

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Błażej Miasojedow

**Oszacowania błędów estymatorów
stosowanych w markowskich metodach
Monte Carlo**

rozprawa doktorska

Promotor rozprawy
dr hab. Wojciech Niemirowicz
Instytut Matematyki Stosowanej i Mechaniki

Kwiecień 2011

Oświadczenie autora rozprawy:
oświadczam, że niniejsza rozprawa została napisana przeze mnie samodzielnie.

.....

data

.....

Błażej Miasojedow

Oświadczenie promotora rozprawy:
niniejsza rozprawa jest gotowa do oceny przez recenzentów.

.....

data

.....

dr hab. Wojciech Niemirow

Streszczenie

W pracy przedstawione są rezultaty dotyczące oszacowania błędów estymatorów uzyskanych za pomocą algorytmów Monte Carlo opartych na łańcuchach Markowa (MCMC). W szczególności są to ograniczenia błędu średniokwadratowego (MSE) techniką regeneracyjną, oszacowania prawdopodobieństwa błędu przy pomocy nierówności Hoeffdinga dla łańcuchów Markowa. Przedstawiamy dowód uogólnionej nierówności Hoeffdinga za pomocą własności spektralnych operatorów w przestrzeni Hilberta. W ostatnim rozdziale pokazujemy, że użyteczne może być rozpatrywanie luki spektralnej w $L^2(\pi)$, również dla nieodwracalnych łańcuchów Markowa. Dzięki takiemu podejściu uzyskujemy rezultaty analogiczne do znanych wyników w przypadku odwracalnym.

Słowa kluczowe:

Łańcuch Markowa, MCMC, regeneracja, warunek dryfu, błąd średniokwadratowy, luka spektralna, nierówności Hoeffdinga.

Klasyfikacja tematyczna wg AMS:

60J10, 60J05, 60F15, 60F05

Tytuł pracy w języku angielskim:

Bounds on the estimation error for MCMC algorithms

Spis treści

1. Wstęp	5
1.1. Markowowskie metody Monte Carlo	5
1.2. Podstawowe algorytmy MCMC	5
1.2.1. Algorytm Metropolisa	7
1.2.2. Próbnik Gibbsa	8
1.3. Przegląd wyników	9
2. Łańcuchy Markowa	11
2.1. Podstawowe definicje i oznaczenia	11
2.2. Małe zbiory	14
2.3. Geometryczna ergodyczność i warunki dryfu	15
2.4. Równanie Poissona	17
3. Regeneracyjne oszacowania błędu średniokwadratowego	19
3.1. Regeneracja	19
3.2. Oszacowania MSE	20
3.3. Oszacowania $\sigma_{as}(f)$ za pomocą warunku dryfu	23
3.4. Oszacowania $\sigma_{as}(f)$ dla jednostajnie ergodycznych łańcuchów Markowa	29
4. Nierówności Hoeffdinga dla łańcuchów Markowa	31
4.1. Nierówności Hoeffdinga	31
4.2. Uogólnienia nierówności Hoeffdinga	37
5. Norma operatora P i jego luka spektralna w $L^2(\pi)$	41
5.1. Oszacowania ρ	41
5.1.1. Oszacowania dla skończonej przestrzeni stanów	43
5.1.2. Oszacowania dla jednostajnie ergodycznych łańcuchów Markowa	43
5.1.3. Oszacowania dla geometrycznie ergodycznych łańcuchów Markowa	44
5.2. Luka spektralna	47
5.3. Funkcjonalne CTG dla funkcji całkowalnych w kwadracie	50
Literatura	53

Rozdział 1

Wstęp

W tym rozdziale przedstawimy krótki opis markowowskich metod Monte Carlo (*Markov chain Monte Carlo*, MCMC) jako punkt wyjścia do rezultatów przedstawionych w pracy. Następnie opiszemy główne wyniki zawarte w pracy.

1.1. Markowowskie metody Monte Carlo

Niech \mathcal{X} będzie przestrzenią, najczęściej wysokiego wymiaru, oraz niech f będzie funkcją na \mathcal{X} o wartościach rzeczywistych. Rozważmy rozkład probabilistyczny π na \mathcal{X} o gęstości p względem miary dx (zazwyczaj jest to miara Lebesgue'a lub miara licząca), czyli $\pi(dx) = p(x)dx$. Ważnym problemem w statystyce bayesowskiej, fizyce statystycznej i wielu innych dziedzinach jest obliczenie wartości całki

$$(1.1) \quad \theta = \mathbb{E}_\pi f = \pi(f) = \int_{\mathcal{X}} f(x)\pi(dx),$$

w sytuacji, gdy gęstość p jest znana tylko z dokładnością do stałej normalizującej, trudnej do obliczenia oraz nie można bezpośrednio losować z π [8, 35]. W tej sytuacji typowym rozwiązaniem problemu jest skonstruowanie łańcucha Markowa $(X_n)_{n \geq 0}$ zbieżnego do rozkładu stacjonarnego π . Wówczas dla dostatecznie dużych n , próbki z łańcucha $(X_n)_{n \geq 0}$ można traktować jak obserwacje z rozkładu π . Odkąd w 1953 roku został przedstawiony prosty i użyteczny algorytm Metropolis'a [39], powstało wiele różnych algorytmów konstruowania łańcuchów oraz metod estymacji nieznannej wartości θ opartych na łańcuchach Markowa [41, 35, 8]. Metody te nazywane są markowowskimi metodami Monte Carlo (MCMC).

Dzięki rozwojowi metod MCMC stało się możliwe generowanie próbek ze skomplikowanych rozkładów *a posteriori* oraz obliczanie całek względem tych rozkładów w złożonych modelach statystyki bayesowskiej [8, 35]. Możliwości jakie dały metody MCMC spowodowały, że modele bayesowskie stały się ważnym i popularnym narzędziem używanym przez statystyków w praktyce. Dlatego ważne jest badanie tempa zbieżności do rozkładu stacjonarnego π przy $n \rightarrow \infty$ oraz oszacowania błędów estymatorów uzyskanych metodami MCMC.

1.2. Podstawowe algorytmy MCMC

Zanim opiszemy algorytmy MCMC przypomnijmy jak wygląda rozwiązanie problemu (1.1) niezależnymi metodami Monte Carlo (MC). Jeśli potrafimy losować z π , to konstruujemy ciąg

i.i.d. obserwacji X_1, X_2, \dots, X_n i estymujemy θ za pomocą

$$(1.2) \quad \hat{\theta}_n := \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Uwaga 1.3. Podstawowe własności estymatorów uzyskanych przy pomocy niezależnego MC są bardzo proste do otrzymania:

- Jeżeli θ istnieje, wówczas $\hat{\theta}_n$ jest nieobciążonym i zgodnym estymatorem θ , na mocy prawa wielkich liczb.
- Jeśli ponadto $\pi(f^2) < \infty$, to zachodzi standardowe Centralne Twierdzenie Graniczne (CTG)

$$\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \pi(f^2) - (\pi f)^2).$$

- Błąd średniokwadratowy (*Mean Square Error*, MSE) jest prosty do oszacowania

$$MSE := \mathbb{E}(\hat{\theta}_n - \theta)^2 = \frac{\pi(f^2) - (\pi f)^2}{n},$$

przy czym wariancję $\pi(f^2) - (\pi f)^2$ z reguły można oszacować.

- Za pomocą znanych nierówności (np. Czebyszewa, Hoeffdinga, Bernsteina,...) możemy otrzymać przedziały ufności.
- Korzystając z CTG otrzymujemy asymptotyczne przedziały ufności, zazwyczaj krótsze niż dokładne przedziały.

Załóżmy teraz, że nie potrafimy losować z π i konstruujemy ergodyczny łańcuch Markowa $(X_n)_{n \geq 0}$ z jądrem przejścia $P(x, A)$ i rozkładem stacjonarnym π . Wówczas otrzymany ciąg zmiennych losowych nie jest niezależny i estymatora (1.2) nie można już badać w tak prosty sposób jak w przypadku i.i.d.

Jest wiele różnych strategii estymowania θ za pomocą MCMC [16, 41, 9, 35, 8]

- *Średnia z próbki.* Używamy tylko jednej trajektorii łańcucha, czyli estymator jest postaci

$$(1.4) \quad \hat{\theta}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(X_i).$$

Często, aby zmniejszyć obciążenie estymatora odrzuca się początek trajektorii ustalonej długości t

$$(1.5) \quad \hat{\theta}_{t,n} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_i),$$

t nazywa się "burn-in".

- *Średnia z przerwami.* Aby zmniejszyć obciążenie odrzucamy początek trajektorii, dodatkowo aby zmniejszyć korelację pomiędzy kolejnymi obserwacjami wybieramy tylko elementy odległe od siebie o s kroków

$$(1.6) \quad \hat{\theta}_{t,n,s} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_{is}).$$

- *Wielokrotna symulacja* Używamy średniej z końcowych stanów n niezależnych łańcuchów. Generujemy m niezależnych trajektorii długości t

$$\begin{array}{c} X_0^{(1)}, X_1^{(1)} \quad \dots, \quad X_t^{(1)}, \\ \vdots \\ X_0^{(n)}, X_1^{(n)} \quad \dots, \quad X_t^{(n)}, \end{array}$$

i za estymator przyjmujemy

$$(1.7) \quad \hat{\theta}_{t,n} = \frac{1}{n} \sum_{m=1}^n f(X_t^{(m)}),$$

gdzie n jest liczbą niezależnych kopii łańcucha, natomiast t powinno być na tyle duże, aby zredukować obciążenie.

- *Mediana ze średnich.* Symulujemy m niezależnych kopii łańcucha. Na podstawie każdej kopii trajektorii obliczamy średnią i za estymator przyjmujemy medianę z tych średnich.
 - Symulujemy m niezależnych kopii łańcucha Markowa o długości n ,

$$X_0^{(k)}, \dots, X_{n-1}^{(k)}, \quad k = 1, \dots, m.$$

- Obliczamy m estymatorów θ na podstawie pojedynczych trajektorii,

$$\hat{\theta}_k = \hat{\theta}_n^{(k)} = \frac{1}{n} \sum_{i=0}^{n-1} f(X_i^{(k)}), \quad k = 1, \dots, m.$$

- Za końcowy estymator przyjmujemy

$$\hat{\theta} = \text{med}\{\hat{\theta}_1, \dots, \hat{\theta}_m\}.$$

Strategie oparte na średnich z pojedynczych trajektorii są trudniejsze do analizy od wielokrotnej symulacji, ponieważ obserwacje nie są niezależne. Jednak wydają się być bardziej efektywne i są zdecydowanie częściej używane w praktyce. Liczba obserwacji jakie musimy wygenerować zależy od parametrów (n, t, s, m) , które wybieramy tak, aby estymator miał "dobre" własności. To zależy oczywiście od definicji dobrych własności. Istnieją wyniki teoretyczne, które porównują różne strategie i z reguły są zgodne z intuicją.

Teraz przedstawimy dwa bardzo ogólne i popularne schematy konstruowania ergodycznych łańcuchów Markowa o zadanym rozkładzie stacjonarnym.

1.2.1. Algorytm Metropolisa

Algorytm Metropolisa został przedstawiony w pracy Metropolisa i in. [39] w roku 1953. Niech $Q(x, dy)$ będzie jądrem przejścia łańcucha Markowa, prostym do symulowania, z gęstością $Q(x, dy) = q(x, y)dy$. Założmy, że rozkład π ma gęstość, $\pi(dx) = p(x)dx$, z nieznaną stałą normującą.

Algorytm Metropolisa

1. Wylosuj X_0 z rozkładu początkowego π_0 (zazwyczaj $\pi_0 = \delta_{x_0}$ dla pewnego $x_0 \in \mathcal{X}$).

2. Będąc w punkcie X_n wylosuj "propozycję" Y_{n+1} z $Q(X_n, \cdot)$.

3. Przyjmij

$$X_{n+1} = \begin{cases} Y_{n+1} & \text{z prawdopodobieństwem } \alpha(X_n, Y_{n+1}), \\ X_n & \text{z prawdopodobieństwem } 1 - \alpha(X_n, Y_{n+1}), \end{cases}$$

gdzie

$$\alpha(x, y) := \min \left\{ 1, \frac{p(y)q(y, x)}{p(x)q(x, y)} \right\}$$

4. Zmień n na $n + 1$ i powróć do 2.

Zauważmy, że w tym algorytmie używamy gęstości p do obliczenia prawdopodobieństwa akceptacji $\alpha(x, y)$, jednak ta gęstość występuje tylko w postaci $\frac{p(x)}{p(y)}$. A ten iloraz możemy obliczyć bez znajomości stałej normującej.

Ważnym elementem algorytmu jest wybór jądra przejścia $Q(x, dy)$. Ze względu na własności tego jądra wyróżnia się kilka klas (zob. [45]):

- *Symetryczny algorytm Metropolisa*. W sytuacji, gdy $q(x, y) = q(y, x)$ mamy $\alpha(x, y) = \min\{1, \frac{\pi(y)}{\pi(x)}\}$.
- *Błądzenie losowe Metropolisa Hastingsa (Random Walk Metropolis-Hastings)*. W sytuacji, gdy $q(x, y) = q(y - x)$, rozkład propozycji odpowiada błądzeniu losowemu.
- *Niezależny próbnik Metropolisa (Independence Sampler)*. W tym algorytmie rozkład propozycji nie zależy od x , czyli $q(x, y) = q(y)$.
- *Algorytm Langevin'a* W tym algorytmie $Q(X_n, \cdot) = N(X_n + (\delta/2)\nabla \log \pi(X_n), \delta)$ dla pewnej $\delta > 0$, gdzie ∇ oznacza gradient.

1.2.2. Próbnik Gibbsa

W statystyce bayesowskiej często mamy do czynienia z sytuacją, gdy mamy przestrzeń produktową tzn. $\mathcal{X} = \prod_{i=1}^d \mathcal{X}_i$, najczęściej jest to \mathbb{R}^d , dodatkowo potrafimy losować ze wszystkich rozkładów warunkowych $\pi(x_i|x_{-i})$, gdzie x_{-i} jest wektorem wszystkich współrzędnych poza i -tą. W tej sytuacji efektywną i prostą metodą konstrukcji łańcucha Markowa o rozkładzie stacjonarnym π jest próbnik Gibbsa. Będąc w stanie $X_n := (X_{1,n}, \dots, X_{d,n})$ losujemy kolejne współrzędne punktu X_{n+1} , przy tym i -tą współrzędną generujemy z rozkładu warunkowego

$$\pi(X_{i,n+1}|X_{1,n+1}, \dots, X_{i-1,n+1}, X_{i+1,n}, \dots, X_{d,n}).$$

Następnie po przejściu przez wszystkie współrzędne otrzymujemy nowy punkt $X_{n+1} := (X_{1,n+1}, \dots, X_{d,n+1})$. Zapiszemy to teraz bardziej formalnie, używając notacji jak w [45], dodatkowo ta notacja umożliwi nam w prosty sposób opisać różne wersje tego algorytmu. Niech

$$S_{x,i,A} = \{y \in \mathcal{X}; y_j = x_j \text{ for } j \neq i, \text{ oraz } a \leq y_i \leq b\}.$$

oraz

$$(1.8) \quad P_i(x, S_{x,i,a,b}) = \frac{\int_a^b p(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d) dt}{\int_{-\infty}^{\infty} p(x_1, \dots, x_{i-1}, t, x_{i+1}, \dots, x_d) dt}.$$

Jądro $P_i(x, \cdot)$ odpowiada zmianie i -tej współrzędnej, czyli jądro przejścia $P(x, \cdot)$ standardowego próbnika Gibbsa jest równe

$$(1.9) \quad P = P_1 \cdots P_{d-1} P_d.$$

Często używa się algorytmu, w którym zamiast ustalonej kolejności współrzędnych stosujemy losowy ich wybór. Taki algorytm nazywa się *próbnikiem Gibbsa z losowym przeglądem* (*Random scan Gibbs Sampler*) i jego jądro przejścia zapisuje się w postaci

$$(1.10) \quad P = \frac{1}{d} \sum_{i=1}^d P_i.$$

W praktyce zdarza się, że nie potrafimy losować z jednego lub kilku rozkładów warunkowych $\pi(x_i|x_{-i})$, wówczas możemy w miejsce P_i wstawić jeden krok odpowiedniego algorytmu Metropolisa. Takie algorytmy nazywa się *Metropolis wewnątrz Gibbsa* (*Metropolis within Gibbs*).

1.3. Przegląd wyników

Dostępna jest bogata literatura dotycząca zarówno algorytmów MCMC jak i łańcuchów Markowa w zastosowaniu do MCMC. Wiele prac dotyczy badania tempa zbieżności do rozkładu stacjonarnego [5, 12, 22, 47, 48], równie wiele prac jest poświęconych badaniu asymptotycznych własności estymatorów [23, 25, 40].

Naszym podstawowym celem będzie otrzymanie nieasymptotycznych oszacowań błędów estymatorów. Tego typu wyniki istnieją tylko przy dość ograniczających założeniach: dla dyskretnej przestrzeni stanów \mathcal{X} i ograniczonej funkcji f [2, 17, 33, 34, 41], w przypadku łańcuchów jednostajnie ergodycznych na ogólnej przestrzeni stanów i ograniczonej funkcji f [18, 26]. Założenie o jednostajnej ergodyczności w praktyce oznacza, że przestrzeń stanów \mathcal{X} jest zwarta, co jest założeniem rzadko spełnionym w zastosowaniach statystycznych.

Niewiele jest wyników przy bardziej realistycznych założeniach. Dla odwracalnych łańcuchów Markowa oszacowania błędu średniokwadratowego $MSE = \mathbb{E}(\hat{\theta}_n - \theta)^2$ za pomocą luki spektralnej w $L^2(\pi)$ można znaleźć w [46]. Dla geometrycznie ergodycznych łańcuchów na ogólnej przestrzeni stanów MSE jest oszacowany w [31, 29] oraz w [14]. My w rozdziale 3 przedstawimy oszacowania MSE przy tych samych założeniach uzyskane inną metodą, dającą w większości przypadków dokładniejsze wyniki. Przedstawione rozumowanie oparte jest na badaniu łańcucha techniką regeneracji. Rozdział ten oparty jest na pracy [32], w której zawarte są oszacowania MSE dla sekwencyjnego estymatora o losowej długości symulacji.

W rozdziale 4 przedstawimy nierówności Hoeffdinga dla łańcuchów Markowa przy założeniu wiedzy o pewnych własnościach spektralnych operatora przejścia w $L^2(\pi)$, które w przypadku odwracalnym sprowadzają się do geometrycznej ergodyczności. Natomiast w przypadku nieodwracalnym prowadzą do węższej klasy łańcuchów o interesujących właściwościach. Wynik ten jest uogólnieniem rezultatu dla skończonej przestrzeni zawartego w [33].

Dla odwracalnych łańcuchów Markowa operator przejścia jest samosprzężonym operatorem w przestrzeni Hilberta $L^2(\pi)$. Ta własność znacznie ułatwia badanie łańcuchów [5, 44, 25, 46, 16]. My w rozdziale 5 pokażemy, że zakładając istnienie luki spektralnej w $L^2(\pi)$

można uzyskać takie same lub zbliżone wyniki jak w przypadku odwracalnym. W szczególności pokażemy nowy dowód faktu (po raz pierwszy zauważonego w pracach [15, 28]), że przy założeniu istnienia luki spektralnej w $L^2(\pi)$ zachodzi CTG dla wszystkich funkcji z tej przestrzeni. Nasze rozumowanie prowadzi do funkcjonalnego CTG. Zamieszczamy również kontrprzykład pokazujący, że CTG nie musi zachodzić dla geometrycznie ergodycznych łańcuchów dla wszystkich funkcji z $L^2(\pi)$. Nasz kontrprzykład jest znacznie prostszy od powszechnie cytowanych przykładów [7, 20].

W rozdziale 2 przedstawimy w skrócie znane fakty, z których będziemy korzystać w dalszej części pracy.

Rozdział 2

Łańcuchy Markowa

W tym rozdziale opiszemy podstawowe własności łańcuchów Markowa oraz fakty, z których będziemy w dalszej części pracy korzystać.

2.1. Podstawowe definicje i oznaczenia

Niech \mathcal{X} będzie przestrzenią polską z borelowskim σ -ciałem $\mathcal{B}(\mathcal{X})$. W całej pracy przez $(X_n)_{n \geq 0}$ będziemy oznaczać jednorodny łańcuch Markowa na przestrzeni \mathcal{X} . Oznaczmy przez $P(x, A)$ jądro przejścia łańcucha $(X_n)_{n \geq 0}$ tzn.

$$\mathbb{P}(X_{n+1} \in A | X_n = x) = P(x, A).$$

Ponadto wprowadzimy następujące oznaczenia

$$\mathbb{P}_x(\cdot) := \mathbb{P}(\cdot | X_0 = x)$$

$$\mathbb{E}_x(\cdot) := \mathbb{E}(\cdot | X_0 = x)$$

Analogicznie \mathbb{P}_ξ będzie oznaczać prawdopodobieństwo, a \mathbb{E}_ξ wartość oczekiwaną przy założeniu, że $X_0 \sim \xi$.

Z jądrem przejścia $P(x, A)$ jest związany operator liniowy P działający z prawej strony na miarach, a z lewej strony na funkcjach. Dla każdej miary ξ oraz funkcji $f : \mathcal{X} \rightarrow \mathbb{R}$, operator P jest zdefiniowany następująco:

$$\xi P(\cdot) := \int_{\mathcal{X}} P(x, \cdot) \xi(dx), \quad Pf(x) := \mathbb{E}_x(f(X_1)) = \int_{\mathcal{X}} f(y) P(x, dy).$$

Przez π będziemy oznaczać rozkład stacjonarny łańcucha Markowa $(X_n)_{n \geq 0}$, tzn. taki rozkład, że spełnione jest następujące równanie

$$(2.1) \quad \pi P = \pi.$$

Jeżeli $X_0 \sim \pi$, to w każdej chwili n rozkład $X_n \sim \pi$. Ze względu na zastosowania do MCMC będziemy rozpatrywać tylko łańcuchy, dla których istnieje rozkład stacjonarny i jest jednoznacznie wyznaczony.

