1 \times \mathcal{G}_2 \ni A_1 \times A_2 \mapsto \mu(A_1)\nu(A_2)$$

można rozszerzyć do miary na $\mathcal{G}_1 \otimes \mathcal{G}_2$ (Rudin, 1986, sekcja 7.7; Billingsley, 1979, sekcja 18).

Miarę $\mu \times \nu$ na $\mathcal{G}_1 \otimes \mathcal{G}_2$ nazywa się miarą produktową.

Podobnie możemy próbować określić warunkową miarę produktową (Swart, 1996).

Definicja A.39. (warunkowa protomiara produktowa) Dla σ -ciał $\mathcal{G}_1, \mathcal{G}_2, \mathcal{G}_3 \subset \mathcal{J}$ i miary $P : \mathcal{J} \rightarrow \mathbb{R}$ warunkową protomiarę produktową $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} : \mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3 \rightarrow \mathbb{R}$ określamy jako

$$P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}(A_1 \times A_2 \times A_3) := \int_{A_3} P(A_1 | \mathcal{G}_3) P(A_2 | \mathcal{G}_3) dP, \quad (\text{A.26})$$

gdzie $A_i \in \mathcal{G}_i$.

Twierdzenie A.40. Równość $P^{(3)}(A) = P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}(A)$ zachodzi dla każdego $A \in \mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$ wtedy i tylko wtedy, gdy $\mathcal{G}_1 \perp\!\!\!\perp \mathcal{G}_2 | \mathcal{G}_3$.

Dowód: Dla $A_i \in \mathcal{G}_i$ mamy $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}(A_1 \times A_2 \times A_3) = \int_{A_3} P(A_1 | \mathcal{G}_3) P(A_2 | \mathcal{G}_3) dP$ oraz $P^{(3)}(A_1 \times A_2 \times A_3) = P(A_1 \cap A_2 \cap A_3) = \int_{A_3} P(A_1 \cap A_2 | \mathcal{G}_3) dP$. Funkcje $P(A_1 \cap A_2 | \mathcal{G}_3)$ i $P(A_1 | \mathcal{G}_3) P(A_2 | \mathcal{G}_3)$ są mierzalne $\mathcal{G}_3 / \mathcal{R}$. Są one równe P -prawie wszędzie wtedy i tylko wtedy, gdy mają one równe całki po wszystkich zbiorach $A_3 \in \mathcal{G}_3$. \square

Funkcja $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ jest skończenie addytywna na $\mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$, ale nie zawsze funkcję tę można rozszerzyć do miary na $\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3$. Kontrprzykłady są znane (Swart, 1996). Jeżeli rozszerzenie jest możliwe, rozszerzoną miarę $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ nazywa się warunkową miarą produktową.

Twierdzenie A.41. Warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ na $\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3$ istnieje, jeżeli zachodzi jeden z poniższych warunków:

- (i) $\mathcal{G}_2 \subset \mathcal{G}_4$ oraz istnieje warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_4 | \mathcal{G}_3}$,
- (ii) $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3$,
- (iii) $\mathcal{G}_3 = \{\emptyset, \Omega\}$,
- (iv) istnieje miara warunkowa $(P(A | \mathcal{G}_i))_{A \in \mathcal{G}_3}$ dla $i = 1$ lub $i = 2$.

Dowód:

- (i) Teza zachodzi, gdyż $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} = P^{\mathcal{G}_1; \mathcal{G}_4 | \mathcal{G}_3} |_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3}$.
- (ii) Teza zachodzi, gdyż $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} = P^{(3)} |_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3}$.
- (iii) Teza zachodzi, gdyż $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}(A \times \Omega) = P \times P |_{\mathcal{G}_1 \otimes \mathcal{G}_2}(A)$ dla $A \in \mathcal{G}_1 \otimes \mathcal{G}_2$.
- (iv) Dowód tej części podał Swart (1996). \square

Twierdzenie A.42. Jeżeli $P^{(3)} |_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} \ll \mu_1 \times \mu_2 \times \mu_3$, gdzie μ_i są miarami na \mathcal{G}_i , to

- (i) istnieje miara warunkowa $(P(A | \mathcal{G}_i))_{A \in \mathcal{G}_3}$ dla $i = 1$ oraz $i = 2$.
- (ii) istnieje warunkowa miara produktowa $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$ i spełnia $P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3} \ll \mu_1 \times \mu_2 \times \mu_3$ oraz

$$\frac{dP^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}}{d\mu_{1,2,3}}(\omega_1 \times \omega_2 \times \omega_3) = \frac{\rho_{1,3}(\omega_1 \times \omega_3) \rho_{2,3}(\omega_2 \times \omega_3)}{\rho_3(\omega_3)}, \quad (\text{A.27})$$

gdzie $\mu_{i_1, \dots, i_n} := \mu_{i_1} \times \dots \times \mu_{i_n}$ oraz $\rho_{i_1, \dots, i_n} := dP^{(n)} |_{\mathcal{G}_{i_1} \otimes \dots \otimes \mathcal{G}_{i_n}} / d\mu_{i_1, \dots, i_n}$.

Dowód:

- (i) Dla dowolnego $A_1 \in \mathcal{G}_1$ i $A_3 \in \mathcal{G}_3$ mamy

$$\begin{aligned} P(A_1 \cap A_3) &= P^{(2)}(A_1 \times A_3) = \int_{A_1 \times A_3} \rho_{1,3} d\mu_{1,3} \\ &= \int_{A_3} \left(\int_{A_1} \frac{\rho_{1,3}(\omega_1 \times \omega_3)}{\rho_3(\omega_3)} d\mu_1(\omega_1) \right) \rho_3(\omega_3) d\mu_3(\omega_3) \\ &= \int_{A_3} \left(\int_{A_1} \frac{\rho_{1,3}(\omega_1 \times \omega_3)}{\rho_3(\omega_3)} d\mu_1(\omega_1) \right) dP(\omega_3) \end{aligned} \quad (\text{A.28})$$

na mocy twierdzenia Fubiniego, gdyż $\rho_{1,3}$ jest nieujemna prawie wszędzie i mierzalna $\mathcal{G}_1 \otimes \mathcal{G}_3 / \mathcal{R}$ (Rudin, 1986, sekcja 7.7). Z mierzalności $\mathcal{G}_1 \otimes \mathcal{G}_3 / \mathcal{R}$ funkcji $\rho_{1,3}$ wynika także mierzalność $\mathcal{G}_1 / \mathcal{R}$ funkcji $\rho_{1,3}(\cdot \times \omega_3)$ oraz mierzalność $\mathcal{G}_3 / \mathcal{R}$ funkcji $\rho_{1,3}(\omega_1 \times \cdot)$ (Rudin, 1986, sekcja 7.5).

