

DUŻE ODCHYLENIA

Notatki do wykładu (wersja 0.45)

PIOTR DYSZEWSKI

2023

Musiabym byc cięzko chory,
zeby pójść na ten numer.
(Po chwili): A wiesz, że ostatnio
coś kiepsko się czuję?

Vabank

Teoria wielkich odchyłeń

Dane dotyczące przedmiotu

Nazwa przedmiotu: Teoria wielkich odchyłeń

Jednostka oferująca przedmiot: Instytut Matematyczny

Założenia: Teoria prawdopodobieństwa 2 (28-MT-S-tTPrawd2)

Strona www: <https://sites.google.com/site/piotrdyszewski/teaching/duÅije-odchylenia>

Forma zajęć: wykład + ćwiczenia

Punkty ECTS: 6

Opis przedmiotu

Celem wykładu jest prezentacja probabilistycznej teorii analizy zdarzeń rzadkich. Zaczniemy od opisu ich prawdopodobieństw. Następnie zbadamy zachowanie miar probabilistycznych warunkowanych zdarzeniami rzadkimi. Pozwala to zrozumieć, skąd biorą się duże odchylenia. Zaczniemy od ugruntowania intuicji poprzez przeliczenie kilku konkretnych przykładów i zbadania zagadnienia na prostej. Pozwoli nam to zaprezentować ogólną teorię wielkich odchyłeń w dalszej części wykładu.

Teoria wielkich odchyłeń jest działem probabilistyki opisującym zdarzenia rzadkie. Interesować nas będzie opis ich prawdopodobieństw, ich pochodzenie oraz ich konsekwencje. Przykładowo, z podstawowego kursu rachunku prawdopodobieństwa wiemy, że dla ciągu niezależnych zmiennych o tym samym rozkładzie X_1, X_2, \dots takich, że $\mathbb{E}[X_1] < \infty$ zachodzi mocne prawo wielkich liczb

$$S_n/n = (X_1 + \dots + X_n)/n \rightarrow \mathbb{E}[X_1] \quad p.w.$$

Oznacza to, że dla każdego $a > \mathbb{E}[X_1]$,

$$\mathbb{P}[S_n > an]$$

IV

zbiega do zera wykładniczo szybko. Innymi słowy $\{S_n > an\}$ jest zdarzeniem rzadkim dla $a > \mathbb{E}[X_1]$. W tym konkretnym ujęciu teoria wielkich odchyłeń odpowiada na pytanie o to, jak szybka jest to zbieżność, t.j. opisuje funkcję tempa I dla której

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}[S_n > an] = I(a).$$

Powyższą zbieżność można zapisać nieformalnie jako

$$\mathbb{P}[S_n > an] \approx e^{-I(a)n}, \quad a > \mathbb{E}[X_1].$$

Chcąc zbadać skąd biorą się rozważane przez nas rzadkie zdarzenia rozważmy ścieżkę S_n zadaną przez $Z_n: [0, 1] \rightarrow \mathbb{R}$,

$$Z_n(t) = (S_{[tn]} + \{tn\}X_{[tn]+1})/n.$$

Okazuje się, że wówczas dla pewnej funkcji I ,

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P}[Z_n \in A] = \inf_{\phi \in A} \int_0^1 I(\phi'(t)) dt.$$

Oznacza to, że jeżeli $S_n > an$ to ścieżka $Z_n(t) = \phi(t)$ minimalizuje

$$\int_0^1 I(\phi'(t)) dt$$

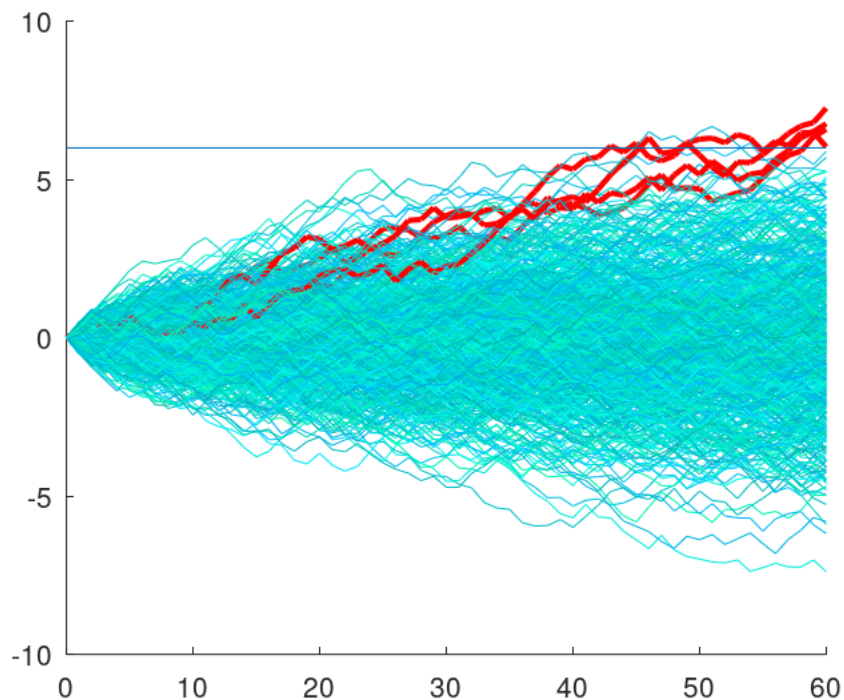
wśród funkcji $\phi: [0, 1] \rightarrow \mathbb{R}$ takich, że $\phi(1) > a$.

W trakcie wykładu opiszemy ogólną teorię stojącą za powyższym przykładem. Zbadanie teorii wielkich odchyłeń na przestrzeniach polskich pozwoli nam na analizę zdarzeń rzadkich w dwóch różnych kontekstach: grafów losowych i równań różniczkowych cząstkowych.

Dla ustalonego $n \in \mathbb{N}$ rozważmy zbiór wierzchołków $V = \{1, \dots, n\}$. Dla ustalonego $p \in (0, 1)$ chcemy badać graf losowy w którym krawędź między wierzchołkami jest rysowana z prawdopodobieństwem p . Problem, którym będziemy chcieli się zająć polega na zliczaniu pewnych określonych struktur (podgrafów) w naszym losowym grafie. Aby nie zamęczać się technicznymi szczegółami ograniczymy się do zliczania trójkątów. Uściślając, dla $j, k \in V$ niech zmienna $X_{j,k}$ będzie dana przez

$$X_{k,j} = \begin{cases} 1 & k \sim j \\ 0 & \text{poza tym} \end{cases}$$

Tutaj $j \sim k$ piszemy, gdy między wierzchołkami j i k jest krawędź. Zgodnie z naszym opisem zmienne $\{X_{k,j}\}_{k < j}$ są iid z rozkładem



Rysunek 0.1. 600 trajektorii spaceru losowego z krokami o rozkładzie jednostajnym na $(-1/2, 1/2)$. Czerwonym kolorem zostały oznaczone spacery S_n dla których $S_{60} > 6$.

$$\mathbb{P}[X_{k,j} = 1] = p.$$

Wówczas rozważana przez nas liczba trójkątów zapisuje się jako

$$T_n = \sum_{1 \leq k \leq j \leq i \leq n} X_{k,j} X_{j,i} X_{i,j}.$$

Z liniowości wartości oczekiwanej widzimy, że

$$\mathbb{E}[T_n] = \binom{n}{3} p^3 \sim p^3 n^3 / 6.$$

Okazuje się, że liczba trójkątów spełnia mocne prawo wielkich liczb

$$T_n / \mathbb{E}[T_n] \rightarrow 1.$$

Zbadamy, jakie jest prawdopodobieństwo, że T_n jest dalekie od swojej średniej. Okazuje się bowiem, że te prawdopodobieństwa maleją bardzo szybko

$$\mathbb{P}[T_n > (1 + \delta)\mathbb{E}[T_n]] \approx e^{-I(\delta)n^2}.$$

VI

W szczególności zastanowimy się, jaki wyglądają grafy losowe, w których jest więcej niż $(1 + \delta)\mathbb{E}[T_n]$ trójkątów.

Nasz ostatni przykład wiąże się z równaniami różniczkowymi cząstkowymi. Niech $G \subseteq \mathbb{R}^d$ będzie ograniczonym, spójnym otwartym zbiorem z gładkim brzegiem ∂G . Dla $\epsilon > 0$ rozważmy zagadnienie Dirichleta

$$\begin{aligned}\frac{\epsilon}{2}\Delta u + b \cdot u &= 0, & \text{w } G \\ u(y) &= f(y) & y \in \partial G.\end{aligned}$$

Tutaj $b: G \rightarrow \mathbb{R}^d$ jest postaci $b = -\nabla V$ dla pewnej gładkiej funkcji V , a \cdot oznacza iloczyn skalarny w \mathbb{R}^d . Interesować nas będzie zachowanie rozwiązania $u = u_\epsilon$ powyższego zagadnienia, gdy $\epsilon \rightarrow 0$. Okazuje się, że kluczowa jest asymptotyka rozwiązań równania różniczkowego zwyczajnego

$$\frac{dx}{dt} = b(x(t)).$$

W sytuacji, kiedy ostatnie zagadnienie ma jedyne rozwiązanie stacjonarne w G przyciągające wszystkie trajektorie zaczęte w G , rozwiązania u_ϵ zagadnienia Dirichleta przejawiają osobliwe zachowanie. Mianowicie

$$\lim_{\epsilon \rightarrow 0^+} u_\epsilon(x) = f(y_0),$$

gdzie y_0 jest punktem, w którym V przyjmuje swoje minimum na ∂G . Okazuje się, że wytłumaczenia powyższego zjawiska należy szukać właśnie w teorii wielkich odchyłeń dla probabilistycznej reprezentacji $u_\epsilon(x)$.

Skrócony plan wykładu

Do opanowania podstaw wykładu niezbędny będzie rachunek prawdopodobieństwa oraz znajomość ruchu Browna. Bardzo przydatna będzie znajomość podstawowych terminów z topologii. Poruszymy zagadnienia takie jak:

- A) Odchylenia na prostej:** twierdzenie Sanova dla skończonych alfabetów, twierdzenie Cramera na prostej;
- B) Ogólna teoria:** Twierdzenie Sanova na przestrzeniach polskich, reguła kontrakcji, lemat Varadhana, minimalizacja entropii;
- C) Duże odchylenia dla ścieżek:** Twierdzenie Cramera na przestrzeniach Banacha, Twierdzenie Schindlera, prawo iterowanego logarytmu Strassena
- D) Zastosowania:** testowanie hipotez, grafy losowe, teoria Freidlina-Wentzella i zagadnienie wyjścia

Podstawową literaturą do wykłady są:

- Frank den Hollander, *Large deviations*, 1999
- Amir Dembo, Ofer Zeitouni, *Large Deviations Techniques and Applications*, Springer New York, NY
- Sourav Chatterjee. *Large Deviations for Random Graphs*, 2015.
- Varadhan, S. R. S. *Large Deviations*, Courant Lecture Notes in Mathematics, 2016

Na literaturę uzupełniającą składają się

- Jean-Dominique Deuschel, Daniel Stroock, *Large Deviations*, Academic Press, Boston 1989.
- Feng, Jin, and Thomas G. Kurtz. *Large deviations for stochastic processes*. No. 131. American Mathematical Soc., 2006.
- Dupuis, Paul, and Richard S. Ellis. *A weak convergence approach to the theory of large deviations*. John Wiley & Sons, 2011.
- Ellis, Richard S. *Entropy, large deviations, and statistical mechanics*. Springer, 2007.
- Varadhan, SR Srinivasa. *Large deviations and applications*. Society for Industrial and Applied Mathematics, 1984.
- Saulis, Leonas, and V. A. Statulevicius. *Limit theorems for large deviations*. Vol. 73. Springer Science & Business Media, 2012.
- Daniel W. Stroock *An introduction to the theory of large deviations*, Springer, 1984

Szczegółowy plan wykładu

Wstępny plan tematów poruszanych na poszczególnych wykładach

1. Problematyka teorii wielkich odchyłeń, rzuty moneta
2. Twierdzenie Sanova dla skończonego alfabetu, reguła kontrakcji
3. Twierdzenie Craméra na prostej
4. Rodziny wykładnicze, transformata Legendre'a
5. Zasada wielkich odchyłeń na przestrzeniach polskich, entropia
6. Twierdzenie Sanova na przestrzeniach polskich
7. Twierdzenie Craméra w przestrzeniach Banacha
8. Duże odchylenia na przestrzeni $C[0, 1]$
9. Twierdzenie Schildera i prawo iterowanego logarytmu Strassena
10. Lemat Varadhana
11. Warunkowanie zdarzeniami rzadkimi, minimalizacja entropii
12. Duże odchylenia dla zmiennych zależnych
13. Grafy losowe: preliminaria
14. Grafy losowe: duże odchylenia
15. Teoria Freidlina-Wentzella, problem wyjścia

Efekty kształcenia

Po wykładzie student:

1. formułuje podstawowe twierdzenia z zakresu teorii wielkich odchyień na prostej (**A**);
2. formułuje prawa iterowanego logarytmu dla ruchu Browna (**C**);
3. analizuje dowody prostych twierdzeń z wykładu z uzasadnieniem poszczególnych ich założeń (**A**);
4. formułuje główne twierdzenia teorii wielkich odchyień na przestrzeniach polskich(**B**);
5. analizuje dowody najważniejszych twierdzeń z wykładu z uzasadnieniem poszczególnych ich założeń (**B, C**);
6. Stosuje teorię wielkich odchyień w przykładach (**D**);

Sposób weryfikacji efektów kształcenia

Na zaliczenie składać się będą:

- aktywność na ćwiczeniach;
- zadania domowe

Metody i kryteria oceniania

Zaliczenie ćwiczeń na podstawie zadań domowych i aktywności w czasie zajęć. Ocena z egzaminu wystawiona jest na podstawie egzaminu ustnego. Warunkiem zaliczenia przedmiotu jest:

- uzyskanie 30% punktów za zadania stanowiące bieżącą weryfikację efektów kształcenia;
- uzyskanie pozytywnej oceny z egzaminu stanowiącego końcową weryfikację efektów kształcenia.

Kryteria ocen:

(dst) student realizuje punkty 1-2 efektów kształcenia

(db) student realizuje punkty 1-4 efektów kształcenia

(bdb) student realizuje punkty 1-6 efektów kształcenia

Spis treści

1	Wstęp	1
1.1	Czym są duże odchylenia?	1
1.2	Rzuty monetą	2
1.3	Twierdzenie Sanova dla skończonego alfabetu	6
1.4	Zasada kontrakcji	10
1.5	Zadania	11
2	Duże odchylenia na prostej	15
2.1	Twierdzenie Craméra	15
2.2	Serie w spacerach losowych	24
3	Ogólna teoria	30
3.1	Lemat Varadhana	35
3.2	Duże odchylenia dla zamiany miary	39
3.3	Reguła kontrakcji	40
3.4	Twierdzenie Sanova	40
3.5	Elementy rachunku prawdopodobieństwa na przestrzeniach Banacha	50
3.5.1	Całka Bochnera	50
3.6	Duże odchylenia na przestrzeniach Banacha	52
3.7	Minimalizacja Entropii	59
3.8	Zadania	64
4	Duże odchylenia dla funkcji ciągłych	65
5	Zmienne zależne	76
5.1	Twierdzenie Gärtnera-Ellisa	76
5.2	Łańcuchy Markowa	85
6	Dyfuzje	89
6.1	Dyfuzje	91
6.2	Problem wyjścia	93

X	Spis treści	
	6.3 Własność Markowa.....	94
	6.4 Problem wyjścia 2	96
	Literatura	99

Wstęp

Streszczenie Zaczniemy od krótkiego opisu problematyki wielkich odchyłeń. Następnie zdobędziemy wstępne intuicje poprzez bezpośrednie wyliczenia dla kilku przykładów.

Teoria wielkich odchyłeń zajmuje się asymptotycznym opisem zdarzeń rzadkich. To, czym jest zdarzenie rzadkie, oraz to, co rozumiemy przez jego asymptotykę, zależy od kontekstu rozważań. W trakcie wykładu postaramy się omówić spacery losowe, grafy losowe oraz procesy dyfuzji. Aby zdobyć podstawowe intuicje skupimy się na pierwszym modelu, który jest niczym innym jak sumą niezależnych zmiennych.

1.1 Czym są duże odchylenia?

Rozważmy przestrzeń probabilistyczną $(\Omega, \mathcal{F}, \mathbb{P})$ na której określone są niezależne zmienne $\{X_k\}_{k \in \mathbb{N}}$, $\mathbb{N} = \{0, 1, 2, \dots\}$ o tym samym rozkładzie, takie, że

$$\mathbb{E}[X_k] = m < \infty, \quad \text{Var}[X_k] = \mathbb{E}[X_k^2] - \mathbb{E}[X_k]^2 = \sigma^2 < \infty.$$

Dla $n \in \mathbb{N}$ rozważmy $S_0, S_n = \sum_{k=1}^n X_k$. Wówczas z podstawowego kursu rachunku prawdopodobieństwa wiemy, że zachodzi mocne prawo wielkich liczb

$$S_n/n \rightarrow m \quad p.w.$$

które może być luźno sformułowane jako „dla dużych n , S_n wynosi w przybliżeniu mn ”. Znamy też centralne twierdzenie graniczne,

$$(S_n - mn)/\sqrt{n} \Rightarrow \mathcal{N}(0, \sigma), \quad (1.1)$$

gdzie \Rightarrow oznacza zbieżność według rozkładu zmiennych losowych. Przez $\mathcal{N}(0, 1)$ oznaczamy rozkład normalny na \mathbb{R} o gęstości danej wzorem

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Zbieżność (1.1) oznacza, że błąd wspomnianego przybliżenia, czyli $S_n - mn$ jest typowo rzędu \sqrt{n} . Jeżeli natomiast $S_n - mn$ jest rzędu większego niż \sqrt{n} takie odchylenia należy uznać za nietypowe. Gdy $S_n - mn$ jest dużo większe niż \sqrt{n} ale mniejsze niż n , będziemy mówili o umiarkowanych odchyleniach. Dla ustalenia uwagi w trakcie wykładu skupimy się na sytuacji, gdy $S_n - mn$ jest rzędu n . Odchylenia tego typu nazywamy dużymi. Przykładem takiego odchylenia jest zdarzenie

$$\{S_n \geq an\}, \quad a > m.$$

Z mocnego prawa wielkich liczb wiemy, że prawdopodobieństwo tego zdarzenia dąży do zera. Naszym celem będzie określenie jak szybko zachodzi ta zbieżność. Zazwyczaj do tego aby $S_n > an$ potrzeba aby dla liniowo wielu k (rzędu n), X_k musi być istotnie większe niż średnia m . To z kolei sprawia, że prawdopodobieństwo tego, że $S_n > an$ maleje wykładniczo szybko, czyli

$$\mathbb{P}[S_n > an] \approx e^{-I(a)n}, \quad a > m.$$

Interesować nas będzie wykładnik tej zbieżności, który ściśle zapisuje się jako

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq an] = -I(a), \quad a > m.$$

Funkcję I charakteryzuje twierdzenie Cramera. Jednak zanim do niego przejdziemy sprawdzimy jak się zachowuje S_n w przypadku rzutów symetryczną monetą.

1.2 Rzuty monetą

Aby nabrać intuicji rozważmy przykład związany z rzutami monetą. Rozważać będziemy zmienne $\{X_k\}_{k \in \mathbb{N}}$ o rozkładzie

$$\mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = 1/2.$$

Wówczas dla $S_n = \sum_{k=1}^n X_k$ mamy $S_n/n \rightarrow \mathbb{E}[X_1] = 1/2$. Badanie zdarzeń rzadkich, $\{S_n > an\}$ dla $a > 1/2$, jest w tym przypadku szczególnie przystępne, ponieważ zmienna S_n ma rozkład Bernoullego.

Twierdzenie 1.1

Rozważmy zmienne losowe $\{X_k\}_{k \in \mathbb{N}}$, które są iid z rozkładem

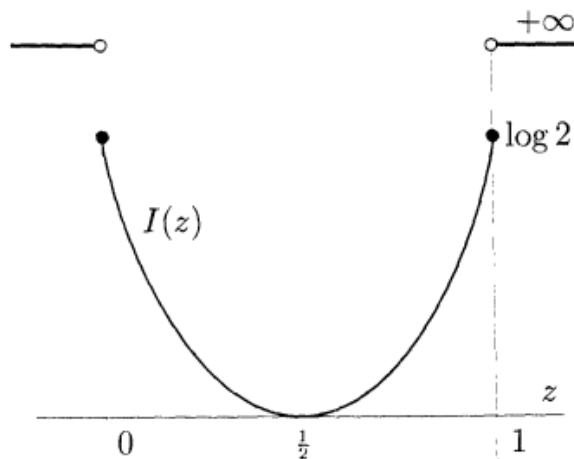
$$\mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = 1/2.$$

Niech $S_n = \sum_{k=1}^n X_k$. Wówczas dla wszystkich $a > 1/2$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq an] = -I(a),$$

gdzie

$$I(z) = \begin{cases} \log(2) + z \log(z) + (1-z) \log(1-z), & z \in [0, 1] \\ +\infty, & \text{poza tym} \end{cases}$$



Rysunek 1.1. Wykres funkcji I w Twierdzeniu 1.1

W dowodzie Twierdzenia 1.1 skorzystamy ze wzoru Stirlinga.

Zadanie

Pokaż, że dla każdego dodatniego $n \in \mathbb{N}$,

$$n \log(n) - n + 1 \leq \log(n!) \leq (n+1) \log(n+1) - n.$$

Wywnioskuj słabą wersję wzoru Stirlinga

$$\log(n!) = n \log(n) - n + O(\log(n)).$$

Dowód Twierdzenia 1.1. Dla $a > 1$ teza jest oczywista. Rozważmy więc $a \in (1/2, 1]$. Mamy

$$\mathbb{P}[S_n \geq an] = 2^{-n} \sum_{k \geq an} \binom{n}{k}.$$

Niech

$$Q_n(a) = \max_{k \geq an} \binom{n}{k} = \binom{n}{\lceil an \rceil}.$$

Korzystając ze słabej wersji wzoru Stirlinga,

$$\log(Q_n(a)) = n \log(n) - an \log(an) - (1-a)n \log((1-a)n) + O(\log(n))$$

otrzymujemy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log Q_n(a) = -a \log(a) - (1-a) \log(1-a).$$

Korzystając teraz z nierówności

$$2^{-n} Q_n(a) \leq \mathbb{P}[S_n \geq an] \leq n 2^{-n} Q_n(a)$$

otrzymujemy tezę. □

Uwaga 1.2

Przez symetrię, dla $a < 1/2$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \leq an] = -I(a).$$

Co z kolei pociąga dla $\delta > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[|S_n/n - 1/2| > \delta] = -I(1/2 + \delta) < 0.$$

Idąc dalej

$$\sum_{n=1}^{\infty} \mathbb{P}[|S_n/n - 1/2| > \delta] < \infty.$$

Co po odwołaniu się do Lematu Borela-Cantelliego i przejściu granicznym z δ do zera daje niezależny dowód $S_n/n \rightarrow 1/2$ p.w.

Zadanie

Pokaż, że dla zmiennej losowej N o standardowym rozkładzie normalnym i dodatniego x zachodzą nierówności

$$\sqrt{2\pi} e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3} \right) \leq \mathbb{P}[N > x] \leq \sqrt{2\pi} e^{-x^2/2} \frac{1}{x}.$$

Zadanie

Rozważmy $S_n = \sum_{j=1}^n X_j$, gdzie $\{X_j\}_{j \in \mathbb{N}}$ są iid o standardowym rozkładzie normalnym. Pokaż, że dla każdego $a > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n > an] = -a^2/2.$$

Przykład 1.3

Założmy, że zmienne $\{X_k\}_{k \in \mathbb{N}}$ modelują rzuty symetryczną monetą, tj. $\mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = 1/2$. Wówczas dla $a > 1/2$,

$$\{S_n \geq an\} = \{\#\{j : X_j = 1\} \geq an\}.$$

Dla $k \geq an \geq n/2$ prawdopodobieństwo

$$\mathbb{P}[\#\{j : X_j = 1\} = k] = \binom{n}{k} 2^{-n}$$

jest maksymalizowane dla $k = \lceil an \rceil$. Oznacza to, że najbardziej prawdopodobnym scenariuszem dla którego $\{S_n \geq an\}$ jest to, że dla $\lceil an \rceil$ wartości j , $X_j = 1$.

W trakcie wykładu będziemy też badali przyczynę dużych odchyłeń. Jednym ze sposobów na jej zrozumienie jest zbadanie zachowania próbki na rzadkim zdarzeniu.

Twierdzenie 1.4

Niech zmienne $\{X_k\}_{k \in \mathbb{N}}$ będą iid z rozkładem

$$\mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = 1/2.$$

Niech $S_n = \sum_{k=1}^n X_k$. Wówczas dla wszystkich $a > 1/2$,

$$\lim_{n \rightarrow \infty} \mathbb{P}[X_1 = 1 \mid S_n \geq an] = a.$$

Powyższy wynik mówi, że jeżeli $S_n \geq an$ to dla dużych n , to $X_1 = 1$ z prawdopodobieństwem $a > 1/2$. Zatem pod warunkiem $\{S_n \geq an\}$ z większym prawdopodobieństwem obserwujemy $X_1 = 1$.