Do badania zbieżności do rozkładu stacjonarnego będziemy używać odległości pełnego wahania. Jeżeli mamy miary ξ_1 oraz ξ_2 to odległością pełnego wahania nazwiemy

$$(2.2) \quad \|\xi_1 - \xi_2\|_{tv} = \sup_{A \subseteq \mathcal{X}} |\xi_1(A) - \xi_2(A)| = \frac{1}{2} \sup_{|f| \leq 1} |\xi_1(f) - \xi_2(f)|.$$

Będziemy rozważać również uogólnienie odległości pełnego wahania, nazywane V - normą. Dla zadanej funkcji $V \geq 1$ definiujemy normę miary ze znakiem ξ następująco

$$(2.3) \quad \|\xi\|_V = \sup_{|f| \leq V} |\xi(f)|.$$

Aby badać własności operatorów przejścia musimy zdefiniować przestrzenie na jakich działają. Będziemy rozpatrywać dwie przestrzenie. Pierwszą oznaczmy przez L_V^∞ . Jest to przestrzeń Banacha L^∞ z "wagą" zadaną przez funkcję $V \geq 1$, z normą $|f|_V := \sup_{\{x \in \mathcal{X}\}} \frac{|f(x)|}{V(x)}$. Zatem mamy:

$$(2.4) \quad L_V^\infty = \{f : \mathcal{X} \rightarrow \mathbb{R} : |f|_V < \infty\}.$$

Inną przestrzenią jaką będziemy rozważać jest przestrzeń Hilberta $L^2(\pi)$, czyli przestrzeń funkcji całkowalnych w kwadracie względem miary π z iloczynem skalarnym

$$\langle f, g \rangle := \int_{\mathcal{X}} f(x)g(x)\pi(dx)$$

i normą $\|f\|_2^2 = \langle f, f \rangle$. Przestrzeń $L^2(\pi)$ jest szczególnie użyteczna przy badaniu łańcuchów odwracalnych ze względu na π .

Definicja 2.5. Łańcuch Markowa $(X_n)_{n \geq 0}$ nazywamy odwracalnym względem π wtedy i tylko wtedy, gdy spełnione jest równanie równowagi szczegółowej

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx).$$

Uwaga 2.6. Jeżeli łańcuch $(X_n)_{n \geq 0}$ jest odwracalny względem π , to π jest rozkładem stacjonarnym:

$$\pi P(A) = \int_{\mathcal{X}} P(x, A)\pi(dx) = \int_A P(y, \mathcal{X})\pi(dy) = \int_A \pi(dy) = \pi(A).$$

Dla odwracalnych łańcuchów operator przejścia P jest operatorem samosprzężonym w $L^2(\pi)$. Z twierdzenia Fubiniiego oraz definicji odwracalności dla każdego $f, g \in L^2(\pi)$ mamy

$$\begin{aligned} \langle f, Pg \rangle &= \int_{\mathcal{X}} f(x)Pg(x)\pi(dx) = \int_{\mathcal{X}} \int_{\mathcal{X}} f(x)g(y)P(x, dy)\pi(dx) \\ &= \int_{\mathcal{X}} \int_{\mathcal{X}} f(x)g(y)P(y, dx)\pi(dy) = \int_{\mathcal{X}} g(y)Pf(y)\pi(dy) = \langle Pf, g \rangle. \end{aligned}$$

Ze względów algebraicznych, aby odróżnić operator od miary, będziemy oznaczać przez Π operator liniowy przypisujący funkcji f jej stacjonarną wartość oczekiwaną $\Pi f := \pi(f)$.

Zdefiniujemy normy operatorowe w przestrzeniach $L^2(\pi)$ oraz L_V^∞ w standardowy sposób:

$$\|T\|_{L^2(\pi)} := \sup_{\|f\|_2=1} \|Tf\|_2$$

oraz

$$\|T\|_{L_V^\infty} := \sup_{|f|_V=1} |Tf|_V.$$

Definicja 2.7. Łańcuch nazywamy ψ -nieprzywiedlnym, jeżeli niezerowa miara ψ na $\mathcal{B}(\mathcal{X})$ spełnia następujący warunek: Dla każdego $x \in \mathcal{X}$ i dla każdego zbioru A miary $\psi(A) > 0$ istnieje n , dla którego zachodzi

$$P^n(x, A) > 0.$$

Uwaga 2.8. W dalszej części pracy rozważać będziemy tylko ψ -nieprzywiedlne łańcuchy Markowa, które posiadają rozkład stacjonarny π . W tej sytuacji wiadomo, że π jest maksymalną miarą nieprzywiedłości, tzn. dla każdej miary ψ spełniającej warunek z definicji 2.7 mamy $\psi \ll \pi$.

Definicja 2.9 (Meyn & Tweedie [40]). Zbiór $A \in \mathcal{B}(\mathcal{X})$ nazywamy powracającym, jeśli

$$\forall x \in A \quad \mathbb{E}_x \left[\sum_{n=0}^{\infty} \mathbb{I}(X_n \in A) \right] = \infty.$$

Zbiór A nazywamy chwilowym, jeśli nie jest powracający. Zbiór A jest powracający w sensie Harrisa, jeśli

$$\forall x \in A \quad \mathbb{P}_x \left(\sum_{n=0}^{\infty} \mathbb{I}(X_n \in A) = \infty \right) = 1,$$

równoważnie $\forall x \in A \quad \mathbb{P}_x(\tau(A) < \infty) = 1$, gdzie

$$(2.10) \quad \tau(A) = \inf \{ n \geq 1 : X_n \in A \}.$$

Uwaga 2.11. Jeśli łańcuch jest π -nieprzywiedlny to albo wszystkie zbiory miary π dodatniej są powracające albo wszystkie są chwilowe [40, Twierdzenie 8.0.1]. W pierwszym przypadku mówimy, że łańcuch jest powracający.

Definicja 2.12. Łańcuch π -nieprzywiedlny nazywamy łańcuchem Harrisa, jeśli każdy zbiór miary π dodatniej jest zbiorem powracającym w sensie Harrisa.

Uwaga 2.13. Jeśli łańcuch jest powracający, to istnieje rozbitcie przestrzeni stanów:

$$\mathcal{X} = N \cup H,$$

takie, że $\pi(N) = 0$ oraz zbiór H jest pochłaniający, czyli $\forall x \in H \quad P(x, H) = 1$. Co więcej łańcuch ograniczony do H jest łańcuchem Harrisa [40, Twierdzenie 9.0.1].

Uwaga 2.14. Każdy łańcuch Harrisa spełnia następujący warunek [40, Twierdzenie 9.1.4]. :

$$\forall x \in \mathcal{X} \quad \forall A \in \mathcal{B}(\mathcal{X}) : \pi(A) > 0 \quad \mathbb{P}_x \left(\sum_{n=0}^{\infty} \mathbb{I}(X_n \in A) = \infty \right) = 1.$$

Wiadomo, że jeżeli istnieje rozkład stacjonarny π to łańcuch jest powracający. Zatem wykorzystując uwagę 2.13 możemy bez straty ogólności (zawężając przestrzeń stanów) zakładać, że łańcuch jest łańcuchem Harrisa.

Definicja 2.15. Łańcuch Markowa π -nieprzywiedlny nazywamy okresowym z okresem $d \geq 2$, jeśli istnieje rozbitcie przestrzeni \mathcal{X} na rozłączne podzbiory $\mathcal{X}_0, \mathcal{X}_1, \dots, \mathcal{X}_{d-1}$ o mierze π dodatniej takie, że

$$\mathbb{P}(x, \mathcal{X}_i) = 1, \quad \text{dla każdego } x \in \mathcal{X}_{i-1 \bmod d}.$$

W przeciwnym przypadku łańcuch jest nieokresowy.

2.2. Małe zbiory

Do badania własności łańcuchów Markowa bardzo użyteczne jest pojęcie małych zbiorów. W literaturze można spotkać wiele różnych definicji małego zbioru. My przedstawimy trzy definicje małego zbioru. Pierwsza, zaczerpnięta z [40], jest najogólniejsza i pozwala na sformułowanie twierdzeń w pełnej ogólności.

Definicja 2.16 (*a_n -mały zbiór ("Petite" set)*). Zbiór $C \in \mathcal{B}(\mathcal{X})$, miary $\pi(C) > 0$, nazywamy a_n -małym zbiorem \iff Istnieją: $\beta > 0$, rozkład prawdopodobieństwa (a_n) na \mathbb{Z}^+ oraz miara probabilistyczna ν na $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ takie, że dla każdego $x \in \mathcal{X}$ zachodzi:

$$\sum_{n=1}^{\infty} a_n P^n(x, A) \geq \beta \mathbb{I}(x \in C) \nu(A)$$

oraz $\nu(C) > 0$

Druga definicja jest szczególnym przypadkiem a_n -małego zbioru, gdy rozkład (a_n) jest skupiony w jednym punkcie.

Definicja 2.17 (*n -mały zbiór (*Small set*)*). Zbiór $C \in \mathcal{B}(\mathcal{X})$, miary $\pi(C) > 0$, nazywamy n -małym zbiorem \iff Istnieją: $\beta > 0$, n oraz miara probabilistyczna ν na $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ takie, że dla każdego $x \in \mathcal{X}$ zachodzi:

$$P^n(x, A) \geq \beta \mathbb{I}(x \in C) \nu(A)$$

oraz $\nu(C) > 0$.

Ostatnia z cyklu definicji małych zbiorów, jest często nazywana warunkiem minoryzacji. Jest to szczególny przypadek n -małego zbioru z $n = 1$, jednak będziemy go rozpatrywać osobno ze względu na to, że łańcuch w tym przypadku ma inne własności.

Definicja 2.18 (*Mały zbiór (*Minorization condition*)*). Zbiór $C \in \mathcal{B}(\mathcal{X})$, miary $\pi(C) > 0$, nazywamy małym zbiorem \iff Istnieją: $\beta > 0$ oraz miara probabilistyczna ν na $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ takie, że dla każdego $x \in \mathcal{X}$ zachodzi:

$$P(x, A) \geq \beta \mathbb{I}(x \in C) \nu(A).$$

Zauważmy, że w ostatniej definicji nie zakładamy, że $\nu(C) > 0$. Zgodnie z (2.10) oznaczamy przez $\tau(C)$ pierwszy moment powrotu do zbioru C . Następne twierdzenie jest uogólnieniem znanego twierdzenia Kaca [40, 43] i pozwala przedstawić rozkład stacjonarny za pomocą czasu powrotu do małego zbioru.

Twierdzenie 2.19 (Meyn & Tweedie 10.0.1 [40]). *Jeżeli $(X_n)_{n \geq 0}$ jest łańcuchem powracającym wówczas istnieje dokładnie jedna miara niezmiennicza π , którą dla każdego zbioru A miary ψ dodatniej możemy przedstawić następująco*

$$\pi(B) = \int_A \mathbb{E}_x \left[\sum_{i=1}^{\tau(A)} \mathbb{I}(X_i \in B) \right] \pi(dx), \quad \forall B \in \mathcal{B}(\mathcal{X}).$$

Ponadto miara π jest skończona, jeśli istnieje a_n -mały zbiór C spełniający:

$$\sup_{x \in C} \mathbb{E}_x \tau(C) < \infty.$$

Z powyższego twierdzenia wynika prosty wniosek.

Wniosek 2.20. *Jeśli łańcuch $(X_n)_{n \geq 0}$ posiada rozkład stacjonarny π oraz jeżeli funkcja $f : \mathcal{X} \rightarrow \mathbb{R}$ spełnia $\pi(|f|) < \infty$, to dla każdego zbioru A , miary π dodatniej, prawdziwa jest równość:*

$$\pi(f) = \int_A \mathbb{E}_x \left[\sum_{i=1}^{\tau(A)} f(X_i) \right] \pi(dx)$$

Za pomocą małych zbiorów można przedstawić warunki konieczne i dostateczne, aby zachodziło \sqrt{n} -CTG [6, 40]. Mówimy, że \sqrt{n} -CTG zachodzi dla funkcji $f : \mathcal{X} \rightarrow \mathbb{R}$, jeżeli istnieje $\sigma_{\text{as}}^2(f) < \infty$ takie, że prawdziwe jest

$$\frac{\bar{S}_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_{\text{as}}^2(f)), \quad \text{gdy } n \rightarrow \infty,$$

gdzie $\bar{S}_n := \sum_{i=0}^{n-1} \bar{f}(X_i)$ oraz $\bar{f} = f - \pi f$. Jeśli zachodzi \sqrt{n} -CTG to $\sigma_{\text{as}}^2(f)$ nazywamy asymptotyczną wariancją.

Dla naszych potrzeb wystarczy nam sformułowanie w przypadku, gdy łańcuch posiada atom. Zbiór $\alpha \in \mathcal{B}(\mathcal{X})$ jest atomem jeśli $\pi(\alpha) > 0$ oraz dla każdego $x \in \alpha$ mamy $P(x, \cdot) = \nu(\cdot)$, gdzie ν jest pewną miarą probabilistyczną.

Twierdzenie 2.21 (Bednorz i in. [6]). *Dla łańcucha $(X_n)_{n \geq 0}$ posiadającego atom α z rozkładem stacjonarnym π zachodzi \sqrt{n} -CTG wtedy i tylko wtedy, gdy*

$$\mathbb{E}_x \left[\left(\sum_{i=1}^{\tau(\alpha)} \bar{f}(X_i) \right)^2 \right] < \infty,$$

ponadto $\sigma_{\text{as}}^2(f) = \pi(\alpha) \mathbb{E}_x \left[\left(\sum_{i=1}^{\tau(\alpha)} \bar{f}(X_i) \right)^2 \right]$, gdzie $x \in \alpha$. Wobec faktu, że α jest atomem, powyższe wzory nie zależą od wyboru x .

2.3. Geometryczna ergodyczność i warunki dryfu

Definicja 2.22. Łańcuch Markowa $(X_n)_{n \geq 0}$ z rozkładem stacjonarnym π oraz z jądrem przejścia $P(x, A)$ jest geometrycznie ergodyczny $\iff \exists \rho < 1 \forall x \in \mathcal{X}, \pi\text{-p.n.} \exists M_x < \infty :$

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{tv} \leq M_x \rho^n.$$

Kres dolny liczb ρ o tej własności nazywamy współczynnikiem ergodyczności.

Do badania geometrycznej ergodyczności ważnym narzędziem jest warunek dryfu do małego zbioru. Istnieje wiele takich warunków różniących się szczegółami. Dokładniej o tym można przeczytać w [45] lub w [40]. My będziemy korzystać z warunku dryfu zaczerpniętego z [5], z niewielkimi modyfikacjami.

Definicja 2.23 (Warunek dryfu). Powiemy, że łańcuch $(X_n)_{n \geq 0}$ spełnia warunek dryfu do zbioru C , jeżeli istnieją: funkcja $V : \mathcal{X} \rightarrow [1, \infty)$, stała $\lambda < 1$ oraz stała $K < \infty$ takie, że zachodzi

$$PV(x) \leq \begin{cases} \lambda V(x) & \text{dla } x \notin C, \\ K & \text{dla } x \in C. \end{cases}$$

Następne twierdzenia przedstawiają związek warunku dryfu z geometryczną ergodycznością. Pierwsze jest złożeniem Twierdzenia 15.0.1 oraz 16.0.1 z [40] i Stwierdzenia 2.1 z [44].

Twierdzenie 2.24. *Niech $(X_n)_{n \geq 0}$ będzie π -nieprzywiedlnym i nieokresowym łańcuchem Markowa z jądrem przejścia $P(x, A)$. Wówczas następujące warunki są równoważne:*

- (i) Łańcuch $(X_n)_{n \geq 0}$ jest geometrycznie ergodyczny.
- (ii) Łańcuch $(X_n)_{n \geq 0}$ spełnia warunek dryfu do pewnego a_n -małego zbioru C z funkcją V .
- (iii) Istnieje n -mały zbiór C , $\rho_C < 1$ oraz stała $M_C < \infty$ takie, że dla każdego $x \in C$ zachodzi:

$$\|\pi|_C P^n(\cdot) - \pi(\cdot)\|_{tv} \leq M_C \rho_C^n,$$

$$\text{gdzie } \pi|_C(A) := \frac{\pi(C \cap A)}{\pi(C)}.$$

- (iv) Istnieje a_n -mały zbiór C oraz $\kappa > 1$ takie, że zachodzi:

$$\sup_{x \in C} \mathbb{E}_x \left[\kappa^{\tau(C)} \right] < \infty.$$

- (v) Istnieje $V_1 \geq 1$, $\rho < 1$ oraz $M < \infty$ takie, że dla każdego n i każdego $x \in \mathcal{X}$ zachodzi:

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{V_1} \leq M V_1(x) \rho^n.$$

- (vi) Istnieje $V_1 \geq 1$ oraz n takie, że zachodzi :

$$\|P^n - \Pi\|_{L^\infty_{V_1}} < 1.$$

Funkcje V (występującą w punkcie iii) oraz V_1 (występującą w punktach v i vi) można wybrać tak, aby były równoważne w następującym sensie: Istnieje stała M_1 taka, że dla każdego $x \in \mathcal{X}$ zachodzi $M_1^{-1} V_1(x) \leq V(x) \leq M_1 V_1(x)$. Ponadto funkcję V można tak dobrać, aby dla każdego ustalonego j zachodziło $\pi(V^j) < \infty$.

Jeżeli dodatkowo założymy, że łańcuch $(X_n)_{n \geq 0}$ jest odwracalny, to zachodzi następujące twierdzenie.

Twierdzenie 2.25. *Jeżeli $(X_n)_{n \geq 0}$ jest odwracalnym ψ -nieprzywiedlnym i nieokresowym łańcuchem Markowa z rozkładem stacjonarnym π oraz jądrem przejścia $P(x, A)$, to następujące warunki są równoważne:*

- (i) Łańcuch $(X_n)_{n \geq 0}$ jest geometrycznie ergodyczny.
- (ii) Dla każdego rozkładu początkowego ξ takiego, że $\frac{d\xi}{d\pi} \in L^2(\pi)$, istnieje $\rho < 1$ oraz stała M_ξ takie, że dla każdego n zachodzi

$$\|\xi P^n(\cdot) - \pi(\cdot)\|_{tv} \leq M_\xi \rho^n.$$

- (iii) Istnieje $\rho < 1$ takie, że zachodzi

$$\|P - \Pi\|_{L^2(\pi)} \leq \rho.$$

Ponadto ρ w punkcie (ii) oraz (iii) jest to samo.

Dowód. Równoważność (ii) \Leftrightarrow (iii) oraz implikacja (i) \Rightarrow (ii) wynikają z Twierdzenia 2.1 z [44].

(ii) \Rightarrow (i) Załóżmy, że zachodzi punkt (ii). Oczywiście dla dowolnego małego zbioru C rozkład stacjonarny obcięty do tego zbioru należy do $L^2(\pi)$. Z punktu (iii) Twierdzenia 2.24 łańcuch $(X_n)_{n \geq 0}$ jest geometrycznie ergodyczny. \square

2.4. Równanie Poissona

W tej części pracy przedstawimy metodę pozwalającą stosować teorię martyngałów do badania własności łańcuchów Markowa. Dla zadanej funkcji f powiemy, że funkcja \hat{f} spełnia równanie Poissona, jeżeli

$$(2.26) \quad \hat{f} - P\hat{f} = f - \pi f$$

Oczywistym rozwiązaniem tego równania jest $\hat{f} = \sum_{n=0}^{\infty} P^n \bar{f}(X_i)$, gdzie $\bar{f} = f - \pi f$. Zauważmy, że sumę $Z_n := \sum_{k=1}^n \bar{f}(X_k)$ można zapisać następująco

$$\begin{aligned} Z_n &= \sum_{k=1}^n \bar{f}(X_k) = \sum_{k=1}^n [\hat{f}(X_k) - P\hat{f}(X_k)] \\ &= \sum_{k=1}^n [\hat{f}(X_k) - P\hat{f}(X_{k-1})] + \sum_{k=1}^n [P\hat{f}(X_{k-1}) - P\hat{f}(X_k)] \\ &= \sum_{k=1}^n [\hat{f}(X_k) - P\hat{f}(X_{k-1})] + P\hat{f}(X_0) - P\hat{f}(X_n) \\ &:= M_n + P\hat{f}(X_0) - P\hat{f}(X_n) \end{aligned}$$

Ciąg M_n jest martyngałem. Dzięki tej prostej obserwacji można stosować teorię martyngałów do ciągu Z_n . Jednym z rezultatów wynikających z tego podejścia jest następujące twierdzenie.

Twierdzenie 2.27 (Meyn & Tweedie 17.4.4 [40]). *Niech $(X_n)_{n \geq 0}$ będzie nieokresowym łańcuchem Harrisa z rozkładem stacjonarnym π oraz jądrem przejścia $P(x, A)$. Niech f będzie funkcją z \mathcal{X} o wartościach rzeczywistych. Jeżeli istnieje rozwiązanie równania Poissona \hat{f} spełniające $\pi(\hat{f}^2) < \infty$ oraz jeśli $\sigma_{\text{as}}^2(f) := \pi[\hat{f}^2 - (P\hat{f})^2] > 0$, to przy $n \rightarrow \infty$ zachodzi*

$$(n\sigma_{\text{as}}^2(f))^{-\frac{1}{2}} s_n(t) \xrightarrow{d} W(t),$$

gdzie $s_n(t) = Z_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)[Z_{\lfloor nt \rfloor + 1} - Z_{\lfloor nt \rfloor}]$ oraz $W(t)$ jest standardowym procesem Wienera na $[0, 1]$.

W szczególnym przypadku, dla $t = 1$ Twierdzenie 2.27 mówi, że zachodzi CTG dla funkcji f z asymptotyczną wariancją $\sigma_{\text{as}}^2(f)$. Podobny wynik dla odwracalnych procesów Markowa można znaleźć w [25].

Rozdział 3

Regeneracyjne oszacowania błędu średniokwadratowego

W tym rozdziale przedstawimy oszacowania błędu średniokwadratowego estymatora $\hat{\theta}_n = \frac{1}{n} \sum_{i=0}^{n-1} f(X_i)$ uzyskanego za pomocą metod MCMC, czyli $MSE := \mathbb{E}(\hat{\theta}_n - \theta)^2$. Oszacowanie to przedstawimy w terminach asymptotycznej wariancji, której oszacowanie w praktyce jest skomplikowane. Dodatkowo przedstawimy oszacowania w terminach warunku dryfu oraz w przypadku jednostajnie ergodycznych łańcuchów Markowa.