Porównując $P(A_1 \cap A_3) = \int_{A_3} P(A_1 | \mathcal{G}_3) dP$ z (A.28) otrzymujemy

$$P(A_1 | \mathcal{G}_3)(\omega_3) = \nu(A_1)(\omega_3) := \int_{A_1} \frac{\rho_{1,3}(\omega_1 \times \omega_3)}{\rho_3(\omega_3)} d\mu_1(\omega_1), \quad (\text{A.29})$$

dla P -prawie wszystkich ω_3 . Z postaci definicji (A.29) wynika, że funkcja $\nu(\cdot)(\omega_3)$ jest σ -addytywna i mierzalna $\mathcal{G}_1/\mathcal{R}$.

Aby wykazać, że ν jest losową miarą prawdopodobieństwa, wystarczy dowieść, że jest ona nieujemna i unormowana. Zauważmy, że

$$\int \mu_1(\{\omega_1 : \rho_{1,3}(\omega_1 \times \omega_3) < 0\}) d\mu_3(\omega_3) = \int \llbracket \rho_{1,3} < 0 \rrbracket d\mu_{1,3} = 0,$$

a więc $\mu_1(\{\omega_1 : \rho_{1,3}(\omega_1 \times \omega_3) < 0\}) = 0$ dla μ_3 -prawie wszystkich ω_3 (Rudin, 1986, sekcja 7.7). Stąd wynika, że dla P -prawie wszystkich ω_3 funkcja $\nu(\cdot)(\omega_3)$ jest nieujemna, gdyż mamy także $P|_{\mathcal{G}_3} \ll \mu_3$ oraz $\rho_3 > 0$ P -prawie na pewno.

Oznaczmy $\tilde{\rho}_3(\omega_3) := \int \rho_{1,3}(\omega_1 \times \omega_3) d\mu_1(\omega_1)$. Z twierdzenia Fubiniego dla każdego $A_3 \in \mathcal{G}_3$ mamy $\int_{A_3} \tilde{\rho}_3 d\mu_3 = \int_{\Omega \times A_3} \rho_{1,3} d\mu_{1,3} = P(A_3) = \int_{A_3} \rho_3 d\mu_3$. Jako że $\tilde{\rho}_3 = \rho_3$ P -prawie na pewno, to $\nu(\cdot)(\omega_3)$ jest unormowana dla P -prawie wszystkich ω_3 .

Ponieważ ν jest losową miarą prawdopodobieństwa, to w świetle tożsamości (A.29) istnieje regularne prawdopodobieństwo warunkowe $(P(A|\mathcal{G}_1))_{A \in \mathcal{G}_3}$. Analogicznie dowodzi się regularności $(P(A|\mathcal{G}_2))_{A \in \mathcal{G}_3}$.

(ii) Weźmy dowolne $A_1 \in \mathcal{G}_1$, $A_2 \in \mathcal{G}_2$ i $A_3 \in \mathcal{G}_3$. Z poprzedniego punktu i twierdzenia A.41 wynika, że warunkowa miara produktowa istnieje i mamy

$$\begin{aligned} & \int_{A_3} P(A_1|\mathcal{G}_3)P(A_2|\mathcal{G}_3) dP \\ &= \int_{A_3} \left(\int_{A_1 \times A_2} \frac{\rho_{1,3}(\omega_1 \times \omega_3)}{\rho_3(\omega_3)} \cdot \frac{\rho_{2,3}(\omega_2 \times \omega_3)}{\rho_3(\omega_3)} d\mu_{1,2}(\omega_1 \times \omega_2) \right) dP(\omega_3) \\ &= \int_{A_1 \times A_2 \times A_3} \frac{\rho_{1,3}(\omega_1 \times \omega_3)\rho_{2,3}(\omega_2 \times \omega_3)}{\rho_3(\omega_3)} d\mu_{1,2,3}(\omega_1 \times \omega_2 \times \omega_3). \end{aligned} \quad (\text{A.30})$$

Z tożsamości (A.30) i jednoznaczności pochodnej Radona-Nikodyma wynika (A.27), gdyż dowolna miara na σ -ciele $\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3$ jest jednoznacznie zadana przez wartości na π -układzie $\mathcal{G}_1 \times \mathcal{G}_2 \times \mathcal{G}_3$ (twierdzenie A.3 pkt. (i)).

□

Istnienie miary produktowej jest ściśle związane z możliwością zakładania niezależności, gdy miara prawdopodobieństwa jest dana tylko częściowo. Jeżeli dane jest $P|_{\mathcal{G}_1}$ oraz $P|_{\mathcal{G}_2}$, to możemy skonstruować $P|_{\mathcal{G}_1 \oplus \mathcal{G}_2}$ przy założeniu, że $\mathcal{G}_1 \perp \mathcal{G}_2$, czyli $P^{(2)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2} = P|_{\mathcal{G}_1} \times P|_{\mathcal{G}_2}$.