Dowód. Dowód przeprowadzimy w oparciu o analizę podobną do tej z dowodu Twierdzenia 1.1. Zauważmy najpierw, że dla $n_a = \lceil an \rceil$ mamy

$$\mathbb{P}[S_n \geq an] = 2^{-n} \sum_{k=n_a}^n \binom{n}{k} = 2^{-n} \binom{n}{n_a} \sum_{k=n_a}^n \frac{(n - n_a) \cdots (n - k)}{(n_a + 1) \cdots k}.$$

Zauważmy, że

$$\sum_{k=n_a}^n \frac{(n - n_a) \cdots (n - k)}{(n_a + 1) \cdots k} \rightarrow \sum_{j=0}^{\infty} \left(\frac{1-a}{a}\right)^j < \infty.$$

Z drugiej strony

$$\begin{aligned} \mathbb{P}[X_1 = 1, S_n \geq n_a] &= \frac{1}{2} \mathbb{P}[S_{n-1} \geq n_a - 1] = 2^{-n} \sum_{k=n_a-1}^n \binom{n-1}{k} \\ &= 2^{-n} \binom{n-1}{n_a-1} \sum_{k=n_a-1}^{n-1} \frac{(n - n_a) \cdot (n - 1 - k)}{n_a \cdots k} \end{aligned}$$

gdzie

$$\sum_{k=n_a-1}^{n-1} \frac{(n-n_a) \cdot (n-1-k)}{n_a \cdots k} \rightarrow \sum_{j=0}^{\infty} \left(\frac{1-a}{a}\right)^j < \infty.$$

Zbierając skrupulatnie zebrane oszacowania dostajemy

$$\mathbb{P}[X_1 = 1 \mid S_n \geq an] = \frac{\binom{n-1}{n_a-1}}{\binom{n}{n_a}} (1 + o(1)) = \frac{n_a}{n} (1 + o(1)) \rightarrow a.$$

□

Zadanie

Założmy, że zmienne $\{X_k\}_{k \in \mathbb{N}}$ są iid o standardowym rozkładzie normalnym. Rozważmy $S_n = X_1 + \dots + X_n$. Pokaż, że dla każdego $a > 0$ rozkład X_1 pod warunkiem zdarzenia $\{S_n > an\}$ zbiega słabo do rozkładu normalnego o średniej a i wariancji jeden.

Zadanie

Założmy, że zmienne $\{X_k\}_{k \in \mathbb{N}}$ są iid o standardowym rozkładzie normalnym. Rozważmy $S_n = X_1 + \dots + X_n$. Pokaż, że dla każdego $\alpha \in (1/2, 1)$ rozkład X_1 pod warunkiem zdarzenia $\{S_n > n^\alpha\}$ zbiega słabo do rozkładu normalnego o średniej zero i wariancji jeden.

1.3 Twierdzenie Sanova dla skończonego alfabetu

Naszym celem jest prezentacja możliwie ogólnych narzędzi, które pozwalają stwierdzić jaka jest najbardziej prawdopodobna konfiguracja X_1, \dots, X_n dla której obserwujemy rzadką wartość S_n/n . W tym celu przejdziemy na nieco bardziej abstrakcyjny język. Zauważmy, że S_n/n jest niczym innym jak średnią empiryczną z liczb X_1, X_2, \dots, X_n . Aby badać wartości ostatniego ciągu rozważać będziemy miarę empiryczną

$$L_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}.$$

Wówczas L_n jest losową miarą na \mathbb{R} . Koncept ten można sformalizować zadając strukturę σ -ciała na przestrzeni miar probabilistycznych na prostej. Tak daleko posunięty formalizm nie będzie nam jednak potrzebny. Miara empiryczna L_n wiąże się ze średnią empiryczną S_n/n poprzez

$$S_n/n = \int s L_n(ds).$$

Powyższa równość będzie dla nas oznaczać, że rzadkie zdarzenia dla S_n pochodzą od zdarzeń rzadkich dla L_n . Jednakże te ostatnie dają dokładniejszą informację o ciągu X_1, X_2, \dots, X_n .

Na początek zakładając będziemy, że zmienne $\{X_k\}_{k \in \mathbb{N}}$ tylko skończenie wiele wartości. Powiemy, że spełniony jest warunek (Γ) jeżeli zmienne $\{X_k\}_{k \in \mathbb{N}}$ przyjmują wartości w skończonym zbiorze Γ takim, że

$$\Gamma = \{1, 2, \dots, r\} \subseteq \mathbb{N}$$

dla pewnego $r \in \mathbb{N}$ przy czym

$$\rho_s = \mathbb{P}[X_1 = s] > 0, \quad s \in \Gamma.$$

Miara empiryczna L_n jest losowym elementem przestrzeni miar probabilistycznych na Γ oznaczanej przez

$$M_1(\Gamma) = \left\{ \nu = (\nu_1, \nu_2, \dots, \nu_r) \in [0, 1]^r : \sum_{j=1}^r \nu_j = 1 \right\}.$$

Podkreślmy, że $\nu = (\nu_1, \nu_2, \dots, \nu_r)$ jest utożsamione z miarą

$$\nu = \sum_{k=1}^r \nu_j \delta_j.$$

Na $M_1(\Gamma)$ rozważać będziemy metrykę pochodzącą od normy całkowitego wahanania,

$$d(\mu, \nu) = \frac{1}{2} \sum_{j=1}^r |\mu_j - \nu_j|.$$

Twierdzenie 1.5

Założmy, że spełniony jest warunek (Γ) . Dla miary empirycznej zadanej przez

$$L_n = \frac{1}{n} \sum_{k=1}^n \delta_{X_k}$$

zachodzi $L_n \rightarrow \rho$ p.w.

Teza powyższego twierdzenia jest równoważna z

$$\lim_{n \rightarrow \infty} d(L_n, \rho) = 0 \text{ p.w.}$$

Dowód. Jest to wniosek z mocnego prawa wielkich liczb. Zauważmy najpierw, że miara empiryczna zapisuje się jako

$$L_n = \sum_{j=1}^r L_{n,j} \delta_j,$$

gdzie

$$L_{n,j} = \frac{1}{n} \#\{k \in \{1, 2, \dots, n\} : X_k = j\}.$$

Wystarczy zauważyć, że skoro Γ jest zbiorem skończonym, to

$$\frac{1}{n} \#\{k \in \{1, 2, \dots, n\} : X_k = j\} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{X_k=j\}} \rightarrow \rho_j$$

pociąga

$$\lim_{n \rightarrow \infty} d(L_n, \rho) = \lim_{n \rightarrow \infty} \frac{1}{2} \sum_{j=1}^r |L_{n,j} - \rho_j| = 0 \text{ p.w.}$$

□

Wobec powyższego dla $\epsilon > 0$ zbiór

$$\{d(L_n, \rho) > \epsilon\}$$

nazwalibyśmy zdarzeniem rzadkim. Chcemy teraz zbadać prawdopodobieństwo z jakim powyższe zdarzenie zachodzi. Dla $a > 0$ przez $\overline{B_a(\rho)}$ oznaczamy będziemy domkniętą kulę w $M_1(\Gamma)$ o środku w punkcie ρ i promieniu $a > 0$, tj.

$$\overline{B_a(\rho)} = \{v \in M_1(\Gamma) : d(\rho, v) \leq a\}.$$

Wówczas

$$\{d(L_n, \rho) > \epsilon\} = \left\{L_n \in \overline{B_\epsilon(\rho)}^c\right\}.$$

Podkreślmy w tym miejscu, że $\overline{B_\epsilon(\rho)}^c$ jest otwartym podzbiorem $M_1(\Gamma)$.

Twierdzenie 1.6

Założmy warunek (Γ) . Wówczas dla każdego otwartego $A \subseteq M_1(\Gamma)$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A] = - \inf_{v \in A} H(v|\rho),$$

gdzie

$$H(v|\rho) = \sum_{j=1}^r v_j \log(v_j/\rho_j).$$

Dowód. Niech

$$K_n = \left\{k = (k_1, k_2, \dots, k_r) \in \mathbb{N}^r : \sum_{j=1}^r k_j = n\right\}.$$

Wówczas $\frac{1}{n}K_n \subseteq M_1(\Gamma)$. Poprzez utożsamienie $k \in K_n$ z miarą $\nu_n(k) = k/n$. Dla każdego $k \in K_n$,

$$\mathbb{P} \left[\forall j \in \Gamma, L_n(j) = \frac{k_j}{n} \right] = n! \prod_{j \in \Gamma} \frac{\rho_j^{k_j}}{k_j!}.$$

Podobnie jak w dowodzie Twierdzenia 1.1 rozważmy

$$Q_n(a) = \max_{k \in K_n : \nu_n(k) \in A} \left(n! \prod_{s=1}^r \frac{\rho_s^{k_s}}{k_s!} \right).$$

Skoro $L_n \in \frac{1}{n}K_n$, to

$$Q_n(a) \leq \mathbb{P}[L_n \in A] \leq |K_n| Q_n(a).$$

Ze wzoru Stirlinga

$$\frac{1}{n} \log \left(n! \prod_{s=1}^r \frac{\rho_s^{k_s}}{k_s!} \right) = \sum_{s=1}^r \frac{k_s}{n} \left(\log \rho_s - \log \frac{k_s}{n} \right) + O \left(\frac{\log(n)}{n} \right)$$

jednostajnie ze względu na $k \in K_n$. Skoro suma po prawej stronie jest równa $-H(\nu_n(k)|\rho)$ oraz $|K_n| = O(n^{r-1})$ mamy

$$\begin{aligned} \frac{1}{n} \log \mathbb{P}[L_n \in A] &= O \left(\frac{\log(n)}{n} \right) + \frac{1}{n} \log Q_n(a) \\ &= O \left(\frac{\log(n)}{n} \right) - \min_{k \in K_n : \nu_n(k) \in A} H(\nu_n(k)|\rho). \end{aligned} \quad (1.2)$$

Aby zakończyć dowód wystarczy pokazać, że

- (i) Zbiór $\bigcup_{n \in \mathbb{N}} \{\nu_n(k) : k \in K_n\}$ jest gęsty w $M_1(\Gamma)$.
- (ii) Funkcja $\nu \mapsto H(\nu|\rho)$ jest ciągła na $M_1(\Gamma)$.

Powyższe warunki gwarantują, że dla każdego $\nu \in A$ istnieje ciąg $\{k_n\}_{n \in \mathbb{N}}$ taki, że $k_n \in K_n$ dla każdego $n \in \mathbb{N}$,

$$\lim_{n \rightarrow \infty} d(\nu_n(k_n), \nu) = 0, \quad \lim_{n \rightarrow \infty} H(\nu_n(k_n)|\rho) = H(\nu|\rho).$$

Skoro A jest zbiorem otwartym powyższe daje

$$\limsup_{n \rightarrow \infty} \min_{k \in K_n : \nu_n(k) \in A} H(\nu_n(k)|\rho) \leq H(\nu|\rho)$$

Skoro powyższa nierówność zachodzi dla każdego $\nu \in A$,

$$\limsup_{n \rightarrow \infty} \min_{k \in K_n : v_n(k) \in A} H(v_n(k)|\rho) \leq \min_{v \in A} H(v|\rho).$$

Z drugiej strony nierówność

$$\liminf_{n \rightarrow \infty} \min_{k \in K_n : v_n(k) \in A} H(v_n(k)|\rho) \geq \min_{v \in A} H(v|\rho)$$

jest prawdziwa dla każdego A . Ostatnie dwa szacowania pokazują, że

$$\lim_{n \rightarrow \infty} \min_{k \in K_n : v_n(k) \in A} H(v_n(k)|\rho) = \min_{v \in A} H(v|\rho).$$

Odwołanie się do (1.2) kończy dowód. \square

Lemat 1.7

$H(\cdot|\rho)$ jest skończona, ciągła i ściśle wypukła na $M_1(\Gamma)$. Dodatkowo $H(v|\rho) \geq 0$ przy czym równość zachodzi tylko dla $v = \rho$.

Dowód. Pozostawiamy jako ćwiczenie. \square

Zadanie

Znajdź $H(v|\rho)$ w przypadku, gdy ρ jest rozkładem jednostajnym na Γ , a v jest rozkładem jednostajnym na $\Gamma' \subset \Gamma$.

Zadanie

Pokaż, że dla każdego zbioru $A \subseteq M_1(\Gamma)$ zachodzi

$$\begin{aligned} - \inf_{v \in A^\circ} H(v|\rho) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{P}[L_n \in A] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{P}[L_n \in A] \leq - \inf_{v \in A} H(v|\rho). \end{aligned}$$

1.4 Zasada kontrakcji

Twierdzenie Sanova pozwala wnioskować o zachowaniu średniej arytmetycznej S_n/n . Zjawisko to nazywa się zasadą kontrakcji. Opierać się będziemy na tożsamości

$$S_n/n = \int s L_n(ds).$$

Dla miary ν na Γ będziemy pisać

$$m(\nu) = \int s \nu(ds).$$

Skoro Γ jest zbiorem skończonym, to $m: M_1(\Gamma) \rightarrow \mathbb{R}$ jest odwzorowaniem ciągłym. Oznacza to dokładnie, że dla otwartego $U \subseteq \mathbb{R}$ zbiór

$$m^{-1}[U] = \{v \in M_1(\Gamma) : m(v) \in U\}$$

jest otwartym podzbiorem $M_1(\Gamma)$.

Twierdzenie 1.8

Założmy warunek (Γ) . Dla zmiennych iid $\{X_k\}_{k \in \mathbb{N}}$ i otwartego $U \subseteq \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n/n \in U] = - \inf \left\{ H(v|\rho) : v \in m^{-1}[U] \right\}$$

Dowód. Zauważmy, że

$$\{S_n/n \in U\} = \{L_n \in m^{-1}[U]\}.$$

Teza wynika zatem z twierdzenia Sanova. \square

Przykład 1.9

Rozważmy $\Gamma = \{0, 1\}$, $U = (a, +\infty)$ oraz $\rho_0 = \rho_1 = 1/2$. Stosując Twierdzenie 1.8 otrzymujemy

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n > an] = - \inf \{H(v|\rho) : m(v) > a\}$$

Dla $v = \alpha\delta_1 + (1 - \alpha)\delta_0$ mamy

$$H(v|\rho) = \log(2) + \alpha \log(\alpha) + (1 - \alpha) \log(1 - \alpha).$$

Stąd, jeżeli $a > 1/2$, to

$$\inf \{H(v|\rho) : m(v) > a\} = \log(2) + a \log(a) + (1 - a) \log(1 - a).$$

1.5 Zadania

Poniżej prezentujemy zadania do tego rozdziału.

Zadanie 1.1

Pokaż, że dla każdego dodatniego $n \in \mathbb{N}$,

$$n \log(n) - n + 1 \leq \log(n!) \leq (n + 1) \log(n + 1) - n.$$

Wywnioskuj słabą wersją wzoru Stirlinga

$$\log(n!) = n \log(n) - n + O(\log(n)).$$

Zadanie 1.2

W tym zadaniu zaprezentujemy probabilistyczny dowód wzoru Stirlinga. Rozważmy zmienne $\{X_k\}_{k \in \mathbb{N}}$ iid z rozkładem wykładniczym ze średnią jeden. Niech $S_n = \sum_{j=1}^n X_j$. Uzasadnij, że

$$\frac{S_n - n}{\sqrt{n}} \Rightarrow \mathcal{N}(0, 1).$$

Pokaż, że $\sup_{n \in \mathbb{N}} \mathbb{E}[(S_n - n)^2/n] < \infty$ i wywnioskuj, że

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[\left| \frac{S_n - n}{\sqrt{n}} \right| \right] = \sqrt{2/\pi}.$$

Pokaż, że dla każdego $n \in \mathbb{N}$ zmienna S_n ma rozkład o gęstości

$$\frac{1}{n!} x^{n-1} e^{-x} \mathbb{1}_{\{x>0\}}$$

i wylicz, że

$$\mathbb{E} \left[\left| \frac{S_n - n}{\sqrt{n}} \right| \right] = \frac{2\sqrt{nn^n}e^{-n}}{n!}.$$

Wywnioskuj, że

$$\lim_{n \rightarrow \infty} \frac{\sqrt{2n\pi}n^n e^{-n}}{n!} = 1.$$

Zadanie 1.3

Pokaż, że dla zmiennej losowej N o standardowym rozkładzie normalnym i dodatniego x zachodzą nierówności

$$\sqrt{2\pi}e^{-x^2/2} \left(\frac{1}{x} - \frac{1}{x^3} \right) \leq \mathbb{P}[N > x] \leq \sqrt{2\pi}e^{-x^2/2} \frac{1}{x}.$$

Zadanie 1.4

Rozważmy $S_n = \sum_{j=1}^n X_j$, gdzie $\{X_j\}_{j \in \mathbb{N}}$ są iid o standardowym rozkładzie normalnym. Pokaż, że dla każdego $a > 0$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n > an] = -a^2/2.$$

Zadanie 1.5

Założmy, że zmienne $\{X_k\}_{k \in \mathbb{N}}$ są iid o standardowym rozkładzie normalnym. Rozważmy $S_n = X_1 + \dots + X_n$. Pokaż, że dla każdego $a > 0$ rozkład X_1 pod warunkiem zdarzenia $\{S_n > an\}$ zbiega słabo do rozkładu normalnego o średniej a i wariancji jeden.

Zadanie 1.6

Niech zmienne $\{X_k\}_{k \in \mathbb{N}}$ będą iid z rozkładem Poissona z parametrem jeden. Rozważmy $S_n = X_1 + \dots + X_n$. Dla $a > 1$ znajdź

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq an].$$

Znajdź słabą granicę dla

$$\mu_n(A) = \mathbb{P}[X_1 \in A \mid S_n \geq an], \quad A \in \mathcal{Bor}(\mathbb{R})$$

przy $n \rightarrow \infty$.

Zadanie 1.7

Niech zmienne $\{X_k\}_{k \in \mathbb{N}}$ będą iid z rozkładem

$$\mathbb{P}[X_k = -1] = \mathbb{P}[X_k = 0] = \mathbb{P}[X_k = 1] = 1/3.$$

Dla $S_n = \sum_{j=1}^n X_j$ i $x > 0$ znajdź

$$\lim_{n \rightarrow \infty} \mathbb{P}[S_n \geq xn].$$

Zadanie 1.8

I_ρ jest skończona, ciągła i ściśle wypukła na $M_1(\Gamma)$. Dodatkowo $I_\rho(v) \geq 0$ przy czym równość zachodzi tylko dla $v = \rho$.

Zadanie 1.9

Znajdź $I_\rho(v)$ w przypadku, gdy ρ jest rozkładem jednostajnym na Γ , a v jest rozkładem jednostajnym na $\Gamma' \subset \Gamma$.

Zadanie 1.10

Pokaż, że dla każdego zbioru $A \subseteq M_1(\Gamma)$ zachodzi

$$\begin{aligned} - \inf_{v \in A^\circ} H(v|\rho) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \mathbb{P}[L_n \in A] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathbb{P}[L_n \in A] \leq - \inf_{v \in A} H(v|\rho). \end{aligned}$$

Zadanie

Udowodnij, że

$$- \lim_{n \rightarrow \infty} \inf_{v \in A \cap K_n} H(v|\rho) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A] = - \inf_{v \in A} H(v|\rho).$$

dla każdego

$$A \subseteq \{v \in M_1(\Gamma) : v \ll \rho\}$$

takiego, że $A \subseteq \text{cl}(\text{int}(A))$. Pokaż, że dla takiego A ,

$$\inf_{v \in A} H(v|\rho) = H(v^*|\rho)$$

dla pewnego $v^* \in \text{cl}(\text{int}(A))$.

Zadanie 1.11

Znajdź domknięty zbiór A dla którego granice

$$\lim_{n \rightarrow \infty} \inf_{\nu \in A \cap K_n} H(\nu | \rho), \quad \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A]$$

nie istnieją.

Zadanie 1.12

Znajdź zbiór domknięty A taki, że

$$\inf_{\nu \in A} H(\nu | \rho) < \infty, \quad A = \text{cl}(\text{int}(A)), \quad \inf_{\nu \in \text{int}(A)} H(\nu | \rho) = \infty.$$

Zadanie 1.13

Niech G będzie otwartym podzbiorem $M_1(\Gamma)$. Niech R będzie miarą wylosowaną jednostajnie z G . Rozważmy $X_1, X_2 \dots$ iid rozkładem R . Pokaż, że dla każdego otwartego A ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A] = - \inf_{\nu \in A} I(\nu),$$

gdzie

$$I(\nu) = \inf_{\mu \in G} H(\nu | \mu).$$

Duże odchylenia na prostej

Streszczenie Zbadamy teraz klasyczny przypadek sum niezależnych zmiennych losowych. Zobaczymy jak techniki oparte na zamianie miary pomagają badać wielkie odchylenia. Rozwinięta w tym rozdziale teoria ułatwi nam obchodzenie się z ogólną teorią w następnym rozdziale.

2.1 Twierdzenie Craméra

Naszym celem jest prezentacja ogólnej wersji Twierdzenia 1.1.

Twierdzenie 2.1 (Cramér 1938 [Cra38])

Niech $\{X_k\}_{k \in \mathbb{N}}$ będzie ciągiem zmiennych iid takim, że dla każdego $t \in \mathbb{R}$,

$$\mathbb{E} \left[e^{tX_1} \right] < \infty.$$

Niech $S_0 = 0$ i $S_n = \sum_{k=1}^n X_k$ dla naturalnego $n \geq 1$. Wówczas dla każdego $a > \mathbb{E}[X_1]$ zachodzi

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq an] = -I(a),$$

gdzie

$$I(a) = \sup_{s \in \mathbb{R}} \left\{ sa - \log \mathbb{E} \left[e^{sX_1} \right] \right\}.$$

Dowód powyższego twierdzenia przeprowadzimy w dwóch częściach, poprzez uzasadnienie dwóch nierówności.

Dowód \leq w Twierdzeniu 2.1. Pokażemy, że dla każdego $x > \mathbb{E}[X_1]$,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq xn] \leq -I(x).$$

Ustalmy $\lambda > 0$. Korzystając z nierówności Czebyszewa

16 2 Duże odchylenia na prostej

$$\mathbb{P}[S_n \geq xn] = \mathbb{P}[e^{\lambda S_n} \geq e^{\lambda nx}] \leq e^{-\lambda nx} \mathbb{E}[e^{\lambda S_n}] = e^{-\lambda nx} \mathbb{E}[e^{\lambda X_1}]^n.$$

Stąd

$$\frac{1}{n} \log \mathbb{P}[S_n \geq xn] \leq -\left(\lambda x - \log \mathbb{E}[e^{\lambda X_1}]\right).$$

Skoro $\lambda > 0$ jest dowolna możemy wywnioskować, że

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n \geq xn] \leq -\sup_{\lambda > 0} \left(\lambda x - \log \mathbb{E}[e^{\lambda X_1}]\right).$$

Aby formalnie zakończyć dowód nierówności pozostaje pokazać, że

$$\sup_{\lambda \in \mathbb{R}} \left\{ \lambda x - \log \mathbb{E}[e^{\lambda X_1}] \right\} = \sup_{\lambda > 0} \left\{ \lambda x - \log \mathbb{E}[e^{\lambda X_1}] \right\} > 0.$$

Dokonyamy tego w niedalekiej przyszłości. □

Uwaga 2.2

Zauważmy, że pokazaliśmy, że dla każdego $x > \mathbb{E}[X_1]$ oraz każdego n ,

$$\mathbb{P}[S_n \geq xn] \leq e^{-nI(x)}. \quad (2.1)$$

Przed prezentacją reszty dowodu Twierdzenia 2.1 przedstawimy nieco więcej teorii. W szczególności przyjrzymy się nieco dokładniej wyrażeniu definiującemu $I(a)$. Zauważmy, że funkcja $s \mapsto \log \mathbb{E}[e^{sX_1}]$ zależy od zmiennej losowej X_1 tylko poprzez jej rozkład.

Definicja 2.3

Niech μ będzie rozkładem prawdopodobieństwa na \mathbb{R} . Funkcja $\Lambda_\mu : \mathbb{R} \rightarrow [0, \infty]$ określona wzorem

$$\Lambda_\mu(\lambda) = \log \left(\int e^{\lambda s} \mu(ds) \right)$$

nazywana jest funkcją tworzącą kumulanty μ .

Jeżeli X_1 ma rozkład μ , to $\Lambda_\mu(\lambda) = \log \mathbb{E}[e^{\lambda X_1}]$.

Przykład 2.4

Jeżeli $\mu = \mathcal{N}(0, 1)$, czyli jest zadana przez gęstość

$$\mu(dx) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx,$$

to funkcja tworząca momenty wylicza się poprzez standardowy rachunek

$$e^{\Lambda_{\mu}(\lambda)} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\lambda s} e^{-s^2/2} ds = \frac{1}{\sqrt{2\pi}} e^{\lambda^2/2} \int_{-\infty}^{\infty} e^{-\frac{(s-\lambda)^2}{2}} ds = e^{\lambda^2/2}.$$

Podsumowując

$$\Lambda_{\mathcal{N}(0,1)}(\lambda) = \lambda^2/2.$$

W przypadku $\mu = \mathcal{N}(m, \sigma^2)$, czyli μ danego przez

$$\mu(dx) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/(2\sigma^2)}$$

poprzez analogiczny rachunek (lub rozważenie odwzorowania $x \mapsto \sigma x + m$ otrzymujemy

$$\Lambda_{\mathcal{N}(m,\sigma^2)}(\lambda) = \lambda m + \lambda^2 \sigma^2 / 2.$$

Przykład 2.5

Rozkład $\mu = \text{Poi}(\gamma)$ dla $\gamma > 0$ zadaje się przez

$$\mu(\{k\}) = \frac{\gamma^k}{k!} e^{-\gamma}, \quad k \in \mathbb{N}.$$

W tym przypadku mamy

$$e^{\Lambda_{\mu}(\lambda)} = \sum_{k=0}^{\infty} e^{\lambda k} \frac{e^{-\gamma} \gamma^k}{k!} = e^{\gamma(e^{\lambda}-1)}.$$

Czyli

$$\Lambda_{\text{Poi}(\gamma)}(\lambda) = \gamma (e^{\lambda} - 1).$$

Przykład 2.6

Dla standardowego rozkładu wykładniczego $\mu = \text{Exp}(1)$ danego przez

$$\mu(dx) = e^{-x} \mathbb{1}_{\{x>0\}} dx$$

mamy

$$e^{\Lambda_{\mu}(\lambda)} = \int_0^{\infty} e^{\lambda x} e^{-x} dx = \begin{cases} 1/(1-\lambda) & \lambda < 1, \\ +\infty & \lambda \geq 1. \end{cases}$$

Lemat 2.7

Niech μ będzie dowolnym rozkładem prawdopodobieństwa na \mathbb{R} . Λ_{μ} jest wypukłą funkcją zmiennej λ . W szczególności,

$$D := \{\lambda : \Lambda_{\mu}(\lambda) < \infty\}$$

jest przedziałem zawierającym zero.