3.1. Regeneracja

Założmy, że istnieje mały zbiór C (definicja 2.18), czyli

$$(3.1) \quad P(x, A) \geq \beta \mathbb{I}(x \in C) \nu(A), \quad \forall x \in \mathcal{X}, \forall A \in \mathcal{B}(\mathcal{X}).$$

Wówczas korzystając z techniki rozszczepiania łańcucha (regeneracji) Nummelina [42] oraz Athreya & Ney [4] skonstruujemy łańcuch Markowa (X_n, Γ_n) na przestrzeni $\mathcal{X} \times \{0, 1\}$. Zmienna losowa Γ_{n-1} zależy od X_{n-1} poprzez $\mathbb{P}(\Gamma_{n-1} = 1 | X_{n-1} = x) = \beta \mathbb{I}(x \in C)$. Natomiast reguła przejścia z (X_{n-1}, Γ_{n-1}) do X_n jest zadana przez

$$\begin{aligned} \mathbb{P}(X_n \in A | \Gamma_{n-1} = 1, X_{n-1} = x) &= \nu(A), \\ \mathbb{P}(X_n \in A | \Gamma_{n-1} = 0, X_{n-1} = x) &= Q(x, A), \end{aligned}$$

gdzie Q jest jądrem przejścia zdefiniowanym przez

$$Q(x, \cdot) := \frac{P(x, \cdot) - \beta \mathbb{I}(x \in C) \nu(\cdot)}{1 - \beta \mathbb{I}(x \in C)}.$$

Zawsze kiedy $\Gamma_{n-1} = 1$, to w chwili n następuje regeneracja łańcucha. Kolejne momenty regeneracji definiujemy następująco

$$\begin{aligned} T &:= T_1 := \min\{n \geq 1 : \Gamma_{n-1} = 1\}, \\ T_k &:= \min\{n \geq T_{k-1} : \Gamma_{n-1} = 1\}. \end{aligned}$$

Oznaczmy przez $\omega_k := T_k - T_{k-1}$ dla $k = 2, 3, \dots$ oraz $\omega_1 := T$. Losowe bloki

$$\begin{aligned} \Xi &:= \Xi_1 := (X_0, \dots, X_{T-1}, T) \\ \Xi_k &:= (X_{T_{k-1}}, \dots, X_{T_k-1}, \omega_k) \end{aligned}$$

są niezależne. Zauważmy, że dla $k = 2, 3, \dots$, każdy blok Ξ_k ma ten sam rozkład co blok Ξ przy rozkładzie początkowym ν , niezależnie od rozkładu początkowego $X_0 \sim \xi$. Jednak w ogólności rozkład pierwszego bloku Ξ jest inny od pozostałych i zależy od rozkładu X_0 . Oznaczmy sumy po blokach następująco:

$$\begin{aligned}\Xi(f) &:= \sum_{i=0}^{T-1} f(X_i), \\ \Xi_k(f) &:= \sum_{i=T_{k-1}}^{T_k-1} f(X_i).\end{aligned}$$

Uwaga 3.2. Ewolucja łańcucha "rozszczonego" $(X_n, \Gamma_n)_{n \geq 0}$ jest całkowicie zdeterminowana przez podanie jądra przejścia $P(x, A)$ oraz rozkładu początkowego ξ zmiennej X_0 , ponieważ $\mathbb{P}(\Gamma_0 = 1 | X_0) = \beta \mathbb{I}(X_0 \in C)$ wyznacza rozkład łączny (X_0, Γ_0) . Ponadto brzegowy rozkład łańcucha $(X_n)_{n \geq 0}$ jest zgodny z jądrem przejścia $P(x, A)$. W tym sensie mamy prawo używać symboli \mathbb{P}_ξ oraz \mathbb{E}_ξ w odniesieniu zarówno do wyjściowego łańcucha jak i do łańcucha rozszczonego.

3.2. Oszacowania MSE

Niech ξ będzie rozkładem początkowym łańcucha Markowa $(X_n)_{n \geq 0}$. Oznaczmy przez ξ_n rozkład w chwili n , tzn $\xi_n = \xi P^n$. Dodatkowo oznaczmy przez \bar{f} scentrowaną wersję funkcji f , tzn. $\bar{f} = f - \pi f$.

Twierdzenie 3.3. *Jeżeli zachodzi (3.1), to*

$$\sqrt{\mathbb{E}_\xi (\hat{\theta}_n - \theta)^2} \leq \frac{\sigma_{\text{as}}(f)}{\sqrt{n}} \left(1 + \frac{C_0}{n}\right) + \frac{C_1(f)}{n} + \frac{C_2(f)}{n},$$

gdzie

$$\begin{aligned}\sigma_{\text{as}}^2(f) &:= \frac{\mathbb{E}_\nu \Xi(\bar{f})^2}{\mathbb{E}_\nu T}, & C_0 &:= \mathbb{E}_\pi T = \frac{\mathbb{E}_\nu T^2}{2\mathbb{E}_\nu T} + \frac{1}{2}, \\ C_1(f) &:= \sqrt{\mathbb{E}_\xi \Xi(\bar{f})^2}, \\ C_2(f) &:= \sqrt{\mathbb{E}_{\xi P^n} \Xi(\bar{f})^2}.\end{aligned}$$

Oszacowania w Twierdzeniu 3.3 mają sens tylko wtedy, gdy $C_0 < \infty$, $\sigma_{\text{as}}^2(f) < \infty$ oraz $C_1(f), C_2(f) < \infty$. Przy założeniu (3.1) wiemy, że $\mathbb{E}_\nu T < \infty$ (patrz dowód twierdzenia 3.3), ale to nie jest wystarczające do uzyskania $\mathbb{E}_\nu T^2 < \infty$. Z drugiej strony wiemy, że $\mathbb{E}_\nu (\Xi(\bar{f}))^2 < \infty$ jest warunkiem koniecznym i dostatecznym, aby zachodziło CTG dla łańcucha Markowa X_n oraz funkcji f (twierdzenie 2.21 zastosowane do atomu $C \times \{1\}$). Z naszego punktu widzenia ważne jest, że $\sigma_{\text{as}}^2(f)$ w Twierdzeniu 3.3 jest faktycznie asymptotyczną wariancją występującą w CTG, tzn.

$$\sqrt{n} (\hat{\theta}_n - \theta) \xrightarrow{d} N(0, \sigma_{\text{as}}^2(f)).$$

Ponadto

$$\lim_{n \rightarrow \infty} n \mathbb{E}_\xi (\hat{\theta}_n - \theta)^2 = \sigma_{\text{as}}^2(f).$$

W tym sensie wiodący składnik $\sigma_{\text{as}}(f)/\sqrt{n}$ z Twierdzenia 3.3 jest “asymptotycznie optymalny” i nie może być poprawiony.

W dalszej części pracy podamy oszacowania $C_0, C_1(f), C_2(f)$ oraz $\sigma_{\text{as}}^2(f)$ za pomocą wielkości łatwych do obliczenia przy sprawdzalnych założeniach. W dowodzie poniżej i również w dalszej części pracy będziemy stosować konwencję, że $\sum_l^u = 0$ kiedy $l > u$.

Dowód Twierdzenia 3.3. Zdefiniujemy

$$R(n) := \min\{r \geq 1 : T_r > n\}, \quad \Delta(n) := T_{R(n)} - n.$$

Innymi słowy, $R(n)$ jest pierwszym momentem regeneracji po przekroczeniu n oraz $\Delta(n)$ jest odległością $R(n)$ od n . Przedstawmy błąd estymatora θ następująco

$$\begin{aligned} \hat{\theta}_n - \theta &= \frac{1}{n} \sum_{i=0}^{n-1} \bar{f}(X_i) \\ &= \frac{1}{n} \left(\sum_{i=0}^{T_1-1} \bar{f}(X_i) + \sum_{i=T_1}^{T_{R(n)}-1} \bar{f}(X_i) - \sum_{i=n}^{T_{R(n)}-1} \bar{f}(X_i) \right) \\ &=: \frac{1}{n} (\mathcal{O}_1 + \mathcal{Z} - \mathcal{O}_2). \end{aligned}$$

Z nierówności trójkąta otrzymujemy

$$\sqrt{\mathbb{E}_\xi (\hat{\theta}_n - \theta)^2} \leq \frac{1}{n} \left(\sqrt{\mathbb{E}_\xi \mathcal{O}_1^2} + \sqrt{\mathbb{E}_\xi \mathcal{Z}^2} + \sqrt{\mathbb{E}_\xi \mathcal{O}_2^2} \right).$$

Zauważmy, że

$$\mathbb{E}_\xi \mathcal{O}_1^2 = \mathbb{E}_\xi \Xi(\bar{f})^2, \quad \mathbb{E}_\xi \mathcal{O}_2^2 = \mathbb{E}_{\xi P^n} \Xi(\bar{f})^2.$$

Zatem został nam do oszacowania środkowy składnik $\mathbb{E}_\xi \mathcal{Z}^2$, który jest główną częścią błędu estymatora. Kluczowym fragmentem dowodu jest następująca nierówność:

$$(3.4) \quad \mathbb{E}_\nu \left(\sum_{i=0}^{T_{R(n)}-1} \bar{f}(X_i) \right)^2 \leq \sigma_{\text{as}}^2(f) (n + 2C_0).$$

Założmy, że już ją udowodniliśmy. Wówczas możemy łatwo zauważyć, że

$$\begin{aligned} \mathbb{E}_\xi \mathcal{Z}^2 &\leq \sum_{j=1}^n \mathbb{E}_\xi \left(\mathcal{Z}^2 | T_1 = j \right) \mathbb{P}_\xi(T_1 = j) \\ &= \sum_{j=1}^n \mathbb{E}_\nu \left(\sum_{i=0}^{T_{R(n-j)}-1} \bar{f}(X_i) \right)^2 \mathbb{P}_\xi(T_1 = j) \\ &\leq \sum_{j=1}^n \sigma_{\text{as}}^2(f) (n - j + 2C_0) \mathbb{P}_\xi(T_1 = j) \\ &\leq \sigma_{\text{as}}^2(f) (n + 2C_0), \end{aligned}$$

czyli $\sqrt{\mathbb{E}_\xi \mathcal{Z}^2} \leq \sqrt{n} \sigma_{\text{as}}(f) (1 + C_0/n)$ i otrzymujemy tezę.

Pozostała nam do udowodnienia nierówność (3.4). Zapiszmy sumę występującą pod wartością oczekiwaną następująco:

$$(3.5) \quad \sum_{i=0}^{T_{R(n)}-1} \bar{f}(X_i) = \sum_{k=1}^{R(n)} \Xi_k(\bar{f}).$$

Jak wcześniej zauważyliśmy, gdy ν jest rozkładem początkowym łańcucha Markowa, to bloki Ξ_k są niezależnymi zmiennymi losowymi o tym samym rozkładzie. A tak jest w naszej sytuacji. Zauważmy, że (3.4) jest nierównością dla losowej sumy zmiennych losowych i.i.d.

Rozkład T przy rozkładzie początkowym $X_0 \sim \nu$ jest identyczny z rozkładem $\tau(C \times \{1\})$ przy założeniu, że punktem startowym łańcucha (X_n, Γ_n) jest $(X_0, \Gamma_0) = (x, 1)$, gdzie $x \in C$. Zatem z twierdzenia 2.19 (w przypadku atomowym) mamy $1/\mathbb{E}_\nu T = 1/\mathbb{E}_{(x,1)} \tau(C \times \{1\}) = \beta\pi(C)$, gdzie $x \in C$. Analogicznie otrzymujemy

$$\mathbb{E}_\nu \Xi(f) = \pi(f) \mathbb{E}_\nu T,$$

czyli $\mathbb{E}_\nu \Xi(\bar{f}) = 0$ i $\text{Var}_\nu \Xi(\bar{f}) = \sigma_{\text{as}}^2(f) \mathbb{E}_\nu T$. Wykorzystamy teraz fakt, że $R(n)$ jest momentem zatrzymania względem filtracji $\mathcal{G}_k = \sigma((\Xi_1(\bar{f}), \omega_1), \dots, (\Xi_k(\bar{f}), \omega_k))$.

Zastosujemy tożsamości Walda. Z drugiej tożsamości Walda otrzymujemy

$$\mathbb{E}_\nu \left(\sum_{k=1}^{R(n)} \Xi_k(\bar{f}) \right)^2 = \text{Var}_\nu \Xi(\bar{f}) \mathbb{E}_\nu R(n) = \sigma_{\text{as}}^2(f) \mathbb{E}_\nu T \mathbb{E}_\nu R(n).$$

Ale na mocy pierwszej tożsamości Walda możemy zastąpić $\mathbb{E}_\nu T \mathbb{E}_\nu R(n)$ przez $\mathbb{E}_\nu T_{R(n)}$

$$\mathbb{E}_\nu T_{R(n)} = \mathbb{E}_\nu \sum_{k=1}^{R(n)} \omega_k = \mathbb{E}_\nu T \mathbb{E}_\nu R(n).$$

Zatem

$$(3.6) \quad \mathbb{E}_\nu \left(\sum_{k=1}^{R(n)} \Xi_k(\bar{f}) \right)^2 = \sigma_{\text{as}}^2(f) \mathbb{E}_\nu T_{R(n)} = \sigma_{\text{as}}^2(f) (n + \mathbb{E}_\nu \Delta(n)),$$

gdzie $\Delta(n) := T_{R(n)} - n$. Do oszacowania pozostała nam już tylko wartość oczekiwana $\mathbb{E}_\nu \Delta(n)$. Przy rozkładzie początkowym ν , proces $T = T_1 < T_2 < \dots < T_k < \dots$ jest klasycznym procesem odnowienia z czasem dyskretnym. Z teorii procesów odnowienia wiadomo [13, Rozdział XV. par 6(c)], że przy dodatkowym założeniu o nieokresowości, $\Delta(n)$ zbiega według rozkładu, przy $n \rightarrow \infty$, do następującego rozkładu granicznego

$$\mathbb{P}_\pi(T = i) = \frac{\mathbb{P}_\nu(T \geq i)}{\mathbb{E}_\nu T} \quad \text{for } i = 1, 2, \dots$$

Przypomnijmy nierówność Lordena [36], której nowy prostszy dowód można znaleźć w [10]. Zauważmy, że ta nierówność nie wymaga nieokresowości.

$$(3.7) \quad \mathbb{E}_\nu \Delta(n) \leq 2 \mathbb{E}_\pi T.$$

Nierówność (3.7) jest ostatnim elementem dowodu (3.4). Za pomocą elementarnych obliczeń otrzymujemy $2\mathbb{E}_\pi T = \mathbb{E}_\nu T^2 / \mathbb{E}_\nu T + 1$.

$$\begin{aligned} \mathbb{E}_\pi T &= \frac{1}{\mathbb{E}_\nu T} \sum_{i=1}^{\infty} i \mathbb{P}_\nu(T \geq i) = \frac{1}{\mathbb{E}_\nu T} \sum_{i=1}^{\infty} i \sum_{j=i}^{\infty} \mathbb{P}_\nu(T = j) \\ &= \frac{1}{\mathbb{E}_\nu T} \sum_{j=1}^{\infty} \mathbb{P}_\nu(T = j) \sum_{i=1}^j i = \frac{1}{\mathbb{E}_\nu T} \sum_{j=1}^{\infty} \mathbb{P}_\nu(T = j) \frac{(j+1)(j)}{2} \\ &= \frac{1}{\mathbb{E}_\nu T} \mathbb{E}_\nu \frac{(T+1)T}{2} = \frac{\mathbb{E}_\nu T^2}{2\mathbb{E}_\nu T} + \frac{1}{2}. \end{aligned}$$

Podstawiamy (3.7) do (3.6) i z (3.5) otrzymujemy (3.4), co kończy dowód. \square

3.3. Oszacowania $\sigma_{\text{as}}(f)$ za pomocą warunku dryfu

W poprzedniej części przedstawiliśmy oszacowania MSE metodą regeneracji. Jednak Twierdzenie 3.3 nie podaje warunków zapewniających, aby stałe w nim występujące były skończone, jak również nie przedstawia sposobu, w jaki można w praktyce je obliczyć lub oszacować. W tej części pracy przedstawimy oszacowania stałych występujących w Twierdzeniu 3.3 za pomocą warunku dryfu.

Najpierw sformułujemy główne wyniki w postaci twierdzenia 3.11 oraz stwierdzenia 3.12, a dowody przedstawimy później. Załóżmy, że istnieje mały zbiór C , czyli spełniający (3.1), dodatkowo załóżmy, że funkcja $V \geq 1$ spełnia następujący warunek dryfu do zbioru C . Istnieją $0 \leq \lambda < 1$ oraz $0 \leq K < \infty$ takie, że

$$(3.8) \quad PV^2(x) \leq \begin{cases} \lambda^2 V^2(x) & \text{dla } x \notin C, \\ K^2 & \text{dla } x \in C. \end{cases}$$

Zastosowaliśmy nietypową notację, aby uprościć zapis sformułowań w dalszej części pracy.

Mając warunek dryfu (3.8) dla funkcji V^2 otrzymujemy, korzystając z nierówności Jense-
na, analogiczny warunek dla funkcji V :

$$(3.9) \quad PV(x) \leq \begin{cases} \lambda V(x) & \text{dla } x \notin C, \\ K & \text{dla } x \in C. \end{cases}$$

Uwaga 3.10. Dla geometrycznie ergodycznych łańcuchów Markowa wiadomo z twierdzenia 2.24, że spełniony jest warunek dryfu do pewnego a_n -małego zbioru (definicja 2.16). Jednak ten zbiór nie musi być mały (definicja 2.18). Jednak w praktyce, jeśli nie potrafimy skonstruować małego zbioru C ze znaną stałą β , to tym bardziej nie będziemy umieli skonstruować ani a_n -małego zbioru ani n -małego zbioru ze znaną stałą β .

Twierdzenie 3.11. *Przy założeniu (3.1) i (3.8) dla każdej funkcji f takiej, że $0 < |\bar{f}|_V < \infty$, zachodzą następujące nierówności:*

$$(i) \quad C_0 \leq \frac{\lambda}{1-\lambda} \pi(V) + \frac{K - \lambda(1+\beta)}{\beta(1-\lambda)},$$

$$(ii) \quad \frac{\sigma_{\text{as}}^2(f)}{|\bar{f}|_V^2} \leq \frac{1+\lambda}{1-\lambda} \pi(V^2) + \frac{2(K-\lambda-\beta)}{\beta(1-\lambda)} \pi(V),$$

$$(iii) \quad \frac{C_1(f)^2}{|\bar{f}|_V^2} \leq \frac{1}{(1-\lambda)^2} \xi(V^2) + \frac{2(K-\lambda-\beta)}{\beta(1-\lambda)^2} \xi(V) + \frac{\beta(3+\lambda)(K^2 - \lambda^2 - \beta) + 2(1+\lambda)(K-\lambda-\beta)^2}{\beta^2(1-\lambda)^2(1+\lambda)},$$

$$(iv) \quad \frac{C_2(f)^2}{|\bar{f}|_V^2} \leq \frac{1}{(1-\lambda)^2} \xi P^n(V^2) + \frac{2(K-\lambda-\beta)}{\beta(1-\lambda)^2} \xi P^n(V) + \frac{\beta(3+\lambda)(K^2 - \lambda^2 - \beta) + 2(1+\lambda)(K-\lambda-\beta)^2}{\beta^2(1-\lambda)^2(1+\lambda)},$$

gdzie $C_0, \sigma_{\text{as}}^2(f), C_1(f), C_2(f)$ były wprowadzone w twierdzeniu 3.3.

W Twierdzeniu 3.11 występują wyrażenia trudne do obliczenia: zależne od rozkładu stacjonarnego π ($\pi(V), \pi(V^2), |\bar{f}|_V^2$) lub od rozkładu łańcucha w chwili n ($\xi P^n(V), \xi P^n(V^2)$). W następnym stwierdzeniu podajemy proste oszacowania tych wyrażeń.

Stwierdzenie 3.12. *Przy założeniach jak w Twierdzeniu 3.11 mamy*

- (i) $\pi(V) \leq \pi(C) \frac{K - \lambda}{1 - \lambda} \leq \frac{K - \lambda}{1 - \lambda},$
- (ii) $\pi(V^2) \leq \pi(C) \frac{K^2 - \lambda^2}{1 - \lambda^2} \leq \frac{K^2 - \lambda^2}{1 - \lambda^2},$
- (iii) *Jeśli $\xi(V) \leq \frac{K}{1 - \lambda}$ to $\xi P^n(V) \leq \frac{K}{1 - \lambda},$*
- (iv) *Jeśli $\xi(V^2) \leq \frac{K^2}{1 - \lambda^2}$ to $\xi P^n(V^2) \leq \frac{K^2}{1 - \lambda^2},$*
- (v) $|\bar{f}|_V$ *możemy zastąpić przez $|f|_V$ za pomocą*
 $|\bar{f}|_V \leq |f|_V \left(1 + \frac{\pi(C)(K - \lambda)}{(1 - \lambda) \inf_{x \in \mathcal{X}} V(x)} \right) \leq |f|_V \left(1 + \frac{K - \lambda}{1 - \lambda} \right).$

Uwaga 3.13. W praktyce w algorytmach MCMC najczęściej punkt początkowy jest wybierany deterministycznie, czyli rozkład początkowy ξ jest równy δ_x dla pewnego $x \in \mathcal{X}$. W tej sytuacji powinniśmy wybrać x tak, aby $V(x)^2 \leq K^2/(1 - \lambda^2)$, wówczas otrzymamy optymalne oszacowania (iii) oraz (iv). Ponadto zauważmy, że nasze oszacowania nie polepszą się, jeśli dodamy do symulacji, jak to zazwyczaj się robi, *burn-in* t i użyjemy estymatora

$$\hat{\theta}_{t,n} := \frac{1}{n} \sum_{i=t}^{n+t-1} f(X_i).$$

Taka konstrukcja ma intuicyjne wyjaśnienie. Im rozkład ξP^t jest bliższy rozkładowi stacjonarnemu π , tym lepiej. Jednak nasze oszacowania zależą tylko od wartości funkcji V w X_0 , co powoduje, że przybliżanie rozkładu początkowego do π ich nie poprawi.

Uwaga 3.14. W szczególnych przypadkach możemy uzyskać ograniczenia lepsze niż w Stwierdzeniu 3.12, szacując $\pi(C)$ dokładniej niż $\pi(C) \leq 1$. Jednak w ogólności zakładamy, że takie oszacowania nie są dostępne.

Zanim przejdziemy do dowodu Twierdzenia 3.11 i Stwierdzenia 3.12 przedstawimy fakty, z których skorzystamy. Następny lemat jest szczególnym przypadkiem Stwierdzenia 4.1 oraz 4.4 z [5], podobny wynik można znaleźć w [37].

Lemat 3.15. *Przy założeniach (3.1) i (3.9) mamy*

$$\mathbb{E}_x \sum_{n=1}^{T-1} V(X_n) \leq \frac{\lambda}{1 - \lambda} V(x) + \frac{K - \lambda - \beta}{\beta(1 - \lambda)}.$$

Podobnie jak poprzednio stosujemy konwencję, że $\sum_k^l = 0$ jeśli $k > l$. Zauważmy, że oszacowanie $\mathbb{E}_x \sum_{n=0}^{T-1} V(X_n)$ zawiera dodatkowy składnik $V(x)$.

Dowód. Ideą dowodu jest rozbiecie sumy na krótsze bloki, dzieląc czas T na kolejne wizyty w zbiorze C . Oznaczmy przez $S := S_0 := \min\{n \geq 0 : X_n \in C\}$ oraz $S_j := \min\{n > S_{j-1} :$

$X_n \in C$ } Dla $j = 1, 2, \dots$. Dodatkowo wprowadźmy oznaczenia:

$$H(x) := \mathbb{E}_x \sum_{n=0}^S V(X_n), \text{ for } x \in \mathcal{X},$$

$$\tilde{H} := \sup_{x \in C} \mathbb{E}_x \left(\sum_{n=1}^{S_1} V(X_n) \mid \Gamma_0 = 0 \right) = \sup_{x \in C} \int Q(x, dy) H(y).$$

Zauważmy, że $H(x) = V(x)$ dla $x \in C$ oraz, że $Q(x, dy)$ jest jądrem przejścia wykorzystywanym w technice regeneracji.