Podobnie, jeżeli dane jest $P|_{\mathcal{G}_1 \oplus \mathcal{G}_3}$ oraz $P|_{\mathcal{G}_2 \oplus \mathcal{G}_3}$, to możemy skonstruować $P|_{\mathcal{G}_1 \oplus \mathcal{G}_2 \oplus \mathcal{G}_3}$ przy założeniu, że $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3$, czyli $P^{(3)}|_{\mathcal{G}_1 \otimes \mathcal{G}_2 \otimes \mathcal{G}_3} = P^{\mathcal{G}_1; \mathcal{G}_2 | \mathcal{G}_3}$. O ile jednak istnienie miary $P|_{\mathcal{G}_1 \oplus \mathcal{G}_2}$ spełniającej niezależność $\mathcal{G}_1 \perp \mathcal{G}_2$ jest zagwarantowane dla dowolnych obcięć $P|_{\mathcal{G}_1}$ oraz $P|_{\mathcal{G}_2}$, to istnieją takie obciążenia $P|_{\mathcal{G}_1 \oplus \mathcal{G}_3}$ oraz $P|_{\mathcal{G}_2 \oplus \mathcal{G}_3}$, że nie istnieje miara prawdopodobieństwa $P|_{\mathcal{G}_1 \oplus \mathcal{G}_2 \oplus \mathcal{G}_3}$, dla której zachodzi niezależność warunkowa $\mathcal{G}_1 \perp \mathcal{G}_2 | \mathcal{G}_3$.

Niech \mathcal{G}^P będzie uzupełnieniem σ -ciała \mathcal{G} względem miary P (definicja A.18). Z własności $\mathcal{G}^P = \{A' \in \mathcal{M}_P : \exists_{A \in \mathcal{G}} P^*(A \ominus A') = 0\}$ łatwo wykazać relacje

$$(\mathcal{G}_1 \cap \mathcal{G}_2)^P = \mathcal{G}_1^P \cap \mathcal{G}_2^P, \quad (\text{A.31})$$

$$(\mathcal{G}_1 \oplus \mathcal{G}_2)^P \supset \mathcal{G}_1^P \oplus \mathcal{G}_2^P \supset \mathcal{G}_1 \oplus \mathcal{G}_2, \quad (\text{A.32})$$

$$(\mathcal{G}_1 \otimes \mathcal{G}_2)^{\tilde{P}} \supset \mathcal{G}_1^{P_1} \otimes \mathcal{G}_2^{P_2} \supset \mathcal{G}_1 \otimes \mathcal{G}_2, \quad (\text{A.33})$$

gdzie w ostatnim związku $P_1(A_1) := \tilde{P}(A_1 \times \Omega)$ i $P_2(A_2) := \tilde{P}(\Omega \times A_2)$ są rzutami miary $\tilde{P} : \mathcal{G}_1 \otimes \mathcal{G}_2 \rightarrow \mathbb{R}$ na \mathcal{G}_1 i \mathcal{G}_2 .

Zauważmy, że dla σ -ciał $\mathcal{G}^P \supset \mathcal{G}' \supset \mathcal{G}$ i miary $\nu \ll P$ prawie na pewno mamy $d\nu|_{\mathcal{G}^P}/dP = d\nu|_{\mathcal{G}'}/dP = d\nu|_{\mathcal{G}}/dP$. Stąd i ze związku (A.33) wynika, że warunkowa informacja wzajemna określona definicją 1.14 spełnia równość (1.31).

A.7. Definiowanie miar przez częstości w ciągach

Twierdzenia A.10 i A.11 zapewniają istnienie szerokiej klasy procesów $X_{\mathbb{T}}$ o wartościach zespolonych i dyskretnych. W niniejszej pracy zajmujemy się głównie procesami o przeliczalnym zbiorze \mathbb{T} , konkretnie $\mathbb{T} = \mathbb{Z}$, nazywanymi także szeregami czasowymi. Nazwa pochodzi stąd, że dla $\mathbb{T} \subset \mathbb{R}$ indeks $t \in \mathbb{T}$ obrazowo nazywa się czasem.

Rozpatrzmy proces $X_{\mathbb{T}}$ taki, że $\tilde{\sigma}(X_t)$ nie zależy od $t \in \mathbb{T}$, a \mathbb{T} jest zamknięte ze względu na dodawanie. Wówczas możemy określić operację $T_{\tau} : X_{\mathbb{T}}(\Omega) \rightarrow X_{\mathbb{T}}(\Omega)$, która przesuwa proces w czasie o odstęp $\tau \in \mathbb{T}$. Dla $X_{\mathbb{T}} = (X_t)_{t \in \mathbb{T}}$ definiujemy $T_{\tau} X_{\mathbb{T}} := (X_{t+\tau})_{t \in \mathbb{T}}$. W szczególności, jeżeli funkcje $\xi_n : X_{\mathbb{T}}(\Omega) \rightarrow X_t(\Omega)$ określimy jako $\xi_t(X_{\mathbb{T}}) := X_t$, to $\xi_t \circ T_{\tau}(X_{\mathbb{T}}) = X_{t+\tau}$.

Niech $P_{X_{\mathbb{T}}} = P \circ X_{\mathbb{T}}^{-1}$ będzie przeniesioną miarą prawdopodobieństwa na σ -ciele $\tilde{\sigma}(X_{\mathbb{T}})$.

Definicja A.43. (proces stacjonarny) Mówimy, że proces $X_{\mathbb{T}}$ jest stacjonarny, gdy $P_{X_{\mathbb{T}}} = P_{X_{\mathbb{T}}} \circ T_{\tau}$ dla wszystkich $\tau \in \mathbb{T}$.

Jeżeli $\mathbb{T} = \mathbb{Z}$ lub $\mathbb{T} = \mathbb{N}$, zaś X_t są zmiennymi rzeczywistymi, to zgodnie z twierdzeniem A.11 proces $X_{\mathbb{T}}$ jest stacjonarny wtedy i tylko wtedy, gdy dla $\mu := P_{X_{\mathbb{T}}}$ zachodzi $\mu_{t,t+1,\dots,t+k-1} = \mu_{t+1,t+2,\dots,t+k}$. Taką miarę μ nazywamy stacjonarną. Jeżeli μ jest stacjonarna, to możemy zdefiniować $\mu^{(k)} := \mu_{t,t+1,\dots,t+k-1}$. Zgodnie z twierdzeniem A.11 mamy

$$\mu^{(1)}(\mathbb{R}) = 1, \quad \forall A \in \sigma(\mathcal{R}^k) \quad \mu^{(k)}(A) = \mu^{(k+1)}(A \times \mathbb{R}) = \mu^{(k+1)}(\mathbb{R} \times A). \quad (\text{A.34})$$

Równości (A.34) są warunkiem koniecznym i dostatecznym tego, aby dla $\mathbb{T} = \mathbb{Z}$ lub $\mathbb{T} = \mathbb{N}$ proces stacjonarny $X_{\mathbb{T}}$ o mierze przeniesionej $\mu = P \circ X_{\mathbb{T}}^{-1}$ istniał na pewnej przestrzeni probabilistycznej (Upper, 1997, twierdzenie B.1.1). Stąd wynika, że dowolny proces stacjonarny $X_{\mathbb{N}}$ o wartościach rzeczywistych lub dyskretnych można rozszerzyć do procesu stacjonarnego $X_{\mathbb{Z}}$ (Kallenberg, 1997, lemat 9.2).