Dowód. Niech $0 < \alpha < 1$. Wówczas dla $\lambda_1, \lambda_2 \in D$,

$$\begin{aligned} \exp \{ \Lambda_\mu(\alpha\lambda_1 + (1-\alpha)\lambda_2) \} &= \int e^{\alpha\lambda_1 s + (1-\alpha)\lambda_2 s} \mu(ds) \\ &= \left(\int e^{\lambda_1 s} \mu(ds) \right)^\alpha \left(\int e^{\lambda_2 s} \mu(ds) \right)^{1-\alpha}. \end{aligned}$$

Stąd

$$\Lambda_\mu(\alpha\lambda_1 + (1-\alpha)\lambda_2) \leq \alpha\Lambda_\mu(\lambda_1) + (1-\alpha)\Lambda_\mu(\lambda_2).$$

Co pokazuje wypukłość i daje w szczególności, że $\alpha\lambda_1 + (1-\alpha)\lambda_2 \in D$. \square

Przypomnijmy, że dla $A \subseteq \mathbb{R}$ przez $\text{int}(A)$ oznaczamy wnętrze zbioru A , czyli największy zbiór otwarty zawarty w A .

Lemat 2.8

Niech μ będzie dowolnym rozkładem prawdopodobieństwa na \mathbb{R} . Λ_μ jest klasy \mathcal{C}^∞ na zbiorze $\text{int}(D)$ oraz

$$\frac{d^n}{d\lambda^n} e^{\Lambda_\mu(\lambda)} = \int s^n e^{\lambda s} \mu(ds), \quad \lambda \in \text{int}(D).$$

Dowód. ćwiczenie. \square

Definicja 2.9

Niech $f: \mathbb{R} \rightarrow [0, +\infty]$ będzie funkcją wypukłą. **Transformatą Legendre'a** f^* funkcji f nazywamy odwzorowanie $f^*: \mathbb{R} \rightarrow [0, \infty]$ dane wzorem

$$f^*(x) = \sup_{\lambda \in \mathbb{R}} (\lambda x - f(\lambda)).$$

Lemat 2.10

Niech μ będzie dowolnym rozkładem prawdopodobieństwa na prostej. Transformata Legendre'a Λ_μ^* funkcji tworzącej kumulanty Λ_μ rozkładu μ ma następujące własności.

- (i) $\Lambda_\mu^* \geq 0$.
- (ii) Λ_μ^* jest półciągła z dołu.
- (iii) Λ_μ^* jest wypukłą.
- (iv) Jeśli $m = \int s \mu(ds) \in \mathbb{R}$ jest dobrze określona, to $\Lambda_\mu^*(m) = 0$. Jeżeli $x > m$, to

$$\Lambda_\mu^*(x) = \sup_{\lambda \geq 0} \{ \lambda x - \Lambda_\mu(\lambda) \}.$$

Jeżeli $x \leq m$, to

$$\Lambda_\mu^*(x) = \sup_{\lambda \leq 0} \{ \lambda x - \Lambda_\mu(\lambda) \}.$$

(v) Jeśli $\Lambda_\mu(\lambda) < \infty$ dla wszystkich $\lambda \in \mathbb{R}$, to

$$\frac{\Lambda_\mu^*(x)}{|x|} \xrightarrow{|x| \rightarrow \infty} \infty.$$

Przypomnijmy, że funkcja $f: \mathbb{R} \rightarrow \mathbb{R}$ jest półciągła z dołu, jeżeli dla dowolnego ciągu $x_n \rightarrow x$ zachodzi

$$\liminf_{n \rightarrow \infty} f(x_n) \geq f(x).$$

Dowód. Zasadność (i) wynika wprost z definicji, biorąc $\lambda = 0$. Aby uzasadnić (ii) zauważmy, że dla dowolnego ciągu $x_n \rightarrow x$ i dowolnej $\lambda \in \mathbb{R}$,

$$\liminf_{n \rightarrow \infty} \Lambda_\mu^*(x_n) \geq \liminf_{n \rightarrow \infty} \lambda x_n - \Lambda_\mu(\lambda) = \lambda x - \Lambda_\mu(\lambda).$$

Optymalizując po λ otrzymujemy tezę. Dla uzasadnienia (iii) rozważmy nieujemne $\alpha + \beta = 1$ i dowolne rzeczywiste x i y . Dla $\lambda \in \mathbb{R}$,

$$\lambda(\alpha x + \beta y) - \Lambda_\mu(\lambda) = \alpha(\lambda x - \Lambda_\mu(\lambda)) + \beta(\lambda y - \Lambda_\mu(\lambda)) \leq \alpha \Lambda_\mu^*(x) + \beta \Lambda_\mu^*(y).$$

Pierwszy postulat (iv) wynika z nierówności Jensena, ponieważ dla $X \sim \mu$,

$$\Lambda_\mu(\lambda) = \log \mathbb{E} \left[e^{\lambda X} \right] \geq \mathbb{E} \left[\log e^{\lambda X} \right] = \lambda m$$

Przy czym równość zachodzi tylko dla $\lambda = 0$. Rozważmy $x \geq m$ oraz $\lambda < 0$. Wówczas

$$\lambda x - \Lambda_\mu(\lambda) \leq \lambda m - \Lambda_\mu(\lambda) \leq \Lambda_\mu^*(m) = 0.$$

Stąd

$$\Lambda_\mu^*(x) = \sup_{\lambda \geq 0} \{ \lambda x - \Lambda_\mu(\lambda) \}.$$

W analogiczny sposób rozumiemy dla $x < m$ oraz $\lambda > 0$. Aby uzasadnić (v) zauważmy, że dla każdego λ ,

$$\frac{\Lambda_\mu^*(x)}{|x|} \geq \lambda \frac{x}{|x|} - \frac{\Lambda_\mu(\lambda)}{|x|}.$$

Jeżeli przykładowo $x \geq 0$, to powyższe dla $\lambda > 0$ implikuje

$$\frac{\Lambda_\mu^*(x)}{|x|} \geq \lambda - \frac{\Lambda_\mu(\lambda)}{|x|} \geq \lambda.$$

Co pociąga tezę. □

Przykład 2.11

Niech $\mu = \mathcal{N}(m, \sigma^2)$. Jak już sprawdziliśmy wcześniej $\Lambda_\mu(\lambda) = \lambda m + \lambda^2 \sigma^2 / 2$. Przez prosty rachunek $\Lambda_\mu^*(x) = (x - m)^2 / (2\sigma^2)$.

Przykład 2.12

Dla $\mu = \text{Poi}(\gamma)$ mamy $\Lambda_\mu(\lambda) = \gamma(e^\lambda - 1)$. Stąd $\Lambda_\mu^*(x) = x \log(x/\gamma) - x + \gamma$.

Przykład 2.13

Dla $\mu = \text{Exp}(1)$ mamy

$$\Lambda_\mu(\lambda) = \begin{cases} -\log(1 - \lambda) & \lambda < 1 \\ +\infty & \lambda \geq 1 \end{cases}$$

oraz

$$\Lambda_\mu^*(x) = \begin{cases} x - 1 - \log x & x > 0 \\ +\infty & x \leq 0 \end{cases}$$

Zauważmy, że (v) poprzedniego lematu nie jest spełnione w tym przypadku.

Przykład 2.14

W przypadku modelowania rzutów monetą wybieramy

$$\mu = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1.$$

Wówczas

$$e^{\Lambda_\mu(\lambda)} = \frac{1}{2}(e^\lambda + 1)$$

oraz

$$\Lambda_\mu^*(x) = \begin{cases} x \log x + (1 - x) \log(1 - x) + \log 2 & x \in [0, 1] \\ +\infty & \text{poza tym.} \end{cases}$$

Przypomnijmy, że w tym przypadku pokazaliśmy już, że dla $x \geq 1/2$,

$$\frac{1}{n} \log \mathbb{P}[S_n \geq xn] \rightarrow -\Lambda_\mu^*(x) \quad (2.2)$$

Aby udowodnić dolne ograniczenie w Twierdzeniu Craméra, tj.

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left[\frac{S_n}{n} \geq x \right] \geq -I(x) = -\Lambda_\mu^*(x),$$

będziemy potrzebowali *rodzin wykładniczych* lub *przechylonych miar*.

Definicja 2.15

Niech μ będzie dowolną miarę na prostej. Przypomnijmy, że $D = \{\lambda : Z(\lambda) < \infty\}$ i zdefiniujmy, dla $\lambda \in D$, miarę probabilistyczną μ_λ poprzez

$$\frac{d\mu_\lambda}{d\mu}(s) = e^{\lambda s - \Lambda_\mu(\lambda)}.$$

Wówczas

$$\{\mu_\lambda \mid \lambda \in D\}$$

nazywamy rodziną wykładniczą dla μ . Definiujemy dla $\lambda \in D$ definiujemy

$$m_\lambda := \int s \mu_\lambda(ds).$$

Zauważmy, że dla $\lambda \in \text{int}(D)$ mamy

$$m_\lambda = \frac{d}{d\lambda} \Lambda_\mu(\lambda).$$

Przykład 2.16

Jeżeli $\mu = \mathcal{N}(0, \sigma^2)$, to $\Lambda_\mu(\lambda) = \lambda^2 \sigma^2 / 2$. Mamy

$$\mu_\lambda(ds) = e^{\lambda s - \lambda^2 \sigma^2 / 2} e^{-s^2 / (2\sigma^2)} \frac{1}{\sqrt{2\pi\sigma^2}} ds.$$

Oznacza to, że μ_λ jest po prostu rozkładem normalnym $\mathcal{N}(\lambda\sigma^2, \sigma^2)$. W szczególności daje to $m_\lambda = \lambda\sigma^2$.

Przykład 2.17

Dla rozkładu Poissona $\mu = \text{Poi}(\gamma)$ mamy $\Lambda_\mu(\lambda) = \gamma(e^\lambda - 1)$ i stąd

$$\mu_\lambda(\{k\}) = \frac{e^{\lambda k} \frac{\gamma^k}{k!} e^{-\gamma}}{e^{\gamma(e^\lambda - 1)}}$$

Innymi słowy $\mu_\lambda = \text{Poi}(e^\lambda \gamma)$. W szczególności $m_\lambda = e^\lambda \gamma$.

Przykład 2.18

Dla standardowego rozkładu wykładniczego $\mu = \text{Exp}(1)$ mamy

$$\Lambda_\mu(\lambda) = \begin{cases} -\log(1 - \lambda) & \lambda < 1 \\ +\infty & \lambda \geq 1 \end{cases}$$

W tym przypadku, dla $s > 0$,

$$\mu_\lambda(ds) = (1 - \lambda) e^{\lambda s} e^{-s} ds = (1 - \lambda) e^{-(1-\lambda)s} ds$$

Czyli $\mu_\lambda = \text{Exp}(1 - \lambda)$ oraz $m_\lambda = \frac{1}{1-\lambda}$.

Przykład 2.19

Dla rzutów monetą:

$$\mu = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_1, \quad e^{\Lambda_\mu(\lambda)} = \frac{1}{2}(1 + e^\lambda).$$

oraz

$$\mu_\lambda(\{0\}) = \frac{e^{\lambda \cdot 0}}{\frac{1}{2}(1 + e^\lambda)} \frac{1}{2} = \frac{1}{e^\lambda + 1}, \quad \mu_\lambda(\{1\}) = \frac{e^\lambda}{e^\lambda + 1}$$

Zatem μ_λ opisuje rzut monetą, na której orzeł wypada z prawdopodobieństwem $e^\lambda / (e^\lambda + 1)$. Stąd $m_\lambda = e^\lambda / (e^\lambda + 1)$.

Przypomnijmy, że przez nośnik miary probabilistycznej μ nazywamy najmniejszy zbiór domknięty o mierze jeden.

Lemat 2.20

Niech μ będzie dowolnym rozkładem prawdopodobieństwa na prostej.

(i) Niech S_μ będzie otoczką wypukłą nośnika μ . Wówczas,

$$\Lambda_\mu^* = +\infty \text{ na } (S_\mu)^c$$

gdzie $(S_\mu)^c$ oznacza dopełnienie S_μ .

(ii) Załóżmy $D = \mathbb{R}$. Wówczas,

$$\Lambda_\mu^*(x) = \lambda x - \Lambda_\mu(\lambda) \Leftrightarrow x = m_\lambda \quad (2.3)$$

(iii) Załóżmy $D = \mathbb{R}$. Wówczas

$$\text{int}(S_\mu) = \{m_\lambda | \lambda \in D\}$$

W szczególności $\Lambda_\mu^*(x)$ jest skończone dla $x \in \text{int}(S_\mu)$.

Dowód. Na potrzeby dowodu wprowadźmy oznaczenie

$$g_x(\lambda) = \int e^{\lambda(s-x)} \mu(ds).$$

Wówczas

$$\Lambda_\mu^*(x) = \sup_{\lambda \in \mathbb{R}} \left[\lambda x - \log \int e^{\lambda s} \mu(ds) \right] = -\log \inf_{\lambda \in \mathbb{R}} g_x(\lambda).$$

Aby uzasadnić (i) rozważmy $x > \sup\{s : s \in S_\mu\}$. Z definicji $s - x \leq 0$ $\mu - p.w.$ Oznacza to, że $g_x(\lambda) \rightarrow 0$ przy $\lambda \rightarrow \infty$. Stąd $\Lambda_\mu^*(x) = +\infty$. Przypadek $x < \inf\{s : s \in S_\mu\}$ jest badany analogicznie. Dla dowodu (ii) i (iii) rozważmy $x \in \text{int}(S_\mu)$. Wówczas

$$g_x(\lambda) \geq \begin{cases} \int_{[x, \infty)} e^{\lambda(s-x)} \mu(ds) \xrightarrow{\lambda \rightarrow \infty} +\infty \\ \int_{(-\infty, x]} e^{\lambda(s-x)} \mu(ds) \xrightarrow{\lambda \rightarrow -\infty} +\infty \end{cases}$$

Skoro $g_x(\cdot)$ jest gładka i wypukła, to g ma minimum w pewnym $\eta \in \mathbb{R}$. Ale $g(\eta) = \inf_\lambda g(\lambda) \Leftrightarrow 0 = g'(\eta) \Leftrightarrow x = m_\eta$. \square

Druga część dowodu Twierdzenia Craméra. Wykorzystując zgromadzony aparat jesteśmy w stanie udowodnić ogólniejszy rezultat. Pokażemy, że dla dowolnego otwartego $U \subset \mathbb{R}$ oraz $x \in U$ zachodzi

$$\liminf_n \frac{1}{n} \log \mathbb{P}(S_n/n \in U) \geq -\Lambda_\mu^*(x).$$

Zgodnie z Lematem 2.20, wystarczy rozważyć $x \in U \cap S_\mu$. W wypukłości Λ_μ^* jest rosnąca dla $x \geq m$ i Λ_μ^* jest malejąca dla $x \leq m$. Wystarczy zatem rozważyć $x \in U \cap \text{int}S_\mu$. Ze względu na nasze założenie $D = \mathbb{R}$, istnieje $\lambda \in \mathbb{R}$ taka, że $\int s\mu_\lambda(ds) = m_\lambda = x$. Niech \mathbb{P}_λ będzie prawdopodobieństwem takim, że X_1, X_2, \dots są niezależnymi i identycznie rozłożonymi zmiennymi losowymi pod \mathbb{P}_λ , wszystkie o rozkładzie μ_λ . Wówczas $S_n/n \rightarrow m_\lambda = x$, \mathbb{P}_λ – prawie na pewno. Ponieważ U jest otwarte, istnieje $\varepsilon > 0$ takie, że $U_\varepsilon(x) = \{s : |s - m_\lambda| < \varepsilon\} \subseteq U$. Mamy

$$\begin{aligned} \mathbb{P}[S_n/n \in U] &\geq \mathbb{P}[|S_n/n - m_\lambda| < \varepsilon] \\ &= \int_{\left\{ \left| \frac{x_1 + \dots + x_n}{n} - m_\lambda \right| < \varepsilon \right\}} \mu(dx_1) \dots \mu(dx_n) \\ &= \int_{\left\{ \left| \frac{x_1 + \dots + x_n}{n} - m_\lambda \right| < \varepsilon \right\}} \prod_{j=1}^n \frac{d\mu}{d\mu_\lambda}(x_j) \mu_\lambda(dx_1) \dots \mu_\lambda(dx_n). \end{aligned}$$

Zauważmy, że skoro

$$\prod_{j=1}^n \frac{d\mu}{d\mu_\lambda}(x_j) = e^{\lambda(x_1 + \dots + x_n) - n\Lambda_\mu(\lambda)}$$

to

$$\begin{aligned} \int_{\left\{ \left| \frac{x_1 + \dots + x_n}{n} - m_\lambda \right| < \varepsilon \right\}} \prod_{j=1}^n \frac{d\mu}{d\mu_\lambda}(x_j) \mu_\lambda(dx_1) \dots \mu_\lambda(dx_n) \\ \geq e^{n\Lambda_\mu(\lambda) - n(\lambda m_\lambda + \varepsilon|\lambda|)} \mathbb{P}_\lambda[|S_n/n - m_\lambda| < \varepsilon] \end{aligned}$$

Przy czym, jak już zauważyliśmy, ostatnie prawdopodobieństwo dąży do 1. Stąd

$$\liminf_n \frac{1}{n} \log \mathbb{P}[S_n/n \in U] \geq -(\lambda m_\lambda + \varepsilon|\lambda|) - \Lambda_\mu(\lambda) = \Lambda_\mu^*(m_\lambda) - \varepsilon|\lambda|$$

Rozważając teraz $\varepsilon \rightarrow 0$ otrzymujemy tezę □

Twierdzenie Craméra można uogólnić w następujący sposób.

Twierdzenie 2.21

Niech X_1, X_2, \dots będą niezależnymi i identycznie rozłożonymi zmiennymi losowymi o rozkładzie μ . Dla domkniętego $A \subseteq \mathbb{R}$, mamy

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n/n \in A] \leq - \inf_{x \in A} \Lambda_\mu^*(x).$$

Dla $U \subseteq \mathbb{R}$ otwartego, mamy

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(S_n/n \in U) \geq - \inf_{x \in U} \Lambda_\mu^*(x).$$

Uwaga 2.22

Jeśli $D = \mathbb{R}$, dla każdego $x \in \mathbb{R}$ istnieje $\lambda = \lambda(x)$ takie, że $m_\lambda = x$. Zobaczmy później, że

$$\mathbb{P}[X_1 \in A | S_n/n \geq x] \rightarrow \mu_{\lambda(x)}(A)$$

Uwaga 2.23

Zakładamy, że $m = \mathbb{E}[X_1] < \infty$ ale $\Lambda_\mu(\lambda) = \infty$ dla każdego $\lambda > 0$, wówczas, dla wszystkich $x > m = \mathbb{E}[Y_1]$

$$\frac{1}{n} \log \mathbb{P}[S_n/n \geq x] \rightarrow 0$$

Wynika to z Twierdzenia 2.21 z $A = [x, \infty)$, zauważając, że $\Lambda_\mu^*(x) = 0$ w tym przypadku. Innymi słowy, $\mathbb{P}[S_n/n \geq x]$ maleje wolniej niż wykładniczo. Okazuje się, że w tym przypadku mamy

$$\mathbb{P}[X_1 \in A | S_n/n \geq x] \rightarrow \mu(A).$$

Uwaga 2.24

Dla dowolnego borelowskiego A , $\frac{1}{n} \log \mathbb{P}(S_n/n \in A)$ nie musi zbiegać (podaj przykład!). Jeśli natomiast A jest otwarty i wypukły, to wówczas $\lim_n \frac{1}{n} \log \mathbb{P}[S_n/n \in A]$ istnieje (ćwiczenie).

2.2 Serie w spacerach losowych

Rozważmy teraz spacer losowy $S_0 = 0$, $S_k = \sum_{i=1}^k X_i$, $k = 1, 2, \dots$, gdzie X_1, X_2, \dots są iid z rozkładem μ , gdzie μ jest miarą prawdopodobieństwa na $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Dla borelowskiego $A \subseteq \mathbb{R}^d$ rozważmy

$$R_n := \max \left\{ l - k : 0 \leq k < l \leq n, \frac{S_l - S_k}{l - k} \in A \right\}$$

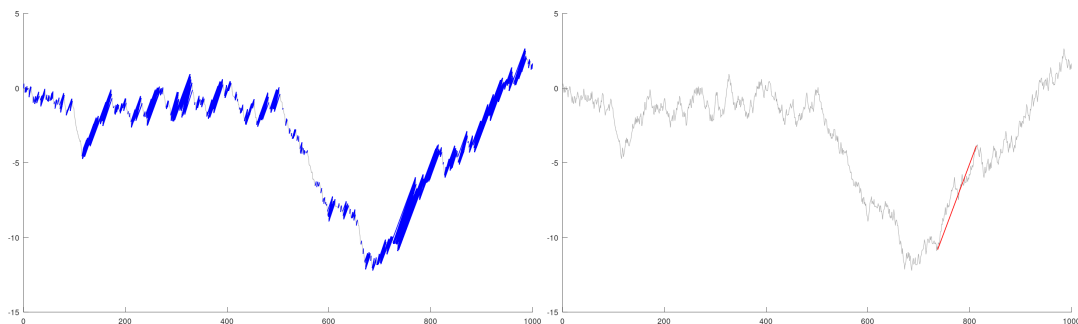
Wówczas R_n jest długością najdłuższego bloku do czasu n , którego średnia należy do zbioru A .

Przykład 2.25

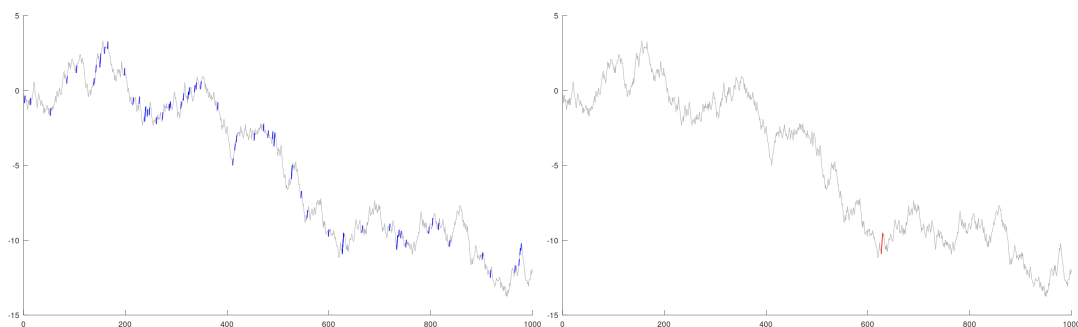
Rozważmy zmienne modelujące niezależne rzuty monetą

$$X_i = \begin{cases} 0 & \text{z prawdopodobieństwem } 1/2 \\ 1 & \text{z prawdopodobieństwem } 1/2 \end{cases}$$

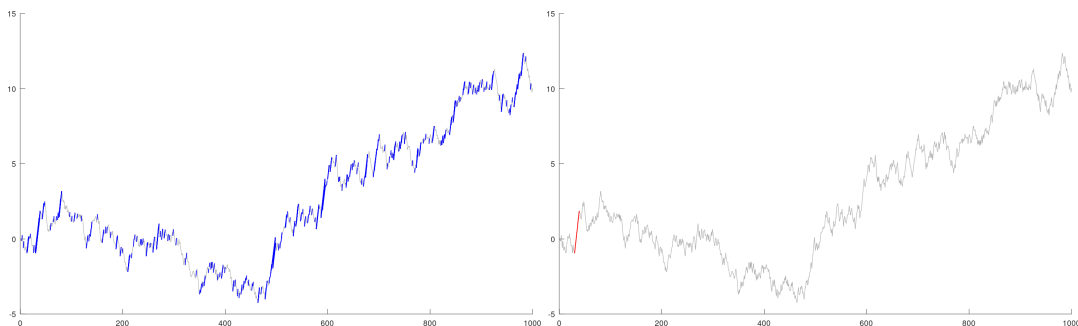
Dla $A = \{1\}$, R_n staje się długością najdłuższej serii jedynek do czasu n .



Rysunek 2.1. Trajektoria spaceru losowego o przyrostach z rozkładem jednostajnym na $(-1/2, 1/2)$. Na niebiesko oznaczone są bloki o średniej w przedziale $[1/11, 1/2]$. Kolorem czerwonym jest oznaczony najdłuższy z nich.



Rysunek 2.2. Trajektoria spaceru losowego o przyrostach z rozkładem jednostajnym na $(-1/2, 1/2)$. Na niebiesko oznaczone są bloki o średniej w przedziale $[5/11, 1/2]$. Kolorem czerwonym jest oznaczony najdłuższy z nich.