Oszacujemy najpierw $H(x)$. Jak łatwo sprawdzić przy założeniu (3.9) dla każdego rozkładu początkowego, $V(X_{n \wedge S})/\lambda^{n \wedge S}$ dla $n = 0, 1, \dots$ jest nadmartyngałem względem filtracji $\mathcal{F}_n := \sigma(X_0, \dots, X_n)$. Istotnie, ponieważ $\{S > n\} \subseteq \{X_n \notin C\}$ to z warunku dryfu dostajemy:

$$\begin{aligned} \mathbb{E} \left(\frac{V(X_{(n+1) \wedge S})}{\lambda^{(n+1) \wedge S}} \mid \mathcal{F}_n \right) &= \mathbb{E} \left(\frac{V(X_{(n+1) \wedge S})}{\lambda^{(n+1) \wedge S}} \mathbb{I}(S \leq n) \mid \mathcal{F}_n \right) + \mathbb{E} \left(\frac{V(X_{(n+1) \wedge S})}{\lambda^{(n+1) \wedge S}} \mathbb{I}(S > n) \mid \mathcal{F}_n \right) \\ &= \mathbb{E} \left(\frac{V(X_S)}{\lambda^S} \mathbb{I}(S \leq n) \mid \mathcal{F}_n \right) + \mathbb{E} \left(\frac{V(X_{n+1})}{\lambda^{n+1}} \mathbb{I}(S > n) \mid \mathcal{F}_n \right) \\ &= \frac{V(X_S)}{\lambda^S} \mathbb{I}(S \leq n) + \mathbb{E} \left(\frac{V(X_{n+1})}{\lambda^{n+1}} \mid \mathcal{F}_n \right) \mathbb{I}(S > n) \\ &\leq \frac{V(X_S)}{\lambda^S} \mathbb{I}(S \leq n) + \frac{\lambda V(X_n)}{\lambda^{n+1}} \mathbb{I}(S > n) \\ &= \frac{V(X_{n \wedge S})}{\lambda^{n \wedge S}}. \end{aligned}$$

Zatem $\mathbb{E}_x V(X_{n \wedge S})/\lambda^{n \wedge S} \leq V(x)$ dla każdego $x \in \mathcal{X}$ oraz $n = 0, 1, \dots$. Mnożymy tę nierówność stronami przez λ^n i otrzymujemy:

$$\mathbb{E}_x V(X_S) \lambda^{n-S} \mathbb{I}(S < n) + \mathbb{E}_x V(X_n) \mathbb{I}(n \leq S) \leq \lambda^n V(x).$$

Sumujemy po $n = 0, 1, \dots$ i mamy

$$\mathbb{E}_x V(X_S) \sum_{n=S+1}^{\infty} \lambda^{n-S} + \mathbb{E}_x \sum_{n=0}^S V(X_n) \leq V(x) \sum_{n=0}^{\infty} \lambda^n$$

lub, równoważnie

$$(3.16) \quad \mathbb{E}_x V(X_S) \frac{\lambda}{1-\lambda} + H(x) \leq V(x) \frac{1}{1-\lambda}.$$

Stąd, ponieważ $\mathbb{E}_x V(X_S) \geq 1$, dla każdego x mamy,

$$(3.17) \quad H(x) \leq \frac{V(x) - \lambda}{1-\lambda}.$$

Z (3.9) wiemy, że $PV(x) = (1-\beta)QV(x) + \beta\nu V \leq K$ dla $x \in C$. Czyli $QV(x) \leq (K-\beta)/(1-\beta)$ zatem z (3.17) otrzymujemy

$$(3.18) \quad \tilde{H} \leq \frac{(K-\beta)/(1-\beta) - \lambda}{1-\lambda} = \frac{K - \lambda - \beta(1-\lambda)}{(1-\lambda)(1-\beta)}.$$

Przypomnijmy, że $T := \min\{n \geq 1 : \Gamma_{n-1} = 1\}$. Stąd, dla $x \in C$, korzystając z (3.18) otrzymujemy

$$\begin{aligned}
& \mathbb{E}_x \sum_{n=1}^{T-1} V(X_n) \\
&= \mathbb{E}_x \sum_{j=1}^{\infty} \sum_{n=S_{j-1}+1}^{S_j} V(X_n) \mathbb{I}(\Gamma_{S_0} = \dots = \Gamma_{S_{j-1}} = 0) \\
&= \sum_{j=1}^{\infty} \mathbb{E} \left(\sum_{n=S_{j-1}+1}^{S_j} V(X_n) \middle| \Gamma_{S_0} = \dots = \Gamma_{S_{j-1}} = 0 \right) (1 - \beta)^j \\
&\leq \sum_{j=1}^{\infty} \tilde{H} (1 - \beta)^j \leq \frac{K - \lambda}{\beta(1 - \lambda)} - 1.
\end{aligned}$$

Dla $x \notin C$ musimy dodać dodatkowy składnik: sumę do momentu pierwszego wejścia do zbioru C :

$$\begin{aligned}
\mathbb{E}_x \sum_{n=1}^{T-1} V(X_n) &= \mathbb{E}_x \sum_{n=1}^{S_0} V(X_n) \\
&+ \mathbb{E}_x \sum_{j=1}^{\infty} \sum_{n=S_{j-1}+1}^{S_j} V(X_n) \mathbb{I}(\Gamma_{S_0} = \dots = \Gamma_{S_{j-1}} = 0).
\end{aligned}$$

Ten dodatkowy składnik jest równy $H(x) - V(x)$ i do jego oszacowania użyjemy (3.17). Ostatecznie mamy

$$(3.19) \quad \mathbb{E}_x \sum_{n=1}^{T-1} V(X_n) \leq \frac{\lambda(V(x) - 1)}{1 - \lambda} \mathbb{I}(x \notin C) + \frac{K - \lambda}{\beta(1 - \lambda)} - 1.$$

□

Z dowodu powyższego lematu wynika prosty wniosek pokazujący zależność pomiędzy miarą π małego zbioru a stałymi występującymi w warunku dryfu.

Stwierdzenie 3.20. *Jeżeli V spełnia warunek dryfu do małego zbioru C to*

$$\pi(C) \geq \frac{1 - \lambda}{K - \lambda}.$$

Dowód. Oznaczmy przez $G(x) := \mathbb{E}_x \left[\sum_{i=1}^{\tau(C)} V(X_i) \right]$. Oczywiście $\mathbb{E}_x \tau(C) \leq G(x)$, ponadto dla $x \in C$ mamy

$$(3.21) \quad G(x) = \int_{\mathcal{X}} P(x, dy) H(y) = PH(x),$$

gdzie $H(x)$ jest takie jak w dowodzie lematu 3.15. Korzystając z nierówności (3.17) oraz z tego, że $PV(x) \leq K$ dla $x \in C$ otrzymujemy $G(x) \leq \frac{K - \lambda}{1 - \lambda}$. Z twierdzenia 2.19 mamy

$$(3.22) \quad 1 = \pi(\mathcal{X}) = \int_C \mathbb{E}_x \tau(C) \pi(dx).$$

Stąd wynika

$$1 \leq \int_C G(x) \pi(dx) \leq \frac{K - \lambda}{1 - \lambda} \pi(C),$$

co kończy dowód. □

Uwaga 3.23. Optymalny ze względu na jakość oszacowań jest warunek dryfu, gdy λ oraz K są małe. Wtedy jednak to stwierdzenie pokazuje, że miara π małego zbioru C nie może być zbyt mała.

Następny fakt jest dobrze znany, np. w [43] jest sformułowany dla ograniczonej funkcji V . Dowód dla nieujemnej funkcji V jest identyczny z wyjątkiem argumentu pozwalającego na skorzystanie z twierdzenia Fubinięgo.

Lemat 3.24. *Jeśli $V \geq 0$ to*

$$\mathbb{E}_\nu \Xi(V)^2 = \mathbb{E}_\nu T \left(\mathbb{E}_\pi V(X_0)^2 + 2 \sum_{n=1}^{\infty} \mathbb{E}_\pi V(X_0)V(X_n)\mathbb{I}(T > n) \right).$$

Dowód Twierdzenia 3.11. (i) Przypomnijmy, że $C_0 = \mathbb{E}_\pi T$,

$$\mathbb{E}_\pi T \leq 1 + \mathbb{E}_\pi \sum_{n=1}^{T-1} V(X_n)$$

i tezę otrzymujemy z Lematu 3.15. W dowodach punktów (ii)-(iv) możemy bez straty ogólności założyć, że $|\bar{f}|_V = 1$.

(ii) Z Lematu 3.24 otrzymujemy

$$\begin{aligned} \sigma_{\text{as}}^2(f) &= \mathbb{E}_\nu \Xi(\bar{f})^2 / \mathbb{E}_\nu T \leq \mathbb{E}_\nu \Xi(V)^2 / \mathbb{E}_\nu T \\ &= \mathbb{E}_\pi V(X_0)^2 + 2\mathbb{E}_\pi \sum_{n=1}^{T-1} V(X_0)V(X_n) =: \text{I} + \text{II}. \end{aligned}$$

Do oszacowania drugiego składnika użyjemy Lematu 3.15.

$$\begin{aligned} \text{II}/2 &= \mathbb{E}_\pi \sum_{n=1}^{T-1} V(X_0)V(X_n) = \mathbb{E}_\pi V(X_0)\mathbb{E}\left(\sum_{n=1}^{T-1} V(X_n) \mid X_0\right) \\ &\leq \mathbb{E}_\pi V(X_0) \left(\frac{\lambda}{1-\lambda} V(X_0) + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \right) \\ &= \frac{\lambda}{1-\lambda} \pi(V^2) + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \pi(V). \end{aligned}$$

Po uproszczeniu wyrażeń otrzymujemy

$$\sigma_{\text{as}}^2(f) \leq \frac{1+\lambda}{1-\lambda} \pi(V^2) + \frac{2(K-\lambda-\beta)}{\beta(1-\lambda)} \pi(V)$$

co kończy dowód punktu (ii).

(iii) Dowód jest podobny do dowodu punktu (ii), ale nie możemy już skorzystać z Lematu 3.24. Rozpiszmy $\mathbb{E}_x \Xi(V)^2$ w następujący sposób

$$\begin{aligned} \mathbb{E}_x \Xi(V)^2 &= \mathbb{E}_x \left(\sum_{n=0}^{T-1} V(X_n) \right)^2 = \mathbb{E}_x \left(\sum_{n=0}^{\infty} V(X_n)\mathbb{I}(n < T) \right)^2 \\ &= \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)^2 \mathbb{I}(n < T) + 2\mathbb{E}_x \sum_{n=0}^{\infty} \sum_{j=n+1}^{\infty} V(X_n)V(X_j)\mathbb{I}(j < T) \\ &=: \text{I} + \text{II}. \end{aligned}$$

Do oszacowania pierwszego wyrażenia skorzystamy z Lematu 3.15 dla funkcji V^2 , który jest prawdziwy przy założeniu (3.8),

$$I = \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)^2 \mathbb{I}(n < T) \leq \frac{1}{1-\lambda^2} V(x)^2 + \frac{K^2 - \lambda^2 - \beta}{\beta(1-\lambda^2)}.$$

Drugą część oszacujemy w dwóch krokach. Najpierw warunkujemy przez X_n i korzystamy z Lematu 3.15 dla funkcji V , następnie ponownie korzystamy z Lematu 3.15 dla funkcji V^2 oraz V .

$$\begin{aligned} II/2 &= \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n) \mathbb{I}(n < T) \mathbb{E} \left(\sum_{j=n+1}^{\infty} V(X_j) \mathbb{I}(j < T) \middle| X_n \right) \\ &\leq \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n) \mathbb{I}(n < T) \left(\frac{\lambda}{1-\lambda} V(X_n) + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \right) \\ &= \frac{\lambda}{1-\lambda} \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n)^2 \mathbb{I}(n < T) + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \mathbb{E}_x \sum_{n=0}^{\infty} V(X_n) \mathbb{I}(n < T) \\ &= \frac{\lambda}{1-\lambda} \left(\frac{1}{1-\lambda^2} V(x)^2 + \frac{K^2 - \lambda^2 - \beta}{\beta(1-\lambda^2)} \right) \\ &\quad + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \left(\frac{1}{1-\lambda} V(x) + \frac{K-\lambda-\beta}{\beta(1-\lambda)} \right). \end{aligned}$$

Po uproszczeniu wyrażen otrzymujemy tezę

$$\begin{aligned} \mathbb{E}_x \Xi(V)^2 &\leq \frac{1}{(1-\lambda)^2} V^2(x) + \frac{2(K-\lambda-\beta)}{\beta(1-\lambda)^2} V(x) \\ &\quad + \frac{\beta(3+\lambda)(K^2 - \lambda^2 - \beta) + 2(1+\lambda)(K-\lambda-\beta)^2}{\beta^2(1-\lambda)^2(1+\lambda)}. \end{aligned}$$

(iv) Dowód taki sam jak punktu (iii) □

Dowód Stwierdzenia 3.12.

(i) Warunek dryfu (3.9) implikuje, że $\pi V = \pi P V \leq \lambda(\pi V - \pi(C)) + K\pi(C)$ stąd w oczywisty sposób otrzymujemy tezę.

(ii) Dowód jest taki sam, jak dowód punktu (i), tylko zamiast (3.9) korzystamy z (3.8).

(iii) Dowodzimy przez indukcję: $\xi P^{n+1} V = \xi P^n(PV) \leq \xi P^n(\lambda V + K) \leq \lambda K/(1-\lambda) + K = K/(1-\lambda)$.

(iv) Taki sam jak (iii).

(v) Obliczamy:

$$\begin{aligned} |\bar{f}|_V &= \sup_{x \in \mathcal{X}} \frac{|f(x) - \pi f|}{V(x)} \leq \sup_{x \in \mathcal{X}} \frac{|f(x)| + |\pi f|}{V(x)} \\ &\leq \sup_{x \in \mathcal{X}} \left(|f|_V \left(1 + \frac{\pi V}{V(x)} \right) \right) \leq |f|_V \left(1 + \frac{\pi(J)(K-\lambda)}{(1-\lambda) \inf_{x \in \mathcal{X}} V(x)} \right). \end{aligned}$$

□

Uwaga 3.25. Warunek dryfu zdefiniowany wzorem (3.8) w połączeniu z warunkiem małego zbioru (3.1) daje się sprawdzić, wraz z jawną postacią występujących tam stałych (K, λ, β) , w wielu modelach statystyki bayesowskiej [29, 3, 19, 24, 22, 49]. Zatem wyniki zawarte w tym rozdziale można zastosować do tych modeli i otrzymać całkowicie jawne oszacowania MSE.

3.4. Oszacowania $\sigma_{\text{as}}(f)$ dla jednostajnie ergodycznych łańcuchów Markowa

Definicja 3.26. Mówimy, że łańcuch Markowa $(X_n)_{n \geq 0}$ z jądrem przejścia $P(x, dy)$ jest jednostajnie ergodyczny wtedy i tylko wtedy, gdy istnieją stałe $M < \infty$ oraz $\rho < 1$ takie, że dla każdego $x \in \mathcal{X}$ i każdego n zachodzi [40, 45]

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{tv} \leq M\rho^n.$$

Jednostajna ergodyczność jest równoważna warunkowi $P^h(x, \cdot) \geq \beta\nu(\cdot)$ dla pewnego $h \geq 1$ i dla wszystkich $x \in \mathcal{X}$ oraz pewnej miary probabilistycznej ν na $\mathcal{B}(\mathcal{X})$ (zob.[45] lub [40]).

Dla ułatwienia rozważmy tylko przypadek kiedy $h = 1$. Jest to typowa sytuacja w zastosoowaniach. Załóżmy zatem, że spełniony jest następujący *warunek Doeblina*: Istnieją $\beta > 0$ oraz miara probabilistyczna ν na $\mathcal{B}(\mathcal{X})$ takie, że

$$(3.27) \quad P(x, \cdot) \geq \beta\nu(\cdot) \quad \text{dla każdego } x \in \mathcal{X}.$$

Innymi słowy spełnione jest założenie (3.1) z małym zbiorem $C = \mathcal{X}$. Oczywistą konsekwencją (3.27) jest, że długość bloków regeneracji ma rozkład geometryczny. Dla każdego $x \in \mathcal{X}$ mamy

$$\mathbb{P}_x(T = i) = \beta(1 - \beta)^{i-1} \quad \text{for } i = 1, 2, \dots$$

Zatem $\mathbb{E}_x T = 1/\beta$, ponadto T jest niezależne od X_0 .

Przejdźmy teraz do oszacowania wielkości występujących w Twierdzeniu 3.3. Obliczenie C_0 jest oczywiste: $C_0 = \mathbb{E}_\pi T = 1/\beta$.

Problem oszacowania asymptotycznej wariancji jest rozważany w [6]. Przy założeniu (3.27), korzystając z wyników tej pracy (wzór (31) w części 5 i dalsze oszacowania), po zsumowaniu szeregu geometrycznego otrzymujemy:

$$(3.28) \quad \sigma_{\text{as}}^2(f) \leq \sigma^2 \left(1 + \frac{2}{1 - \sqrt{1 - \beta}}\right) = \sigma^2 \left(1 + 2\frac{1 + \sqrt{1 - \beta}}{\beta}\right) \leq 4\sigma^2/\beta,$$

gdzie $\sigma^2 = \pi \bar{f}^2$ jest wariancją stacjonarną.

Jeżeli dodatkowo założymy odwracalność, to możemy uzyskać lepsze oszacowanie. Dla odwracalnych łańcuchów spełniających (3.27) wiadomo, że spektrum operatora $P - \Pi$, oznaczmy je przez \mathcal{S} , jest zawarte w $[-1 + \beta, 1 - \beta]$ [44]. Z twierdzenia o rozkładzie spektralnym operatorów samosprężonych (zob. [16, 25]) mamy

$$(3.29) \quad \sigma_{\text{as}}^2(f) = \int_{\mathcal{S}} \frac{1+s}{1-s} E_{f,P}(ds) \leq \frac{2-\beta}{\beta} \sigma^2,$$

gdzie $E_{f,P}$ jest miarą spektralną związaną z f oraz P . Oszacowania $\sigma_{\text{as}}^2(f)$ w (3.28) i (3.29) zależą od β w sposób optymalny. Ponadto nierówność (3.29) jest optymalna.

Przykład 3.30. Niech $\beta < 1$ zdefiniujemy łańcuch Markowa na przestrzeni stanów $\{0, 1\}$ przez macierz przejścia

$$P = \begin{bmatrix} 1 - \frac{\beta}{2} & \frac{\beta}{2} \\ \frac{\beta}{2} & 1 - \frac{\beta}{2} \end{bmatrix}.$$

Wówczas rozkład stacjonarny π jest równy $\pi := [\frac{1}{2}, \frac{1}{2}]$. Niech $f(x) = x$ wówczas $\sigma^2 = \frac{1}{4}$. Obliczmy wariancję asymptotyczną

$$\begin{aligned} \sigma_{\text{as}}^2(f) &= \sigma^2 + 2 \sum_{i=1}^{\infty} \text{Cov}[f(X_0), f(X_i)] \\ &= \sigma^2 + 2\sigma^2 \sum_{i=1}^{\infty} (1 - \beta)^i = \frac{2 - \beta}{\beta} \sigma^2. \end{aligned}$$

Uwaga 3.31. W rozdziale 5 przedstawimy analogiczne oszacowania $\sigma_{\text{as}}^2(f)$ przy założeniu istnienia luki spektralnej w $L^2(\pi)$ bez zakładania odwracalności.

Oszacowanie $C_1(f)$ oraz $C_2(f)$ przy założeniu (3.27) nie jest takie proste. W przypadku kiedy nie mamy warunku dryfu, nie jest dostępny analog Lematu 3.15. Użyjemy innego podejścia i ograniczymy się do szczególnego przypadku kiedy (3.27) jest spełniony z $\nu = \pi$. Ważną klasą odwracalnych łańcuchów jest Niezależny algorytm Metropolisa-Hastingsa (n.p. [45]). Wiadomo, że ten algorytm prowadzi do jednostajnie ergodycznego łańcucha wtedy i tylko wtedy, gdy prawdopodobieństwo odrzucenia propozycji jest jednostajnie ograniczone przez $1 - \beta$. Ten warunek jest równoważny temu, że rozkład propozycji jest ograniczony z dołu przez $\beta\pi$ ([38, 3]). To oznacza, że spełniony jest (3.27) z $\nu = \pi$, dodatkowo w [3] pokazane jest, że $\mathcal{S} \subseteq [0, 1 - \beta]$. W tym przypadku, jeśli dodatkowo założymy, że $\xi = \nu = \pi$, to łatwo otrzymamy $C_1(f)^2 = C_2(f)^2 = \mathbb{E}_\nu \Xi(\bar{f})^2 = \sigma_{\text{as}}^2(f)/\beta$. Stąd

$$(3.32) \quad \sqrt{\mathbb{E}_\pi(\hat{\theta}_n - \theta)^2} \leq \sigma \left(\sqrt{\frac{2 - \beta}{\beta n}} \left(1 + \frac{1}{\beta n} \right) + \frac{2\sqrt{2 - \beta}}{\beta n} \right).$$

W większej ogólności, jeżeli ξ jest dowolnym rozkładem początkowym, to standardową metodą uzyskujemy proste oszacowanie $\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2 \leq \|d\xi/d\pi\|_\infty \mathbb{E}_\pi(\hat{\theta}_n - \theta)^2$, gdzie $\|\cdot\|_\infty$ jest normą supremum i $d\xi/d\pi$ jest pochodną Radona-Nikodyma. Ostatecznie otrzymujemy oszacowanie MSE, czyli $\mathbb{E}_\xi(\hat{\theta}_n - \theta)^2$ tylko za pomocą σ , β oraz $\|d\xi/d\pi\|_\infty$.

Rozdział 4

Nierówności Hoeffdinga dla łańcuchów Markowa

W tym rozdziale przedstawimy uogólnienie dla łańcuchów Markowa dobrze znanych nierówności Hoeffdinga [21]. Jeżeli mamy ciąg X_n niezależnych zmiennych losowych o tym samym rozkładzie taki, że $0 \leq X_1 \leq 1$ oraz o wartości oczekiwanej $\mathbb{E}X_1 = \theta$, to dla każdego $\varepsilon > 0$ takiego, że $\theta + \varepsilon < 1$, zachodzą następujące nierówności:

$$\begin{aligned} \mathbb{P}(\bar{X} - \theta > \varepsilon) &\leq \left\{ \left(\frac{\theta}{\theta + \varepsilon} \right)^{\theta + \varepsilon} \left(\frac{1 - \theta}{1 - \theta - \varepsilon} \right)^{1 - \theta - \varepsilon} \right\}^n \\ &\leq e^{-2n\varepsilon^2}, \end{aligned}$$

gdzie \bar{X} oznacza średnią arytmetyczną z n pierwszych elementów ciągu X_n .

Tego typu nierówności dla łańcuchów Markowa są dość często rozważane. W przypadku jednostajnie ergodycznym można je znaleźć np. w [18] później poprawione w [26]. Dla łańcuchów na dyskretnej przestrzeni stanów podobne wyniki zostały otrzymane w [33] oraz w [34].