Interesującej konstrukcji niektórych procesów stacjonarnych dostarcza częstościowa interpretacja prawdopodobieństwa (von Mises, 1919). Dla ciągu $c_{\mathbb{N}}$, gdzie $c_i \in \mathbb{R}$, oraz $A \in \sigma(\mathcal{R}_0^{\mathbb{N}})$ określimy

$$\mathbf{F}(A)(c_{\mathbb{N}}) := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I_A(c_{k:\infty}), \quad (\text{A.35})$$

gdzie $I_A(x) := \llbracket x \in A \rrbracket$, zaś $\lim_{n \rightarrow \infty}$ dane jest wzorem (4.9). W zasadzie wzoru (A.35) można użyć do zdefiniowania funkcji zdarzeń $\mathbf{F}(\cdot)(c_{\mathbb{N}})$ również wtedy, gdy ciąg $c_{\mathbb{N}}$ nie jest ciągiem liczbowym.

Dla pewnych ciągów $c_{\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ oraz ciał $\mathcal{G} \subset \sigma(\mathcal{R}_0^{\mathbb{N}})$ (niekoniecznie σ -ciał) można wykazać, że $\mathbf{F}(\cdot)(c_{\mathbb{N}}) : \mathcal{G} \rightarrow [0, 1]$ jest stacjonarną miarą prawdopodobieństwa. Najtrudniejsza do pokazania jest określoność (tzn. $\mathbf{F}(\cdot)(c_{\mathbb{N}}) \neq \perp$) i σ -addytywność $\mathbf{F}(\cdot)(c_{\mathbb{N}})$. Skończona addytywność i własność $\mathbf{F}(A)(c_{\mathbb{N}}) = \mathbf{F}(T_{\tau} A)(c_{\mathbb{N}})$ wynikają ze wzoru (A.35) bezpośrednio. Zatem jeżeli $c_i \in U$, gdzie U jest zbiorem o skończonej liczbie wartości, a $\mathbf{F}(A)(c_{\mathbb{N}}) \neq \perp$ dla każdego $A \in \mathcal{G} := \mathcal{R}_0^{\mathbb{N}}|_U$, to z twierdzenia A.11 funkcja $\mathbf{F}(\cdot)(c_{\mathbb{N}}) : \mathcal{G} \rightarrow [0, 1]$ jest istotnie stacjonarną miarą prawdopodobieństwa. (Miarą $\mu = \mathbf{F}(\cdot)(c_{\mathbb{N}})|_{\mathcal{G}}$ generowana jest przez

rozkłady $\mu^{(k)}$. Rozkłady te są miarami o skończonej dziedzinie, dla których σ -addytywność sprowadza się do skończonej addytywności.)

Ogólniej, jeżeli proces stacjonarny $X_{\mathbb{Z}}$ jest zmienną borelowską, to istnieje stacjonarna i ergodyczna miara losowa F_X na $\tilde{\sigma}(X_{\mathbb{Z}})$ taka, że dla każdego $A \in \sigma(X_{\mathbb{N}})$ z osobną równość $F_X(X_{\mathbb{Z}}(A))(\omega) = \mathbf{F}(X_{\mathbb{N}}(A))(X_{\mathbb{N}}(\omega))$ zachodzi dla prawie wszystkich ω (zob. dowód twierdzenia 4.7). Pomimo tego istnieje nieprzeliczalnie wiele $x_{\mathbb{N}} \in X_{\mathbb{N}}(\Omega)$, dla których $\mathbf{F}(\cdot)(x_{\mathbb{N}})$ jest miarą stacjonarną lecz nieergodyczną. Na przykład, jeżeli określone są ergodyczne miary $\mathbf{F}(\cdot)(a_{\mathbb{N}}) \neq \mathbf{F}(\cdot)(b_{\mathbb{N}})$, to dla

$$c_n := \begin{cases} a_n, & 2^{k-1} < n \leq 2^{k-1} + 2^{k-2}, \\ b_n, & 2^{k-1} + 2^{k-2} < n \leq 2^k \end{cases} \quad (\text{A.36})$$

mamy $\mathbf{F}(\cdot)(c_{\mathbb{N}}) = [\mathbf{F}(\cdot)(a_{\mathbb{N}}) + \mathbf{F}(\cdot)(b_{\mathbb{N}})]/2$, a więc miara $\mathbf{F}(\cdot)(c_{\mathbb{N}})$ nie jest ergodyczna na mocy twierdzenia 4.6 pkt. (vi).

Transformacja \mathbf{F} nie jest bijekcją. Co więcej, aby proces $X_{\mathbb{Z}}$ cechował się probabilistyczną losowością (tzn. miał niezerową entropię $\bar{H}(X_{\mathbb{Z}})$), sam ciąg $c_{\mathbb{N}}$ nie musi być algorytmicznie losowy (tzn. może istnieć efektywna procedura generująca $c_{\mathbb{N}}$). Dla przykładu rozpatrzmy ciąg Champernowne'a $c_{\mathbb{N}} = (1, 2, 3, 4, 5, 6, 7, 8, 9, 1, 0, 1, 1, \dots)$ równy kolejnym cyfrom rozwinięcia dziesiętnego kolejnych liczb naturalnych. Dla takiego $c_{\mathbb{N}}$ proces $X_{\mathbb{Z}}$ jest ciągiem niezależnych zmiennych losowych (Li i Vitányi, 1993, przykład 1.9.1).

A.8. Procesy Markowa i ukryte procesy Markowa

Ważną klasą procesów stochastycznych definiowanych w oparciu o założenia warunkowej niezależności są procesy Markowa.