Rysunek 2.3. Trajektoria spaceru losowego o przyrostach z rozkładem jednostajnym na $(-1/2, 1/2)$. Na niebiesko oznaczone są bloki o średniej w przedziale $[3/11, 1/2]$. Kolorem czerwonym jest oznaczony najdłuższy z nich.

Aby zbadać asymptotyczne zachowanie R_n rozważmy

$$T_r := \inf \left\{ l : \frac{S_l - S_k}{l - k} \in A \text{ dla pewnego } k \text{ takiego, że } 0 \leq k \leq l - r \right\}$$

czas oczekiwania na "blok A " o długości co najmniej r . Wówczas

$$\{R_n \geq r\} = \{T_r \leq n\}.$$

Twierdzenie 2.26

Założmy, że

$$\Lambda(A) := - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[S_n/n \in A] \in [0, \infty] \quad (2.4)$$

istnieje. Wtedy

$$\lim_{n \rightarrow \infty} \frac{R_n}{\log n} = \lim_{r \rightarrow \infty} \frac{r}{\log T_r} = \frac{1}{\Lambda(A)}.$$

Uwaga 2.27

Granica (2.4) istnieje, jeżeli A jest otwarty i wypukły (ćwiczenie).

Przykład 2.28

Dla rzutów monetą i $A = \{1\}$ mamy

$$\mathbb{P}[S_n/n \in A] = 2^{-n}.$$

Stąd $\Lambda(A) = \log 2$. Zatem: $R_n / \log n \rightarrow 1 / \log 2$ prawie na pewno. Oznacza to, że $R_n \approx \log_2 n$. Na przykład, dla $n = 256$, mamy $R_n \approx 8$.

Dowód Twierdzenia 2.26. **Krok 1** Pokażemy, że jeżeli $\log(T_r)/r \rightarrow \Lambda(A)$, to

$$\lim_{n \rightarrow \infty} \frac{R_n}{\log n} = \lim_{r \rightarrow \infty} \frac{r}{\log T_r}.$$

W tym celu pokażemy dwie nierówności

$$\limsup_{n \rightarrow \infty} \frac{R_n}{\log n} \leq \frac{1}{\Lambda(A)} \quad (i)$$

oraz

$$\liminf_{n \rightarrow \infty} \frac{R_n}{\log n} \geq \frac{1}{\Lambda(A)}. \quad (ii)$$

Zauważmy najpierw, że $T_{R_n} \leq n$ a co za tym idzie

$$\limsup_{n \rightarrow \infty} \frac{R_n}{\log(n)} \leq \limsup_{n \rightarrow \infty} \frac{R_n}{\log(T_{R_n})} = \lim_{r \rightarrow \infty} \frac{r}{T_r} = \frac{1}{\Lambda(A)}.$$

Co dowodzi nierówności (i). Zauważmy też, że nierówność (ii) jest spełniona automatycznie jeżeli $\Lambda(A) = \infty$. Możemy więc bez zmniejszania ogólności założyć, że $\Lambda(A) < \infty$. Przypomnijmy, że $\{R_n < r\} = \{T_r > n\}$. Dla dowolnego $c < 1/\Lambda(A) \in (0, +\infty]$ oraz $n_r = \lfloor c \log(n) \rfloor$ mamy

$$\{R_n < n_r\} = \{T_{n_r} > n\} = \left\{ \frac{r_n}{\log(T_{r_n})} < \frac{r_n}{\log(n)} \right\}.$$

Skoro $r_n/\log(n) \rightarrow c < 1/\Lambda(A)$, to z prawdopodobieństwem jeden powyższe zdarzenia zachodzą jedynie dla skończenie wielu n . Stąd

$$\mathbb{P} \left[\liminf_{n \rightarrow \infty} \frac{R_n}{\log(n)} \geq c \right] = 1$$

dla dowolnego $c < 1/\Lambda(A)$. Wybierając c_n rosnące do $1/\Lambda(A)$ wnioskujemy, że

$$\liminf_{n \rightarrow \infty} \frac{R_n}{\log(n)} \geq 1/\Lambda(A)$$

prawie wszędzie. To dowodzi (ii).

Krok 2 Pokażemy, że

$$\frac{\log T_r}{r} \rightarrow \Lambda(A) \quad \mathbb{P} - \text{prawie na pewno}$$

W tym celu, jak poprzednio, pokażemy dwie nierówności. Będziemy pisać

$$\mu_j(A) = \mathbb{P}[S_j/j \in A].$$

Aby pokazać $\liminf_{r \rightarrow \infty} \log(T_r)/r \geq \Lambda(A)$ prawie na pewno napiszmy

$$\{T_r \leq n\} \subset \bigcup_{k=0}^{n-r} \bigcup_{l=k+r}^n C_{k,l}, \quad \text{gdzie } C_{k,l} = \left\{ \frac{S_l - S_k}{l-k} \in A \right\}.$$

Stąd, ponieważ $\mathbb{P}(C_{k,l}) = \mu_{l-k}(A)$ dla $l-k \geq r$ otrzymujemy

$$\mathbb{P}[T_r \leq n] \leq \sum_{k=0}^{n-r} \sum_{l=k+r}^n \mu_{l-k}(A) \leq \sum_{k=0}^{n-r} \sum_{j=r}^{\infty} \mu_j(A) \leq n \sum_{j=r}^{\infty} \mu_j(A).$$

Założmy, że $0 < \Lambda(A) < \infty$ i dla dowolnego $\epsilon > 0$ rozważmy $n = \lfloor e^{r(\Lambda(A)-\epsilon)} \rfloor$. Skoro dla dostatecznie dużych j , $\mu_j(A) \leq e^{-j(\Lambda(A)-\epsilon/2)}$, to dla dostatecznie dużych r

$$\mathbb{P}[T_r \leq n_r] \leq n_r \sum_{j=r}^{\infty} e^{-j(\Lambda(A)-\epsilon/2)} \leq e^{-\epsilon r/2}.$$

Lemat Borela-Cantelliego daje $\liminf_r \log(T_r)/r \geq \Lambda(A)$ \mathbb{P} -prawie na pewno. Jeżeli $\Lambda(A) = \infty$, to ten sam argument z $n = \Lambda(A) - \epsilon$ zastąpionym przez dowolnie dużą stałą daje $\liminf_r \log(T_r)/r \geq \Lambda(A)$ \mathbb{P} -prawie na pewno. W przypadku, gdy $\Lambda(A) = 0$ ostatnie nierówność jest oczywista. Pokażemy wreszcie

$$\limsup_{r \rightarrow \infty} \log(T_r)/r \leq \Lambda(A)$$

Rozważmy w tym celu zmienną losową

$$\sigma_r = \inf\{l : (S_{lr} - S_{(l-1)r})/r \in A\}.$$

Skoro zmienne $\{S_{lr} - S_{(l-1)r}\}_{l \in \mathbb{N}}$ są iid widzimy, że σ_r ma rozkład geometryczny parametrem $\mu_r(A)$. Stąd dla dowolnego n ,

$$\mathbb{P}[\sigma_r > n] = (1 - \mu_r(A))^n \leq \exp\{-n\mu_r(A)\}.$$

Rozważmy teraz $n = n_r = \lfloor e^{r(\Lambda(A)+\epsilon)} \rfloor$. Skoro $\mu_r(A) \geq e^{-r(\Lambda(A)+\epsilon/2)}$ dla dostatecznie dużych r , to

$$\mathbb{P}[\sigma_r > n] \leq \exp\{-e^{r\epsilon/2}\}$$

Powołując się raz jeszcze na Lemat Borella-Cantelliego $\sigma_r \geq n_r$ dla dostatecznie dużych r . Stąd

$$T_r \leq r\sigma_r \leq rn_r$$

dla dostatecznie dużych r a co za tym idzie

$$\limsup_{r \rightarrow \infty} \log(T_r)/r \leq \Lambda(A) + \epsilon.$$

Rozważając ostatnie przejście graniczne $\epsilon \rightarrow \infty$ kończymy dowód. \square

Przykład 2.29

Rozważmy ciąg niezależnych zmiennych losowych X_1, X_2, \dots o wspólnym rozkładzie

$$\mathbb{P}[X_1 = 1/2] = \mathbb{P}[X_1 = 3/2] = 1/2.$$

Chcemy zbadać asymptotyczne zachowanie $\mathbb{E}[(S_n/n)^n]$. Z jednej strony, patrząc na prawo wielkich liczb $S_n/n \rightarrow 1$, można przewidywać, że $\mathbb{E}[(S_n/n)^n]$ zbiega do jedynek. Nie jest to jednak prawdą. Z Twierdzenia 1.1 funkcją intensywności jest

$$I(a) = \begin{cases} \log(2) + \left(a - \frac{1}{2}\right) \log\left(a - \frac{1}{2}\right) + \left(\frac{3}{2} - a\right) \log\left(\frac{3}{2} - a\right), & a \in \left[\frac{1}{2}, \frac{3}{2}\right] \\ +\infty, & \text{poza tym} \end{cases}.$$

Mamy

$$\begin{aligned} \mathbb{E}[(S_n/n)^n] &= \int_0^\infty na^{n-1}\mathbb{P}[S_n/n \geq a]da = \int_{1/2}^{3/2} na^{n-1}\mathbb{P}[S_n \geq na]da \\ &\simeq \int_{1/2}^{3/2} e^{n(\log(a)-I(a))}da \simeq \exp\left\{n \sup_{a>0}[\log(a) - I(a)]\right\}. \end{aligned}$$

Stąd

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E}[(S_n/n)^n] = b, \quad \text{gdzie } b = \sup_{a>0}[\log(a) - I(a)].$$

Jeżeli oznaczymy przez a^* oznaczymy argument dla którego przyjęty jest ostatni kres to można sprawdzić, że (zadanie) $a^* > 1$ oraz $b > 0$. Okazuje się więc, że zachowanie $\mathbb{E}[(S_n/n)^n]$ jest dyktowane nie przez typowe zdarzenia, ale przez zdarzenia rzadkie dla których $S_n \approx a^*n$.

Ogólna teoria

Streszczenie Przedstawimy teorię wielkich odchyłeń na przestrzeniach polskich.

Niech S będzie ośrodkową przestrzenią topologiczną metryzowalną w sposób zupełny. Przez d oznaczajmy zupełną metrykę na S oraz dla $x \in S$, $\epsilon > 0$,

$$B_\epsilon(x) = \{y \in S : d(x, y) < \epsilon\}.$$

Definicja 3.1

Funkcję $f: S \rightarrow [-\infty, \infty]$ nazywamy półciągłą z dołu jeżeli spełniony jest jeden z następujących warunków

(i) Dla każdego $x_n \rightarrow x$ w S zachodzi

$$\liminf_{n \rightarrow \infty} f(x_n) \geq f(x).$$

(ii) Dla każdego $x \in S$,

$$\lim_{\epsilon \rightarrow 0^+} \inf_{y \in B_\epsilon(x)} f(y) = f(x).$$

(iii) f ma domknięte podpoziomice: dla każdego $c \in \mathbb{R}$ zbiór

$$f^{-1}[-\infty, c] = \{y \in S : f(y) \leq c\}$$

jest domknięty.

Lemat 3.2

Warunki obecne w Definicji 3.1 są równoważne. Dodatkowo każda półciągła z dołu funkcja przyjmuje swój dolny kres na niepustych zbiorach zwartych.

Dowód. Zadanie.

□

Definicja 3.3

Powiemy, że $I: S \rightarrow [0, \infty]$ jest funkcją tempa, jeżeli

(D1) $I \neq \infty$,

(D2) I jest półciągłe z dołu.

Powiemy, że funkcja tempa I jest dobra, jeżeli dodatkowo

(D3) I ma zwarte podpoziomice.

Rozważać będziemy miary probabilistyczne na S . W tym celu omówimy pokrótce strukturę borelowską na S . Przez Σ oznaczają będziemy σ -ciało borelowskich podzbiorów S , czyli

$$\Sigma = \sigma(U : U \subseteq \text{Sotwarty}).$$

Przypomnijmy, że U jest otwarty, jeżeli może być zapisany jako dowolna suma kul otwartych $B_\epsilon(x)$. Zauważmy, że wobec zakładanej ośrodkowości S każdy zbiór otwarty może być wyrażony jako przeliczalna suma otwartych kul. W konsekwencji Σ jest generowane przez kule, czyli

$$\Sigma = \sigma(B_\epsilon(x) : x \in S, \epsilon > 0).$$

Miarą probabilistyczną na S nazywamy będziemy funkcję $\mathbf{P}: \Sigma \rightarrow [0, 1]$ taką, że $\mathbf{P}[S] = 1$ oraz

$$\mathbf{P} \left[\bigcup_{n \geq 0} A_n \right] = \sum_{n \geq 0} \mathbf{P}[A_n]$$

dla dowolnych parami rozłącznych $\{A_n\}_{n \in \mathbb{N}}$ z Σ .

Definicja 3.4

Powiemy, że ciąg miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ i funkcją tempa I , jeżeli

(D1') Dla każdego domkniętego $C \subseteq S$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[C] \leq -I(C) = - \inf_{x \in C} I(x).$$

(D2') Dla każdego otwartego $O \subseteq S$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[O] \geq -I(O) = -\inf_{x \in O} I(x).$$

Przykład 3.5

Niech $S = \mathbb{R}$. Dla zmiennych iid $\{X_k\}_{k \in \mathbb{N}}$ połóżmy $S_n = X_1 + \dots + X_n$. Wówczas ciąg

$$\mathbf{P}_n[\cdot] = \mathbb{P}[S_n/n \in \cdot]$$

spełnia zasadę wielkich odchyłeń z prędkością n i funkcją intensywności $I(a) = \Lambda^*(a)$, gdzie

$$\Lambda^*(a) = \sup_{\lambda \in \mathbb{R}} \left(\lambda a - \log \mathbb{E} \left[e^{\lambda X_1} \right] \right).$$

Zauważmy, że jeżeli $\mathbb{E}[|X_1|] < \infty$, ale $\mathbb{E}[e^{\lambda X}] = \infty$ dla każdej $\lambda > 0$, to dla $x > \mathbb{E}[X_1]$,

$$\Lambda^*(x) = \sup_{\lambda \geq 0} \left(\lambda x - \mathbb{E} \left[e^{\lambda X_1} \right] \right) = 0.$$

Oznacza to, że $(\mathbb{E}[X_1], \infty) \subseteq (\Lambda^*)^{-1}[-\infty, 0]$. Funkcja tempa nie jest zatem dobra.

Przykład 3.6

Założmy, że Γ jest zbiorem skończonym. Dla $S = M_1(\Gamma)$. Dla $\{X_k\}_{k \in \mathbb{N}}$ iid z rozkładem ρ ciąg miar

$$\mathbf{P}_n[\cdot] = \mathbb{P} \left[\frac{1}{n} \sum_{k=1}^n \delta_{X_k} \in \cdot \right]$$

spełnia zasadę wielkich odchyłeń z prędkością n oraz funkcją tempa $I(\nu) = H(\nu|\rho)$. Jeżeli $\rho(\{s\}) > 0$ dla każdego $s \in \Gamma$, to I jest dobrą funkcją tempa.

Twierdzenie 3.7

Założmy, że $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyłeń. Wówczas stwarzająca funkcja intensywności jest jedyna.

Dowód. Założmy, że I oraz J są funkcjami intensywności dla ciągu $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$. Chcemy pokazać, że $J(x) = I(x)$ dla każdego $x \in S$. Rozważmy ciąg kul otwartych $B_N = B_{1/N}(x)$, $N \in \mathbb{N}$. Wówczas

$$\begin{aligned} -I(x) &\leq -I(B_{N+1}) \leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[B_{N+1}] \\ &\leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[\text{cl}(B_{N+1})] \leq -J(\text{cl}(B_{N+1})) \leq -J(B_N), \end{aligned}$$

gdzie w ostatniej nierówności skorzystaliśmy z zawierania $\text{cl}(B_{N+1}) \subseteq B_N$. Przechodząc z N do nieskończoności i korzystając z półciągłości J z dołu

$$\lim_{N \rightarrow \infty} J(B_N) = J(x).$$

Otrzymujemy w ten sposób nierówność $I(x) \geq J(x)$. Przez symetryczny argument otrzymujemy $J(x) \geq I(x)$ a co za tym idzie $I(x) = J(x)$. \square

Fakt 3.8

Ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyłeń z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ oraz funkcją tempa I wtedy i tylko wtedy, gdy spełnione są dwa warunki

(D1'') Dla każdej $\alpha < \infty$ i każdego mierzalnego $\Gamma \subseteq S$ takiego, że $\text{cl}(\Gamma) \subset I^{-1}(\alpha, +\infty)$

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n[\Gamma] \leq -\alpha$$

(D2'') Dla każdego $x \in S$ takiego, że $I(x) < \infty$ i dowolnego mierzalnego $\gamma \subseteq S$ takiego, że $x \in \gamma$,

$$\liminf_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n[\gamma] \geq -I(x).$$

Dowód. Ćwiczenie. \square

Jedną z możliwych strategii przy dowodzeniu, że ciąg miar probabilistycznych spełnia zasadę wielkich odchyłeń jest uzasadnienie górnego ograniczenia dla zbiorów zwartych. To uzasadnia poniższą definicję.

Definicja 3.9

Powiemy, że ciąg miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia słabą zasadę wielkich odchyłeń z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ i funkcją tempa I , jeżeli

(D1') Dla każdego zwartego $K \subseteq S$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[K] \leq -I(K) = -\inf_{x \in K} I(x).$$

(D2') Dla każdego otwartego $O \subseteq S$,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbf{P}_n[O] \geq -I(O) = -\inf_{x \in O} I(x).$$

Przykład 3.10

Rozważmy ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ na $S = \mathbb{R}$ zadany wzorem

$$\mathbf{P}_n = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_n.$$

Wówczas $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia słabą zasadę wielkich odchyień z dobrą funkcją tempa $I(0) = 0$ i $I(x) = \infty$. Dla zwartego $K \subseteq \mathbb{R}$ mamy bowiem $\mathbf{P}_n[K] = \delta_0(K)/2$ dla dostatecznie dużych n . Stąd

$$\log \mathbf{P}_n[K] = - \inf_{x \in K} I(x)$$

dla dostatecznie dużych n . Dla otwartego $O \subseteq \mathbb{R}$ mamy

$$\log \mathbf{P}_n[O] = \begin{cases} -\infty & 0, n \in O \\ 0 & \text{poza tym} \end{cases} \geq \begin{cases} -\infty & 0 \in O \\ 0 & \text{poza tym} \end{cases} = - \inf_{x \in O} I(x).$$

Zauważmy, że dla $C = [1, +\infty)$,

$$\log \mathbf{P}_n[C] = 0 > -\infty = - \inf_{x \in C} I(x).$$

Ciąg ten nie spełnia więc (silnej) zasady wielkich odchyień.

Widzimy, że przejście ze słabej do silnej zasady wielkich odchyień wymaga uzasadnienia, że ciąg miar koncentruje większość swych mas na zbiorach zwartych.

Definicja 3.11

Powiemy, że ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ na S jest wykładniczo ciasny z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$, jeżeli dla każdej $\alpha > 0$ istnieje zwarty $K_\alpha \subseteq S$ taki, że

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n[S \setminus K_\alpha] \leq -\alpha.$$

Fakt 3.12

Jeżeli ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ na S jest wykładniczo ciasny z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ spełnia słabą zasadę wielkich odchyień (z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$), to spełnia (silną) zasadę wielkich odchyień (z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$).

Dowód. Zadanie. □

3.1 Lemat Varadhana

Sformułowanie zasady wielkich odchyień jak i związana z nią wykładnicza ciasność są pewną wykładniczą analogią dla słabej zbieżności. Przypomnijmy, że ciąg miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ zbiega słabo do miary probabilistycznej \mathbf{P} jeżeli dla każdego domkniętego $C \subseteq S$,

$$\limsup_{n \rightarrow \infty} \mathbf{P}_n[C] \leq \mathbf{P}[C]$$

lub równoważnie dla każdego otwartego $O \subseteq S$,

$$\liminf_{n \rightarrow \infty} \mathbf{P}_n[O] \geq \mathbf{P}[O].$$

Zbadamy te analogie nieco dokładniej. Przypomnijmy, że słaba zbieżność $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ do \mathbf{P} jest równoważna

$$\lim_{n \rightarrow \infty} \mathbf{P}_n[A] = \mathbf{P}[A]$$

dla każdego $A \subseteq S$ takiego, że $\mathbf{P}[\partial A] = 0$. Podobny warunek można sformułować dla wielkich odchyień. Mianowicie, jeżeli ciąg $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień, to

$$\lim_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n[A] = I(A)$$

dla każdego A dla którego

$$I(\text{cl}(A)) = \inf_{x \in \text{cl}(A)} I(x) = I(\text{int}(A)) = \inf_{x \in \text{int}(A)} I(x).$$

Zauważmy, że dla ciągłej funkcji I ostatnie warunek jest spełniony dla każdego A takiego, że $A \subseteq \text{cl}(\text{int}(A))$ (porównaj z zadaniem ...).

Jest jeszcze jedno sformułowanie słabej zbieżności $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ do \mathbf{P} . Dla dowolnej ograniczonej funkcji ciągłej $F: S \rightarrow \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \int_S F(x) \mathbf{P}_n[dx] = \int_S F(x) \mathbf{P}[dx].$$

Nasz następny rezultat jest wykładniczym analogonem powyższego warunku.

Twierdzenie 3.13 (Lemat Varadhana)

Załóżmy, że rodzina miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ oraz funkcją intensywności I . Niech $F: S \rightarrow \mathbb{R}$ będzie ciągłą funkcją ograniczoną z góry. Wówczas

$$\lim_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbf{P}_n(dx) = \sup_{x \in S} [F(x) - I(x)].$$

Dowód. Niech

$$J_n(A) = \int_A e^{\gamma_n F(x)} \mathbf{P}_n(dx), \quad A \in \Sigma$$

oraz

$$a = \sup_{x \in S} F(x), \quad b = \sup_{x \in S} [F(x) - I(x)].$$

Zauważmy, że $b \leq a$.

Oszacowanie górne: Zastosujemy rozbitcie S . Niech $C = F^{-1}([b, a])$. Dla $N \in \mathbb{N}$ połączmy

$$C_j^N = F^{-1} [c_{j-1}^N, c_j^N], \quad j = 1, 2, \dots, N,$$

gdzie $c_j^N = b + j(a - b)/N$. Wówczas

$$C = \bigcup_{j=1}^N C_j^N.$$

Skoro F jest ciągła, to wszystkie zbiory C_j^N są domknięte. Z zasady wielkich odchyłeń dla każdego j ,

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n [C_j^N] \leq -I(C_j^N).$$

Biorąc pod uwagę, że $F(x) \leq c_j^N$ dla $x \in C_j^N$ otrzymujemy

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(C) \leq \max_{1 \leq j \leq N} [c_j^N - I(C_j^N)].$$

Oszacowanie to można udoskonalić poprzez wykorzystanie nierówności

$$c_j^N \leq \inf_{x \in C_j^N} F(x) + (a - b)/N$$

w następujący sposób

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(C) &\leq \max_{1 \leq j \leq N} \left\{ \inf_{x \in C_j^N} F(x) - \inf_{x \in C_j^N} I(x) \right\} + \frac{1}{N}(a - b) \\ &\leq \max_{1 \leq j \leq N} \sup_{x \in C_j^N} [F(x) - I(x)] + \frac{1}{N}(b - a) \\ &= \sup_{x \in C} [F(x) - I(x)] + \frac{1}{N}(a - b) \leq b + \frac{1}{N}(a - b). \end{aligned}$$

Przechodząc z $N \rightarrow \infty$ otrzymujemy

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(C) \leq b.$$

Z drugiej strony skoro

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(S \setminus C) \leq b,$$

otrzymujemy

$$\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(S) \leq b.$$

Oszacowanie dolne: Ustalmy $x \in S$ i rozważmy $\epsilon > 0$. Zbiór

$$O_{x,\epsilon} = \{y \in S : F(y) > f(x) - \epsilon\}$$

jest otwartym otoczeniem x . Z zasady wielkich odchyień

$$\liminf_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \mathbf{P}_n[O_{x,\epsilon}] \geq -I(O_{x,\epsilon}).$$

Skoro $I(O_{x,\epsilon}) > I(x)$, to

$$\liminf_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n[O_{x,\epsilon}] \geq F(x) - I(x) - \epsilon.$$

Skoro $J_n(S) \geq J_n(O_{x,\epsilon})$, to wymierając $n \rightarrow \infty$ a następnie $\epsilon \rightarrow 0^+$ i wreszcie supremum po wszystkich $x \in S$,

$$\liminf_{n \rightarrow \infty} \frac{1}{\gamma_n} \log J_n(S) \geq b.$$

□

Wniosek 3.14

Założmy, że rodzina miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ oraz funkcją intensywności I . Niech $F: S \rightarrow \mathbb{R}$ będzie ciągłą funkcją taką, że

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbb{1}_{\{F(x) > M\}} \mathbf{P}_n(dx) = \sup_{x \in S} [F(x) - I(x)].$$

Wówczas

$$\lim_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbf{P}_n(dx) = \sup_{x \in S} [F(x) - I(x)].$$

Dowód. Niech

$$A(M) = \limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbb{1}_{\{F(x) > M\}} \mathbf{P}_n(dx) = \sup_{x \in S} [F(x) - I(x)].$$

Wówczas $A(M) \rightarrow 0$ przy $M \rightarrow \infty$. Z nierówności

$$e^{n(F(x) \wedge M)} \leq e^{nF(x)} \leq e^{n(F(x) \wedge M)} + e^{nF(x)} \mathbb{1}_{\{F(x) > M\}}$$

i Lematu Varadhana otrzymujemy

$$\begin{aligned} \sup_{x \in S} [F(x) \wedge M - I(x)] &\leq \liminf_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbf{P}_n(dx) \leq \\ &\limsup_{n \rightarrow \infty} \frac{1}{\gamma_n} \log \int_S e^{\gamma_n F(x)} \mathbf{P}_n(dx) \leq \sup_{x \in S} [F(x) \wedge M - I(x)] + A(M). \end{aligned}$$

Przechodząc z M do nieskończoności i korzystając, że z monotoniczności $\sup_{x \in S} [F(x) \wedge M - I(x)] \rightarrow \sup_{x \in S} [F(x) - I(x)]$ otrzymujemy tezę. \square

Przykład 3.15

Niech $\{X_k\}_{k \in \mathbb{N}}$ będą iid z rozkładem takim, że

$$\Lambda(t) = \log \mathbb{E} \left[e^{tX_1} \right] < \infty$$

dla każdego $t \in \mathbb{R}$. Wówczas z Twierdzenia Craméra $\mathbf{P}_n[\cdot] = \mathbb{P}[S_n/n \in \cdot]$, gdzie $S_n = X_1 + \dots + X_n$, spełnia zasadę wielkich odchyłeń z prędkością $\gamma_n = n$ oraz funkcją tempa

$$I(x) = \sup_{t \in \mathbb{R}} [tx - \Lambda(t)] = \Lambda^*(x).$$

Ustalmy $\lambda \in \mathbb{R}$. Niech $F(x) = \lambda x$. Skoro

$$\int_S e^{nF(x)} \mathbb{1}_{\{F(x) > M\}} \mathbf{P}_n(dx) = \mathbb{E} \left[e^{\lambda S_n} \mathbb{1}_{\{S_n > Mn\}} \right] \leq e^{-Mn + n\Lambda(\lambda + 1)}.$$

Z wniosku otrzymujemy zatem

$$\frac{1}{n} \log \mathbb{E} \left[e^{\lambda S_n} \right] = \frac{1}{n} \log \int_S e^{nF(x)} \mathbf{P}_n(dx) \rightarrow \sup_{x \in \mathbb{R}} [\lambda x - \Lambda^*(x)].$$

Z drugiej strony powyższy ciąg jest niewątpliwie stałe równy $\Lambda(\lambda)$. Oznacza to, że

$$\Lambda(\lambda) = \sup_{x \in \mathbb{R}} [\lambda x - \Lambda^*(x)] = \Lambda^{**}(\lambda).$$

Zadanie

Niech $\{Z_n\}_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych takich, że

$$\mathbb{P}[Z_n = 0] = 1 - 2p_n, \quad \mathbb{P}[Z_n = \pm m_n] = p_n.$$

Pokaż, że jeżeli

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log p_n = -\infty,$$

to rodzina rozkładów $\{Z_n\}$ jest wykładniczo ciasna i spełnia zasadę wielkich odchyień z funkcją tempa $I(0) = 0$ i $I(x) = \infty$. Załóżmy dodatkowo, że $m_n = -\log(p_n)/n$. Pokaż, że

$$\Lambda(\lambda) = \lim_{n \rightarrow \infty} \log \mathbb{E} \left[e^{\lambda n Z_n} \right] = \begin{cases} 0 & |\lambda| \leq 1 \\ \infty & |\lambda| > 1 \end{cases}$$

oraz $\Lambda^*(x) = |x|$. Zauważ wreszcie, że $\Lambda(\lambda) \neq \sup_x [\lambda x - I(x)]$ oraz, że $\Lambda^* \neq I$.