W dalszej części tego rozdziału wykażemy nierówności typu Hoeffdinga dla łańcuchów na ogólnej przestrzeni stanów przy założeniu wiedzy o własnościach operatora przejścia w przestrzeni $L^2(\pi)$. Następnie przy dodatkowych założeniach pokażemy jak z tych nierówności można otrzymać oszacowania podwykładnicze na prawdopodobieństwa wielkich odchyień bez zakładania ograniczoności.

4.1. Nierówności Hoeffdinga

Przypomnijmy, że $L^2(\pi)$ jest przestrzenią funkcji całkowalnych w kwadracie względem miary π z iloczynem skalarnym $\langle f, g \rangle = \int_{\mathcal{X}} f(x)g(x)\pi(dx)$. Ponadto $\|\cdot\|_2$ oznacza normę w przestrzeni $L^2(\pi)$, zaś standardowo zdefiniowaną normę operatorów działających na tej przestrzeni oznaczmy przez $\|\cdot\|_{L^2(\pi)}$. Do końca rozdziału zakładamy, że związany z łańcuchem $(X_n)_{n \geq 0}$ operator przejścia P posiada własność:

$$(4.1) \quad \|P - \Pi\|_{L^2(\pi)} = \rho < 1.$$

Uwaga 4.2. Przy tym założeniu łańcuch $(X_n)_{n \geq 0}$ jest geometrycznie ergodyczny (patrz stw. 5.3). Implikacja w drugą stronę pozostaje prawdziwa przy założeniu odwracalności (twierdzenie 2.25). W ogólności niestety nie jest to prawdą, co utrudnia sprawdzanie warunku (4.1). Dokładniej o problemach związanych z tym założeniem napiszemy w następnym rozdziale.

Niech f będzie funkcją $f : \mathcal{X} \rightarrow [0, 1]$, oznaczmy $\theta := \pi f$ oraz $S_n := \sum_{i=0}^{n-1} f(X_i)$.

Twierdzenie 4.3 (Nierówności Hoeffdinga). *Przy powyższych założeniach i oznaczeniach dla każdego $0 < \varepsilon < 1 - \theta$ zachodzą następujące nierówności:*

$$\begin{aligned} \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) &\leq \left[\frac{\theta + \bar{\theta}\rho}{1 - 2(\bar{\theta} - \varepsilon)/(1 + \sqrt{\Delta})} \right]^{n(\theta + \varepsilon)} \left[\frac{\bar{\theta} + \theta\rho}{1 - 2(\theta + \varepsilon)/(1 + \sqrt{\Delta})} \right]^{n(\bar{\theta} - \varepsilon)} \\ &\leq \exp \left\{ -2 \frac{1 - \rho}{1 + \rho} \varepsilon^2 n \right\}, \end{aligned}$$

gdzie $\bar{\theta} = 1 - \theta$ oraz $\Delta = 1 + \frac{4\rho(\theta + \varepsilon)(\bar{\theta} - \varepsilon)}{\theta(1 - \rho)^2}$.

Uwaga 4.4. W przypadku kiedy łańcuch jest odwracalny, możemy otrzymać analogiczny rezultat jak w twierdzeniu 4.3 z $\rho := \max\{\lambda, 0\}$, gdzie $\lambda := \sup\{z : z \in \sigma(P - \Pi)\}$. Wystarczy minimalna modyfikacja lematu 4.7, aby to zauważyć. W tej formie sformułowany jest wynik w [33].

Dowód tego twierdzenia przeprowadzimy w sposób analogiczny do dowodu zawartego w [33]. W pierwszym kroku skonstruujemy nowy łańcuch Markowa na przestrzeni \mathcal{X} i zredukujemy problem do własności operatora liniowego związanego z funkcją tworzącą momenty nowego łańcucha. Następnie w drugim kroku zredukujemy problem do łańcucha na dwustanowej przestrzeni i skorzystamy z nierówności Hoeffdinga z [33].

Na początku, gdy skorzystamy z nierówności Markowa otrzymamy:

$$(4.5) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq e^{-tn(\theta + \varepsilon)} \mathbb{E}_\pi e^{tS_n}.$$

Dalej będziemy szacować funkcję tworzącą momenty $\mathbb{E}_\pi e^{tS_n}$. W tym celu użyjemy nowego łańcucha Markowa $(X'_n)_{n \geq 0}$ z jądrem przejścia $Q(x, A)$ zdefiniowanym następująco: dla każdego zbioru $A \in \mathcal{B}(\mathcal{X})$ mamy:

$$Q(x, A) = (1 - \rho)\pi(A) + \rho\mathbb{I}(x \in A),$$

gdzie ρ jest dane równaniem 4.1. Przez Q oznaczmy operator liniowy związany z jądrem przejścia $Q(x, A)$, czyli $Q(g)(x) := \int g(y)Q(x, dy)$. Jak łatwo zauważyć $Q = (1 - \rho)\Pi + \rho\mathbb{I}$, gdzie \mathbb{I} jest operatorem identycznościowym. Dla ułatwienia zapisu wprowadzimy jeszcze jedno oznaczenie: jeżeli T jest operatorem liniowym, to dla każdego $t > 0$ przez \widehat{T}_t będziemy oznaczać następujący operator $\widehat{T}_t(g) := e^{\frac{t}{2}f} T(e^{\frac{t}{2}f} g)$. Funkcja f występuje w twierdzeniu 4.3, zaś dla każdego $t > 0$, e^{tf} jest operatorem działającym na funkcje g zgodnie ze wzorem $e^{tf}g(x) := e^{tf(x)}g(x)$. Operator Q jest dodatni (tzn. Q przeprowadza funkcje nieujemne na funkcje nieujemne) oraz samosprężony, czyli operator \widehat{Q}_t jest również dodatni i samosprężony. Z tej prostej obserwacji wynika, że

$$(4.6) \quad \left\| \widehat{Q}_t \right\|_{L^2(\pi)} = \sup_{\{g : \|g\|_2 = 1\}} \left| \langle g, \widehat{Q}_t g \rangle \right| = \sup_{\{g : \|g\|_2 = 1\}} \langle g, \widehat{Q}_t g \rangle.$$

Lemat 4.7. *Przy powyższych oznaczeniach zachodzi nierówność:*

$$\mathbb{E}_\pi \exp(tS_n) \leq \left\| \widehat{Q}_t \right\|_{L^2(\pi)}^n.$$

Dowód. Wiadomo, że $\mathbb{E}_\pi \exp(tS_n)$ można przedstawić za pomocą następującego iloczynu skalarnego $\mathbb{E}_\pi \exp(tS_n) = \langle 1, (e^{tf}P)^{n1} \rangle$ [27, 11]. Istotnie pokażemy, że $\mathbb{E}_x \exp(tS_n) = (e^{tf}P)^{n1}(x)$. Dowód tego faktu jest indukcyjny: warunkując przez X_1 , mamy

$$\mathbb{E}_x \exp(tS_n) = \exp(tf(x)) \mathbb{E}_x \left[\mathbb{E}_{X_1} \exp \left(t \sum_{i=1}^{n-1} f(X_i) \right) \right] = e^{tf}P \left[(e^{tf}P)^{n-1} 1 \right] (x) = (e^{tf}P)^n 1(x).$$

Z faktu, że $P1 = 1$ otrzymujemy

$$\mathbb{E}_\pi \exp(tS_n) = \langle 1, (e^{tf}P)^{n1} \rangle = \langle 1, (e^{tf}P)^{n-1} e^{tf}1 \rangle = \langle e^{\frac{t}{2}f}1, (\widehat{P}_t)^{n-1} (e^{\frac{t}{2}f}1) \rangle.$$

Zastosujemy do lewej strony tej równości nierówność Cauchy'ego - Schwarz

$$(4.8) \quad \mathbb{E}_\pi \exp(tS_n) \leq \|e^{\frac{t}{2}f}1\|_2^2 \|\widehat{P}_t\|_{L^2(\pi)}^{n-1} = \pi(e^{tf}) \|\widehat{P}_t\|_{L^2(\pi)}^{n-1}.$$

Dla dowolnej funkcji g przez g_C oznaczymy jej scentrowaną wersję, czyli $g_C := g - \pi g$. Dodatkowo operator $P - \Pi$ oznaczymy przez P_0 . Przy tych oznaczeniach korzystając z definicji normy operatorowej w przestrzeni $L^2(\pi)$, z założenia 4.1 oraz ze wzorów (4.6), (4.8) otrzymujemy

$$\begin{aligned} \|\widehat{P}_t\|_{L^2(\pi)} &= \sup_{\{g, h : \|g\|_2 = \|h\|_2 = 1\}} \langle h, \widehat{P}_t g \rangle \\ &= \sup_{\{g, h : \|g\|_2 = \|h\|_2 = 1\}} \langle e^{\frac{t}{2}f}h, P e^{\frac{t}{2}f}g \rangle \\ &= \sup_{\{g, h : \|g\|_2 = \|h\|_2 = 1\}} \left\{ \pi(e^{\frac{t}{2}f}h) \pi(e^{\frac{t}{2}f}g) + \langle (e^{\frac{t}{2}f}h)_C, P_0(e^{\frac{t}{2}f}g)_C \rangle \right\} \\ &\leq \sup_{\{g, h : \|g\|_2 = \|h\|_2 = 1\}} \left\{ \pi(e^{\frac{t}{2}f}h) \pi(e^{\frac{t}{2}f}g) + \rho \|(e^{\frac{t}{2}f}h)_C\|_2 \|(e^{\frac{t}{2}f}g)_C\|_2 \right\} \\ &\leq \sup_{\{g, h : \|g\|_2 = \|h\|_2 = 1\}} \left\{ \sqrt{[\pi(e^{\frac{t}{2}f}g)]^2 + \rho \|(e^{\frac{t}{2}f}g)_C\|_2^2} \sqrt{[\pi(e^{\frac{t}{2}f}h)]^2 + \rho \|(e^{\frac{t}{2}f}h)_C\|_2^2} \right\} \\ &\leq \sup_{\{g : \|g\|_2 = 1\}} \left\{ [\pi(e^{\frac{t}{2}f}g)]^2 + \rho \|(e^{\frac{t}{2}f}g)_C\|_2^2 \right\} \\ &= \sup_{\{g : \|g\|_2 = 1\}} \langle e^{\frac{t}{2}f}g, \pi(e^{\frac{t}{2}f}g) + \rho(e^{\frac{t}{2}f}g)_C \rangle \\ &= \|\widehat{Q}_t\|_{L^2(\pi)}. \end{aligned}$$

Ponadto mamy

$$\pi(e^{tf}) \|\widehat{Q}_t\|_{L^2(\pi)} \geq \langle e^{\frac{t}{2}f}1, \widehat{Q}_t e^{\frac{t}{2}f}1 \rangle = \langle e^{tf}1, Q e^{tf}1 \rangle = \pi(e^{tf})^2 + \rho \pi((e^{\frac{t}{2}f})_C^2) \geq \pi(e^{tf})^2.$$

Co kończy dowód. \square

Do tej pory otrzymaliśmy oszacowanie funkcji tworzącej momenty wyjściowego łańcucha poprzez normę operatora związanego z nowym prostszym łańcuchem. Teraz tak jak w [33] będziemy porównywać łańcuch Markowa $(X'_n)_{n \geq 0}$ o jądrze przejścia Q z łańcuchem Markowa na przestrzeni $\{0, 1\}$. Niech $(Y_n)_{n \geq 0}$ będzie łańcuchem Markowa na przestrzeni $\{0, 1\}$ z macierzą przejścia $M_{\theta, \rho}$ zdefiniowaną następująco

$$M_{\theta, \rho} = \rho \mathbb{I} + (1 - \rho) 1\theta^\top.$$

Zatem ρ jest drugą wartością własną macierzy przejścia oraz wektor $\theta = [1 - \theta, \theta]^\top = [\bar{\theta}, \theta]^\top$ jest rozkładem stacjonarnym. Przez \mathbb{E}_θ oznaczmy wartość oczekiwaną liczoną dla łańcucha $(Y_n)_{n \geq 0}$ w stanie stacjonarnym. Niech D_t będzie macierzą diagonalną o przekątnej $(1, e^t)$, oznaczmy przez η_t największą wartość własną macierzy $D_t^2 M_{\theta, \rho}$. Przy tych oznaczeniach możemy zapisać Stwierdzenie 1 z [33] w następujący sposób.

Stwierdzenie 4.9. *Przy oznaczeniach jak w twierdzeniu 4.3 zachodzą następujące nierówności*

$$\begin{aligned} \eta_t^n e^{-nt(\theta + \varepsilon)} &\leq \left[\frac{\theta + \bar{\theta}\rho}{1 - 2(\bar{\theta} - \varepsilon)/(1 + \sqrt{\Delta})} \right]^{n(\theta + \varepsilon)} \left[\frac{\bar{\theta} + \theta\rho}{1 - 2(\theta + \varepsilon)/(1 + \sqrt{\Delta})} \right]^{n(\bar{\theta} - \varepsilon)} \\ &\leq \exp \left\{ -2 \frac{1 - \rho}{1 + \rho} \varepsilon^2 n \right\}. \end{aligned}$$

Następne twierdzenie pozwala porównać funkcje tworzące momenty łańcuchów $(X'_n)_{n \geq 0}$ oraz $(Y_n)_{n \geq 0}$.

Lemat 4.10. *Przy powyższych oznaczeniach dla każdej wypukłej funkcji $G : \mathbb{R} \rightarrow \mathbb{R}$ mamy*

$$\mathbb{E}_\pi [G(f(X'_0) + \dots + f(X'_{n-1}))] \leq \mathbb{E}_\theta [G(Y_0 + \dots + Y_{n-1})].$$

Dowód tego lematu jest identyczny do dowodu Twierdzenia 2 z [33]. Jediną różnicą jest wymiar przestrzeni stanów \mathcal{X} , co nie wpływa na rozumowanie.

Zdefiniujmy funkcję $g(x) := \frac{e^{\frac{t}{2}f(x)}}{r_t - \rho e^{tf(x)}}$. Jeżeli r_t jest rozwiązaniem równania

$$(4.11) \quad 1 = \int_{\mathcal{X}} \frac{(1 - \rho)e^{tf(x)}}{r_t - \rho e^{tf(x)}} d\pi(x)$$

takim, że $r_t > \rho \operatorname{ess\,sup}_{x \in \mathcal{X}}(e^{tf(x)})$, to funkcja g jest dodatnią funkcją własną dla operatora \widehat{Q}_t . Niestety w ogólności równanie (4.11) może nie mieć rozwiązań. Aby uniknąć tego problemu przybliżymy funkcję f funkcjami prostymi. Dla każdego $k \in \mathbb{Z}^+$ definiujemy funkcje f_k następująco

$$f_k(x) = \sum_{i=1}^k \frac{i}{k} 1(x \in A_{i,k}),$$

gdzie $A_{i,k} := \{x \in \mathcal{X} : \frac{i-1}{k} < f(x) \leq \frac{i}{k}\}$. Oznaczmy przez θ_k oraz $\widehat{Q}_{t,k}$ wyrażenia powstałe przez zamianę f przez f_k w definicji θ i \widehat{Q}_t odpowiednio. Korzystając z dodatniości i samosprężoności operatora \widehat{Q}_t oraz z faktu, że $f_k \geq f$ otrzymujemy

$$(4.12) \quad \left\| \widehat{Q}_t \right\|_{L^2(\pi)} = \sup_{\{h \geq 0 : \|h\|_2 = 1\}} \langle h, \widehat{Q}_t h \rangle \leq \sup_{\{h \geq 0 : \|h\|_2 = 1\}} \langle h, \widehat{Q}_{t,k} h \rangle = \left\| \widehat{Q}_{t,k} \right\|_{L^2(\pi)}.$$

Ponadto zauważmy, że na mocy twierdzenia Lebesgue'a o zmajorzowanym przejściu z granicą pod znak całki, zachodzi $\lim_{k \rightarrow \infty} \theta_k = \theta$. Oznaczmy przez $\eta_t(x)$ największą wartość własną macierzy $D_t^2 M_{x, \rho}$. Jest to ciągła funkcja zmiennej x , czyli $\eta_{t,k} := \eta_t(\theta_k)$ zbiega do η_t , gdy k dąży do nieskończoności.

Rozważmy funkcję

$$F_k(r) := \int_{\mathcal{X}} \frac{(1 - \rho)e^{tf_k(x)}}{r - \rho e^{tf_k(x)}} d\pi(x).$$

Dla $r > \rho \operatorname{ess\,sup}_{x \in \mathcal{X}}(e^{tf_k(x)})$ funkcja F_k jest ciągła i malejąca. Jeżeli r dąży do nieskończoności to $F_k(r) \rightarrow 0$. Zauważmy, że istnieje $j \in \mathbb{Z}^+$ takie, że dla każdego $i > j$ zachodzi $\pi(A_{i,k}) = 0$

oraz $\pi(A_{j,k} > 0)$. Czyli $\text{ess sup}_{x \in \mathcal{X}} f_k(x) = \frac{j}{k}$ oraz jeśli r zbiega do $\rho e^{\frac{j}{k}}$, to $F_k(r) \rightarrow \infty$. Stąd oraz z monotoniczności F_k wiemy, że dla każdego k istnieje dokładnie jedno $r_{t,k}$ spełniające równanie $F_k(r_{t,k}) = 1$. Zatem, jak łatwo sprawdzić, funkcja $g_k(x) := \frac{e^{\frac{t}{2} f_k(x)}}{r_{t,k} - \rho e^{t f_k(x)}}$ jest funkcją własną dla operatora $\widehat{Q}_{t,k}$ o wartości własnej $r_{t,k}$.

Następny lemat kończy przygotowania do dowodu Twierdzenia 4.3.

Lemat 4.13. *Przy oznaczeniach jak powyżej, dla każdego $k \in \mathbb{Z}^+$ zachodzi*

i)

$$r_{t,k} = \left\| \widehat{Q}_{t,k} \right\|_{L^2(\pi)}$$

ii)

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\pi \exp \left(t \sum_{i=0}^{n-1} f_k(X'_i) \right) = \log(r_{t,k})$$

Dowód. Ad. i) Wiemy, że g_k jest jego funkcją własną operatora $\widehat{Q}_{t,k}$ z wartością własną $r_{t,k}$. Zatem norma operatora $\widehat{Q}_{t,k}$ wynosi conajmniej $r_{t,k}$. Ponadto operator $\widehat{Q}_{t,k}$ jest samosprężony. Więc jeśli jego norma jest większa niż $r_{t,k}$, to jest osiągana na przestrzeni prostopadłej do funkcji g_k . Załóżmy, że tak jest, wówczas istnieje ciąg funkcji h_n , być może stały, spełniający:

1. dla każdego $n = 1, 2, \dots$ $\|h_n\|_2 = 1$ i

$$(4.14) \quad h_n \perp g_k$$

2. Ciąg iloczynów skalarnych $\langle h_n, \widehat{Q}_{t,k} h_n \rangle$ jest niemalejący oraz dla każdego $n = 1, 2, \dots$ zachodzi

$$(4.15) \quad \left\| \widehat{Q}_{t,k} \right\|_{L^2(\pi)} \leq \langle h_n, \widehat{Q}_{t,k} h_n \rangle + \frac{1}{n}$$

Funkcja g_k jest dodatnia oraz ograniczona przez

$$(4.16) \quad C_1 := \frac{1}{r_{t,k}} \leq g_k \leq \text{ess sup}_{x \in \mathcal{X}} g_k(x) =: C_2 \quad \pi \text{ p.n.}$$

Z (4.14) wiemy, że funkcje $h_n^+ = \max(h_n, 0)$ i $h_n^- = \max(-h_n, 0)$ spełniają

$$\pi h_n^+ \leq \frac{C_2}{C_1} \pi h_n^- =: C_3 \pi h_n^-$$

oraz

$$(4.17) \quad \pi h_n^- \leq \frac{C_2}{C_1} \pi h_n^+ =: C_3 \pi h_n^+.$$

Dla każdej z funkcji h_n mamy

$$(4.18) \quad \begin{aligned} \langle h_n, \widehat{Q}_{t,k} h_n \rangle &= (1 - \rho) \left[\pi(e^{\frac{t}{2} f_k} h_n) \right]^2 + \rho \pi(e^{t f_k} h_n^2) = \\ &= (1 - \rho) \left[\left(\pi(e^{\frac{t}{2} f_k} h_n^+) \right)^2 - 2 \pi(e^{\frac{t}{2} f_k} h_n^+) \pi(e^{\frac{t}{2} f_k} h_n^-) + \left(\pi(e^{\frac{t}{2} f_k} h_n^-) \right)^2 \right] + \rho \pi(e^{t f_k} h_n^2). \end{aligned}$$

Operator $\widehat{Q}_{t,k}$ jest dodatni, czyli dla każdej funkcji h mamy $\langle |h|, \widehat{Q}_{t,k} |h| \rangle \geq \langle h, \widehat{Q}_{t,k} h \rangle$. Funkcje h_n^+ , h_n^- są nieujemne zatem z (4.15) oraz (4.18) otrzymujemy

$$(4.19) \quad \frac{1}{n} \geq \langle |h_n|, \widehat{Q}_{t,k} |h_n| \rangle - \langle h_n, \widehat{Q}_{t,k} h_n \rangle = 4(1-\rho)\pi(e^{\frac{t}{2}f_k} h_n^+) \pi(e^{\frac{t}{2}f_k} h_n^-) \geq 4(1-\rho)\pi(h_n^+) \pi(h_n^-)$$

Ostatecznie z (4.17), (4.18), (4.19) oraz definicji $r_{t,k}$ mamy

$$\begin{aligned} \langle h_n, \widehat{Q}_{t,k} h_n \rangle &\leq (1-\rho) \left[\left(\pi(e^{\frac{t}{2}f_k} h_n^+) \right)^2 + \left(\pi(e^{\frac{t}{2}f_k} h_n^-) \right)^2 \right] + \rho \pi(e^{tf_k} h_n^2) \\ &\leq e^t (1-\rho) \left[\left(\pi(h_n^+) \right)^2 + \left(\pi(h_n^-) \right)^2 \right] + \pi(h_n^2) \rho \operatorname{ess\,sup}_{x \in \mathcal{X}} e^{tf_k}(x) \\ &\leq e^t (1-\rho) \left[\left(\pi(h_n^+) \right)^2 + \left(\pi(h_n^-) \right)^2 \right] + r_{t,k} \\ &\leq \frac{e^t C_3}{2n} + r_{t,k} \end{aligned}$$

Zbiegając z n do nieskończoności otrzymujemy $\|\widehat{Q}_{t,k}\|_{L^2(\pi)} \leq r_{t,k}$. Co kończy dowód punktu i).