Definicja A.44. (łańcuch Markowa) *Mówimy, że proces $X_{\mathbb{T}}$ jest procesem (lub łańcuchem) Markowa, jeżeli \mathbb{T} jest zbiorem liniowo uporządkowanym i zachodzi $(X_i)_{i < j, i \in \mathbb{T}} \perp\!\!\!\perp (X_i)_{i > j, i \in \mathbb{T}} | X_j$ dla wszystkich $j \in \mathbb{T}$.*

Proces $X_{\mathbb{Z}}$ jest łańcuchem Markowa wtedy i tylko wtedy, gdy zachodzi $X_{i:j-1} \perp\!\!\!\perp X_{j+1} | X_j$ dla wszystkich $i, j \in \mathbb{Z}$ takich, że $i < j$. Proces $X_{\mathbb{N}}$ jest łańcuchem Markowa wtedy i tylko wtedy, gdy zachodzi $X_{1:j-1} \perp\!\!\!\perp X_{j+1} | X_j$ dla wszystkich $j \in \mathbb{N}$. Dla procesów $X_{\mathbb{Z}}$ i $X_{\mathbb{N}}$ używa się dodatkowo pojęcia łańcucha Markowa skończonego rzędu.

Definicja A.45. (łańcuch Markowa n -tego rzędu) *Mówimy, że proces $X_{\mathbb{Z}}$ lub $X_{\mathbb{N}}$ jest łańcuchem Markowa n -tego rzędu, jeżeli dla $Y_i := X_{i:i+n-1}$ proces $Y_{\mathbb{Z}}$ jest łańcuchem Markowa.*

Teoria dyskretnych łańcuchów Markowa jest dobrze rozwinięta, zarówno dla zmiennych przybierających skończoną (Iosifescu, 1988) jak i nieskończoną przeliczalną liczbę wartości (Hernandez-Lerma i Lasserre, 2003; Kemeny *et al.*, 1966). Pierwsza klasa procesów jest znana pod nazwą skończonych, a druga pod nazwą przeliczalnych łańcuchów Markowa.

Pewnym uogólnieniem łańcuchów Markowa są ukryte łańcuchy Markowa.

Definicja A.46. (ukryty łańcuch Markowa) *Mówimy, że proces $X_{\mathbb{T}}$ jest ukrytym łańcuchem Markowa, jeżeli istnieje proces $Y_{\mathbb{T}}$ taki, że $(X_i \times Y_i)_{i \in \mathbb{T}}$ jest łańcuchem Markowa. Proces $Y_{\mathbb{T}}$ nazywa się procesem ukrytym, proces $X_{\mathbb{T}}$ — procesem jawnym.*

Każdy łańcuch Markowa n -tego rzędu $X_{\mathbb{Z}}$ jest ukrytym łańcuchem Markowa, gdyż za proces ukryty możemy wziąć $Y_{\mathbb{Z}}$, gdzie $Y_i = X_{i+1:i+n-1}$. Aby pojęcie ukrytego łańcucha Markowa nie było trywialne (tzn. aby nie obejmowało wszystkich procesów stochastycznych w zadanej klasie), istotne bywa narzucanie dodatkowych warunków na własności

procesu ukrytego. W przeciwnym razie możemy reprezentować za pomocą pojedynczej wartości zmiennej Y_j całą przeszłość $(X_i)_{i < j, i \in \mathbb{T}}$.

Twierdzenie A.47. *Dowolny proces $X_{\mathbb{Z}}$ o wartościach dyskretnych lub rzeczywistych jest ukrytym łańcuchem Markowa takim, że zmienne Y_i procesu ukrytego $Y_{\mathbb{Z}}$ są zmiennymi rzeczywistymi.*

Dowód: Na mocy twierdzenia A.17 istnieją funkcje f_i bijektywnie mierzalne i zmienne rzeczywiste $Y_i = f_i(X_{-\infty:i})$. Mamy wówczas $\sigma(X_{-\infty:i}) = \sigma(Y_i)$, a zatem $\sigma((X_k \times Y_k)_{k \leq i}) = \sigma(X_i \times Y_i)$, czyli na mocy twierdzenia A.35 proces $(X_i \times Y_i)_{i \in \mathbb{T}}$ jest łańcuchem Markowa. \square

Twierdzenie A.48. *Dowolny proces $X_{\mathbb{N}}$ o wartościach dyskretnych jest ukrytym łańcuchem Markowa takim, że zmienne Y_i procesu ukrytego $Y_{\mathbb{N}}$ są zmiennymi dyskretnymi.*

Dowód: Zdefiniujmy zmienne dyskretne $Y_n = X_{1:n}$. Mamy wówczas $\sigma(X_{1:i}) = \sigma(Y_i)$, a zatem $\sigma((X_k \times Y_k)_{k \leq i}) = \sigma(X_i \times Y_i)$, czyli na mocy twierdzenia A.35 proces $(X_i \times Y_i)_{i \in \mathbb{T}}$ jest łańcuchem Markowa. \square

Analogicznie jak w przypadku zwykłych łańcuchów Markowa, proces $X_{\mathbb{Z}}$ nazwiemy skończonym ukrytym łańcuchem Markowa, gdy wszystkie zmienne $X_i \times Y_i$ przybierają wartości z tego samego skończonego zbioru. Z kolei $X_{\mathbb{Z}}$ nazwiemy przeliczalnym ukrytym łańcuchem Markowa, gdy zbiór wartości zmiennych $X_i \times Y_i$ jest przeliczalny.

Pomimo dość ogólnych twierdzeń A.47 i A.48 nie każdy dyskretny stacjonarny proces $X_{\mathbb{N}}$ o skończonej liczbie wartości jest skończonym ukrytym łańcuchem Markowa, por. twierdzenie 2.11 oraz Crutchfield i Feldman (2003). Fakt ten wynika z finitaryności skończonych łańcuchów ukrytych oraz istnienia procesów niefinitarnych. Jako że istnieją niefinitarne przeliczalne ukryte łańcuchy Markowa, nie jest jasne, czy dowolny dyskretny proces stacjonarny można przedstawić jako przeliczalny łańcuch ukryty.