Lemat Varahdana można odwrócić. Połóżmy

$$\Theta_n(F) = \frac{1}{n} \log \int_S e^{nF(x)} \mathbf{P}_n[dx]$$

dla ciągłej i ograniczonej F .

Twierdzenie 3.16 (Bryc 1990)

Załóżmy, że ciąg miar $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ jest wykładniczo ciasny z prędkością n . Jeżeli granica

$$\Theta(F) = \lim_{n \rightarrow \infty} \Theta_n(F) \in \mathbb{R}$$

istnieje dla każdej ciągłej i ograniczonej F , to $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z funkcją prędkości

$$I(x) = \sup_{F \in C_b(S)} [F(x) - \Theta(F)].$$

Dowód. Zadanie. □

3.2 Duże odchylenia dla zamiany miary

Twierdzenie 3.17

Załóżmy, że rodzina miar probabilistycznych na S , $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z funkcją intensywności I . Niech $F: S \rightarrow \mathbb{R}$ będzie ciągłą funkcją ograniczoną z góry. Połóżmy

$$J_n(D) = \int_D e^{nF(x)} \mathbf{P}_n(dx).$$

Wówczas ciąg miar probabilistycznych na S zadany przez

$$\mathbf{P}_n^F(D) = \frac{J_n(D)}{J_n(S)}$$

spełnia zasadę wielkich odchyłeń z funkcją intensywności

$$I^F(x) = \sup_{y \in S} [F(y) - I(y)] - [F(x) - I(x)].$$

3.3 Reguła kontrakcji

Widzieliśmy już, że zasady wielkich odchyłeń przenoszą się przez odwzorowania ciągłe. Poniżej prezentujemy ogólne sformułowanie tego zjawiska.

Twierdzenie 3.18

Niech S i R będą przestrzeniami polskimi i niech $f: S \rightarrow R$. Rozważmy dobrą funkcję tempa $I: S \rightarrow [0, +\infty]$.

(a) Dla $r \in R$ połóżmy

$$I'(y) = \inf\{I(s) : s \in S, f(s) = r\}.$$

Wówczas $I': R \rightarrow [0, +\infty]$ jest dobrą funkcją tempa.

(b) Jeżeli $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ jest ciągiem miar probabilistycznych na S spełniających zasadę wielkich odchyłeń s prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ oraz funkcją tempa I , to ciąg miar $\{\mathbf{Q}_n\}_{n \in \mathbb{N}}$ zadany przez $\mathbf{Q}_n = \mathbf{P}_n \circ f^{-1}$ spełnia zasadę wielkich odchyłeń z prędkością $\{\gamma_n\}_{n \in \mathbb{N}}$ oraz funkcją tempa I' .

Dowód. i

□

3.4 Twierdzenie Sanova

Przypomnijmy, że zmienną losową nazywamy mierzalne odwzorowanie $(\Omega, \mathcal{F}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. W podobny sposób definiujemy losową funkcję, losową miarę czy ogólnie losowy element przestrzeni mierzalnej. Przypomnijmy, że przez Σ oznaczamy σ -ciało podzbiorów borelowskich S .

Definicja 3.19

Elementem losowym przestrzeni polskiej S nazywamy dowolne mieralne odwzorowanie $X: (\Omega, \mathcal{F}) \rightarrow (S, \Sigma)$. Rozkładem elementu losowego X nazywamy miarę probabilistyczną $\mathbf{P}_X = \mathbb{P} \circ X^{-1}$ na S zadaną przez

$$\mathbf{P}_X[A] = \mathbb{P} \left[X^{-1}[A] \right] = \mathbb{P}[X \in A], \quad A \in \Sigma.$$

Niech $\{Y_n\}_{n \in \mathbb{N}}$ będzie ciągiem niezależnych elementów losowych S . Dla $n \in \mathbb{N}$ rozkład empiryczny $\{Y_k\}_{k \leq n}$ zadaje się jako

$$L_n := \frac{1}{n} \sum_{k=1}^n \delta_{Y_k}, \quad n = 1, 2, 3, \dots$$

gdzie δ_x oznacza miarę Diraca w punkcie x , określoną przez

$$\delta_x(A) = \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases}$$

Podkreślmy w tym miejscu, że $L_n = L_n(\omega)$ jest losową miarą probabilistyczną. Będziemy rozważać L_n jako losowy element przestrzeni

$$M_1(S) = \{ \mathbf{P} : \mathbf{P} \text{ jest miarą probabilistyczną na } (S, \Sigma) \}.$$

Przestrzeń $M_1(S)$ wyposażymy w topologię słabej zbieżności. Przypomnijmy raz jeszcze, że ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ na S zbiega słabo do miary probabilistycznej \mathbf{P} na S jeżeli

$$\lim_{n \rightarrow \infty} \int_S f(s) \mathbf{P}_n(ds) = \int_S f(s) \mathbf{P}(ds)$$

dla każdej funkcji f z klasy $C_b(S)$ ciągłych i ograniczonych odwzorowań $S \rightarrow \mathbb{R}$. Przestrzeń $M_1(S)$ wyposażona w topologię słabej zbieżności, jest ponownie przestrzenią polską (patrz [?]).

Twierdzenie 3.20

Niech $\{Y_n\}_{n \in \mathbb{N}}$ będą niezależnymi i jednakowo rozłożonymi zmiennymi losowymi o rozkładzie \mathbf{P} . Wtedy z prawdopodobieństwem jeden L_n zbiega słabo do \mathbf{P} .

Dowód. Należy pokazać, że

$$\mathbb{P} \left[\forall f \in C_b(S) \lim_{n \rightarrow \infty} \int f(x) L_n(dx) = \int f(x) \mathbf{P}(dx) \right] = 1.$$

Niech $f \in C_b(S)$. Mamy

$$\int_S f(s) L_n(ds) = \frac{1}{n} \sum_{k=1}^n f(Y_k).$$

Skoro $f(Y_k)$ są ograniczonymi zmiennymi losowymi, to z mocnego prawa wielkich liczb ostatnie wyrażenie jest zbieżne z p-stewm jeden do

$$\mathbb{E}[f(Y)] = \int_S f(s) \mathbf{P}(ds).$$

Pokazaliśmy właśnie, że

$$\forall f \in C_b(S) \quad \mathbb{P} \left[\lim_{n \rightarrow \infty} \int f(x) L_n(dx) = \int f(x) \mathbf{P}(dx) \right] = 1.$$

Należy teraz uzasadnić, że z prawdopodobieństwem jeden powyższa zbieżność zachodzi dla wszystkich funkcji $f \in C_b(S)$. Skoro $M_1(S)$ jest ośrodkowa, to dla każdej miary istnieje przeliczalna baza jej otoczeń generująca słabą topologię. Skoro zbiory bazowe z definicji są postaci

$$U(f, \epsilon) = \left\{ \mathbf{Q} \in M_1(S) : \left| \int_S f(s) \mathbf{P}(ds) - \int_S f(s) \mathbf{Q}(ds) \right| < \epsilon \right\}$$

to oznacza to, że istnieje przeliczalna kolekcja funkcji $f_k \in C_b(S)$ o następującej własności: \mathbf{P}_n zbiega słabo do \mathbf{P} wtedy i tylko wtedy, gdy

$$\lim_{n \rightarrow \infty} \int_S f_k(s) \mathbf{P}_n(ds) = \int_S f_k(s) \mathbf{P}(ds)$$

dla każdego $k \in \mathbb{N}$. Stosując tę obserwację dla L_n otrzymujemy tezę, ponieważ

$$\begin{aligned} \mathbb{P} \left[\forall f \in C_b(S) \lim_{n \rightarrow \infty} \int f(x) L_n(dx) = \int f(x) \mathbf{P}(dx) \right] = \\ \mathbb{P} \left[\forall k \in \mathbb{N} \lim_{n \rightarrow \infty} \int f_k(x) L_n(dx) = \int f_k(x) \mathbf{P}(dx) \right] = 1. \end{aligned}$$

Ostatnia równość wynika z tego, że zdarzenie pod prawdopodobieństwem to przeliczalny przekrój zdarzeń o prawdopodobieństwie jeden. \square

Chcemy sformułować twierdzenie Sanowa na S . W tym celu konieczne jest zdefiniowanie entropii.

Definicja 3.21

Dla $\nu, \mu \in M_1(S)$, entropia względna, $H(\nu|\mu)$, jest dana przez

$$H(\nu|\mu) = \begin{cases} \int \log \frac{d\nu}{d\mu} d\nu, & \nu \ll \mu \\ +\infty, & \text{w przeciwnym przypadku.} \end{cases}$$

Tutaj $\frac{d\nu}{d\mu}$ oznacza pochodną Radona-Nikodyma ν względem μ .

Przykład 3.22

$S = \mathbb{R}$, $\mu = \mathcal{N}(0, \sigma^2)$, $\nu = \mathcal{N}(m, \sigma^2)$. Wówczas $\nu \ll \mu$ i dla $x \in \mathbb{R}$,

$$\frac{d\nu}{d\mu}(x) = \frac{\frac{d\nu}{d\lambda}(x)}{\frac{d\mu}{d\lambda}(x)} = \frac{e^{-\frac{(x-m)^2}{2\sigma^2}}}{e^{-\frac{x^2}{2\sigma^2}}} = e^{\frac{2xm - m^2}{2\sigma^2}}.$$

Zatem $H(\nu|\mu) = m^2/(2\sigma^2)$, ponieważ

$$\int \log \frac{d\nu}{d\mu}(x) \nu(dx) = \mathbb{E} \left[\frac{2Ym - m^2}{2\sigma^2} \right] = \frac{2\mathbb{E}[Y]m - m^2}{2\sigma^2}$$

gdzie $Y \stackrel{d}{=} \mathcal{N}(m, \sigma^2)$.

Twierdzenie 3.23 (Sanov)

Założmy, że $\{Y_k\}_{k \in \mathbb{N}}$ są iid z rozkładem μ . Ciąg miar probabilistycznych $\{\mathbb{P} \circ L_n^{-1}\}_{n \in \mathbb{N}}$ spełniają zasadę wielkich odchyłeń z prędkością n i dobrą, wypukłą funkcją intensywności $I(\nu) = H(\nu|\mu)$.

Lemat 3.24

Ustalmy $\mu \in M_1(S)$. Funkcja $H(\cdot|\mu)$ jest nieujemna i ściśle wypukła na $M_1(S)$.

Dowód. Rozważmy wypukłą funkcję $u: \mathbb{R} \rightarrow \mathbb{R}$ daną przez $u(x) = x \log(x)$. Mamy

$$H(\nu|\mu) = \int \log \left(\frac{d\nu}{d\mu} \right) d\nu = \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu = \int u \left(\frac{d\nu}{d\mu} \right) d\mu.$$

Z nierówności Jensena wynika, że $H(\cdot|\mu)$ jest nieujemna, bo

$$H(\nu|\mu) \geq u \left(\int \frac{d\nu}{d\mu} d\mu \right) = u(1) = 0.$$

Dodatkowo równość zachodzi wtedy i tylko wtedy, gdy $dv/d\mu$ jest μ -p.w. stała co ma miejsce wtedy i tylko wtedy, gdy $\nu = \mu$ (obie miary są probablistyczne). Aby pokazać, że $H(\cdot|\mu)$ jest (ściśle) wypukła napiszmy

$$\begin{aligned} H(\alpha\nu_1 + (1-\alpha)\nu_2|\mu) &= \int u \left(\alpha \frac{d\nu_1}{d\mu} + (1-\alpha) \frac{d\nu_2}{d\mu} \right) d\mu \\ &\leq \alpha \int u \left(\frac{d\nu_1}{d\mu} \right) d\mu + (1-\alpha) \int u \left(\frac{d\nu_2}{d\mu} \right) d\mu \\ &= \alpha H(\nu_1|\mu) + (1-\alpha) H(\nu_2|\mu). \end{aligned}$$

Zauważmy, że ponieważ u jest ściśle wypukła, nierówność jest ściśła dla $\nu_1 \neq \nu_2$. Dlatego dla $\nu_1 \neq \nu_2$ mamy

$$H(\alpha\nu_1 + (1-\alpha)\nu_2|\mu) < \alpha H(\nu_1|\mu) + (1-\alpha) H(\nu_2|\mu) \quad (3.1)$$

□

Będzie nam potrzebny jeszcze jeden pomocniczy rezultat.

Lemat 3.25

Ustalmy $\nu, \mu \in M_1(S)$. Wówczas

$$H(\nu|\mu) = \sup_{f \in C_b(S)} \left(\int f d\nu - \log \int e^f d\mu \right).$$

Dowód. Załóżmy najpierw, że $\nu \ll \mu$, czyli istnieje $A \in \mathcal{S}$ takie, że $\nu(A) > 0$ i $\mu(A) = 0$. Istnieje ciągła, nieujemna funkcja ciągła f taka, że $\int f d\nu = 1$ oraz $\mu(\text{supp}(f)) = 0$. Wtedy dla każdego $M > 0$

$$\int Mf d\nu - \log \int e^f d\mu = M.$$

Stąd

$$H(\nu|\mu) = +\infty = \sup_{f \in C_b(S)} \left(\int f d\nu - \log \int e^f d\mu \right).$$

Załóżmy teraz, że $\nu \ll \mu$. Niech $f \in C_b(S)$. Zdefiniujmy $\mu_f \in M_1(S)$ przez

$$\frac{d\mu_f}{d\mu} = \frac{e^f}{\int e^f d\mu}.$$

Mamy

$$\begin{aligned} H(\nu|\mu) &= \int \log \frac{d\nu}{d\mu} d\nu = \int \log \left(\frac{d\nu}{d\mu_f} \cdot \frac{d\mu_f}{d\mu} \right) d\nu \\ &= H(\nu|\mu_f) + \int \log \left(\frac{e^f}{\int e^f d\mu} \right) d\nu \\ &\geq \int f d\nu - \log \int e^f d\mu \end{aligned}$$

Skoro f jest dowolna, to daje

$$H(\nu|\mu) \geq \sup_{f \in \mathcal{C}_b(S)} \left(\int f d\nu - \log \int e^f d\mu \right).$$

Aby pokazać, że $H(\nu|\mu) \leq \sup(\dots)$, weźmy

$$f = \log \frac{d\nu}{d\mu}$$

dla której

$$\int f d\nu - \log \int e^f d\mu = H(\nu|\mu)$$

i przybliżamy funkcję f funkcjami z $\mathcal{C}_b(S)$ odpowiednio. \square

Skoro funkcja

$$\nu \mapsto \int f d\nu - \log \int e^f d\mu$$

jest ciągła dla wszystkich $f \in \mathcal{C}_b(S)$, to $H(\cdot|\mu)$, jako supremum funkcji ciągłych, jest półciągła z dołu.

Dowód Twierdzenia Sanowa: szacowanie dolne. Niech $U \subseteq M_1(S)$ będzie otwarty. Musimy pokazać:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in U] \geq - \inf_{\nu \in U} H(\nu|\mu).$$

Wystarczy zatem uzasadnić, że dla każdego $\nu \in U$ mamy

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in U] \geq -H(\nu|\mu).$$

Ustalmy $\nu \in U$ i bez straty ogólności załóżmy, że $\nu \ll \mu$. Dowód przeprowadzimy w dwóch krokach. Załóżmy najpierw, że $\mu \ll \nu$. Dla $\mathbf{y} = (y_1, \dots, y_n) \in S^n$ niech

$$\ell_n(\mathbf{y}) = \frac{1}{n} \sum_{k=1}^n \delta_{y_k}$$

Wówczas

$$\ell_n(Y_1, \dots, Y_n) = L_n.$$

Mamy

$$\begin{aligned} \mathbb{P}[L_n \in U] &= \int_{\{\ell_n(\mathbf{y}) \in U\}} \mu(d\mathbf{y}_1) \dots \mu(d\mathbf{y}_n) \\ &= \int_{\{\ell_n(\mathbf{y}) \in U\}} \prod_{j=1}^n \frac{d\mu}{d\nu}(y_j) \nu(d\mathbf{y}_1) \dots \nu(d\mathbf{y}_n). \end{aligned}$$

Rozważmy teraz $\{Y'_k\}_{k \in \mathbb{N}}$ iid o rozkładzie ν . Niech

$$L'_n = \frac{1}{n} \sum_{k=1}^n \delta_{Y'_k} = \ell_n(Y'_1, \dots, Y'_n).$$

Nasza reprezentacja dla $\mathbb{P}[L_n \in U]$ zapisuje się w terminach Y'_k jako

$$\mathbb{P}[L_n \in U] = \mathbb{E} \left[\prod_{j=1}^n \frac{d\mu}{d\nu}(Y_j) \mathbb{1}_{\{L'_n \in U\}} \right].$$

Zauważmy, że skoro $\nu \in U$ a $L'_n \rightarrow \nu$ z prawdopodobieństwem jeden, to

$$\lim_{n \rightarrow \infty} \mathbb{P}[L'_n \in U] = 1.$$

Aby zrozumieć zachowanie produktu napiszmy z mocnego prawa wielkich liczb

$$\frac{1}{n} \log \prod_{j=1}^n \frac{d\mu}{d\nu}(Y_j) = \frac{1}{n} \sum_{j=1}^n \log \frac{d\mu}{d\nu}(Y_j) \rightarrow \int \log \left(\frac{d\mu}{d\nu} \right) d\nu = -H(\nu|\mu).$$

Oznacza to, że dla $\epsilon > 0$ zdarzenie

$$B_n(\epsilon) = \left\{ \prod_{j=1}^n \frac{d\mu}{d\nu}(Y_j) \geq e^{-n(H(\nu|\mu) + \epsilon)} \right\}$$

jest asymptotycznie pewne, czyli $\mathbb{P}[B_n(\epsilon)] \rightarrow 1$. Podsumowując

$$\mathbb{P}[L_n \in U] \geq e^{-n(H(\nu|\mu) + \epsilon)} \mathbb{P}[L'_n \in U, B_n]$$

Stąd

$$\liminf_n \frac{1}{n} \log \mathbb{P}(\rho_n \in U) \geq -(H(\nu|\mu) + \epsilon)$$

Przechodząc z $\epsilon \rightarrow 0$ otrzymujemy tezę. Jeśli μ nie jest absolutnie ciągła względem ν , zdefiniujemy ciąg $\nu_k = (1 - 1/k)\nu + \mu/k$ i zauważmy, że $\mu \ll \nu_k$ i, ze względu na wypukłość entropii, $H(\nu_k|\mu) \leq (1 - 1/k)H(\nu|\mu)$. Z pierwszej części dowodu mamy

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in U] \geq -H(\nu_k|\mu) \geq -(1 - 1/k)H(\nu|\mu).$$

Biorąc granicę dla $k \rightarrow \infty$ otrzymujemy tezę. □

Lemat 3.26

Załóżmy, że $\{Y_k\}_{k \in \mathbb{N}}$ są iid z rozkładem μ . Ciąg miar probabilistycznych $\{\mathbb{P} \circ L_n^{-1}\}_{n \in \mathbb{N}}$ spełnia słabą zasadę wielkich odchyłeń z prędkością n i dobrą, wypukłą funkcją intensywności $I(\nu) = H(\nu|\mu)$.

Dowód. Wystarczy uzasadnić górne szacowanie. Niech $A \subseteq M_1(S)$ będzie zbiorem zwartym. Załóżmy bez straty ogólności, że $\inf_{\nu \in A} H(\nu|\mu) > 0$. Wybierzmy $0 < \alpha < \inf_{\nu \in A} H(\nu|\mu)$. Dla każdego $\nu \in A$ istnieje $f_\nu \in \mathcal{C}_b(S)$ takie, że

$$\int f_\nu d\nu - \log \int e^{f_\nu} d\mu > \alpha.$$

Rozważmy otwarte otoczenie ν zadane przez

$$U(\nu) = \left\{ \eta \in M_1(S) : \int f_\nu d\eta - \log \int e^{f_\nu} d\mu > \alpha \right\}.$$

Zauważmy, że skoro Y_k mają rozkład μ , to nierówność Markowa

$$\begin{aligned} \mathbb{P}[L_n \in U(\nu)] &= \mathbb{P} \left[\int f_\nu dL_n > \alpha + \log \int e^{f_\nu} d\mu \right] \\ &= \mathbb{P} \left[e^{n \int f_\nu d\rho_n} > e^{n\alpha + n \log \int e^{f_\nu} d\mu} \right] \\ &\leq e^{-n\alpha} \left[\int e^{f_\nu} d\mu \right]^{-n} \mathbb{E} \left[e^{n \int f_\nu d\rho_n} \right] = e^{-n\alpha}. \end{aligned}$$

Skoro A jest zbiorem zwartym, to z jego pokrycia $\{U(\nu)\}_{\nu \in A}$ możemy wybrać podpokrycie skończone $\{U(\nu_j)\}_{j \leq N}$, $n \in \mathbb{N}$. Stąd

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A] \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \sum_{i=1}^N \mathbb{P}[L_n \in U(\nu_i)] \leq -\alpha$$

Skoro $\alpha < \inf_{\nu \in A} H(\nu|\mu)$ jest dowolna, to

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}[L_n \in A] \leq - \inf_{\nu \in A} H(\nu|\mu).$$

□

Definicja 3.27

Powiemy, że miara probabilistyczna \mathbf{P} na (S, Σ) jest *ciasna* jeżeli dla każdego $\varepsilon > 0$ istnieje zbiór zwarty $K \subset X$ taki, że $\mathbf{P}[K] \geq 1 - \varepsilon$.

Twierdzenie 3.28

Jeżeli S jest ośrodkowa i zupełna, to każda miara probabilistyczna na (S, Σ) jest ciasna.

Lemat 3.29

Każdy całkowicie ograniczony podzbiór przestrzeni metrycznej zupełnej ma zwarte domknięcie.

Przypomnijmy, że $A \subseteq S$ nazywamy całkowicie ograniczonym, jeżeli dla każdego $\varepsilon > 0$, istnieje skończone pokrycie kulami otwartymi o promieniu ε .