Ad. ii) Z punktu i) dla każdego n mamy

$$\begin{aligned} \|\widehat{Q}_{t,k}\|_{L^2(\pi)} &= r_{t,k} \\ &= \left(\frac{\langle g_k, \widehat{Q}_{t,k}^{n-1} g_k \rangle}{\|g_k\|_\pi^2} \right)^{\frac{1}{n-1}} \\ &\leq \left(\frac{C_2^2}{\|g_k\|_\pi^2} \right)^{\frac{1}{n}} \langle e^{\frac{t}{2}f_k} 1, \widehat{Q}_{t,k}^{n-1} e^{\frac{t}{2}f_k} 1 \rangle^{\frac{1}{n-1}} \\ &\leq \left(\frac{e^t C_2^2}{\|g_k\|_\pi^2} \right)^{\frac{1}{n-1}} \langle 1, \widehat{Q}_{t,k}^{n-1} 1 \rangle^{\frac{1}{n-1}} \\ &\leq \left(\frac{e^t C_2^2}{\|g_k\|_\pi^2} \right)^{\frac{1}{n-1}} \|\widehat{Q}_{t,k}\|_{L^2(\pi)} \\ &=: C_4^{\frac{1}{n-1}} \|\widehat{Q}_{t,k}\|_{L^2(\pi)}, \end{aligned}$$

ale $\mathbb{E}_\pi \exp\left(t \sum_{i=0}^{n-1} f_k(X'_i)\right) = \langle e^{\frac{t}{2}f_k} 1, \widehat{Q}_{t,k}^{n-1} e^{\frac{t}{2}f_k} 1 \rangle$. Zatem, po zlogarytmowaniu stronami, z twierdzenia o trzech ciągach otrzymujemy tezę. \square

Teraz możemy dokończyć dowód Twierdzenia 4.3

Dowód Twierdzenia 4.3. Z nierówności Markowa dla każdego $t > 0$ otrzymujemy

$$(4.20) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq e^{-tn(\theta + \varepsilon)} \mathbb{E}_\pi e^{tS_n}.$$

Z Lematu 4.7 oraz wzoru (4.12) mamy

$$\mathbb{E}_\pi e^{tS_n} \leq \|\widehat{Q}_t\|_{L^2(\pi)}^n \leq \|\widehat{Q}_{t,k}\|_{L^2(\pi)}^n.$$

Niech $(Y_n^k)_{n \geq 0}$ będzie łańcuchem Markowa z macierzą przejścia $M_{\theta_k, \rho}$. Wówczas z Lematu 4.10 oraz Lematu 4.13 otrzymujemy

$$\log(\|\widehat{Q}_{t,k}\|_{L^2(\pi)}) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\pi \exp \left(t \sum_{i=0}^{n-1} f_k(X_i') \right) \leq \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}_\theta \exp \left(t \sum_{i=0}^{n-1} Y_i^k \right) = \log(\eta_{t,k}),$$

czyli

$$\mathbb{E}_\pi e^{tS_n} \leq \eta_{t,k}^n.$$

Zbiegając z k do nieskończoności otrzymujemy

$$\mathbb{E}_\pi e^{tS_n} \leq \eta_t^n.$$

Zatem ze Stwierdzenia 4.9 otrzymujemy tezę. □

4.2. Uogólnienia nierówności Hoeffdinga

W poprzednim podrozdziale udowodniliśmy nierówności Hoeffdinga dla łańcuchów Markowa. Te nierówności są prawdziwe tylko w stanie stacjonarnym oraz dla funkcji ograniczonych. W tej części pokażemy jak można ich użyć w sytuacji, kiedy łańcuch nie jest w stanie stacjonarnym oraz funkcja f jest nieograniczona.

Zacniemy od prostszego problemu, czyli innego rozkładu początkowego niż rozkład stacjonarny π . Załóżmy, że rozkład początkowy ξ jest absolutnie ciągły względem π , czyli posiada gęstość względem π . Oznaczmy ją przez $\frac{d\xi}{d\pi}$ oraz dla $1 \leq p \leq \infty$ określmy jej p -tą normę przez

$$\left\| \frac{d\xi}{d\pi} \right\|_p = \begin{cases} \left(\int_{\mathcal{X}} \left| \frac{d\xi}{d\pi}(x) \right|^p d\pi(x) \right)^{\frac{1}{p}} & \text{gd}y \quad p < \infty \\ \text{ess sup}_{\{x \in \mathcal{X}\}} \left| \frac{d\xi}{d\pi}(x) \right| & \text{gd}y \quad p = \infty. \end{cases}$$

Przy tak przyjętych oznaczeniach możemy sformułować prosty wniosek z Twierdzenia 4.3.

Wniosek 4.21. *Przy założeniach jak w Twierdzeniu 4.3, dla dowolnego rozkładu początkowego ξ absolutnie ciągłego względem π oraz dla dowolnych liczb $p, q \geq 1$, spełniających $\frac{1}{p} + \frac{1}{q} = 1$, zachodzi*

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) \leq \left\| \frac{d\xi}{d\pi} \right\|_p \exp \left\{ -2 \frac{1 - \rho}{q(1 + \rho)} \varepsilon^2 n \right\}.$$

Dowód. Rozpiszmy lewą stronę nierówności

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) = \int_{\mathcal{X}} \mathbb{P}_x(S_n \geq n(\theta + \varepsilon)) \frac{d\xi}{d\pi}(x) \pi(dx),$$

korzystamy z nierówności Höldera i otrzymujemy

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) \leq \left\| \frac{d\xi}{d\pi} \right\|_p \left(\int_{\mathcal{X}} \mathbb{P}_x(S_n \geq n(\theta + \varepsilon))^q \pi(dx) \right)^{\frac{1}{q}}.$$

Ponieważ prawdopodobieństwo jest zawsze nie większe niż jeden to wyrażenie pod całką możemy oszacować następująco

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) \leq \left\| \frac{d\xi}{d\pi} \right\|_p \left(\int_{\mathcal{X}} \mathbb{P}_x(S_n \geq n(\theta + \varepsilon)) \pi(dx) \right)^{\frac{1}{q}}.$$

Zatem

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) \leq \left\| \frac{d\xi}{d\pi} \right\|_p \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon))^{\frac{1}{q}},$$

czyli korzystając z Twierdzenia 4.3 otrzymujemy tezę. \square

Dzięki temu prostemu wnioskowi uwolniliśmy się od niepraktycznego w zastosowaniach założenia o stacjonarności łańcucha. Jednak algorytmy używane w MCMC najczęściej stosowane są w przypadku, gdy funkcja f jest nieograniczona. Teraz, korzystając z nierówności Hoeffdinga, pokażemy oszacowania na prawdopodobieństwa błędów estymatora dla pewnej klasy funkcji nieograniczonych. Niech f będzie funkcją na przestrzeni \mathcal{X} o wartościach rzeczywistych, dodatkowo o f będziemy zakładać, że ma wykładnicze ogony względem rozkładu stacjonarnego. To znaczy, że istnieją stałe $a > 0$ i $b > 0$ oraz $M_0 > 0$ takie, że dla każdego $M > M_0$ zachodzi

$$(4.22) \quad \mathbb{P}_\pi(|f(X)| > \frac{M}{2}) \leq ae^{-bM}.$$

Załóżmy, że łańcuch $(X_n)_{n \geq 0}$ spełnia założenie (4.1). Podobnie jak wcześniej, oznaczmy przez S_n sumę $S_n := \sum_{i=0}^{n-1} f(X_i)$. Dodatkowo przez $S_{n,M}$ oznaczmy sumę funkcji obciętych przez $\frac{M}{2}$, czyli $S_{n,M} := \sum_{i=0}^{n-1} f(X_i) \mathbb{I}(|f(X_i)| \leq \frac{M}{2})$ oraz niech $\theta_M := \mathbb{E}_\pi(f(X_0) \mathbb{I}(|f(X_0)| \leq \frac{M}{2}))$. Przy tych oznaczeniach możemy oszacować interesujące nas prawdopodobieństwo następująco:

$$(4.23) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq \mathbb{P}_\pi(S_{n,M} \geq n(\theta_M + \varepsilon)) + \mathbb{P}_\pi(\exists_{1 \leq j \leq n} : |f(X_j)| > \frac{M}{2})$$

Wiadomo, że zachodzą następujące nierówności:

$$\begin{aligned} \mathbb{P}_\pi(\exists_{1 \leq j \leq n} : |f(X_j)| > \frac{M}{2}) &= \mathbb{P}_\pi\left(\bigcup_{i=1}^n \left\{ |f(X_i)| > \frac{M}{2} \right\}\right) \\ &\leq \sum_{i=1}^n \mathbb{P}_\pi\left(|f(X_i)| > \frac{M}{2}\right) \\ &= n\mathbb{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right). \end{aligned}$$

Możemy przekształcić nierówność (4.23) i otrzymamy

$$(4.24) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq \mathbb{P}_\pi\left(\frac{S_{n,M}}{M} + \frac{n}{2} \geq n\left(\frac{\theta_M}{M} + \frac{1}{2} + \frac{\varepsilon}{M}\right)\right) + n\mathbb{P}_\pi\left(|f(X_1)| > \frac{M}{2}\right).$$

Załóżmy, że $M > M_0$ oraz, że $\varepsilon + \theta < \frac{M}{2}$ wówczas do pierwszej części (4.24) możemy zastosować Twierdzenie 4.3. Natomiast drugą część oszacujemy korzystając z faktu, że f ma ogony wykładnicze. Otrzymujemy

$$(4.25) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq e^{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{M^2}} + e^{-bM+\log(an)}.$$

Teraz zoptymalizujemy to wyrażenie względem M . Oznaczmy przez $W(M)$ wielomian

$$(4.26) \quad W(M) := M^3 - M^2 \frac{\log(an)}{b} - 2 \frac{1-\rho}{b(1+\rho)} n \varepsilon^2.$$

Zauważmy, że wielomian $W(M) < 0$ dla $M \leq 0$ oraz $\frac{W(M)}{M^2}$ jest funkcją rosnącą dla $M > 0$, zatem $W(M)$ posiada tylko jeden pierwiastek rzeczywisty M^* i nierówność (4.25) możemy przekształcić następująco:

$$(4.27) \quad \mathbb{P}_\pi(S_n \geq n(\theta + \varepsilon)) \leq 2e^{-2\frac{1-\rho}{1+\rho}n\frac{\varepsilon^2}{M^{*2}}}.$$

Ponadto wiemy, że $W\left(\frac{\log(an)}{b}\right) \leq 0$ oraz $W\left(\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}\right) \leq 0$, czyli $M^* \geq \max\left\{\frac{\log(an)}{b}, \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}\right\}$.

Przez całe rozumowanie zakładaliśmy, że $M^* > M_0$ oraz, że $\varepsilon + \theta < \frac{M^*}{2}$, czyli nierówność (4.27) jest prawdziwa dla każdego $n \in \mathbb{Z}^+$ oraz dla każdego $\varepsilon > 0$ spełniających:

$$(4.28) \quad \log(an) > bM_0 \quad \text{oraz} \quad \varepsilon < \frac{\log(an)}{2b} - \theta,$$

albo

$$(4.29) \quad \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}} > M_0, \quad \text{oraz} \quad \varepsilon + \theta < \frac{1}{2}\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}.$$

Przyjrzyjmy się przypadkowi, gdy $\frac{\log(an)}{b} \geq \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}$. Wówczas M^* jest mniejsze lub równe od jedynego rzeczywistego rozwiązania równania

$$M^3 - \frac{\log(an)}{b}M^2 - \frac{\log(an)^3}{b^3} = 0.$$

Podstawiamy do tego równania $D = \frac{bM}{\log(an)}$ i po podzieleniu stronami przez $\frac{\log(an)^3}{b^3}$ otrzymujemy

$$D^3 - D^2 - 1 = 0.$$

Rozwiązaniem tego równania jest liczba $D^* \approx 1.465571$, czyli $M^* \leq D^*\frac{\log(an)}{b}$. Analogicznie postępujemy w przypadku, gdy $\frac{\log(an)}{b} < \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}$ i otrzymujemy następujący wniosek z Twierdzenia 4.3.

Wniosek 4.30. Niech X_n będzie łańcuchem Markowa z rozkładem stacjonarnym π i operatorem przejścia P spełniającym $\|P - \Pi\|_{L^2(\pi)} \leq \rho$. Niech f będzie funkcją $f : \mathcal{X} \rightarrow \mathbb{R}$ taką, że dla każdego $M > M_0$ zachodzi $\mathbb{P}_\pi(|f(X_0)| > \frac{M}{2}) \leq ae^{-bM}$. Wówczas dla dowolnego rozkładu początkowego ξ absolutnie ciągłego względem π , dla dowolnego $1 \leq p, q \leq \infty$ takich, że $\frac{1}{p} + \frac{1}{q} = 1$, oraz dla każdego n i $\varepsilon > 0$ spełniających (4.28) oraz (4.29) zachodzi:

$$\mathbb{P}_\xi(S_n \geq n(\theta + \varepsilon)) \leq \left\| \frac{d\xi}{d\pi} \right\|_p 2 \exp \left\{ -\frac{2}{q} \frac{1-\rho}{1+\rho} \varepsilon^2 \frac{n}{M^{*2}} \right\},$$

gdzie M^* jest rozwiązaniem

$$M^3 - M^2 \frac{\log(an)}{b} - 2 \frac{1-\rho}{b(1+\rho)} n \varepsilon^2 = 0.$$

Ponadto

$$\max \left\{ \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}, \frac{\log(an)}{b} \right\} \leq M^* \leq 1.465571 \max \left\{ \left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}, \frac{\log(an)}{b} \right\}.$$

Uwaga 4.31. W stwierdzeniu 5.25 pokażemy, że jeśli V spełnia podwójny warunek dryfu (definicja 5.14) implikujący istnienie ρ z warunku (4.1), to funkcje $|f| \leq \log(V)$ mają wykładnicze ogony względem rozkładu π .

Uwaga 4.32. W przypadku zmiennych losowych niezależnych, własność (4.22) implikuje istnienie funkcji tworzącej momenty dla sumy S_n . W sytuacji gdy rozpatrujemy łańcuchy Markowa warunkiem koniecznym i dostatecznym istnienia funkcji tworzącej momenty sumy $S_{\tau(C)}$ (dla nieograniczonej f) jest silniejszy warunek dryfu (zob. [1]):

$$\exp(-V(x)) P(\exp(V))(x) \leq \exp(-|f(x)| + b\mathbb{I}(x \in C)),$$

gdzie $V > 0$, $b > 0$ oraz C jest a_n -małym zbiorem. Warunek ten wydaje się być trudnym do sprawdzenia i nie są znane nietrywialne przykłady łańcuchów go spełniających. Czyli przy założeniach (4.22) oraz (4.1) nie możemy użyć standardowej metody szacowania funkcji tworzącej momenty.

Rozdział 5

Norma operatora P i jego luka spektralna w $L^2(\pi)$

W poprzednim rozdziale wykazaliśmy nierówności Hoeffdinga oraz wnioski z nich wynikające korzystając z istnienia $\rho < 1$ spełniającego (4.1), czyli $\|P - \Pi\|_{L^2(\pi)} \leq \rho$. W tej części pracy przyjrzymy się dokładniej temu założeniu. Z Twierdzenia 2.25 wiemy, że w przypadku odwracalnym istnienie ρ spełniającego (4.1) jest równoważne geometrycznej ergodyczności. W pierwszej części tego rozdziału przeanalizujemy to założenie w przypadku łańcuchów nieodwracalnych. W szczególności sformułujemy pewien nowy warunek dryfu pozwalający na oszacowanie ρ .

Dla odwracalnych łańcuchów użytecznym narzędziem jest luka spektralna w przestrzeni $L^2(\pi)$ [16, 25, 28, 46, 44]. Wykorzystując lukę spektralną można uzyskać lepsze oszacowania błędów estymatorów [46] lub też osłabić założenia dla CTG [44]. W drugiej części tego rozdziału pokażemy, że można rozpatrywać pojęcie luki spektralnej w $L^2(\pi)$ również dla nieodwracalnych łańcuchów. Okaże się, że można wówczas uzyskać podobne rezultaty jak w przypadku odwracalnym. Jednak w przypadku nieodwracalnym własność posiadania luki spektralnej w $L^2(\pi)$ nie jest już, tak jak dla odwracalnych, równoważna geometrycznej ergodyczności. W związku z tym przedstawimy również założenia gwarantujące istnienie luki w języku warunków dryfu.

5.1. Oszacowania ρ .

Zanim sformułujemy odpowiednie fakty dla przypadku ogólnego, potrzebujemy wprowadzić nowe oznaczenia. Jeżeli $P(x, dy)$ jest jądrem przejścia z rozkładem stacjonarnym π , to zdefiniujemy jądro przejścia $P^*(y, dx)$ sprzężone do $P(x, dy)$ poprzez równanie:

$$\int_A \pi(dx)P(x, B) = \int_B \pi(dy)P^*(y, A), \quad \forall A, B \in \mathcal{B}(\mathcal{X}),$$

w skrócie:

$$(5.1) \quad \pi(dx)P(x, dy) = \pi(dy)P^*(y, dx), \quad \forall x, y \in \mathcal{X}.$$

Istnienie jądra spełniającego warunek (5.1) wynika z faktu, że \mathcal{X} jest przestrzenią polską i możemy określić $P^*(y, A)$ jako regularną wersję prawdopodobieństwa warunkowego $\mathbb{P}_\pi(X_0 \in A | X_1 = y)$. Operator P^* , związany z jądrem $P^*(x, dy)$ jest operatorem sprzężonym do P .

Uwaga 5.2. Można wyznaczyć $P^*(y, dx)$ jeśli jądro $P(x, dy)$ oraz rozkład π mają gęstości względem tej samej miary. Jednak najważniejszym przykładem nieodwracalnego łańcucha Markowa jest próbnik Gibbsa. Operator przejścia dla próbnika Gibbsa jest postaci $P = P_1 P_2 \cdots P_d$, gdzie jak łatwo sprawdzić wszystkie P_i są samosprężonymi operatorami. Zatem, w tym szczególnym przypadku, operator sprzężony P^* jest równy $P^* = P_d \cdots P_2 P_1$. Sprzężone jądro przejścia $P^*(y, dx)$ jest również próbnikiem Gibbsa, tylko z odwrotną kolejnością zamiany współrzędnych.

Stwierdzenie 5.3. *Jeżeli łańcuch Markowa z jądrem przejścia $P(x, dy)$ spełnia (4.1), to jest geometrycznie ergodyczny.*

Dowód. Wiadomo, że jeśli łańcuch jest π -nieprzywiedlny, to istnieje n -mały zbiór C ([40]). Z (4.1) wiemy, że istnieje jednoznacznie wyznaczony rozkład stacjonarny π , więc łańcuch $(X_n)_{n \geq 1}$ jest π -nieprzywiedlny. Stąd korzystając z (4.1), nierówności Cauchy'ego-Schwartz'a oraz z faktu, że miara ξP^n ma gęstość względem π postaci $\frac{d\xi P^n}{d\pi}(x) = P^{*n} \left(\frac{d\xi}{d\pi} \right)(x)$ otrzymujemy

$$\begin{aligned} \|\pi|_C P^n(\cdot) - \pi(\cdot)\|_{tv} &= \frac{1}{2} \sup_{|f| \leq 1} |\pi|_C P^n(f) - \pi(f)| = \frac{1}{2} \sup_{|f| \leq 1} \left| \int_{\mathcal{X}} f(x) \left(\frac{d\pi|_C P^n}{d\pi}(x) - 1 \right) \pi(dx) \right| \\ (5.4) \qquad \qquad \qquad &\leq \frac{1}{2} \left\| (P^{*n} - \Pi) \frac{d\pi|_C}{d\pi}(x) \right\|_2 \leq \frac{1}{2} \left\| \frac{d\pi|_C}{d\pi}(x) \right\|_2 \|P^{*n} - \Pi\|_{L^2(\pi)} \end{aligned}$$

Wiadomo, że norma operatora $P^{*n} - \Pi$ jest równa normie operatora sprzężonego $P^n - \Pi$, czyli z (4.1) przekształcamy (5.4) w

$$\|\pi|_C P^n(\cdot) - \pi(\cdot)\|_{tv} \leq \frac{1}{2} \left\| \frac{d\pi|_C}{d\pi}(x) \right\|_2 \|P^n - \Pi\|_{L^2(\pi)} \leq \frac{1}{2} \left\| \frac{d\pi|_C}{d\pi}(x) \right\|_2 \rho^n$$

Zatem na mocy Twierdzenia 2.24 otrzymujemy, że łańcuch z jądrem $P(x, dy)$ jest geometrycznie ergodyczny. \square

Implikacja w przeciwną stronę w ogólnym przypadku nie jest prawdziwa, ale można przedstawić następującą charakteryzację warunku (4.1).

Twierdzenie 5.5. *Łańcuch Markowa z jądrem przejścia $P(x, dy)$ spełnia (4.1) wtedy i tylko wtedy, gdy łańcuch Markowa z jądrem przejścia $PP^*(x, dy)$ jest geometrycznie ergodyczny. Ponadto γ spełnia*

$$\|(PP^*)^n(x, \cdot) - \pi(\cdot)\|_{tv} \leq M_x \gamma^n$$

wtedy i tylko wtedy, gdy $\sqrt{\gamma} \geq \rho = \|P - \Pi\|_{L^2(\pi)}$.

Dowód. Dowodzimy twierdzenie poprzez ciąg równoważności. Przypomnijmy dobrze znany fakt z analizy funkcjonalnej: Jeżeli T jest operatorem liniowym na przestrzeni Hilberta \mathcal{H} , to

$$\|T\|^2 = \|T^*T\| = \|TT^*\| = \|T^*\|^2.$$

Zatem $\|P - \Pi\|_{L^2(\pi)} \leq \rho < 1$ wtedy i tylko wtedy, gdy $\|(P - \Pi)(P - \Pi)^*\|_{L^2(\pi)} \leq \rho^2$. Operator Π jest rzutem prostym komutującym zarówno z P jak i z P^* , czyli

$$(P - \Pi)(P - \Pi)^* = PP^* - \Pi P^* - P\Pi + \Pi = PP^* - \Pi.$$

Z Twierdzenia 2.25 wiemy, że $\|PP^* - \Pi\|_{L^2(\pi)} \leq \rho^2 < 1$ jest równoważne geometrycznej ergodyczności łańcucha z jądrem przejścia $PP^*(x, dy)$. Ponadto z Twierdzenia 2.25 wiemy, że ρ^2 jest równe współczynnikowi geometrycznej ergodyczności łańcucha z jądrem $PP^*(x, A)$, co kończy dowód. \square

Przedstawiona powyżej charakteryzacja warunku (4.1) w typowych przypadkach nie pozwala na oszacowanie ρ wprost. Teraz przedstawimy oszacowania ρ dla podstawowych klas łańcuchów Markowa.

5.1.1. Oszacowania dla skończonej przestrzeni stanów

W ogólności nie wszystkie ergodyczne łańcuchy na skończonej przestrzeni stanów spełniają (4.1).