Teoria łańcuchów ukrytych jest słabiej rozwinięta niż teoria zwykłych łańcuchów Markowa. Pojęcie łańcucha ukrytego zrodziło się z praktycznej potrzeby komputerowego modelowania procesu przetwarzania danych symbolicznych w lingwistyce komputerowej (Jelinek, 1997; Charniak, 1993) i bioinformatyce (Ewens i Grant, 2001; Baxevanis i Ouellette, 2004). Głównym zastosowaniem łańcuchów ukrytych w tych dziedzinach jest obliczanie najlepszego predyktora zmiennych ukrytych $Y_{1:n}$ przy danych zmiennych $X_{1:n}$ (Dębowski, 2004b). Dla łańcuchów skończonych predyktor zmiennych $Y_{1:n}$ można obliczyć w czasie liniowym w n za pomocą programowania dynamicznego, zwanego też algorytmem Viterbiego (Ewens i Grant, 2001; Jelinek, 1997, Viterbi search).

Teoretyczne własności ukrytych procesów Markowa, w zasadzie tylko dla łańcuchów skończonych, badane są od stosunkowo niedawna (Kehagias, 1992; Ephraim i Merhav, 2002). Pewne kryterium istnienia procesu ukrytego $Y_{\mathbb{Z}}$ oraz algorytm minimalizacji liczby wartości zmiennych tego procesu zostały podane przez Uppera (1997). Algorytm minimalizacji opiera się na ortogonalizacji skończenie wymiarowej przestrzeni liniowej rozpinanej przez miary $(P(A|X_{-\text{len } w+1:0} = w))_{A \in \sigma(X_{1:\infty})}$, gdzie $w \in [X_1(\Omega)]^*$. Z koncepcji tej wywodzi się także idea nowatorskiego algorytmu estymacji parametrów ukrytego łańcucha Markowa z jego realizacji. Algorytm ten nie wykorzystuje przeszukiwania lokalnego, a przez to może być lepszy niż algorytm EM (Ewens i Grant, 2001; Jelinek, 1997, expectation-maximization). Przypuszczamy, że zastosowanie idei ortogonalizacji przestrzeni miar warunkowych do przeliczalnych łańcuchów ukrytych może przynieść rozstrzygnięcia odnośnie istnienia nieskończonego dyskretnego procesu ukrytego $Y_{\mathbb{Z}}$ dla dowolnego procesu stacjonarnego $X_{\mathbb{Z}}$.

Bibliografia

- R. Ahlswede, B. Balkenhol, L. H. Khachatrian (1996) *Some properties of fix-free codes*, w: *Proceedings of the First INTAS International Seminar on Coding Theory and Combinatorics, 1996, Thakadzor, Armenia*, s. 20–33.
- J.-P. Allouche, J. Shallit (2003) *Automatic Sequences. Theory, Applications, Generalizations*, Cambridge University Press.
- A. D. Baxevanis, B. F. F. Ouellette (red.) (2004) *Bioinformatyka. Podręcznik do analizy genów i białek*, PWN.
- B. Baxter (1962) *An asymptotic result for the finite predictor*, *Mathematica Scandinavica*, t. 10, s. 137–144.
- T. C. Bell, J. G. Cleary, I. H. Witten (1990) *Text Compression*, Prentice Hall.
- J. Beran (1994) *Statistics for Long-Memory Processes*, Chapman & Hall.
- V. Berthé (1994) *Conditional entropy of some automatic sequences*, *Journal of Physics A*, t. 27, s. 7993–8006.
- P. Billingsley (1965) *Ergodic Theory and Information*, J. Wiley.
— (1979) *Probability and Measure*, J. Wiley.
— (1987) *Prawdopodobieństwo i miara*, PWN.
- P. Bloomfield, N. P. Jewell, E. Hayashi (1983) *Characterizations of completely nondeterministic stochastic process*, *Pacific Journal of Mathematics*, t. 107, s. 307–317.
- L. Breiman (1992) *Probability*, Philadelphia: SIAM.
- P. J. Brockwell, R. A. Davis (1987) *Time Series: Theory and Methods*, Springer.
- M. Charikar, E. Lehman, D. Liu, R. Panigrahy, M. Prabhakaran, A. Rasala, A. Sahai, A. Shelat (2002) *Approximating the smallest grammar: Kolmogorov complexity in natural models*, w: *Proceedings of the 29th Symposium on Theory of Computing*, s. 792–801.
- E. Charniak (1993) *Statistical Language Learning*, The MIT Press.
- Y. S. Chow, H. Teicher (1978) *Probability Theory. Independence, Interchangeability, Martingales*, Springer.
- T. M. Cover, P. Gacs, R. M. Gray (1989) *Kolmogorov's contributions to information theory and algorithmic complexity*, *Annals of Probability*, t. 17, s. 840–865.
- T. M. Cover, J. A. Thomas (1991) *Elements of Information Theory*, J. Wiley.
- J. P. Crutchfield, D. P. Feldman (2003) *Regularities unseen, randomness observed: The entropy convergence hierarchy*, *Chaos*, t. 15, s. 25–54.
- Ł. Dębowski (1999) *Teoria funkcyjności gęstości dla monowarstw zaadsorbowanych na podłożu krystalicznym*, Praca magisterska, Wydział Fizyki, Uniwersytet Warszawski.
— (2004a) *Entropic Subextensivity in Language and Learning*, w: C. Tsallis, M. Gell-Mann (red.), *Nonextensive Entropy—Interdisciplinary Applications*, s. 335–345, Oxford University Press.
— (2004b) *Trigram morphosyntactic tagger for Polish*, w: M. A. Kłopotek, S. T. Wierchoń, K. Trojanowski (red.), *Intelligent Information Processing and Web Mining. Proceedings of the International IIS:IIPWM'04 Conference held in Zakopane, Poland, May*