Dowód. Pokażemy, że dowolny ciąg elementów z $\text{cl}(A)$ można wybrać podciąg zbieżny. Dla $n \in \mathbb{N}$ niech $\{B_{n,j}\}$ będzie skończonym pokryciem A kulami o promieniu $1/n$. Zauważmy, że wówczas (ze skończoności) $\{B_{n,j}\}$ jest pokryciem $\text{cl}(A)$. Niech x_m będzie dowolnym ciągiem elementów $\text{cl}(A)$. Ponownie powołując się na skończoność pokryć pokazujemy, istnieje k_n taki, że $B_{n+1,k_{n+1}} \subseteq B_{n,k_n}$ oraz B_{n,k_n} zawiera nieskończenie wiele wyrazów ciągu $\{x_m\}$. Wybieramy teraz podciąg $x_{m_n} \in B_{n,k_n}$ z konstrukcji jest to ciąg Cauchy'ego ($\{x_{m_k}\}_{k \geq n} \subseteq B_{n,k_n}$), który jest zbieżny z założenia o zupełności. \square

Dowód Twierdzenia 3.28. Skoro S jest ośrodkowa, to dla każdego $k \in \mathbb{N}$ istnieje przeliczalne pokrycie S kulami $\{A_{k,j}\}_{j \in \mathbb{N}}$ o promieniu $1/k$. Ustalmy $\varepsilon > 0$. Dla każdego ustalonego k niech $n_k \in \mathbb{N}$ będzie na tyle duże aby

$$\mathbf{P} \left[\bigcup_{j \geq n_k} A_{k,j} \right] > 1 - \varepsilon 2^{-k}.$$

Zbiór

$$A = \bigcup_{k \geq 1} \bigcup_{j \leq n_k} A_{k,j}$$

jest całkowicie ograniczony. Z zupełności S , domknięcie $\text{cl}(A) = K$ jest zbiorem zwartym. Z konstrukcji wynika, że

$$\mathbf{P}[S \setminus K] \leq \sum_{k=1}^{\infty} \mathbf{P} \left[S \setminus \bigcup_{j \neq n_k} A_{k,j} \right] \leq \varepsilon.$$

\square

Dowód Twierdzenia Sanowa: szacowanie górne. Pozostaje udowodnić, że rozkłady $\{L_n\}_{n \in \mathbb{N}}$ są eksponencjalnie zwarte, tj. dla każdego $L > 0$ istnieje zwarty $\exists K_L$, taki że

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} [L_n \in K_L^c] \leq -L.$$

Ustalmy L . Skoro S jest przestrzenią Polską, to dla każdego $\varepsilon > 0$ istnieje $B = B_\varepsilon$ zwarty, taki że $\mu(B^c) = b < \varepsilon$. Niech

$$A = \{\nu : \nu(B^c) \leq \varepsilon\} \subseteq M_1(S).$$

Wówczas A jest domknięty w $M_1(S)$. Istotnie, jeżeli $\nu_n \rightarrow \nu$ w $M_1(S)$ to z testowania zbieżności na zbiorze otwartym

$$\epsilon \geq \liminf_{n \rightarrow \infty} \nu_n(B^c) \geq \nu(B^c).$$

Mamy

$$\mathbb{P}[L_n \notin A] = \mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n 1_{B^c}(Y_i) > \epsilon\right]$$

Ale $1_{B^c}(Y_i), i = 1, 2, \dots$ są niezależnymi zmiennymi losowymi o rozkładzie

$$1_{B^c}(Y_1) = \begin{cases} 1 & \text{z prawdopodobieństwem } b = \mu(B^c) \\ 0 & \text{z prawdopodobieństwem } 1 - b = 1 - \mu(B^c) \end{cases}$$

Twierdzenie Craméra dla iid zmiennych losowych Bernoulliego $W_i = 1_{B^c}(Y_i)$ implikuje

$$\frac{1}{n} \log \mathbb{P}\left[\frac{1}{n} \sum_{i=1}^n 1_{B^c}(Y_i) > \epsilon\right] \leq -h(\epsilon|b),$$

gdzie

$$h(\epsilon|b) = \sup_{\lambda \geq 0} \left(\lambda \epsilon - \log \mathbb{E} \left[e^{\lambda W_1} \right] \right) = \epsilon \log \frac{\epsilon}{b} + (1 - \epsilon) \log \frac{1 - \epsilon}{1 - b}.$$

Dla dowolnego ciągu $0 < b_m < \epsilon_m \rightarrow 0^+, m = 1, 2, \dots$, dobieramy B_{ϵ_m} zwarty jak wyżej oraz $A_{\epsilon_m} = \{\nu : \nu(B_{\epsilon_m}^c) \leq \epsilon_m\}$. Niech

$$K = \bigcap_m A_{\epsilon_m}.$$

Rodzina miar $\{\nu\}_{\nu \in K}$ jest ciasna: dla każdego $\epsilon > 0$ istnieje $\epsilon_m \leq \epsilon$. Wtedy $\nu(B_{\epsilon_m}^c) \leq \epsilon$ dla każdego $\nu \in K$. Z twierdzenie Prohorowa K ciasność jest równoważna z warunkową zwartością. Skoro K jest niewątpliwie domknięty, to musi być zwarty. Mamy dalej

$$\mathbb{P}[L_n \in K^c] \leq \sum_m \mathbb{P}[L_n \in A_{\epsilon_m}^c] \leq \sum_m e^{-nh(\epsilon_m|b_m)}$$

Dla każdego ϵ_m możemy wybrać b_m tak, aby $h(\epsilon_m|b_m) \geq 2Lm$ (ponieważ $\lim_{b \rightarrow 0} h(\epsilon|b) = +\infty$). Stąd

$$\mathbb{P}(L_n \in K^c) \leq \sum_{m=1}^{\infty} e^{-n2Lm} \leq e^{-nL}.$$

Rozkłady (L_n) są eksponencjalnie zwarte. □

3.5 Elementy rachunku prawdopodobieństwa na przestrzeniach Banacha

Niech E będzie przestrzenią Banacha, tj. przestrzenią liniową unormowaną w której metryka zadana przez normę jest zupełna.

Przykład 3.30

Jednym z głównych przykładów, które będziemy badać są funkcje ciągłe $E = C[0, 1]$ z normą $\|f\|_\infty = \sup\{|f(t)| : t \in [0, 1]\}$.

Przez \mathcal{E} oznaczać będziemy σ -ciało podzbiorów borelowskich E . Rozważać będziemy ciąg elementów losowych $\{Y_k\}_{k \in \mathbb{N}}$ w E niezależnych o jednakowym rozkładzie. Chcemy badać asymptotyczne własności ciągu $\{S_n\}_{n \in \mathbb{N}}$ elementów losowych w E zadanego przez

$$S_n = \sum_{k=1}^n Y_k,$$

gdzie sumę należy interpretować jako dodawanie wektorów w E . W pierwszej kolejności skupimy się na prawie wielkich liczb.

3.5.1 Całka Bochnera

Całka Bochnera stanowi naturalne uogólnienie całki Lebesgue'a do przestrzeni Banacha. Opiszemy pokrótce jej konstrukcję. Niech μ będzie dowolną miarą probabilistyczną na E .

Definicja 3.31

Element losowy $X: (\Omega, \mathcal{F}) \rightarrow (E, \mathcal{E})$ nazywamy prostym, jeżeli jest postaci

$$X(\omega) = \sum_{j=1}^n e_j \mathbb{1}_{A_j}$$

dla pewnych $A_j \in \mathcal{F}$ oraz $e_j \in E$ i $n \in \mathbb{N}$. Wartością oczekiwaną X nazywamy

$$\mathbb{E}[X] = \sum_{j=1}^n e_j \mathbb{P}[A_j].$$

W tym momencie należy się na chwilę zatrzymać i upewnić, czy powyższa definicja wartości oczekiwanej jest dobra. Dokładniej, czy wartość $\mathbb{E}[X]$ nie zależy od reprezentacji X w postaci skończonej sumy.

Definicja 3.32

Powiemy, że element losowy $X \in E$ jest całkowalny w sensie Bochnera, jeżeli istnieją proste elementy losowe $X_n \in E$ takie, że

$$\lim_{n \rightarrow \infty} \mathbb{E}[\|X - X_n\|] = 0.$$

Definiujemy wówczas wartość oczekiwaną X poprzez

$$\mathbb{E}[X] = \lim_{n \rightarrow \infty} \mathbb{E}[X_n].$$

Dodatkowo X jest całkowalny w sensie Bochnera wtedy i tylko wtedy, gdy $\mathbb{E}[\|X\|] < \infty$

Fakt 3.33

Powyższa definicja jest poprawna. Dokładniej wartość $\mathbb{E}[X]$ nie zależy od wyboru ciągu aproksymującego $\{X_n\}_{n \in \mathbb{N}}$.

Dowód. Zadanie. □

Fakt 3.34

Niech \mathbf{P}_X będzie rozkładem elementu losowego X . Wówczas $\mathbb{E}[X]$ jest jedynym elementem E takim, że

$$\lambda(\mathbb{E}[X]) = \int \lambda(x) \mathbf{P}_X(dx).$$

Dowód. Zadanie. □

Przykład 3.35

Niech $E := C[0, 1]$. Niech B będzie ruchem Browna. Rozważmy $m = \mathbb{E}[B] \in C[0, 1]$. Dla każdego $t \in [0, 1]$, definiujemy $g_t(x) := x(t)$. Wówczas $g_t \in E^*$. Dla każdego $t \in [0, 1]$,

$$m(t) = g_t(m) = \int g_t(x) \mathbf{P}_B(dx) = \mathbb{E}[B_t] = 0.$$

Twierdzenie 3.36 (Prawo wielkich liczb)

Niech $\{Y_k\}_{k \in \mathbb{N}}$ będzie ciągiem elementów losowych E niezależnych o jednakowym rozkładzie takich, że

$$\mathbb{E}[e^{\alpha \|Y\|}] < \infty$$

dla każdego $\alpha > 0$. Wówczas

$$S_n/n \rightarrow \mathbb{E}[Y].$$

Dowód. Dowód będzie wynikał z zasady wielkich odchyień. \square

3.6 Duże odchylenia na przestrzeniach Banacha

Chcemy sformułować i udowodnić zasadę wielkich odchyień dla średnich empirycznych w przestrzeniach Banacha. W tym celu będziemy chcieli posłużyć się twierdzeniem Sanowa i regułą kontrakcji. Dla $f \in \mathcal{C}_b(E)$ definiujemy $F : M_1(E) \rightarrow \mathbb{R}$ jako $F(\nu) = \int f(x)\nu(dx)$. Obserwujemy

$$\Phi\left(\frac{1}{n}(\delta_{X_1} + \dots + \delta_{X_n})\right) = \int f(y)\left(\frac{1}{n}(\delta_{X_1} + \dots + \delta_{X_n})\right)(dy) = \frac{1}{n} \sum_{i=1}^n f(X_i).$$

Średnia empiryczna odpowiada $f(x) = x$. Jest to funkcja ciągła, lecz nieograniczona, więc $\mu_n \rightarrow \mu$ nie implikuje $\int x d\mu_n \rightarrow \int x d\mu$.

Przykład 3.37

Weźmy $\mu_n = (1 - 1/n)\delta_0 + 1/n\delta_n \rightarrow \delta_0$. Jednak dla $f(x) = x$ mamy $\int f d\mu_n = 1$ dla każdego $n \in \mathbb{N}$ oraz $0 = f(0) = \int f d\delta_0$.

Zatem $\mu \mapsto \int x d\mu(x)$ nie jest ciągła względem topologii słabej zbieżności. Jak się okaże, jest to tylko techniczna trudność.

Twierdzenie 3.38 (Donsker-Varadhan)

Niech E będzie ośrodkową przestrzenią Banacha, $(Y_i)_{i \in \mathbb{N}}$ będzie ciągiem elementów iid o wartościach w E i rozkładzie $\mu \in M_1(E)$. Załóżmy, że dla każdego $\alpha > 0$,

$$\int e^{\alpha\|y\|} \mu(dy) < \infty \tag{3.2}$$

Wówczas rozkłady $\sum_{k=1}^n Y_k/n$ spełniają zasadę wielkich odchyień z dobrną, wypukłą funkcją tempa

$$I(x) = \inf \left\{ H(\nu|\mu) : \nu \in M_1(E), \int y \nu(dy) = x \right\}.$$

Dowód Twierdzenia Donskera-Varadhana dla miar o zwartym nośniku. Załóżmy, że istnieje wypukły, zwarty $K \subseteq E$ taki, że $\mu(K) = 1$. Istnieje funkcja $f \in \mathcal{C}_b(E)$ taka, że $f(x) = x$ dla $x \in K$. Wówczas teza wynika z reguły kontrakcji zastosowanej do omawianej wcześniej funkcji $F(\nu) = \int f(x)\nu(dx)$. \square

Aby udowodnić Twierdzenie Donskera-Varadhana w ogólnym przypadku będzie nam potrzebna ogólniejsza wersja reguły kontrakcji.

Twierdzenie 3.39 (Uogólniona reguła kontrakcji)

Niech $\Gamma, \tilde{\Gamma}$ będą przestrzeniami polskimi i załóżmy, że $(\mu_n)_n \subseteq M_1(\Gamma)$ spełnia zasadę wielkich odchyień z dobrą funkcją tempa I na Γ . Niech $(F_L)_{L>0}$ będzie rosnącym ciągiem zbiorów domkniętych w Γ , niech $F := \bigcup_{L>0} F_L$, i załóżmy, że

$$\mu_n(F) = 1 \quad \text{dla wszystkich } n \in \mathbb{N}$$

i

$$\limsup_n \frac{1}{n} \log \mu_n(F_L^c) \leq -L \quad \text{dla wszystkich } L.$$

Niech $\Phi : F \rightarrow \tilde{\Gamma}$ będzie odwzorowaniem takim, że $\Phi|_{F_L}$ jest ciągłe dla każdego L . Połóżmy

$$\begin{aligned} \tilde{\mu}_n(\tilde{A}) &= \mu_n(x \in F : \Phi(x) \in \tilde{A}) \quad \text{dla } \tilde{A} \subseteq \tilde{\Gamma}. \\ \tilde{I}(\tilde{x}) &= \inf\{I(x) : x \in F, \Phi(x) = \tilde{x}\} \quad \text{dla } \tilde{x} \in \tilde{\Gamma}. \end{aligned}$$

Wówczas $(\tilde{\mu}_n)$ spełnia zasadę wielkich odchyień z dobrą funkcją tempa \tilde{I} .

Dowód. Pierwszy krok: \tilde{I} dobrą funkcją tempa.

Weźmy $\tilde{x} \in \tilde{E}$ takie, że $\tilde{I}(\tilde{x}) < \infty$. Wtedy istnieje $x_0 \in F$ takie, że $\Phi(x_0) = \tilde{x}$ i $I(x_0) < \infty$. Rozważmy $L := I(x_0) + 1$ i zauważmy, że z dolnym ograniczeniem LDP mamy

$$-L \geq \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(F_L^c) \geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(F_L^c) \geq -I(F_L^c).$$

Oznacza to, że $I^{-1}((-\infty, I(x_0)]) \subseteq F_L$, a przez nasz wybór L otrzymujemy

$$\begin{aligned} \tilde{I}(\tilde{x}) &= \inf\{I(x) : x \in F, \Phi(x) = \tilde{x}\} \\ &= \inf\{I(x) : \Phi|_{F_L}^{-1}(\{\tilde{x}\}) \cap F_L \cap I^{-1}((-\infty, I(x_0)])\}. \end{aligned}$$

Infimum jest brane na zbiorze zwartym: $\Phi|_{F_L}^{-1}(\{\tilde{x}\})$ jest obrazem zbioru domkniętego, zatem jest domknięty w topologii podprzestrzeni, a ponadto F_L sam w sobie domknięty, a zbiory podpoziomicowe I są zwarte. Ale to oznacza (patrz zadanie domowe 4), że istnieje pewne $x \in E$ takie, że $I(x) = \tilde{I}(\tilde{x})$. Zatem

$$\tilde{I}((-\infty, \alpha]) = \Phi|_{F_L}(I^{-1}(-\infty, \alpha])$$

gdzie, powiedzmy, $L := \alpha + 1$, a zbiór po lewej stronie jest zwarty jako obraz zbioru zwartego przez funkcję ciągłą.

Drugi krok: Dolne ograniczenie LDP.

Niech \tilde{A} będzie zbiorem otwartym, a $\tilde{x} \in \tilde{A}$ takim, że $\tilde{I}(\tilde{x}) < \infty$. Istnieje wtedy $x \in E$ takie, że $I(x) = \tilde{I}(\tilde{x})$ i $x \in F_L$ dla pewnego L . Poprzez zwiększenie L możemy założyć, że

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(F_L^c) \leq -I(x) - 1. \quad (3.3)$$

Teraz $\Phi|_{F_L}$ jest ciągła, a $(\Phi|_{F_L})^{-1}(\tilde{A})$ jest otwarty w topologii podprzestrzeni. To oznacza, że możemy znaleźć zbiór otwarty A taki, że $(\Phi|_{F_L})^{-1}(\tilde{A}) = A \cap F_L$. Ponadto $x \in A$, więc

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \geq -I(A) \geq -I(x).$$

Ale to wraz z (3.3) implikuje

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A \cap F_L) &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \\ &\leq \liminf_{n \rightarrow \infty} \frac{1}{n} \log [\mu_n(A \cap F_L) + \mu_n(F_L^c)] \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A \cap F_L) \end{aligned}$$

Zatem musimy mieć równość wszędzie. Ostatecznie mamy

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\mu}_n(\tilde{A}) &\geq \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A \cap F_L) = \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(A) \\ &\geq -I(x) = -\tilde{I}(\tilde{x}). \end{aligned}$$

Teraz wystarczy rozważyć supremum po $x \in \tilde{A}$.

Ostatni krok: Górne ograniczenie LDP.

Niech \tilde{B} będzie zbiorem domkniętym. Załóżmy najpierw, że $0 < \tilde{I}(\tilde{B}) < \infty$, i niech $L := \tilde{I}(\tilde{B})$. Wtedy

$$\limsup_n \frac{1}{n} \log \tilde{\mu}_n(\tilde{B}) \leq \limsup_n \frac{1}{n} \log [\mu_n(\Phi|_{F_L}^{-1}(\tilde{B}) \cap F_L) + \mu_n(F_L^c)]$$

Ponieważ $\Phi|_{F_L}$ jest ciągła, a F_L jest domknięty, to $\Phi|_{F_L}^{-1}(\tilde{B}) \cap F_L$ jest także domknięty. Z górnym ograniczeniem LDP mamy

$$\limsup_n \frac{1}{n} \log \tilde{\mu}_n(\tilde{B}) \leq -\inf\{I(x) : x \in F_L, \Phi(x) \in \tilde{B}\} \vee (-L)$$

$$\leq -\inf\{I(x) : x \in F, \Phi(x) \in \tilde{B}\} \vee (-L) = -\tilde{I}(\tilde{B})$$

Na koniec rozważmy przypadek $\tilde{I}(\tilde{B}) = \infty$. Ponieważ $I^{-1}((-\infty, L]) \subseteq F_L$ dla wszystkich L , jasne jest, że $I(x) = \infty$ implikuje $x \notin F$, więc $\tilde{B} \subseteq F^c$ w tym przypadku. Ale mamy $\mu_n(F^c) = 0$ dla wszystkich n . \square

Skorzystamy z Twierdzenia Sanova oraz uogólnionej zasady kontrakcji z odwzorowaniem

$$\Phi : \tilde{M}_1(E) \rightarrow E, \nu \mapsto m(\nu)$$

gdzie $\tilde{M}_1(E)$ to zbiór wszystkich μ w $M_1(E)$ takich, że $\int \|x\| \mu(dx) < \infty$, a $m(\nu) = \int x \nu(dx)$. Musimy zatem skonstruować zbiory F_L jak w sformułowaniu zadany odbicia. Połóżmy

$$g(\alpha) = \log \int e^{\alpha \|x\|} \mu(dx)$$

oraz

$$g^*(x) = \sup_{\alpha > 0} \{\alpha \|x\| - g(\alpha)\}.$$

Zauważmy, że (3.2) implikuje

$$\lim_{\|x\| \rightarrow \infty} \frac{g^*(x)}{\|x\|} = \infty. \quad (3.4)$$

Definiujemy

$$F_L := \left\{ \nu \in M_1(E) : \int g^*(x) \nu(dx) \leq L \right\}.$$

Oznaczmy przez $F = \bigcup_{L > 0} F_L$, a μ_n rozkład

$$L_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}.$$

Jasne jest, że

$$\mu_n(F) = \mathbb{P}[L_n \in F] = \mathbb{P}[g^*(Y_i) < \infty \text{ dla wszystkich } i \leq n] = 1.$$

Lemat 3.40

Zbiór F_L jest domknięty dla każdego $L > 0$.

Dowód. Ponieważ $g^*(\cdot)$ jest półciągła z dołu, istnieje ciąg $(g_m)_m \subseteq \mathcal{C}_b(E)$ tak, że $g_m \nearrow g^*$. Niech $(\nu_n)_n \subseteq F_L$, będą zbieżne słabo do ν . Sprawdzamy, że $\nu \in F_L$:

$$\int g^* d\nu = \sup_m \int g_m d\nu = \sup_m \lim_{n \rightarrow \infty} \int g_m d\nu_n \leq \sup_m \int g^* d\nu_n \leq L$$

\square

Lemat 3.41

Załómy (3.2). Funkcja $m : M_1(E) \rightarrow E, \nu \mapsto m(\nu)$ jest ciągła na F_L , dla dowolnego L .

Dowód Lematu 3.41. Niech $(\nu_n) \subseteq F_L$ będą zbieżne słabo do ν . Musimy pokazać, że $m(\nu_n) \rightarrow m(\nu) \in E$. Szacujemy

$$\|m(\nu_n) - m(\nu)\| = \sup_{\lambda \in E^*, \|\lambda\| \leq 1} \langle \lambda, m(\nu_n) - m(\nu) \rangle \quad (3.5)$$

$$\begin{aligned} &= \sup_{\lambda \in E^*, \|\lambda\| \leq 1} \int \langle \lambda, x \rangle (\nu_n(dx) - \nu(dx)) \\ &\leq \sup_{\lambda \in E^*, \|\lambda\| \leq 1} \int_{\|x\| > R} \langle \lambda, x \rangle (\nu_n(dx) - \nu(dx)) \\ &\quad + \sup_{\lambda \in E^*, \|\lambda\| \leq 1} \int_{\|x\| \leq R} \langle \lambda, x \rangle (\nu_n(dx) - \nu(dx)) \quad (3.6) \end{aligned}$$

Pierwszy składnik w (3.6) możemy oszacować przez

$$\leq 2 \sup_{\eta \in F_L} \int_{\|x\| > R} \|x\| \eta(dx) \leq 2 \underbrace{\sup_{\|x\| > R} \frac{\|x\|}{g^*(x)}}_{\xrightarrow{R \rightarrow \infty} 0 \text{ ze względu na (3.4)}} \underbrace{\sup_{\eta \in F_L} \int g^*(x) \eta(dx)}_{\leq L \text{ bo } \eta \in F_L} \rightarrow 0$$

Aby oszacować drugi składnik wybierzmy h_R ciągłą taką, że $h_R(x) = 1$ jeśli $\|x\| \leq R$, $h_R(x) = 0$ jeśli $\|x\| \geq R + 1$, oraz $0 \leq h_R(x) \leq 1$ w przeciwnym razie. Dla $\lambda \in E^*$ będziemy pisać $h_R^\lambda(x) := h_R(x) \cdot \langle \lambda, x \rangle$. Wtedy rodzina

$$\left\{ h_R^\lambda : \lambda \in E^*, \|\lambda\| \leq 1 \right\}$$

jest jednostajnie ograniczona i jednostajnie ciągła na każdym zbiorze zwartym. Drugi człon w (3.6) szacuje się przez

$$\leq \sup_{\lambda \in E^*, \|\lambda\| \leq 1} \int h_R^\lambda(x) (\nu_n(dx) - \nu(dx)) \xrightarrow{n \rightarrow \infty} 0.$$

Zbieżność do zera wynika z następującego **ogólnego faktu**: Dla $\nu_n \xrightarrow{w} \nu$ oraz rodziny funkcji $(g_s)_{s \in S}$ jednostajnie ograniczonej i jednostajnie ciągłej na każdym zbiorze zwartym, zachodzi

$$\sup_{s \in S} \left| \int g_s d\nu_n - \int g_s d\nu \right| \xrightarrow{n \rightarrow \infty} 0.$$

□

Lemat 3.42

Założmy (3.2). Dla dostatecznie dużego L mamy

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(F_L^c) \leq -L.$$

Dowód Lematu 3.42. Niech $\alpha \in (0, 1)$. Pokażemy, że

$$\int e^{\alpha g^*(x)} \mu(dx) < \infty. \quad (3.7)$$

Wtedy mamy

$$\mathbb{P}(L_n \in F_L^c) = \mathbb{P}\left(\frac{1}{n} \sum_{i=1}^n g^*(Y_i) > L\right) \leq \mathbb{E}\left[e^{\alpha \sum_{i=1}^n g^*(Y_i)}\right] e^{-n\alpha L}.$$

Stąd

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(\rho_n \in F_L^c) \leq \log \int e^{\alpha g^*(y)} \mu(dy) - \alpha L.$$

□

Aby uzasadnić brakujący krok ostatniego dowodu rozważmy

$$\bar{g}(x) = \sup_{\alpha \geq 0} \left[\alpha x - \log \int e^{\alpha \|y\|} \mu(dy) \right], \quad x \in \mathbb{R}$$

i zauważmy, że $g^*(x) = \bar{g}(\|x\|)$. Skorzystamy z następującego lematu.