Przykład 5.6. Niech $(X_n)_{n \geq 1}$ będzie łańcuchem na przestrzeni stanów $\{0, 1, 2\}$ z następującymi prawdopodobieństwami przejścia

$$\begin{aligned} P(0, 0) &= \frac{1}{2}, \\ P(0, 1) &= \frac{1}{2}, \\ P(1, 2) &= 1, \\ P(2, 0) &= 1, \end{aligned}$$

(pozostałe prawdopodobieństwa przejścia są równe 0). Ten łańcuch jest nieokresowy i nieprzywiedlny, a co za tym idzie jest ergodyczny. Jego rozkład stacjonarny jest równy $\pi = [1/2, 1/4, 1/4]$, czyli na podstawie definicji (5.1) mamy $P^*(2, 1) = 1$. Stąd $PP^*(1, 1) = 1$, zatem łańcuch z macierzą przejścia PP^* jest przywiedlny, czyli nie jest ergodyczny. Zatem z Twierdzenia 5.5 $(X_n)_{n \geq 1}$ nie spełnia założenia (4.1).

Jednak dla przestrzeni stanów \mathcal{X} skończonej, $|\mathcal{X}| = d$, zachodzi następujący fakt, który jest szczególnym przypadkiem twierdzenia 5.5.

Stwierdzenie 5.7. *Niech $(X_n)_{n \geq 0}$ będzie ergodycznym łańcuchem Markowa na skończonej przestrzeni stanów \mathcal{X} z macierzą przejścia P . Wówczas, jeżeli druga co do wielkości wartość własna macierzy PP^* jest mniejsza niż 1, to ρ występujące w warunku (4.1) jest równe jej pierwiastkowi. W przeciwnym przypadku, nie jest spełniony warunek (4.1).*

5.1.2. Oszacowania dla jednostajnie ergodycznych łańcuchów Markowa

W ogólności jednostajnie ergodyczne łańcuchy mogą nie spełniać warunku (4.1) (patrz. Przykład 5.6). Jednak jeśli spełniony jest warunek Doeblina (3.27), a tak jest w typowych sytuacjach, to warunek (4.1) jest spełniony. Przypomnijmy, że warunek Doeblina oznacza

$$P(x, \cdot) \geq \beta \nu(\cdot) \quad \text{dla każdego } x \in \mathcal{X}.$$

Stwierdzenie 5.8. *Jeśli spełniony jest warunek Doeblina, to spełniony jest warunek (4.1) z $\rho \leq \sqrt{1 - \frac{1}{4}\beta^2}$.*

Następny fakt jest niezbędny do dowodu tego stwierdzenia. Przedstawimy go w większej ogólności, aby móc skorzystać z niego również w dalszej części pracy.

Lemat 5.9. *Jeśli dla pewnego zbioru C miary π dodatniej, $\beta > 0$ oraz miary probabilistycznej ν zachodzi:*

$$P(x, A) \geq \beta \mathbb{I}(x \in C) \nu(A) \quad A \in \mathcal{B}(\mathcal{X}), x \in \mathcal{X},$$

to

$$PP^*(x, A) \geq \frac{1}{4} \beta^2 \mathbb{I}(x \in C) \pi(C \cap A) \quad A \in \mathcal{B}(\mathcal{X}), x \in \mathcal{X}.$$

Dowód. Dla zbiorów A takich, że $A \cap C = \emptyset$ to lemat jest oczywisty. Niech $A \cap C \neq \emptyset$ oraz niech $x \in C$, wówczas z założenia mamy:

$$(5.10) \quad PP^*(x, A) = \int_{\mathcal{X}} P^*(z, A)P(x, dz) \geq \beta \int_{\mathcal{X}} P^*(z, A \cap C)\nu(dz)$$

Łatwo, zauważyć, że miara ν posiada gęstość $\frac{d\nu}{d\pi}$ względem π . Oznaczmy, przez $D(\varepsilon) := \{x \in \mathcal{X} : \frac{d\nu}{d\pi} \geq \varepsilon\}$. Ponieważ $\int_{\mathcal{X}} \frac{d\nu}{d\pi} \pi(dx) = 1$ oraz $\frac{d\nu}{d\pi} \geq 0$, to dla każdego $0 \leq \varepsilon \leq 1$ zachodzi $\pi(D(\varepsilon)) > 0$. Skorzystamy teraz z definicji P^* (5.1) oraz z własności zbiorów $D(\varepsilon)$ i przekształcając (5.10) otrzymujemy:

$$(5.11) \quad PP^*(x, A) \geq \beta \int_{D(\varepsilon)} \int_{A \cap C} \varepsilon P^*(z, dy)\pi(dz) = \beta\varepsilon \int_{D(\varepsilon)} \int_{A \cap C} \pi(dy)P(y, dz).$$

Korzystamy ponownie z założenia i z (5.11) otrzymujemy

$$(5.12) \quad PP^*(x, A) \geq \beta^2\varepsilon\nu(D(\varepsilon))\pi(A \cap C).$$

Pozostało nam do oszacowania $\nu(D(\varepsilon))$, ale

$$\nu(D(\varepsilon)^C) = \int_{D(\varepsilon)^C} \frac{d\nu}{d\pi}(x)\pi(dx) \leq \varepsilon \int_{\mathcal{X}} \pi(dx) = \varepsilon,$$

więc $\nu(D(\varepsilon)) \geq 1 - \varepsilon$. Wstawiając to do (5.12) i optymalizując względem ε otrzymujemy

$$(5.13) \quad PP^*(x, A) \geq \frac{1}{4}\beta^2\pi(A \cap C).$$

□

Dowód Stwierdzenia 5.8. Z lematu 5.9 wynika, że łańcuch o jądrze przejścia $PP^*(x, dy)$ również spełnia warunek Doeblina, ponadto łańcuch ten jest odwracalny. Stąd wiadomo, że spektrum $PP^* - \Pi$ jest zawarte w $[-1 + \frac{1}{4}\beta^2, 1 - \frac{1}{4}\beta^2]$ [44]. Korzystając z Twierdzenia 5.5 oraz 2.25 otrzymujemy, że $\rho \leq \sqrt{1 - \frac{1}{4}\beta^2}$ □

W przypadku odwracalnych jednostajnie ergodycznych łańcuchów Markowa warunek (4.1) jest zawsze spełniony (twierdzenie 2.25). Dodatkowo jeśli $(X_n)_{n \geq 1}$ spełnia

$$P^h(x, \cdot) \geq \beta\nu(\cdot) \quad \text{dla każdego } x \in \mathcal{X}.$$

to $\rho \leq (1 - \beta)^{\frac{1}{h}}$. Jest tak, ponieważ $\|P^h - \Pi\|_{L^2(\pi)} = \|P - \Pi\|_{L^2(\pi)}^h$ oraz spektrum $P^h - \Pi$ jest zawarte w $[-1 + \beta, 1 - \beta]$.

5.1.3. Oszacowania dla geometrycznie ergodycznych łańcuchów Markowa

Wiadomo, że geometryczna ergodyczność w ogólności nie gwarantuje spełnienia warunku (4.1). Najpierw podamy zmodyfikowany warunek dryfu zapewniający warunek (4.1). Następnie przedstawimy oszacowania ρ w terminach wielkości występujących w tym warunku.

Definicja 5.14. Powiemy, że $V \geq 1$ spełnia podwójny warunek dryfu, jeśli spełnione są następujące warunki:

1. Zbiory $C_r := \{x \in \mathcal{X} : V(x) \leq r\}$ są małymi zbiorami dla każdego $r \geq 1$, tzn.

$$P(x, dy) \geq \beta_r \mathbb{I}(x \in C_r) \nu_r(dy),$$

gdzie $\beta_r > 0$ oraz miara probabilistyczna ν_r mogą zależeć od r .

2. Istnieją λ_1 i λ_2 takie, że $\lambda_1 \lambda_2 < 1$ oraz istnieje stała $K < \infty$, dla których spełniony jest warunek

$$\begin{aligned} PV(x) &\leq \lambda_1 V(x) + K \\ P^*V(x) &\leq \lambda_2 V(x) + K \end{aligned}$$

Twierdzenie 5.15. *Jeżeli V spełnia podwójny warunek dryfu, to łańcuch $(X_n)_{n \geq 1}$ spełnia warunek (4.1).*

Dowód. Z lematu 5.9 otrzymujemy, że dla każdego $r \geq 1$ zachodzi

$$(5.16) \quad PP^*(x, A) \geq \frac{1}{4} \beta_r^2 \pi(A \cap C_r) \mathbb{I}(x \in C_r)$$

Czyli warunek małego zbioru dla P implikuje warunek małego zbioru dla PP^* . Korzystając z Definicji 5.14 obliczamy

$$(5.17) \quad PP^*V(x) \leq \lambda_1 \lambda_2 V(x) + (1 + \lambda_2)K.$$

Dla $\lambda_2 > 1$ oraz $r = \frac{2(1+\lambda_2)K}{1-\lambda_1\lambda_2}$ mamy

$$(5.18) \quad PP^*V(x) \leq \frac{1 + \lambda_1 \lambda_2}{2} V(x) \quad x \notin C_r$$

$$(5.19) \quad PP^*V(x) \leq \frac{(1 + \lambda_1 \lambda_2)(1 + \lambda_2)K}{1 - \lambda_1 \lambda_2} \quad x \in C_r$$

Dla $\lambda_2 \leq 1$ oraz $r = \frac{2(1+\lambda_2)K}{1-\lambda_1}$ mamy

$$(5.20) \quad PP^*V(x) \leq \frac{1 + \lambda_1(2\lambda_2 - 1)}{2} V(x) \quad x \notin C_r$$

$$(5.21) \quad PP^*V(x) \leq \frac{(1 + 2\lambda_1\lambda_2 - \lambda_1)(1 + \lambda_2)K}{1 - \lambda_1} \quad x \in C_r$$

Zatem z Twierdzenia 2.24 otrzymujemy, że łańcuch Markowa z jądrem PP^* jest geometrycznie ergodyczny, więc z Twierdzenia 5.5 otrzymujemy tezę. \square

Dla ułatwienia zapisu w następnym stwierdzeniu wprowadźmy oznaczenia.

Dla $\lambda_2 > 1$:

$$\begin{aligned} r^* &:= \frac{2(1 + \lambda_2)K}{1 - \lambda_1 \lambda_2} \\ \tilde{\lambda} &:= \frac{2\lambda_1 + \lambda_1 \lambda_2 + 1}{2(1 + \lambda_2)} \\ \tilde{K} &:= \frac{2\lambda_1 + \lambda_1 \lambda_2 + 1}{1 - \lambda_1 \lambda_2} K \\ \beta_0 &:= \frac{1}{4} \beta^2 \frac{1 - \tilde{\lambda}}{\tilde{K} - \tilde{\lambda}} \\ \lambda_0 &:= \frac{1 + \lambda_1 \lambda_2}{2} \\ K_0 &:= \frac{(1 + \lambda_1 \lambda_2)(1 + \lambda_2)K}{1 - \lambda_1 \lambda_2}, \end{aligned}$$

oraz dla $\lambda_2 \leq 1$:

$$\begin{aligned} r^* &:= \frac{2(1 + \lambda_2)K}{1 - \lambda_1} \\ \tilde{\lambda} &:= \lambda_1 + \frac{1 - \lambda_1}{2(1 + \lambda_2)} \\ \tilde{K} &:= \frac{\lambda_1 + 2\lambda_1\lambda_2 + 1}{1 - \lambda_1}K \\ \beta_0 &:= \frac{1}{4}\beta^2 \frac{1 - \tilde{\lambda}}{\tilde{K} - \tilde{\lambda}} \\ \lambda_0 &:= \frac{1 + \lambda_1(2\lambda_2 - 1)}{2} \\ K_0 &:= \frac{(1 + 2\lambda_1\lambda_2 - \lambda_1)(1 + \lambda_2)K}{1 - \lambda_1}. \end{aligned}$$

Stwierdzenie 5.22. *Jeżeli V spełnia podwójny warunek dryfu z $\lambda_1 < 1$, to*

$$\rho \leq R_0^{-\frac{1}{2}},$$

gdzie

$$R_0 := \min\{\lambda_0^{-1}, (1 - \beta_0)^{-\frac{1}{\alpha_1}}\}$$

oraz

$$\alpha_1 := 1 + \frac{\log\left(\frac{K_0 - \beta_0}{1 - \beta_0}\right)}{\log(\lambda_0^{-1})}.$$

Dowód. Przypomnijmy, że z (5.16) wiemy, że C_{r^*} jest małym zbiorem (dla PP^*)

$$PP^*(x, dy) \geq \frac{1}{4}\beta^2\pi(C_{r^*})\pi|_{C_{r^*}}(dy)\mathbb{I}(x \in C_{r^*}).$$

Jeżeli $\lambda_1 < 1$, to łańcuch z jądrem przejścia $P(x, A)$ spełnia warunek dryfu do małego zbioru C_{r^*} z funkcją V oraz z $\tilde{\lambda}$ i \tilde{K} . Stąd ze stwierdzenia 3.20 oraz z (5.16) otrzymujemy

$$PP^*(x, dy) \geq \beta_0\pi|_{C_{r^*}}(dy)\mathbb{I}(x \in C_{r^*}).$$

Z (5.18) oraz (5.19) gdy $\lambda_2 > 1$ lub z (5.20) oraz (5.21) gdy $\lambda_2 \leq 1$ otrzymujemy, że łańcuch z jądrem $PP^*(x, A)$ spełnia warunek dryfu do zbioru C_{r^*} :

$$PP^*V(x) \leq \begin{cases} \lambda_0 V(x) & \text{dla } x \notin C_{r^*}, \\ K_0 & \text{dla } x \in C_{r^*}. \end{cases}$$

Łańcuch z jądrem przejścia $PP^*(x, dy)$ jest odwracalny oraz operator PP^* jest dodatnio określony i z Twierdzenia 5.5, Twierdzenia 2.24 oraz Twierdzenia 1.3 z [5] otrzymujemy, że

$$\rho \leq R_0^{-\frac{1}{2}},$$

gdzie

$$R_0 := \min\{\lambda_0^{-1}, (1 - \beta_0)^{-\frac{1}{\alpha_1}}\}$$

oraz

$$\alpha_1 := 1 + \frac{\log\left(\frac{K_0 - \beta_0}{1 - \beta_0}\right)}{\log(\lambda_0^{-1})}.$$

□

Uwaga 5.23. W definicji 5.14, możemy zastąpić warunek małego zbioru dla wszystkich r przez warunek małego zbioru tylko dla r^* . Jednak typowa sytuacja jest taka, że potrafimy wykazać, że dla każdego r zbiory C_r są małe.

Uwaga 5.24. W przypadku odwracalnym z Twierdzenia 2.25 wiemy, że warunek (4.1) jest równoważny geometrycznej ergodyczności. Jeśli spełniony jest standardowy warunek dryfu do małego zbioru C oraz łańcuch jest silnie nieokresowy, tzn. miara ν występująca w definicji małego zbioru C spełnia $\nu(C) > 0$, to ρ możemy oszacować z Twierdzenia 1.2 z [5].

Następne stwierdzenie pokazuje jak przy spełnieniu podwójnego warunku dryfu można określić klasę funkcji nieograniczonych spełniających założenia wniosku 4.30.

Stwierdzenie 5.25. *Jeżeli spełniony jest warunek dryfu 5.14, to wszystkie funkcje $|f| \leq \log(V)$ mają wykładnicze ogony, ponadto dla każdego $M > 0$:*

$$\mathbb{P}_\pi \left(|f(X_0)| > \frac{M}{2} \right) \leq \frac{K - \lambda_{\min}}{1 - \lambda_{\min}} e^{-\frac{M}{2}},$$

czyli stałe występujące we wniosku 4.30 są postaci: $M_0 = 0$, $a = \frac{K - \lambda_{\min}}{1 - \lambda_{\min}}$ oraz $b = \frac{1}{2}$, gdzie $\lambda_{\min} = \min\{\lambda_1, \lambda_2\}$.

Dowód. Załóżmy, że $\lambda_{\min} = \lambda_1$, wówczas z nierówności Markowa oraz powtarzając rozumowanie z dowodu stwierdzenia 3.12 (i) otrzymujemy

$$\mathbb{P}_\pi \left(|f(X_1)| > \frac{M}{2} \right) \leq e^{-\frac{M}{2}} \pi V \leq \frac{K - \lambda_1}{1 - \lambda_1} e^{-\frac{M}{2}}.$$

W przeciwnym przypadku, jeśli $\lambda_{\min} = \lambda_2$ to zastępujemy łańcuch z jądrem $P(x, dy)$ przez łańcuch z jądrem $P^*(x, dy)$ i otrzymujemy tezę. \square

5.2. Luka spektralna

Definicja 5.26. Powiemy, że łańcuch Markowa z jądrem przejścia $P(x, dy)$ i rozkładem stacjonarnym π ma lukę spektralną $(1 - \varrho) > 0$, jeżeli

$$\varrho := \sup\{|\lambda| : \lambda \in \sigma(P - \Pi)\} < 1,$$

gdzie $\sigma(P - \Pi)$ oznacza spektrum operatora $P - \Pi$.

Taka sama definicja luki spektralnej znajduje się również w [15, 16, 27, 28, 44, 46]. Przytoczmy znany fakt o promieniu spektralnym:

$$(5.27) \quad \varrho = \lim_{n \rightarrow \infty} \|(P - \Pi)^n\|_{L^2(\pi)}^{\frac{1}{n}} = \lim_{n \rightarrow \infty} \|P^n - \Pi\|_{L^2(\pi)}^{\frac{1}{n}},$$

przy czym ostatnia równość wynika z faktu, że Π jest rzutem prostopadłym komutującym z P . Zauważmy jeszcze, że istnienie luki spektralnej jest równoważne własności ϱ -mieszania dla stacjonarnych łańcuchów Markowa. W następnym stwierdzeniu podajemy inne podstawowe właściwości luki spektralnej.

Stwierdzenie 5.28. *Dla łańcucha Markowa z jądrem przejścia $P(x, dy)$ i rozkładem stacjonarnym π , następujące własności są równoważne:*

(i) *Łańcuch ma lukę spektralną $1 - \varrho$ w $L^2(\pi)$.*

(ii) Istnieje $n_0 \in \mathbb{Z}^+$ oraz $\gamma_0 < 1$ takie, że

$$\|P^{n_0} - \Pi\|_{L^2(\pi)} \leq \gamma_0.$$

(iii) Istnieje stała $M < \infty$ oraz $\gamma < 1$ takie, że dla każdego n zachodzi

$$\|P^n - \Pi\|_{L^2(\pi)} \leq M\gamma^n.$$

Dowód.

(i) \Rightarrow (ii) Z (5.27) i z własności (i) wiemy, że granica $\lim_{n \rightarrow \infty} \|P^n - \Pi\|_{L^2(\pi)}^{\frac{1}{n}} \leq \varrho < 1$, stąd istnieje n_0 takie, że $\|P^{n_0} - \Pi\|_{L^2(\pi)} \leq \gamma_0 < 1$.

(ii) \Rightarrow (iii) Obliczmy dla dowolnego n , pamiętając, że $\|P - \Pi\|_{L^2(\pi)} \leq 1$:

$$\|P^n - \Pi\|_{L^2(\pi)} \leq \|P^{n_0} - \Pi\|_{L^2(\pi)}^{\lfloor \frac{n}{n_0} \rfloor} \leq \gamma_0^{\lfloor \frac{n}{n_0} \rfloor} \leq \gamma_0^{\frac{n-n_0}{n_0}},$$

podstawmy $\gamma := \gamma_0^{\frac{1}{n_0}}$ oraz $M := \gamma_0^{-1}$ i otrzymamy (iii)

(iii) \Rightarrow (i) Z (5.27) mamy

$$\varrho = \lim_{n \rightarrow \infty} \|P^n - \Pi\|_{L^2(\pi)}^{\frac{1}{n}} \leq \lim_{n \rightarrow \infty} M^{\frac{1}{n}} \gamma = \gamma.$$

□

Uwaga 5.29. Na ogół zachodzi nierówność $\varrho \leq \rho$, gdzie $1 - \varrho$ jest luką spektralną zaś $\rho = \|P - \Pi\|_{L^2(\pi)}$. Dla łańcuchów odwracalnych mamy równość.

Wniosek 5.30. Jeżeli łańcuch Markowa z jądrem przejścia $P(x, dy)$ ma lukę spektralną w $L^2(\pi)$ to jest geometrycznie ergodyczny.

Dowód jest identyczny jak dowód Stwierdzenia 5.3. Dla nieodwracalnych łańcuchów implikacja w drugą stronę nie jest prawdziwa [28]. My przedstawimy prostszy przykład oparty tylko na faktach dotyczących łańcuchów Markowa.

Przykład 5.31. Niech przestrzeń stanów $\mathcal{X} := \mathbb{Z}^+ \times \mathbb{Z}^+$ oraz niech p będzie rozkładem prawdopodobieństwa na \mathbb{Z}^+ takim, że istnieje $\kappa > 1$ spełniająca

$$(5.32) \quad \sum_{n=0}^{\infty} \kappa^n p(n) < \infty$$

oraz dla każdego n mamy $p(n) > 0$. Na przykład $p(n) = \frac{1}{2^n}$ spełnia te założenia. Zdefiniujemy łańcuch Markowa na \mathcal{X} poprzez prawdopodobieństwa przejścia:

$$\begin{aligned} P((0, 0), (0, 0)) &= p(0), \\ P((0, 0), (n, n)) &= p(n), \\ P((n, k), (n, k-1)) &= 1 \quad \text{dla } k = 2, 3, \dots, n, \\ P((n, 1), (0, 0)) &= 1, \\ P((n, k), (j, i)) &= 0, \quad \text{w pozostałych przypadkach.} \end{aligned}$$

Oznaczmy przez τ pierwszy moment powrotu do $(0, 0)$, czyli $\tau := \inf\{n > 0 : X_n = (0, 0)\}$. Zauważmy, że

$$(5.33) \quad \mathbb{P}(\tau = n) = p(n-1) \quad n = 1, 2, \dots$$

Zatem z (5.32) istnieje $\kappa > 1$, spełniająca $\mathbb{E}_{(0,0)}\kappa^T < \infty$. Stąd na mocy Twierdzenia 2.24 mamy, że nasz łańcuch jest geometrycznie ergodyczny, w szczególności istnieje rozkład stacjonarny π . Z Twierdzenia 2.19 oczywistym jest, że $\pi(n, k) = \pi(n, j)$ dla $j, k \in \{1, 2, \dots, n\}$. Jeśli $P^*(x, A)$ jest sprzężonym jądrem, to dla każdego n mamy

$$P^{*n}((n+1, 1), (n+1, n+1)) = 1.$$

Stąd dla każdego n mamy

$$P^n P^{*n}((n+1, n+1), (n+1, n+1)) = 1.$$

Stąd dla każdego n wiemy, że łańcuch Markowa z jądrem przejścia $P^n P^{*n}(x, dy)$ jest przywiedlny, więc na podstawie Twierdzenia 5.5 dla każdego n

$$\|P^n - \Pi\|_{L^2(\pi)} = 1$$

Z wzoru (5.27) otrzymujemy, że $\varrho = 1$, czyli nie istnieje luka spektralna.