- 17-20, 2004, s. 409–413, Springer Verlag.
- (2005) *On Hilberg's law and its links with Guiraud's law*. Preprint.
- S. Dégerine (1982) *Partial autocorrelation function for a scalar stationary discrete time series*, w: *Proceedings of the 3rd Franco Belgian Meeting of Statisticians*, s. 79–94, Bruxelles: Publication des Facultes Universitaire Saint-Louis.
- S. Dégerine, S. Lambert-Lacroix (2003) *Characterization of the partial autocorrelation function of a nonstationary time series*, *Journal of Multivariate Analysis*, t. 87, s. 46–59.
- L. Devroye (1987) *A Course in Density Estimation*, Birkhäuser.
- J. L. Doob (1953) *Stochastic processes*, J. Wiley.
- J. Durbin (1960) *The fitting of time series models*, *Review of the International Statistical Institute*, t. 28, s. 233–244.
- W. Ebeling, T. Pöschel (1994) *Entropy and long-range correlations in literary English*, *Europhysics Letters*, t. 26, s. 241–246.
- P. Elias (1975) *Universal codeword sets and representations for the integers*, *IEEE Transactions on Information Theory*, t. IT-21, s. 194–203.
- Y. Ephraim, N. Merhav (2002) *Hidden Markov processes*, *IEEE Transactions on Information Theory*, t. 48, s. 1518–1569.
- W. J. Ewens, G. R. Grant (2001) *Statistical Methods in Bioinformatics*, Springer.
- P. D. Finch (1960) *On the covariance determinants of autoregressive and moving average models*, *Biometrika*, t. 47, s. 194–211.
- I. M. Gel'fand, A. N. Kolmogorov, A. M. Jaglom (1956) *K obščemu operedeleniju količestva informaciji*, *Doklady Akademii Nauk SSSR*, t. 111, s. 745–748.
- D. Gillman, R. L. Rivest (1995) *Complete variable-length "fix-free" codes*, *Designs, Codes and Cryptography*, t. 5, s. 109–114.
- R. L. Graham, D. E. Knuth, O. Patashnik (1994) *Concrete Mathematics. A Foundation for Computer Science*, Addison-Wesley.
- T. Gramss (1994) *Entropy of the symbolic sequence for critical circle maps*, *Physical Review E*, t. 50, s. 2616–2620.
- C. W. J. Granger, R. Joyeux (1980) *An introduction to long memory time series models and fractional differencing*, *Journal of Time Series Analysis*, t. 1, s. 15–29.
- R. M. Gray (1987) *Probability, Random Processes, and Ergodic Properties*, Springer.
- R. M. Gray, L. D. Davisson (1974) *The ergodic decomposition of stationary discrete random processes*, *IEEE Transactions on Information Theory*, t. IT-20, s. 625–636.
- R. M. Gray, J. C. Kieffer (1980) *Asymptotically mean stationary measures*, *The Annals of Probability*, t. 8, s. 962–973.
- U. Grenander, G. Szegö (1958) *Toeplitz Forms and Their Applications*, Berkeley: University of California Press.
- H. Guiraud (1954) *Les caractères statistiques du vocabulaire*, Paris: Presses Universitaires de France.
- J. Haag (1928) *Sur un problème général de probabilités et ses diverses applications*, w: *Proceedings of the International Congress of Mathematicians, Toronto, 1924*, s. 629–674.
- H. Helson, D. Sarason (1967) *Past and future*, *Mathematica Scandinavica*, t. 21, s. 5–16.
- O. Hernandez-Lerma, J. B. Lasserre (2003) *Markov Chains and Invariant Probabilities*, Birkhäuser.
- W. Hilberg (1990) *Der bekannte Grenzwert der redundanzfreien Information in Texten — eine Fehlinterpretation der Shannonschen Experimente?*, *Frequenz*, t. 44, s. 243–248.
- J. R. M. Hosking (1981) *Fractional differencing*, *Biometrika*, t. 68, s. 165–176.
- G. Hu (1962) *On the amount of information*, *Teorija verovatnostej i ee primenenija*, t. 4,

- s. 439–447.
- A. Inoue (2000) *Asymptotics for the partial autocorrelation function of a stationary process*, Journal d'Analyse Mathématique, t. 81, s. 65–109.
- (2005) *What does the partial autocorrelation function look like for large lags?* <http://www.math.sci.hokudai.ac.jp/~inoue/>.
- A. Inoue, Y. Kasahara (2004) *Partial autocorrelation functions of the fractional ARIMA processes with negative degree of differencing*, Journal of Multivariate Analysis, t. 89, s. 135–147.
- (2005) *Explicit representation of finite predictor coefficients and its applications*. <http://www.math.sci.hokudai.ac.jp/~inoue/>.
- M. Iosifescu (1988) *Skończone procesy Markowa i ich zastosowania*, PWN.
- F. Jelinek (1997) *Statistical Methods for Speech Recognition*, The MIT Press.
- T. Kailath (1987) *Signal processing applications of some moment problems*, w: H. J. Landau (red.), *Moments in Mathematics. Proceedings of Symposia in Applied Mathematics*, t. 37, s. 71–109, American Mathematical Society.
- O. Kallenberg (1997) *Foundations of Modern Probability*, Springer.
- A. Kehagias (1992) *Approximation of Stochastic Processes by Hidden Markov Models*, Rozprawa doktorska, Brown University.
- J. G. Kemeny, J. L. Snell, A. W. Knapp (1966) *Denumerable Markov Chains*, D. Van Nostrand.
- J. C. Kieffer, E. Yang (2000) *Grammar-based codes: A new class of universal lossless source codes*, IEEE Transactions on Information Theory, t. 46, s. 737–754.
- (2002) *Structured grammar-based codes for universal lossless data compression*, Communications in Information and Systems, t. 2, s. 29–52.
- J. F. C. Kingman (2002) *Procesy Poissona*, PWN.
- K. Knopp (1956) *Szeregi nieskończone*, PWN.
- I. P. Kornfel'd, S. V. Fomin, J. G. Sinaj (1982) *Ergodic Theory*, Springer.
- T. Kowalczyk, E. Pleszczyńska, F. Ruland (2004) *Grade Models and Methods for Data Analysis*, Springer.
- S. Kullback, R. A. Leibler (1951) *On information and sufficiency*, Annals of Mathematical Statistics, t. 22, s. 79–86.
- E. Lehman, A. Shelat (2002) *Approximation algorithms for grammar-based compression*, w: *Proceedings of the Thirteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, s. 205–212, ACM/SIAM.
- M. Li, P. M. B. Vitányi (1993) *An Introduction to Kolmogorov Complexity and Its Applications*, Springer.
- B. B. Mandelbrot (1983) *The Fractal Geometry of Nature*, New York: W. H. Freeman.
- C. G. de Marcken (1996) *Unsupervised Language Acquisition*, Rozprawa doktorska, Massachusetts Institute of Technology.
- A. I. McLeod (1998) *Hyperbolic decay time series*, The Journal of Time Series Analysis, t. 19, s. 473–483.
- R. von Mises (1919) *Fundamentalsätze der Wahrscheinlichkeitsrechnung*, Mathematische Zeitschriften, t. 5, s. 52–99.
- C. Nevill-Manning (1996) *Inferring Sequential Structure*, Rozprawa doktorska, University of Waikato.
- C. G. Nevill-Manning, I. H. Witten (1997) *Compression and explanation using hierarchical grammars*, The Computer Journal, t. 40, s. 103–116.
- E. Pap (2002) *Some Elements of Classical Measure Theory*, w: E. Pap (red.), *Handbook*