Lemat 3.43

Niech X będzie zmienną losową o wartościach w $[0, \infty)$. Założmy, że $Z(\lambda) = \mathbb{E}[e^{\lambda X}] < \infty$ dla wszystkich $\lambda \in \mathbb{R}$. Niech $m_0 = \mathbb{E}[X]$ i niech $\bar{g}(z) = \sup_{\lambda \geq 0} [\lambda z - \log Z(\lambda)]$: dla $z > m_0$, $\bar{g}(z) = I(z)$ jest funkcją stopy Craméra, dla $z \leq m_0$, $\bar{g}(z) = 0$. Wówczas

$$\mathbb{E}[e^{a\bar{g}(X)}] < \infty \quad \text{dla wszystkich } a \in (0, 1).$$

Dowód. Zauważmy najpierw, że dla $t > m_0 = \mathbb{E}[X]$,

$$\mathbb{P}(X > t) \leq e^{-I(t)}$$

Teraz mamy

$$\begin{aligned}
\mathbb{E}[e^{a\bar{g}(X)}] &= \int_0^\infty \mathbb{P}[e^{a\bar{g}(X)} > t] dt \\
&= \int_0^{e^{am_0}} \mathbb{P}[e^{a\bar{g}(X)} > t] dt + \int_{e^{am_0}}^\infty \mathbb{P}[e^{a\bar{g}(X)} > t] dt \\
&\leq e^{am_0} + \int_{e^{am_0}}^\infty \mathbb{P}[\bar{g}(X) > \frac{1}{a} \log t] dt \\
&= e^{am_0} + \int_{e^{am_0}}^\infty \mathbb{P}[X > I^{-1}(\frac{1}{a} \log t)] dt \\
&\leq C_1 + \int_{C_2}^\infty e^{-\frac{1}{a} \log t} dt \text{ gdzie skorzystaliśmy z (3.2)} \\
&\leq C_1 + \int_{C_2}^\infty t^{-\frac{1}{a}} dt < \infty \text{ ponieważ } \alpha \in (0,1).
\end{aligned}$$

□

Aby pokazać (3.6), zastosuj lemat z $X = \|Y_1\|$.

Pozostaje udowodnić, że odwzorowanie $x \mapsto I(x) = \inf_{v:m(v)=x} H(v|\mu)$ jest funkcją wypukłą.

Lemat 3.44

Zdefiniujmy $\Gamma_L := \{v : H(v|\mu) \leq L\}$ i załóżmy, że (3.2) zachodzi. Wtedy odwzorowanie $v \mapsto m(v)$ jest ciągłe na Γ_L .

Dowód Lematu 3.44: Ponieważ g^* jest półciągła z dołu, Lemat ?? i (3.7) implikują, że

$$\alpha \int g^*(x) v(dx) \leq H(v|\mu) + \log \int e^{\alpha g^*(x)} \mu(dx) \leq H(v|\mu) + C_\alpha(\mu) < \infty$$

dla $\alpha \in (0,1)$. Stąd wynika, że

$$\int g^*(x) v(dx) \leq \frac{1}{\alpha} H(v|\mu) + C_\alpha(\mu) \frac{1}{\alpha}$$

Ale to oznacza, że $\Gamma_L \subseteq F_K$ dla pewnego $K = K(L)$, i już udowodniliśmy, że m jest ciągłe na F_K . □

Założmy, że $I(x) < \infty$. Pokażemy, że istnieje v_x , dla której $I(x) = H(v_x|\mu)$. Mamy $I(x) = \inf_{v:m(v)=x} H(v|\mu)$. Niech (v_n) będzie ciągiem takim, że $m(v_n) = x$ dla wszystkich n i $H(v_n|\mu) \rightarrow I(x)$. Ponieważ $v \mapsto H(v|\mu)$ jest dobrą funkcją tempa, zbiór $\{v : H(v|\mu) \leq C\}$ jest zwarty dla każdego C . Dlatego istnieje v i podciąg (v_{n_k}) taki, że $v_{n_k} \xrightarrow{w} v$. Ale ponieważ $m(\cdot)$ jest ciągłe na $\{v : H(v|\mu) \leq C\}$, musi zachodzić $m(v) = \lim_{k \rightarrow \infty} m(v_{n_k}) = x$.

Aby pokazać wypukłość, bez utraty ogólności, że $I(x) < \infty$ i $I(y) < \infty$, więc znajdziemy v_x, v_y takie, że $I(x) = H(v_x|\mu)$ i $I(y) = H(v_y|\mu)$. Wówczas z wypukłości $H(\cdot|\mu)$,

$$\begin{aligned}\alpha I(x) + (1 - \alpha)I(y) &= \alpha H(v_x|\mu) + (1 - \alpha)H(v_y|\mu) \\ &\geq H(\alpha v_x + (1 - \alpha)v_y|\mu) \\ &\geq I(\alpha x + (1 - \alpha)y)\end{aligned}$$

3.7 Minimalizacja Entropii

Niech Y_1, Y_2, \dots będą niezależnymi i identycznie rozłożonymi elementami losowymi przestrzeni polskiej S . Niech

$$\rho_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i} \in M_1(S).$$

Niech $B \subset M_1(S), A \in \mathcal{S}$. Chcemy zbadać $\mathbb{P}(Y_1 \in A | \rho_n \in B)$ przy $n \rightarrow \infty$.

Przykład 3.45

$S = \mathbb{R}, Y_1 \stackrel{d}{=} \mathcal{N}(0, 1)$ oraz $B = \nu \in M_1(\mathbb{R}) : \int x\nu(dx) \geq m$ dla pewnego $m > 0$. Wówczas

$$\mathbb{P}(Y_1 \leq t | \rho_n \in B) \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-\frac{(s-m)^2}{2}} ds = \mathbb{P}(Z \leq t),$$

gdzie Z ma rozkład normalny o średniej m i wariancji 1.

Twierdzenie 3.46 (I. Csiszar)

Niech $\mu \in M_1(S), B \subseteq M_1(S)$ będzie zbiorem wypukłym i domkniętym. Załóżmy dodatkowo, że istnieje $\eta \in B$ takie, że $H(\eta|\mu) < \infty$. Wówczas istnieje dokładnie jedno $\bar{\nu} \in B$ takie, że $H(\bar{\nu}|\mu) = \min_{\eta \in B} H(\eta|\mu)$ oraz dla każdego $\eta \in B$ mamy:

$$H(\eta|\mu) \geq H(\eta|\bar{\nu}) + H(\bar{\nu}|\mu) \quad (3.8)$$

Fakt 3.47 (Tożsamość równoległoboku)

Założmy $\mu, \nu, \eta \in M_1(S)$. Wówczas

$$H(\mu|\eta) + H(\nu|\eta) = 2H\left(\frac{\mu + \nu}{2} \middle| \eta\right) + H\left(\mu \middle| \frac{\mu + \nu}{2}\right) + H\left(\nu \middle| \frac{\mu + \nu}{2}\right)$$

Dowód. Ćwiczenie. □

Określmy na $M_1(S)$ metrykę całkowitego wahania $\|\nu - \mu\| = \sup_{A \in \mathcal{S}} |\nu(A) - \mu(A)|$. Przypomnijmy, że na ćwiczeniach udowodniliśmy następujący fakt.

Fakt 3.48 (nierówność Pinskera)

$$\|v - \mu\|^2 \leq \frac{1}{2}H(v|\mu) \quad (3.9)$$

Fakt 3.49

Niech $B \subseteq M_1(S)$ będzie zwarty i domknięty. Załóżmy, że $(v_n) \subseteq B$ takie, że $H(v_n|\mu) \rightarrow \inf_{\eta \in B} H(\eta|\mu)$. Wówczas (v_n) zbiega w normie całkowitego wahania.

Dowód. B jest wypukły, więc $(v_n + v_m)/2 \in B$, co oznacza

$$H\left(\frac{v_n + v_m}{2} \middle| \mu\right) \geq \inf_{\eta \in B} H(\eta|\mu).$$

Stąd, z Faktu 3.47

$$\begin{aligned} H(v_m|\mu) + H(v_n|\mu) - 2 \inf_{\eta \in B} H(\eta|\mu) \\ \geq H(v_m|\mu) + H(v_n|\mu) - 2H\left(\frac{v_n + v_m}{2} \middle| \mu\right) \\ = H\left(v_m \middle| \frac{v_m + v_n}{2}\right) + H\left(v_n \middle| \frac{v_m + v_n}{2}\right) \rightarrow 0. \end{aligned}$$

Zatem,

$$H\left(v_m \middle| \frac{v_m + v_n}{2}\right) \xrightarrow{n,m \rightarrow \infty} 0$$

Stąd,

$$\|v_m - v_n\| \leq \left\|v_m - \frac{v_m + v_n}{2}\right\| + \left\|v_n - \frac{v_m + v_n}{2}\right\| \xrightarrow{n,m \rightarrow \infty} 0.$$

(v_n) jest ciągiem Cauchy'ego względem $\|\cdot\|$. Z zupełności istnieje $\bar{v} := \lim_{n \rightarrow \infty} v_n \in B$, ponieważ B jest domknięty. Zauważmy że skoro zbieżność zachodzi w sensie całkowitego wahania, to

$$\frac{dv_n}{d\mu} \rightarrow \frac{d\bar{v}}{d\mu}$$

μ -prawie wszędzie Mamy

$$H(\bar{v}|\mu) = \int \left(\log \frac{d\bar{v}}{d\mu}\right) \frac{d\bar{v}}{d\mu} d\mu \leq \liminf_{n \rightarrow \infty} H(v_n|\mu).$$

Wiemy, że $v \rightarrow H(v|\mu)$ jest półciągła z dołu, więc

$$H(\bar{v}|\mu) \leq \liminf_{n \rightarrow \infty} H(v_n|\mu) = \inf_{\eta \in B} H(\eta|\mu).$$

Jedynność \bar{v} wynika ze ścisłej wypukłości entropii. □

Lemat 3.50

Niech $\bar{\nu} \in B$. Wówczas

$$H(\bar{\nu}|\mu) = \inf_{\eta \in B} H(\eta|\mu).$$

Wtedy i tylko wtedy, gdy

$$H(\bar{\nu}|\mu) \leq \int \log \left(\frac{d\bar{\nu}}{d\mu} \right) d\eta \quad \forall \eta \in B \text{ takie, że } H(\eta|\mu) < \infty.$$

Dowód. Załóżmy, że $\eta \in B$, $H(\eta|\mu) < \infty$. Niech $\nu_\alpha = \alpha\eta + (1-\alpha)\bar{\nu}$. Wtedy $\nu_\alpha \in B$. Wzważmy teraz funkcję $f_\alpha := \frac{d\nu_\alpha}{d\mu} \log \left(\frac{d\nu_\alpha}{d\mu} \right)$. Odwzorowanie $\alpha \mapsto f_\alpha$ jest funkcją wypukłą oraz

$$\frac{\partial}{\partial \alpha} f_\alpha|_{\alpha=0} = \left(\frac{d\eta}{d\mu} - \frac{d\bar{\nu}}{d\mu} \right) \left(\log \frac{d\bar{\nu}}{d\mu} + 1 \right)$$

Stąd,

$$\begin{aligned} \frac{\partial}{\partial \alpha} H(\nu_\alpha|\mu)|_{\alpha=0} &= \frac{\partial}{\partial \alpha} \left(\int f_\alpha d\mu \right) \Big|_{\alpha=0} \\ &\stackrel{(*)}{=} \int \left(\frac{d\eta}{d\mu} - \frac{d\bar{\nu}}{d\mu} \right) \left(\log \frac{d\bar{\nu}}{d\mu} + 1 \right) d\mu \\ &= \int \log \frac{d\bar{\nu}}{d\mu} d\eta - H(\bar{\nu}|\mu) \end{aligned}$$

A co za tym idzie

$$H(\bar{\nu}|\mu) = H(\nu_\alpha|\mu)|_{\alpha=0} = \min_{\eta \in B} H(\eta|\mu)$$

wtedy i tylko wtedy, gdy dla każdej $\eta \in B$,

$$\frac{\partial}{\partial \alpha} H(\nu_\alpha|\mu)|_{\alpha=0} \geq 0$$

Ostatni warunek jest równoważny

$$\int \log \frac{d\bar{\nu}}{d\mu} d\eta \geq H(\bar{\nu}|\mu)$$

Pozostaje uzasadnić (*): (*) wynika z monotonicznej zbieżności, ponieważ

$$\alpha \mapsto f_\alpha \text{ jest wypukłe} \Rightarrow \frac{f_{\alpha+h} - f_\alpha}{h} \text{ maleje dla } h \rightarrow 0 \quad (3.10)$$

Aby udowodnić (3.10), zauważmy, że wypukłość implikuje

$$\frac{1}{2}(f_{\alpha+h} + f_{\alpha}) \geq f_{\alpha+h/2}$$

stąd

$$\frac{1}{2}(f_{\alpha+h} - f_{\alpha}) \geq f_{\alpha+h/2} - f_{\alpha} \Rightarrow \frac{f_{\alpha+h} - f_{\alpha}}{h} \geq \frac{f_{\alpha+h/2} - f_{\alpha}}{h/2},$$

i ten sam argument działa zamiast $1/2$ dla $\gamma \in (0, 1)$. \square

Definicja 3.51

Zbiór $B \subseteq M_1(S)$ jest **całkowicie wypukły** jeśli dla każdej $m \in M_1(B)$ mamy

$$\int_B v m(dv) \in B.$$

Całkowicie wypukły podzbiór $M_1(S)$ jest oczywiście wypukły. Rozważmy ν_1, ν_2 należące do całkowicie wypukłego zbioru B . Wówczas dla dowolnej $\alpha \in (0, 1)$ poprzez rozważenie miary $m = \alpha\delta_{\nu_1} + (1 - \alpha)\delta_{\nu_2}$ otrzymujemy

$$\int_B v m(dv) = \alpha\nu_1 + (1 - \alpha)\nu_2.$$

Zgodnie z definicją powyższe jest elementem B . Istnieją zbiory wypukłe, które nie są całkowicie wypukłe. Ćwiczenie: znajdź przykład!

Twierdzenie 3.52 (I. Csiszar)

Niech Y_1, Y_2, \dots będą iid o rozkładzie μ dla pewnego $\mu \in M_1(S)$. Załóżmy, że $B \subseteq M_1(S)$ jest całkowicie wypukły i domknięty. Załóżmy również

$$\inf_{\eta \in \text{int}(B)} H(\eta|\mu) = \inf_{\eta \in B} H(\eta|\mu) < \infty.$$

Niech $L_n = \frac{1}{n} \sum_{i=1}^n \delta_{Y_i}$ oraz Wówczas

(i) $\exists! \bar{\nu} \in M_1(S)$ takie, że $H(\bar{\nu}|\mu) = \min_{\eta \in B} H(\eta|\mu)$.

(ii) Niech $\nu_n(\cdot) = \mathbb{P}[(Y_1, \dots, Y_n) \in \cdot | L_n \in B]$. Wówczas $H(\mu_n|\bar{\nu}^{\otimes n}) = o(n)$.

(iii) Niech $\mu_n(\cdot) = \mathbb{P}[Y_1 \in \cdot | L_n \in B]$. Wówczas $H(\mu_n|\bar{\nu}) \rightarrow 0$

a w szczególności, z (3.9), $\|\mu_n - \bar{\nu}\| \rightarrow 0$.

Dowód. Dla $\eta \in M_1(S^n)$ przez η_i oznaczmy i ty rozkład brzegowy η . Niech

$$\tilde{B}_n = \{\eta \in M_1(S^n) : \eta_i \in B, i \leq n\}$$

Pokażemy, że

$$\min_{\eta \in \mathcal{B}_n} H(\eta | \mu^{\otimes n}) = H(\bar{\nu} | \mu^{\otimes n}).$$

Dla dowolnego $\eta \in B_n$, takiego, że $\eta \ll \mu^{\otimes n}$ mamy $\prod_{j=1}^n \eta_j \ll \mu^{\otimes n}$. Mamy

$$\begin{aligned} H(\eta | \mu^{\otimes n}) &= H\left(\eta \left| \prod_{i=1}^n \eta_i\right.\right) + \int \log\left(\frac{d\prod_{j=1}^n \eta_j}{d\mu^{\otimes n}}\right) d\eta \\ &= H\left(\eta \left| \prod_{i=1}^n \eta_i\right.\right) + \int \sum_{j=1}^n \log\left(\frac{d\eta_j}{d\mu}\right) d\eta \\ &= H\left(\eta \left| \prod_{i=1}^n \eta_i\right.\right) + \sum_{j=1}^n H(\eta_j | \mu) \geq nH(\bar{\nu} | \mu) = H(n\bar{\nu}^{\otimes n} | \mu^{\otimes n}). \end{aligned}$$

W szczególności z(3.8) wnioskujemy, że

$$H(\eta | \mu^{\otimes n}) \geq H(\eta | \bar{\nu}^{\otimes n}) + H(\bar{\nu}^{\otimes n} | \mu^{\otimes n}). \quad (3.11)$$

Pokażemy teraz, że $\mu_n \in B$.

$$\begin{aligned} \int g d\mu_n &= \int g(Y_1) d\mathbb{P}(\cdot | L_n \in B) = \int \frac{1}{n} (g(Y_1) + \dots + g(Y_n)) d\mathbb{P}(\cdot | L_n \in B) \\ &= \int \left(\int g dL_n \right) d\mathbb{P}(\cdot | L_n \in B) \end{aligned}$$

Innymi słowy

$$\mu_{B,n} = \int L_n(\omega) \mathbb{P}(d\omega | L_n \in B) \in B$$

ponieważ B jest całkowicie wypukły. Zauważmy, że skoro $\mu_n \in B$, to $\nu_n \in B_n$. Mamy

$$H(\nu_n | \nu^{\otimes n}) \leq H(\nu_n | \mu^{\otimes n}) - H(\bar{\nu}^{\otimes n} | \mu^{\otimes n}) = -\log \mathbb{P}(L_n \in B) - nH(\bar{\nu} | \mu) \quad (3.12)$$

Ale

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(L_n \in B) \geq - \inf_{\eta \in \text{int}(B)} H(\eta | \mu) = - \inf_{\eta \in B} H(\eta | \mu) = -H(\bar{\nu} | \mu),$$

i wnioskujemy stąd, że

$$\limsup_{n \rightarrow \infty} \frac{1}{n} H(\nu_{B,n} | \bar{\nu}^{\otimes n}) = 0.$$

Aby z tego miejsca uzasadnić (iii) zauważmy, że

$$H(\nu_n | \bar{\nu}^{\otimes n}) = H(\nu_n | \mu_n^{\otimes n}) + H(\mu_n^{\otimes n} | \bar{\nu}^{\otimes n})$$

Stąd $H(\nu_n | \bar{\nu}^{\otimes n}) \geq nH(\mu_n | \bar{\nu})$. Teraz (ii) implikuje, że $H(\mu_{B,n} | \bar{\nu}) \rightarrow 0$. \square

3.8 Zadania

Zadanie 3.1

Załóżmy, że S jest lokalnie zwarta (każdy punkt posiada bazę otoczeń, których domknięcia są zwarte). Pokaż, że jeżeli ciąg miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia (mocną) zasadę wielkich odchyłeń z dobrą funkcją tempa, to jest on wykładniczo ciasny.

Zadanie 3.2

Podaj przykład ciągu miar probabilistycznych $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$, który zbiega słabo do pewnej miary probabilistycznej, spełnia słabą zasadę wielkich odchyłeń, ale nie spełnia silnej zasady wielkich odchyłeń.

Duże odchylenia dla funkcji ciągłych

Streszczenie Zaczniemy

Przez \mathbb{W} oznaczać będziemy miarę Wienera na $C_0[0, 1] = \{x \in C[0, 1] : x(0) = 0\}$. Niech $\{B_j\}_{j \in \mathbb{N}}$ będzie ciągiem niezależnych ruchów Browna. Przez \mathbf{P}_n oznaczać będziemy rozkład losowej funkcji ciągłej $S_n: [0, 1] \rightarrow \mathbb{R}$ danej przez

$$S_n(t) = \frac{1}{n} \sum_{j=1}^n B_j(t).$$

Zatem $\mathbf{P}_n \in M_1(C_0[0, 1])$. Zauważmy, że przez skalowanie procesu Browna, \mathbf{P}_n jest także rozkładem B_1/\sqrt{n} . Aby pokazać zasadę wielkich odchyłeń dla $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ wykorzystamy twierdzenie Cramera na przestrzeniach Banaacha. Musimy sprawdzić, czy spełnione są założenia. Przypomnijmy, że na $C_0[0, 1]$ rozważamy normę $\|x\| = \max_{t \in [0, 1]} |x(t)|$, $x \in C_0[0, 1]$. Należy pokazać, że dla każdej $a > 0$,

$$\mathbb{E} \left[e^{a\|B\|_\infty} \right] = \int e^{a\|f\|_\infty} \mathbb{W}(df) < \infty.$$

Z zasady odbicia wiemy, że $\max_{t \in [0, 1]} B(t)$ ma ten sam rozkład co $|B(1)|$. Mamy

$$\max_{t \in [0, 1]} |XBt| \leq \max_{t \in [0, 1]} B(t) + \left(- \min_{t \in [0, 1]} B(t) \right)$$

przy czym oczywiście, $-\min_{t \in [0, 1]} X(t)$ także ma ten sam rozkład co $|X(1)|$. Stosując zatem nierówność Cauchy'ego-Schwarza,

$$\mathbb{E} \left[e^{a\|B\|_\infty} \right] \leq \mathbb{E} \left[e^{2a|B(1)|} \right] < \infty.$$

Zatem $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełniają zasadę wielkich odchyłeń z prędkością n i funkcją tempa

$$I(x) = \min \left\{ H(\nu|\mathbb{W}) : \int f \mathbb{W}(df) = x \right\}, \quad x \in C_0[0, 1].$$

Powyższa reprezentacja I nie jest jednak użyteczna. Aby wyznaczyć bardziej rozważamy przestrzeń Camerona-Martina daną przez

$$H = \left\{ x \in C_0[0,1] : \exists g \in L^2[0,1] \text{ takie, że } x(t) = \int_0^t g(s) ds, t \in [0,1] \right\},$$

Dla $x \in H$ będziemy pisali $x' = g$. Należy jednak zauważyć, że $x \in H$ nie musi być różniczkowalna wszędzie.

Twierdzenie 4.1 (Schilder)

Ciąg $\{\mathbf{P}_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień z dobrą, wypukłą funkcją tempa

$$I(x) = \begin{cases} \frac{1}{2} \int_0^1 |x'(s)|^2 ds & x \in H \\ \infty & \text{poza tym.} \end{cases}$$

Z dyskusji poprzedzającej sformułowanie powyższego twierdzenia wynika, że musimy tylko pokazać:

$$\inf \left\{ H(v|\mathbb{W}) : \int f \mathbb{W}(df) = x \right\} = \begin{cases} \frac{1}{2} \int_0^1 |x'(s)|^2 ds & x \in H \\ +\infty & \text{poza tym.} \end{cases}$$

Aby zrozumieć entropię względem \mathbb{W} będziemy musieli odwołać się do konstrukcji ruchu Browna. W tym celu przypomnijmy układ funkcji Haara, $(\varphi_{n,k})$ danych przez

$$\varphi_0(t) \equiv 1 \quad \varphi_{n,k} = \begin{cases} 2^{\frac{n-1}{2}} & \frac{k-1}{2^{n-1}} \leq t \leq \frac{k-\frac{1}{2}}{2^{n-1}} \\ -2^{\frac{n-1}{2}} & \frac{k-\frac{1}{2}}{2^{n-1}} \leq t \leq \frac{k}{2^{n-1}} \\ 0 & \text{poza tym} \end{cases}$$

Do konstrukcji ruchu Browna posłużą nam funkcje $\{e_{n,k}\}$ dane przez

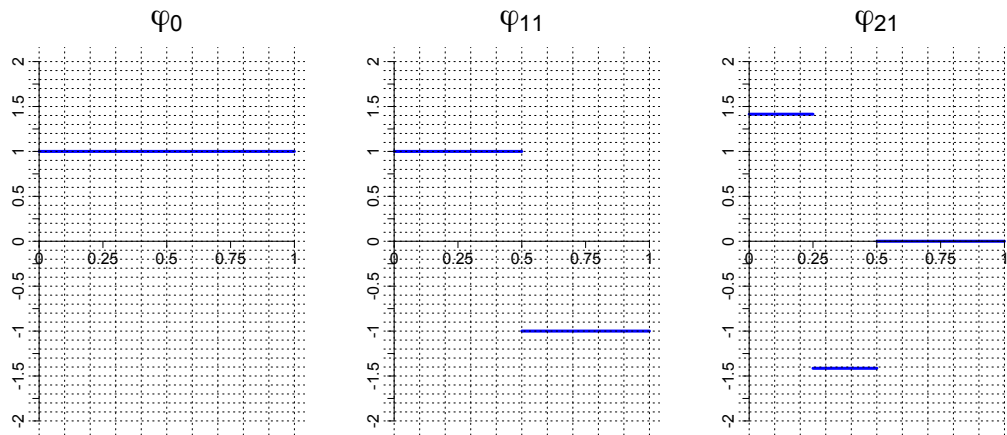
$$e_{n,k}(t) = \int_0^t \varphi_{n,k}(s) ds.$$

Twierdzenie 4.2 (Twierdzenie Itô-Nisio)

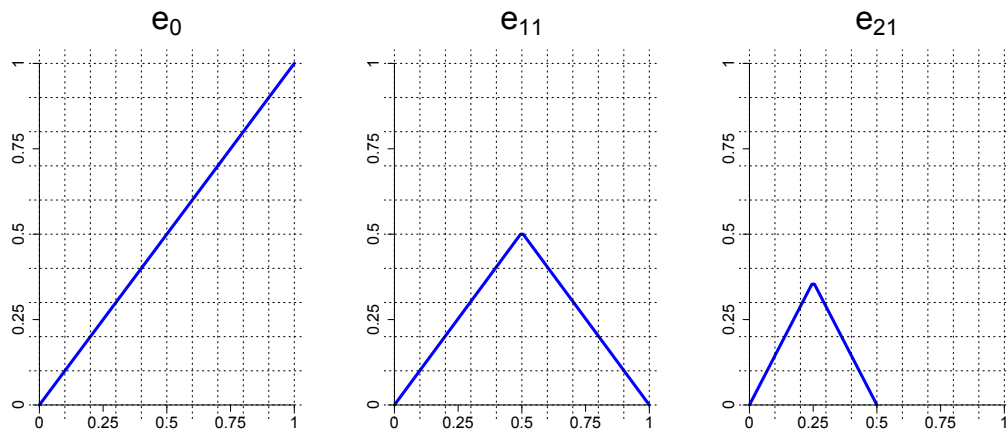
Rozważmy $\{Z_{n,k}\}_{n,k \in \mathbb{N}}$ iid. z rozkładem $\mathcal{N}(0,1)$. Wtedy

$$X_t(\omega) = \sum_{n \geq 1, k \leq 2^{n-1}} Z_{n,k}(\omega) e_{n,k}(t),$$

jest ruchem Browna na $[0,1]$.