Uwaga 5.34. Powyższy przykład jest procesem odnowienia z niewielką modyfikacją powodującą, że powroty do zera dla różnych n mają rozłączne trajektorie. Używając tych samych argumentów możemy skonstruować zwyczajny geometrycznie ergodyczny proces odnowienia nie posiadający luki spektralnej. Wystarczy dobrać tak prawdopodobieństwa $p(n)$, aby były niezerowe tylko dla $n = a_k$ dla pewnego k , gdzie a_n jest ciągiem określonym rekurencyjnie przez $a_n = n + a_{n-1}$ oraz $a_0 = 0$.

Uwaga 5.35. Łańcuch z jądrem sprzężonym do jądra przejścia z Przykładu 5.31 jest również geometrycznie ergodyczny. Stąd wniosek, że geometryczna ergodyczność obu łańcuchów (z jądrem P oraz z jądrem P^* nie gwarantuje istnienia luki spektralnej.

Przedstawimy teraz warunek dryfu implikujący istnienie luki spektralnej. Niestety ten warunek, w przeciwieństwie do warunku 5.14, nie pozwala na jawne obliczenie ϱ , ale spełnia go znacznie szersza klasa łańcuchów.

Definicja 5.36. Powiemy, że łańcuch Markowa spełnia słaby podwójny warunek dryfu, jeśli spełnione są następujące warunki:

1. Istnieją zbiory C i C^* , które są a_n -małe względem jądra $P(x, dy)$ i $P^*(x, dy)$ odpowiednio. Tzn. istnieją $\beta, \beta^* > 0$, miary probabilistyczne ν, ν^* oraz rozkłady prawdopodobieństwa na \mathbb{Z}^+ $(a_n), (a_n^*)$ takie, że

$$\begin{aligned} \sum_{n=1}^{\infty} a_n P^n(x, dy) &\geq \beta \mathbb{I}(x \in C) \nu(dy) \quad \forall x \in \mathcal{X} \\ \sum_{n=1}^{\infty} a_n^* P^{*n}(x, dy) &\geq \beta^* \mathbb{I}(x \in C^*) \nu^*(dy) \quad \forall x \in \mathcal{X} \end{aligned}$$

2. Istnieją funkcje $V, V^* \geq 1$ takie, że $\frac{1}{B}V \leq V^* \leq BV$ dla pewnego $B < \infty$ oraz spełnione są warunki dryfu dla jądra $P(x, dy)$ i $P^*(x, dy)$ odpowiednio. Czyli istnieją stałe $\lambda, \lambda^* < 1$ oraz $K, K^* < \infty$ takie, że

$$PV(x) \leq \begin{cases} \lambda V(x) & \text{dla } x \notin C, \\ K & \text{dla } x \in C \end{cases}$$

oraz

$$P^*V^*(x) \leq \begin{cases} \lambda^* V^*(x) & \text{dla } x \notin C^*, \\ K^* & \text{dla } x \in C^*. \end{cases}$$

Uwaga 5.37. Jeżeli zachodzi podwójny warunek dryfu (definicja 5.14) z $\lambda_1, \lambda_2 < 1$ to implikuje słaby warunek dryfu (definicja 5.36).

Twierdzenie 5.38. *Jeżeli łańcuch Markowa z jądrem przejścia $P(x, dy)$ spełnia słaby podwójny warunek dryfu 5.36, to ma lukę spektralną w $L^2(\pi)$.*

Dowód. Z Twierdzenia 2.24 wiemy, że zarówno łańcuch z jądrem $P(x, A)$ jak i ten z jądrem $P^*(x, A)$ są geometrycznie ergodyczne. Ponadto istnieją funkcje $V_1, V_1^* \geq 1$ oraz n_0, n_1 takie, że

$$(5.39) \quad \|P^{n_0} - \Pi\|_{L_{V_1}^\infty} < 1$$

$$(5.40) \quad \|P^{*n_1} - \Pi\|_{L_{V_1^*}^\infty} < 1$$

Ponadto V_1 i V_1^* są równoważne z V i V^* odpowiednio. Ponieważ V i V^* są równoważne, to V_1 i V_1^* są równoważne. Czyli odpowiadające im normy są równoważne, tzn. istnieje stała $M < \infty$ taka, że $\|\cdot\|_{L_{V_1^*}^\infty} \leq M \|\cdot\|_{L_{V_1}^\infty}$. Z (5.40) otrzymujemy, że istnieje k spełniające

$$(5.41) \quad \|P^{*kn_1} - \Pi\|_{L_{V_1}^\infty} < 1.$$

Weźmy $n = n_0 n_1 k$, wówczas z (5.39) oraz (5.41) otrzymujemy

$$(5.42) \quad \|P^n P^{*n} - \Pi\|_{L_{V_1}^\infty} \leq \|P^{*kn_1} - \Pi\|_{L_{V_1}^\infty}^{n_0} \|P^{n_0} - \Pi\|_{L_{V_1}^\infty}^{kn_1} < 1.$$

Z Twierdzenia 2.24 mamy, że łańcuch z jądrem $P^n P^{*n}(x, dy)$ jest geometrycznie ergodyczny. Z Twierdzenia 5.5 otrzymujemy $\|P^n - \Pi\|_{L^2(\pi)} < 1$, zatem ze stwierdzenia 5.28 otrzymujemy, że łańcuch z jądrem przejścia $P(x, dy)$ ma lukę spektralną. \square

5.3. Funkcjonalne CTG dla funkcji całkowalnych w kwadracie

Wiadomo, że dla odwracalnych geometrycznie ergodycznych łańcuchów Markowa warunkiem dostatecznym, aby zachodziło \sqrt{n} -CTG jest istnienie stacjonarnego drugiego momentu funkcji f ([44]), tzn. $\pi(f^2) < \infty$. Żeby \sqrt{n} -CTG zachodziło dla nieodwracalnych geometrycznie ergodycznych łańcuchów potrzeba, aby istniał wyższy niż drugi moment, tzn. dla pewnego $\delta > 0$ mamy $\pi(f^{2+\delta}) < \infty$. Istnieją przykłady pokazujące, że to założenie jest istotne [7] oraz [20]. Okazuje się również, że dla łańcucha z Przykładu 5.31 można bardzo prosto skonstruować funkcję f taką, że $\pi(f^2) < \infty$ oraz nie zachodzi \sqrt{n} -CTG. Ten przykład jest prosty i opiera się tylko na podstawowych własnościach łańcuchów Markowa.

Przykład 5.43. Niech $(X_n)_{n \geq 0}$ będzie łańcuchem Markowa zdefiniowanym w Przykładzie 5.31 z rozkładem $p(n) = \frac{1}{2}^n$. Zdefiniujmy funkcję f następująco:

$$\begin{aligned} f(n, k) &= \frac{2^{n/2}}{n^{3/2}} \quad \text{dla } k = 1, 2, \dots, n \\ f(i, j) &= 0 \quad \text{w pozostałych przypadkach} \end{aligned}$$

Najpierw pokażmy, że $\pi(f^2) < \infty$. Z Wniosku 2.20 mamy

$$(5.44) \quad \pi(f^2) = \pi(0) \mathbb{E}_{(0,0)} \sum_{i=0}^{\tau-1} f^2(X_i)$$

Zauważmy, że $\tau = n$ tylko wtedy, gdy $X_1 = (n-1, n-1)$, czyli

$$\begin{aligned}
(5.45) \quad \mathbb{E}_{(0,0)} \sum_{i=0}^{\tau-1} f(X_i) &= \sum_{n=1}^{\infty} \sum_{i=1}^{n-1} f^2(n-1, n-1-i) p(n-1) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{n-1} \frac{2^n}{n^3} 2^{-n+1} \\
&= \sum_{n=1}^{\infty} \frac{2(n-1)}{n^3} < \infty
\end{aligned}$$

Zatem z (5.44) oraz (5.45) otrzymujemy $\pi(f^2) < \infty$. Teraz wykażemy, że dla tej funkcji nie jest spełnione CTG. Z Twierdzenia 2.21 wiemy, że warunkiem koniecznym i dostatecznym, aby spełnione było CTG jest $\mathbb{E}_{(0,0)} \left[\left(\sum_{i=1}^{\tau} f(X_i) - \pi(f) \right)^2 \right] < \infty$. Policzmy to wyrażenie

$$(5.46) \quad \mathbb{E}_{(0,0)} \left[\left(\sum_{i=1}^{\tau} f(X_i) - \pi(f) \right)^2 \right] = \mathbb{E}_{(0,0)} \left[\left(\sum_{i=1}^{\tau} f(X_i) \right)^2 \right] + 2\pi(f) \mathbb{E}_{(0,0)} \left[\tau \sum_{i=1}^{\tau} f(X_i) \right] + [\pi(f)]^2 \mathbb{E}_{(0,0)} \tau^2$$

Trzeci składnik jest w sposób oczywisty skończony. Musimy sprawdzić pierwsze i drugie wyrażenie. Zaczniemy od drugiego, $\pi(|f|) < \infty$ oraz

$$\begin{aligned}
(5.47) \quad \mathbb{E}_{(0,0)} \left[\tau \sum_{i=1}^{\tau} f(X_i) \right] &= \sum_{n=1}^{\infty} n \sum_{i=1}^{n-1} f(n-1, n-i) p(n-1) \\
&= \sum_{n=1}^{\infty} n \sum_{i=1}^{n-1} \frac{2^{n/2}}{n^{3/2}} 2^{-n+1} \\
&= \sum_{n=1}^{\infty} \frac{2n(n-1)}{n^{3/2}} 2^{-n/2} < \infty
\end{aligned}$$

Pozostała nam pierwsza część prawej strony (5.46)

$$(5.48) \quad \mathbb{E}_{(0,0)} \left[\left(\sum_{i=1}^{\tau} f(X_i) \right)^2 \right] = \mathbb{E}_{(0,0)} \sum_{i=1}^{\tau} f^2(X_i) + 2\mathbb{E}_{(0,0)} \sum_{i < j}^{\tau} f(X_i) f(X_j)$$

Z (5.45) wiemy, że $\mathbb{E}_{(0,0)} \sum_{i=1}^{\tau} f^2(X_i) < \infty$, natomiast

$$\begin{aligned}
(5.49) \quad \mathbb{E}_{(0,0)} \sum_{i < j}^{\tau} f(X_i) f(X_j) &= \sum_{n=1}^{\infty} \sum_{i=1}^{n-1} \sum_{j=i}^n f(n-1, n-i) f(n-1, n-j) p(n-1) \\
&= \sum_{n=1}^{\infty} \sum_{i=1}^{n-2} \sum_{j=i}^{n-1} \frac{2^n}{n^3} 2^{-n+1} \\
&= \sum_{n=1}^{\infty} \frac{(n-1)(n-2)}{n^3} = \infty
\end{aligned}$$

Czyli z (5.46), (5.47), (5.48), (5.49) otrzymujemy, że

$$\mathbb{E}_{(0,0)} \left[\left(\sum_{i=1}^{\tau} f(X_i) - \pi(f) \right)^2 \right] = \infty,$$

czyli na mocy Twierdzenia 2.21 nie zachodzi CTG dla funkcji f i łańcucha X_n .

Powyższy przykład sugeruje, że istnieje związek pomiędzy istnieniem luki spektralnej a CTG dla funkcji całkowalnych w kwadracie. Rzeczywiście tak jest, w [28, dowód Twierdzenia 1.4] wykazano (wcześniej sformułowany w [15, Twierdzenie 4.11]) fakt, że dla łańcuchów z luką spektralną w $L^2(\pi)$ zachodzi CTG dla wszystkich funkcji całkowalnych w kwadracie. My przedstawimy dowód nieco ogólniejszego rezultatu.

Twierdzenie 5.50. *Jeśli łańcuch Markowa $(X_n)_{n \geq 0}$ z jądrem przejścia $P(x, dy)$ i rozkładem stacjonarnym π ma lukę spektralną $1 - \varrho$ w $L^2(\pi)$, to dla każdej funkcji f takiej, że $\pi(f^2) < \infty$ zachodzi funkcyjalne CTG:*

$$(n\sigma_{\text{as}}^2(f))^{-\frac{1}{2}} s_n(t) \xrightarrow{d} W(t),$$

gdzie $s_n(t) = Z_{\lfloor nt \rfloor} + (nt - \lfloor nt \rfloor)[Z_{\lfloor nt \rfloor + 1} - Z_{\lfloor nt \rfloor}]$, $Z_n := \sum_{i=0}^{n-1} \bar{f}(X_i)$ oraz $W(t)$ jest standardowym procesem Wienera na $[0, 1]$. Ponadto

$$(5.51) \quad \sigma_{\text{as}}^2(f) = \pi(\bar{f}^2) + 2\mathbb{E}_\pi \sum_{n=1}^{\infty} \bar{f}(X_0)\bar{f}(X_n)$$

Dowód. Jeśli istnieje luka spektralna to $1 \notin \sigma(P - \Pi)$, zatem z definicji spektrum operator $\mathbb{I} - P + \Pi$ jest odwracalny w $L^2(\pi)$. Stąd dla każdego $f \in L^2(\pi)$ istnieje $\tilde{f} \in L^2(\pi)$ takie, że

$$\tilde{f} - P\tilde{f} + \pi(\tilde{f}) = f.$$

Ponieważ $\pi(f) = \pi(\tilde{f})$ więc $\hat{f} := \tilde{f} - \pi(f)$ spełnia równanie Poissona (2.26). Stąd korzystając z Twierdzenia 2.27 otrzymujemy, że zachodzi funkcyjalne CTG przy czym, $\sigma_{\text{as}}^2(f) := \pi[\hat{f}^2 - (P\hat{f})^2]$. Jeżeli istnieje całkowalne względem π rozwiązanie równania Poissona to jest jedyne z dokładnością do stałej [40], zatem $\hat{f} = \sum_{n=0}^{\infty} P^n \bar{f}(X_i)$. Podstawiamy $\hat{f} - \bar{f}$ za $P\hat{f}$ i otrzymujemy

$$\sigma_{\text{as}}^2(f) = \pi \left[\hat{f}^2 - (\hat{f} - \bar{f})^2 \right] = \pi \left(-\bar{f}^2 + 2\pi(\bar{f}\hat{f}) \right).$$

Co daje (5.51). □

Uwaga 5.52. W przypadku odwracalnym asymptotyczną wariancję można oszacować przez

$$\sigma_{\text{as}}^2(f) \leq \frac{1 + \varrho}{1 - \varrho} \pi(\bar{f}^2),$$

gdzie $1 - \varrho$ jest luką spektralną [16, 25]. W przypadku nieodwracalnym otrzymujemy podobne oszacowanie. Weźmy γ i M ze Stwierdzenia 5.28, wówczas korzystając z twierdzenia Fubiniiego otrzymujemy

$$\sigma_{\text{as}}^2(f) = \pi(\bar{f}^2) + 2 \sum_{n=1}^{\infty} \mathbb{E}_\pi \bar{f}(X_0)\bar{f}(X_n) \leq \pi(\bar{f}^2) M \left(1 + 2 \sum_{n=1}^{\infty} \gamma^n \right) = \pi(\bar{f}^2) M \frac{1 + \gamma}{1 - \gamma}.$$

W szczególności jeśli spełniony jest warunek (4.1) mamy

$$\sigma_{\text{as}}^2(f) \leq \frac{1 + \rho}{1 - \rho} \pi(\bar{f}^2),$$

ale $\rho := \|P - \Pi\|_{L^2(\pi)}$ nie musi być równe ϱ .

Bibliografia

- [1] R. Adamczak, W. Bednorz, *Exponential Inequalities for Markov Chains*, w przygotowaniu.
- [2] D. Aldous, *On the Markov Chain Simulation Method for Uniform Combinatorial Distributions and Simulated Annealing*, Probability in the Engineering and Informational Sciences 1, 33-46, 1987.
- [3] Y.F. Atchade, F. Perron, *On the geometric ergodicity of Metropolis-Hastings algorithms*, Statistics 41, 77–84, 2007.
- [4] K.B. Athreya and P. Ney, *A new approach to the limit theory of recurrent Markov chains*, Trans. Amer. Math. Soc. 245, 493–501, 1978.
- [5] P.H. Baxendale, *Renewal Theory and Computable Convergence Rates for Geometrically Ergodic Markov Chains*, Ann. Appl. Probab. 15, 700–738, 2005.
- [6] W. Bednorz, R. Latała and K. Łatuszyński, *A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains*, Elect. Comm. in Probab. 13, 85–98, 2008.
- [7] R.C. Bradley, Jr. *Information regularity and the central limit question*. Rocky Mountain J. Math., 13(1), 77–97, 1983.
- [8] G. Casella, C. P. Robert, *Monte Carlo Statistical Methods*. Springer-Verlag, New York, 1999.
- [9] K. S. Chan, H. Yue, *Asymptotic Efficiency of the Sample Mean in Markov Chain Monte Carlo Schemes*, Journal of the Royal Statistical Society, Series B. 58 (3), 525-539, 1996.
- [10] J.T. Chang *Inequalities for the overshoot*. Ann. Appl. Probab. 4, 1223–1233, 1994.
- [11] I. H. Dinwoodie, *A probability inequality for the occupation measure of a reversible Markov chain*, Ann. Appl. Probab. 5 37–43. 1995
- [12] R. Douc, E. Moulines, J. S. Rosenthal, *Quantitative bounds on convergence of time-inhomogeneous Markov Chains*. Ann. Appl. Probab. 14, 1643-1665, 2003.
- [13] W. Feller, *Wstęp do rachunku prawdopodobieństwa, I*, PWN Warszawa, 1977.
- [14] G. Fort, E. Moulines, *Convergence of the Monte Carlo expectation maximization for curved exponential families*, Ann. Statist. Volume 31, Number 4 (2003), 1220–1259, 2003.
- [15] C. J. Geyer, *Markov Chain Monte Carlo Lecture Notes*, 2005.

- [16] C. J. Geyer, *Practical Markov Chain Monte Carlo*. Stat. Sci. 7 (4), 473-511, 1992.
- [17] D. Gillman, *A Chernoff Bound for Random Walks on Expander Graphs*. SIAM J. Comput. 27 (4), 1203-1220, 1998.
- [18] P. W. Glynn, D. Ormoneit, *Hoeffding's inequality for uniformly ergodic Markov chains*. Statist. Probab. Lett., 56(2):143-146, 2002.
- [19] J. P. Hobert, C.J. Geyer, *Geometric Ergodicity of Gibbs and Block Gibbs Samplers for a Hierarchical Random Effects Model*, J. Multiv. Analysis 67, 414-430, 1998.
- [20] O. Häggström, *On the Central Limit Theorem for Geometrically Ergodic Markov Chains*, Probability Theory and Related Fields, 132, 74-82, 2005.
- [21] W. Hoeffding, *Probability inequalities for sums for bounded random variables*. J. Amer. Statist. Assoc. 58 13-30, 1963.
- [22] G.L. Jones, J.P. Hobert, *Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random Effects Model*. The Annals of Statistics 32 (2), 784-817, 2004.
- [23] G. L. Jones, M. Haran, B. S. Caffo, R. Neath, *Fixed-Width Output Analysis for Markov Chain Monte Carlo*, Journal of the American Statistical Association, 101, 1537-1547, 2006.
- [24] G.L Jones, A.A. Johnson, *Gibbs Sampling for a Bayesian Hierarchical General Linear Model*, arXiv: 0712.3056v5, 2010.
- [25] C. Kipnis, S.R.S. Varadhan, *Central Limit Theorem for Additive Functionals of Reversible Markov Processes and Applications to Simple Exclusions*, Commun. Math. Phys. 104, 1-19, 1986.
- [26] I. Kontoyiannis, L. Lastras-Montano, S. P. Meyn, *Relative Entropy and Exponential Deviation Bounds for General Markov Chains*, 2005 IEEE International Symposium on Information Theory, 2005.
- [27] I. Kontoyiannis, S.P. Meyn *Large deviation asymptotics and the spectral theory of multiplicatively regular Markov processes*, Electron. J. Probab., 10(3):61-123, 2005.
- [28] I. Kontoyiannis, S.P. Meyn *Geometric ergodicity and the spectral gap of non-reversible Markov chains.*, preprint
- [29] K. Łatuszyński, *Regeneration and Fixed-Width Analysis of Markov Chain Monte Carlo Algorithms*. PhD Dissertation, 2008. Available at arXiv:0907.4716v1
- [30] K. Łatuszyński, W. Niemirow, *$(\varepsilon - \alpha)$ -MCMC approximation under drift condition*, in: Proceedings of the 6th International Workshop on Rare Event Simulation (RESIM 2006), 2006.
- [31] K. Łatuszyński, W. Niemirow, *Rigorous confidence bounds for MCMC under a geometric drift condition*, 2010, ukaże się w J. Complex. arXiv:0908.2098
- [32] K.Łatuszyński, B.Miasojedow, W. Niemirow, *Nonasymptotic bounds on the estimation error for regenerative MCMC algorithms*, dostępna w arXiv:0907.4915v1 .

- [33] C. A. León, F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains*. Ann. Appl. Probab. Volume 14, Number 2, 958-970, 2004.
- [34] P. Lézaud, *Chernoff-type bound for finite Markov chains*. Ann. Appl. Probab. 8 849-867, 1998.
- [35] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Springer, 2001.
- [36] G. Lorden, *On excess over the boundary*. Ann. Math. Statist. 41, 520–527, 1970.
- [37] R.B. Lund, R.L. Tweedie, *Geometric convergence rates for stochastically ordered Markov chains*, Mathematics of Operations Research 21, 182–194, 1996.
- [38] K.L. Mengersen, L.R. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, Ann. Statist. 24, 1, 101–121, 1996.
- [39] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *Equations of state calculations by fast computing machines*. J. Chem. Phys. 21, 1087-1091, 1953.
- [40] S.P. Meyn, R.L. Tweedie: *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.
- [41] W. Niemi, P. Pokarowski, *Fixed Precision MCMC Estimation by Median of Products of Averages*. Journal of Applied Probability, 46(2), 309-329, 2009.
- [42] E. Nummelin, *A splitting technique for Harris recurrent Markov chains*, Z. Wahr. Verw. Geb. 43, 309–318, 1978.
- [43] E. Nummelin, *MC's for MCMC'ists*, International Statistical Review, 70, 215–240, 2002.
- [44] G.O. Roberts, J.S. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Elec. Comm. Prob. 2 (2), 1997.
- [45] G.O. Roberts, J.S. Rosenthal, *General state space Markov chains and MCMC algorithms*, Probability Surveys 1, 20–71, 2004.
- [46] D. Rudolf, *Explicit error bounds for lazy reversible Markov chain Monte Carlo*, J. of Complexity. 25, 11–24, 2008.
- [47] G. O. Roberts, R. L. Tweedie, *Bounds on Regeneration Times and Convergence Rates for Markov Chains*, Stochastic Process. Appl. 91, 337-338, 1999.
- [48] J. S. Rosenthal, *Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo*, Journal of the American Statistical Association, 90, 558-566, 1995.
- [49] V. Roy, J.P. Hobert, *On Monte Carlo methods for Bayesian multivariate regression models with heavy-tailed errors*, Journal of Multivariate Analysis 101 1190–1202, 2010.