- of Measure Theory*, s. 27–82, Elsevier Science.
- J. Pearl (2000) *Causality*, Cambridge University Press.
- K. Polański (red.) (1999) *Encyklopedia językoznawstwa ogólnego*, Zakład Narodowy im. Ossolińskich.
- M. Pourahmadi (2001) *Foundations of Time Series Analysis and Prediction Theory*, J. Wiley.
- F. L. Ramsey (1974) *Characterization of the partial autocorrelation function*, The Annals of Statistics, t. 2, s. 1296–1301.
- W. Rudin (1986) *Analiza rzeczywista i zespolona*, PWN.
- I. Schur (1917) *Über Potenzreihen, die im Innern des Einheitskreises beschränkt sind*, Journal für die Reine und Angewandte Mathematik, t. 147, s. 205–232.
- C. R. Shalizi (2001) *Causal Architecture, Complexity and Self-Organization for Time Series and Cellular Automata*, Rozprawa doktorska, University of Wisconsin-Madison.
- C. Shannon (1950) *Prediction and entropy of printed English*, Bell System Technical Journal, t. 30, s. 50–64.
- J. M. Swart (1996) *A conditional product measure theorem*, Statistics & Probability Letters, t. 28, s. 131–135.
- M. Taniguchi, Y. Kakizawa (2000) *Asymptotic Theory of Statistical Inference for Time Series*, Springer.
- D. R. Upper (1997) *Theory and Algorithms for Hidden Markov Models and Generalized Hidden Markov Models*, Rozprawa doktorska, University of California.
- S. R. S. Varadhan (2001) *Probability Theory*, American Mathematical Society. Courant Institute of Mathematical Sciences.
- T. Weissman (2004) *Not all universal source codes are pointwise universal*. <http://www.stanford.edu/~tsachy/interest.htm>.
- H. Wold (1938) *A Study in the Analysis of Stationary Time Series*, Stockholm: Almqvist & Wiksell.
- J. G. Wolff (1980) *Language acquisition and the discovery of phrase structure*, Language and Speech, t. 23, s. 255–269.
- R. W. Yeung (2002) *First Course in Information Theory*, Kluwer Academic Publishers.
- J. Ziv, A. Lempel (1977) *A universal algorithm for sequential data compression*, IEEE Transactions on Information Theory, t. 23, s. 337–343.
- (1978) *Compression of individual sequences via variable-rate coding*, IEEE Transactions on Information Theory, t. 24, s. 530–536.

English summary

Łukasz Dębowski, *Excess entropy for stochastic processes over various alphabets*. Doctoral dissertation supervised by Jan Mielniczuk, Institute of Computer Science, Polish Academy of Science, Warsaw 2005.

Excess entropy is defined as mutual information between the σ -field of infinite past and the σ -field of infinite future. It constitutes a natural measure of long memory especially for discrete processes whose complex dependence cannot be described in the language of autocorrelations and linear predictors.

Excess entropy assumes nonnegative values. Stationary processes exhibiting infinite excess entropy are called infinitary (Crutchfield and Feldman, 2003). Gaussian infinitary processes cannot be represented as ARMA processes (Finch, 1960). On the other hand, discrete infinitary processes cannot be represented as finite hidden Markov models. The latter fact is of importance to natural language engineering. Although language engineers use finite hidden Markov models intensively, there is theoretical and empirical evidence that better probabilistic language models could be provided by some infinitary processes (Hilberg, 1990; Dębowski, 2005).

In this dissertation, we present some links between the value of excess entropy and a wide range of asymptotic properties for discrete and Gaussian stationary processes. We discuss the following topics:

1. Measure-theoretic approach to information theory.
 - a) Conditional mutual information as a function of σ -fields (Gel'fand *et al.*, 1956).
 - b) Equivalent integral and series expressions for entropy rate and excess entropy.
 - c) Ergodic decompositions of entropy rate and excess entropy.
2. The properties of stationary discrete processes.
 - a) The process is infinitary if there is a Bernoulli process whose values can be predicted arbitrarily well given any sufficiently long section of the process.
 - b) For any universal code, the expected excess code lengths are infinitely often greater than finite-order excess entropies. Nevertheless, for almost all stationary processes, the expected excess code lengths diverge to infinity.
 - c) We define mappings called spelling and segmentation (*przeliterowanie/segmentacja*) which map stationary processes onto differently-valued stationary processes. The mappings preserve the value of excess entropy under moderate conditions.
3. The properties of stationary Gaussian processes.
 - a) Expressing conditional mutual information in terms of partial autocorrelation.
 - b) Sufficient conditions for the existence of MA(∞) and AR(∞) representations.
 - c) Ergodicity and pure nondeterminism of finitary and quasifinitary processes.
 - d) Expressing the sum of autocorrelations as a product of partial autocorrelations.
 - e) Simple bounds imposed by the values of autocorrelations onto the values of partial autocorrelations and converse.

- f) Examples of ergodic infinitary processes with rapidly decaying autocorrelations.