Rysunek 4.1. kilka pierwszych funkcji z układu Haara

Rysunek 4.2. kilka pierwszych funkcji z układu $\{e_{n,k}\}$

Dowód. Sprawdźmy funkcję korelacji $(X_t)_{t \in [0,1]}$. Mamy

$$\begin{aligned} \mathbb{E}[X_t X_s] &= \sum_{n,k} e_{n,k}(t) e_{n,k}(s) = \sum_{n,k} \int_0^t \varphi_{n,k}(u) du \int_0^s \varphi_{n,k}(v) dv \\ &= \sum_{n,k} \langle \varphi_{n,k}, \mathbb{1}_{[0,t]} \rangle \langle \varphi_{n,k}, \mathbb{1}_{[0,s]} \rangle = \langle \mathbb{1}_{[0,t]}, \mathbb{1}_{[0,s]} \rangle = t \wedge s, \end{aligned}$$

gdzie $\langle f, g \rangle = \int_0^1 f(s)g(s)ds$ oznacza iloczyn skalarny w $L^2[0,1]$.

Wystarczy teraz pokazać, że $\sum_{n \leq N, k \leq 2^{n-1}} Z_{n,k}(\omega) e_{n,k}(t)$ zbiega jednostajnie dla $N \rightarrow \infty$. Wówczas wiemy, że $(X(t))_{0 \leq t \leq 1}$ jest ciągłym procesem Gaussowskim o kowariancji $s \wedge t$. \square

Dla $x \in C_0[0,1]$ niech

$$\Delta_{n,k}(x) := x \left(\frac{k - \frac{1}{2}}{2^{n-1}} \right) - \frac{1}{2} \left(x \left(\frac{k}{2^{n-1}} \right) + x \left(\frac{k-1}{2^{n-1}} \right) \right)$$

oraz $z_{n,k}(x) := 2^{\frac{n+1}{2}} \Delta_{n,k}(x)$ Ponadto, niech $z_0(x) = x(1)$. Dla $x \in C_0[0,1]$ będziemy pisać

$$x^N(t) = z_0(x)t + \sum_{n=1}^N \sum_{k=1}^{2^{n-1}} z_{n,k}(x)e_{n,k}(t).$$

Wówczas x^N jest interpolacją liniową x na N -tym podziale diadycznym $[0,1]$. W szczególności, mamy

$$\lim_{N \rightarrow \infty} \sup_{t \in [0,1]} |x^N(t) - x(t)| = 0$$

Zauważmy, że dla dowolnego ruchu Browna $B = (B_t)_{t \in [0,1]}$ zmienne $Z_{n,k} = z_{n,k}(B)$ są iid ze standardowym rozkładem normalnym. **Od tej pory będziemy pisać $Z_{n,k} = Z_j$.**

Lemat 4.3

Dla $x \in C_0[0,1]$ mamy

$$\sum_{j=1}^{\infty} z_j(x)^2 = \begin{cases} \int_0^1 |x'(s)|^2 ds & x \in H \\ +\infty & x \notin H. \end{cases}$$

Dowód. Mamy

$$x(t) = \sum_{j=1}^{\infty} z_j(x)e_j(t)$$

przy czym szereg po prawej stronie jest zbieżny jednostajnie dla $[0,1]$. Rozważmy

$$g(t) = \sum_{j=1}^{\infty} z_j(x)\varphi_j(t).$$

Skoro φ_j są bazą $L^2[0,1]$, to $\|g\|_{L^2[0,1]}^2 = \sum_{j=1}^{\infty} z_j(x)^2$. Jeżeli ostatni szereg jest zbieżny, to szereg definiujący g jest zbieżny w L^2 . Wtedy też

$$\int_0^t g(s)ds = x(t),$$

czyli $g = x'$. □

Lemat 4.4

Dla $x \in C_0[0,1]$,

$$\inf \left\{ H(\nu|\mathbb{W}) : \int f \nu(df) = x \right\} = \frac{1}{2} \sum_{j=1}^{\infty} z_j(x)^2$$

Dowód. Dla $x \in C_0[0,1]$, $x = \sum_{j=1}^{\infty} z_j(x)e_j$. Skoro z_j są liniowe, to dla $\nu \in M_1(C_0[0,1])$, mamy

$$\int f \nu df = x \iff \int z_j(f) \nu(dy) = z_j(x).$$

Powyższa obserwacja pozwoli nam Przerzucić problem minimalizacji entropii na przestrzeń ciągów. Rozważmy funkcję $\Phi: C_0[0,1] \rightarrow \mathbb{R}^{\mathbb{N}}$ zadane wzorem

$$\Phi(x) = \{z_j(x)\}_{j \in \mathbb{N}}$$

Korzyścią z takiego podejścia jest to, że miara Wienera się trywializuje:

$$\bar{\mathbb{W}} = \mathbb{W} \circ \Phi^{-1} = \bigotimes_{j \in \mathbb{N}} \mathcal{N}(0,1).$$

Dla dowolnej miary $\nu \in M_1(C_0[0,1])$ będziemy pisać $\bar{\nu} = \nu \circ \Phi^{-1}$. Wówczas dla dowolnej nieujemnej, mierzalnej funkcji $F: \mathbb{R}^{\mathbb{N}} \rightarrow \mathbb{R}$,

$$\int F(a) \bar{\nu}(da) = \int F(\Phi(f)) \nu(df).$$

W szczególności $\nu \ll \mu$ wtedy i tylko wtedy, gdy $\bar{\nu} \ll \bar{\mu}$ oraz

$$\frac{d\bar{\nu}}{d\bar{\mu}}(\Phi(f)) = \frac{d\nu}{d\mu}(f).$$

Stąd też

$$H(\nu|\mu) = H(\bar{\nu}|\bar{\mu}).$$

Dokonałiśmy zatem następującej redukcji

$$\inf \left\{ H(\nu|\mu) : \int f \nu(df) = x \right\} = \inf \left\{ H(\bar{\nu}|\bar{\mu}) : \int a_j \bar{\nu}(da) = z_j(x) \right\}.$$

Dla dowolnej miary $\bar{\nu} \in M_1(\mathbb{R}^{\mathbb{N}})$ i naturalnego j przez $\bar{\nu}_j$ oznaczać będziemy j ty rozkład brzegowy $\bar{\nu}$. Zauważmy, że skoro $\bar{\mathbb{W}}$ jest miarą produktową, to

$$\mathbb{W} = \bigotimes_{j \in \mathbb{N}} \bar{\mathbb{W}}_j.$$

Powyższa własność jest kluczowa przy minimalizacji entropii. Zastosujemy następującą obserwację (zadanie): dla $\bar{\nu} \ll \bar{\mu}$,

$$\bar{\nu} \ll \bigotimes_j \bar{\mu}_j \implies \bar{\nu} \ll \bigotimes_j \bar{\nu}_j.$$

Wybierzmy $\tilde{\nu} = \bigotimes_j \bar{\nu}_j$. Wówczas $\bar{\nu} \ll \tilde{\nu} \ll \bar{\mathbb{W}}$. Wtedy mamy

$$\begin{aligned} H(\tilde{\nu}|\tilde{\mathbb{W}}) &= \int \log \frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}} d\tilde{\nu} = \int \log \left(\frac{d\tilde{\nu}}{d\tilde{\nu}} \frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}} \right) d\tilde{\nu} \\ &= \int \log \frac{d\tilde{\nu}}{d\tilde{\nu}} d\tilde{\nu} + \int \log \frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}} d\tilde{\nu} = H(\tilde{\nu}|\tilde{\nu}) + \int \log \frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}} d\tilde{\nu}. \end{aligned} \quad (4.1)$$

Z racji, że obie miary $\tilde{\nu}$ oraz $\tilde{\mathbb{W}}$ są produktowe, to dla $a = (a_1, a_2, \dots) \in \mathbb{R}^{\mathbb{N}}$ mamy

$$\frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}}(a) = \prod_{j \in \mathbb{N}} \frac{d\nu_j}{d\tilde{\mathbb{W}}_j}(a_j).$$

Stąd drugi składnik w (4.1) zapisuje się jako

$$\int \log \frac{d\tilde{\nu}}{d\tilde{\mathbb{W}}} d\tilde{\nu} = \sum_{j \geq 1} \int \log \frac{d\nu_j}{d\tilde{\mathbb{W}}_j} d\tilde{\nu} = \sum_{j \geq 1} \int \log \frac{d\nu_j}{d\tilde{\mathbb{W}}_j} d\nu_j.$$

Ostatnia równość wynika z tego, że dla każdego j , $d\nu_j/d\tilde{\mathbb{W}}_j$ zależy tylko od j -tej osi. Otrzymaliśmy zatem

$$H(\tilde{\nu}|\tilde{\mu}) = H(\tilde{\nu}|\tilde{\nu}) + \sum_j H(\tilde{\nu}_j|\tilde{\mathbb{W}}_j) \quad (4.2)$$

przy czym pierwszy wyraz po prawej jest nieujemny. Z ćwiczeń wiemy, że

$$\inf \left\{ H(\tilde{\nu}_j|\mathcal{N}(0,1)) : \int s \tilde{\nu}_k(ds) = z_k(x) \right\} = H(\mathcal{N}(\bar{z}_k, 1)|\mathcal{N}(0,1)) = \frac{1}{2} z_k(x)^2.$$

Stosując to do (4.2) otrzymujemy

$$\inf \left\{ H(\nu|\mathbb{W}) : \int f \nu(df) = x \right\} = \frac{1}{2} \sum_{j=1}^{\infty} z_j(x)^2.$$

□

Teraz pokażemy, jak twierdzenie Schildera może być wykorzystane do wyprowadzenia funkcjonalnego prawa iterowanego logarytmu. Przypomnijmy, że

$$H = \left\{ x \in C_0[0,1] : \exists g \in L^2[0,1] \text{ takie, że } x(t) = \int_0^t g(s) ds \quad \forall t \in [0,1] \right\}.$$

Dla $x \in H$ piszemy $x' = g$. Niech

$$K := \left\{ x \in H : \int_0^1 |x'(s)|^2 ds \leq 1 \right\}.$$

Twierdzenie 4.5 (Prawo iterowanego logarytmu (LIL) Strassena)
Niech $(B_t)_{t \geq 0}$ będzie ruchem Browna. Rozważmy ciąg funkcji losowych

$$X_t^{(n)} = \frac{B_{nt}}{\sqrt{2n \log \log n}}, \quad 0 \leq t \leq 1.$$

Wtedy z prawdopodobieństwem jeden $\{X^{(n)} : n \geq 3\}$ zwarty w $C_0[0, 1]$ i zbiór jego punktów granicznych to K .

W szczególności, dla każdej ciągłej funkcji $F: C_0[0, 1] \rightarrow \mathbb{R}$,

$$\limsup_{n \rightarrow \infty} F(X^{(n)}) = \sup_{x \in K} F(x) \quad p.w.$$

Przykład 4.6

Weźmy $F(x) = x(1)$. Wtedy $\sup_{x \in K} F(x) = 1$, ponieważ

$$x(1) = \int_0^1 x'(s) ds \leq \left(\int_0^1 |x'(s)|^2 ds \right)^{1/2} \leq 1,$$

ponieważ $x \in K$. Supremum jest osiągnięte dla $x(s) = s, 0 \leq s \leq 1$. Otrzymujemy stąd klasyczne prawo iterowanego logarytmu

$$\limsup_{n \rightarrow \infty} \frac{B_n}{\sqrt{2n \log \log n}} = 1 \quad \mathbb{P}\text{-a.s.}$$

Przykład 4.7

Jeżeli rozważymy $F(x) = \|x\|_\infty$, to $\sup_{x \in K} F(x) = 1$. Jak poprzednio dla $x \in K$ mamy

$$x(t) = \int_0^t x'(s) ds \leq \sqrt{t} \left(\int_0^t x'(s)^2 ds \right)^{1/2} \leq 1.$$

Przy czym dla $x(s) = s$ dostajemy $F(x) = 1$. Stąd

$$\limsup_{n \rightarrow \infty} \sup_{0 \leq t \leq n} \frac{B_t}{\sqrt{2n \log \log n}} = 1 \quad \mathbb{P}\text{-a.s.}$$

Przykład 4.8

Rozważmy teraz następujący ciąg

$$\int_0^n B_s ds.$$

Chcemy znaleźć optymalne asymptotyczne oszacowanie górne dla powyższego wyrażenia. Niech

$$F(x) = \int_0^1 x(s) ds$$

Wtedy $\sup_{x \in K} F(x) = 1/\sqrt{3}$. Istotnie, mamy

$$\begin{aligned} \int_0^1 x(u) du &= \int_0^1 \int_0^u x'(s) ds du \\ &= \int_0^1 \int_s^1 x'(s) du ds = \int_0^1 (1-s)x'(s) ds \\ &\leq \left(\int_0^1 (1-s)^2 ds \right)^{\frac{1}{2}} \left(\int_0^1 |x'(s)|^2 ds \right)^{\frac{1}{2}} \leq \frac{1}{\sqrt{3}}. \end{aligned}$$

Przy czym równość w nierówności Cauchy'ego-Schwarza zachodzi dla

$$\begin{aligned} x(s) &= \sqrt{3} \left(s - \frac{1}{2} s^2 \right) \quad 0 \leq s \leq 1, \\ x'(s) &= \sqrt{3}(1-s). \end{aligned}$$

Zatem

$$\limsup_{n \rightarrow \infty} \frac{1}{\sqrt{2n \log \log n}} \frac{1}{n} \int_0^n B_t dt = \frac{1}{\sqrt{3}}.$$

Aby udowodnić prawo iterowanego logarytmu będzie nam potrzebna następująca modyfikacja zasady wielkich odchyłeń dla ruchu Browna.

Twierdzenie 4.9

Niech $(B_t)_{0 \leq t \leq 1}$ będzie procesem Browna, a (a_n) będzie ciągiem takim, że $a_n \rightarrow \infty$. Wtedy, ciąg rozkładów $(B_t/a_n)_{0 \leq t \leq 1}$ spełniają zasadę wielkich odchyłeń z szybkością a_n^2 i funkcją tempa

$$I(x) = \begin{cases} \frac{1}{2} \int_0^1 |x'(s)|^2 ds & x \in H \\ +\infty & \text{w przeciwnym przypadku.} \end{cases}$$

Poprzednie sformułowanie ZWO jest szczególnym przypadkiem dla $a_n = \sqrt{n}$. Dowód jest analogiczny.

Lemat 4.10

Dla każdego $\lambda > 0$ ciąg $(\xi_{[\lambda^n]})$ jest warunkowo zwarty w $C[0, 1]$. Każdy punkt graniczny należy do K .

Dowód. Oznaczmy $k_n = \lfloor \lambda^n \rfloor$ dla pewnego $\lambda > 1$. Niech $K_\delta = \bigcup_{x \in K} U_\delta(x)$, gdzie $U_\delta(x) = \{y : \|y - x\| < \delta\}$. Wtedy zbiór K_δ jest otwarty, ponieważ jest to suma zbiorów otwartych. Dodatkowo

$$\inf_{x \in K_\delta^c} I(x) > \frac{1}{2}. \quad (4.3)$$

Istotnie, ponieważ $I(\cdot)$ jest dobrą funkcją tempa, dla dowolnego $(x_n) \subseteq K_\delta^c$ takiego, że $I(x_n) \rightarrow \frac{1}{2}$ i dowolnego $c > 0$ to $I(x_n) \leq 1$ dla dostatecznie dużych n . Skoro podpoziomice I są zwarte, to istnieje podciąg zbieżny (x_{n_k}) taki, że $x_{n_k} \rightarrow x$ oraz

$$I(x) \leq \liminf_{k \rightarrow \infty} I(x_{n_k}) \leq 1/2.$$

Ostatnia nierówność oznacza, że $x \in K$, ale jest to sprzeczność, ponieważ K_δ^c jest domknięty. Możemy wybrać zatem γ takie, że

$$\frac{1}{2} < \gamma < \inf_{x \in K_\delta^c} I(x) \quad (4.4)$$

Stąd

$$\begin{aligned} \mathbb{P}(\xi_{k_n} \in K_\delta^c) &= \mathbb{P}\left(\left(\frac{B_{\lfloor \lambda^n \rfloor t}}{\sqrt{2k_n \log \log k_n}}\right)_{0 \leq t \leq 1} \in K_\delta^c\right) \\ &= \mathbb{P}\left(\left(\frac{B_t}{\sqrt{2 \log \log k_n}}\right)_{0 \leq t \leq 1} \in K_\delta^c\right) \leq e^{-2(1+o(1))\gamma \log \log k_n} \\ &\leq \frac{1}{2(\log \lambda)^{2\gamma}} \frac{1}{n^{2\gamma+o(1)}} \end{aligned}$$

Zatem

$$\sum_n \mathbb{P}(\xi_{k_n} \in K_\delta^c) < \infty$$

Wynika stąd, że lemat Borela-Cantellego implikuje, że dla każdego $\delta > 0$ mamy \mathbb{P} -prawie na pewno, że $\xi_{k_n} \in K_\delta$ dla dostatecznie dużych n . Skoro $\delta > 0$ jest dowolna, to dla każdego $n_m \rightarrow \infty$ takiego, że $\xi_{k_{n_m}} \rightarrow \xi$ w $C[0, 1]$,

$$\xi \in \bigcap_{\delta > 0} K_\delta = \text{cl}K = K.$$

□

Lemat 4.11

Ciąg (ξ_n) jest warunkowo zwarty w $C[0, 1]$. Każdy punkt graniczny należy do K .

Dowód. Niech $M = \sup_{h \in K} \|h\|$. Zauważmy, że $0 \geq s < t \leq 1$,

$$\sup_{h \in K} |h(t) - h(s)| \leq \sqrt{t - s}.$$

Dla $\lambda > 1$ i naturalnego n istnieje m takie, że $k_m \leq n < k_{m+1}$. Oznaczmy $N = k_{m+1}$. Chcemy pokazać, że

$$\lim_{n \rightarrow \infty} d(\xi_n, K) = 0.$$

Zrobimy to przybliżając ξ_n przez ξ_N . Oznaczmy $g(n) = \sqrt{2n \log \log n}$. Ustalmy $\delta < 0$. Wybieramy $\lambda > 1$ na tyle blisko 1 aby $1/\lambda - 1 \leq \delta^2$. Z poprzedniego lematu możemy wybrać n na tyle duże aby

$$d(\xi_N, K) \leq \delta.$$

Stąd $\|\xi_N\| \leq M + \delta$. Dodatkowo istnieje $f \in K$ taka, że $\|\xi_N - f\| < \delta$. Mamy

$$d(\xi_n, K) \leq d(\xi_N, K) + d(\xi_n, \xi_N) \leq \delta + d(\xi_n, \xi_N).$$

Mamy

$$\begin{aligned} \|\xi_N - \xi_n\| &\leq \sup_t \left| \frac{g(n)}{g(N)} \xi_N(nt/N) - \xi_n(t) \right| \\ &= |g(n)/g(N) - 1| \|\xi_N\| + \sup_{t \in [0,1]} |\xi_N(nt/N) - \xi_n(t)| \\ &= |g(n)/g(N) - 1| (M + \delta) + \sup_{t \in [0,1]} |f(nt/N) - f(t)| + 2\delta \\ &= |g(n)/g(N) - 1| (M + \delta) + \sqrt{n/N - 1} + 2\delta \leq (\lambda - 1)(M + \delta) + 3\delta. \end{aligned}$$

□

Lemat 4.12

Każdy $x \in K$ jest punktem granicznym ciągu (ξ_n) .

Dowód. Wystarczy pokazać, że dla każdego $x \in K$ takiego, że $I(x) > 1/2$ mamy \mathbb{P} -prawie na pewno

$$\liminf_{n \rightarrow \infty} \|\xi_n - x\|_\infty = 0.$$

Niech, dla $k \geq 2$,

$$x_k(t) = \begin{cases} 0 & 0 \leq t \leq 1/k \\ x(t) - x(1/k) & 1/k \leq t \leq 1 \end{cases}$$

oraz

$$\eta_{n,k} = \begin{cases} 0 & 0 \leq t \leq \frac{1}{k} \\ \frac{B_{k^{nt}} - B_{k^{n-1}}}{\sqrt{2k^n \log \log k^n}} & \frac{1}{k} \leq t \leq 1 \end{cases}$$

Zauważmy, że

$$\left(\frac{B_{k^{nt}} - B_{k^{n-1}}}{\sqrt{2k^n \log \log k^n}} \right)_{0 \leq t \leq 1} \stackrel{d}{=} \left(\frac{\overbrace{\tilde{B}_{k^{nt}}}^{\text{znów BM}}}{\sqrt{2k^n \log \log k^n}} \right)_{0 \leq t \leq 1} \quad (4.5)$$

Mamy więc

$$\begin{aligned} \|\zeta_{k^n} - x\| &\leq \|\zeta_{k^n}\|_{[0, \frac{1}{k}]} + \|x\|_{[0, \frac{1}{k}]} + \|\zeta_{k^n} - x\|_{[\frac{1}{k}, 1]} \\ &\leq 2\|\zeta_{k^n}\|_{[0, \frac{1}{k}]} + 2\|x\|_{[0, \frac{1}{k}]} + \|\eta_{n,k} - x_k\| \end{aligned}$$

Ponieważ (ζ_n) jest \mathbb{P} -prawie na pewno całkowicie zwarty, dla każdego $\delta > 0$ mamy \mathbb{P} -prawie na pewno pewne $k = k(\delta) \geq 2$, takie że dla $n \geq 1$

$$\|\zeta_{k^n} - x\| \leq \delta + \|\eta_{n,k} - x_k\|.$$

Pozostaje pokazać, że dla każdego $k \geq 2$ mamy

$$\liminf_{n \rightarrow \infty} \|\eta_{n,k} - x_k\| = 0 \quad \mathbb{P}\text{-prawie na pewno.}$$

Ponieważ $(\eta_{n,k})_{n=1,2,\dots}$ są niezależne, wystarczy pokazać, zgodnie z lematem Borela-Cantelliego,

$$\sum_{n=1}^{\infty} \mathbb{P}(\|\eta_{n,k} - x\| < \varepsilon) = \infty. \quad (4.6)$$

Ale

$$\begin{aligned} \mathbb{P}(\|\eta_{n,k} - x\| < \varepsilon) &\stackrel{(4.5)}{=} \mathbb{P}\left(\left(\frac{B_{k^n t}}{\sqrt{2k^n \log \log k^n}}\right)_{0 \leq t \leq 1} \in U_\varepsilon(x_k)\right) \\ &= \mathbb{P}\left(\left(\frac{B_t}{\sqrt{2 \log \log k^n}}\right)_{0 \leq t \leq 1} \in U_\varepsilon(x_k)\right) \\ &\stackrel{Tw.??}{\geq} e^{-2(\bar{\gamma} + \bar{\varepsilon}) \log \log k^n} \quad \text{dla } n \text{ dostatecznie dużego,} \end{aligned}$$

gdzie $\bar{\gamma} = \inf_{y \in U_\varepsilon(x)} I(y) \leq I(x)$ i $\bar{\varepsilon}$ taka, że $\kappa = 2(\bar{\gamma} + \bar{\varepsilon}) \leq 2I(x) + 2\bar{\varepsilon} < 1$.
Wtedy

$$\mathbb{P}(\|\eta_{n,k} - x\| < \varepsilon) \geq \frac{1}{(n \log k)^\kappa} \quad \text{dla } n \text{ dostatecznie dużego}$$

i (4.6) wynika.

□

Zmienne zależne

Streszczenie Zaprezentujemy ogólne narzędzie służące do badania wielkich odchyłeń dla systemów złożonych ze zmiennych zależnych. Zastosujemy to podejście do badania łańcuchów Markowa.

Do tej pory badaliśmy duże odchylenia dla procesów złożonych ze zmiennych niezależnych i jednakowo rozłożonych takich jak spacery losowe czy miary empiryczne pochodzące od próbek iid. Okazuje się jednak, że teoria wielkich odchyłeń nie ogranicza się do zmiennych iid. Przedstawimy ogólną teorię którą następnie zastosujemy do łańcuchów Markowa. Pod koniec wykładu zobaczymy jak ta teoria stosuje się do grafów losowych.

5.1 Twierdzenie Gärtnera-Ellisa

Niech $\{X_n\}_{n \in \mathbb{N}}$ będzie ciągiem zmiennych losowych przyjmujących wartości w \mathbb{R}^d dla $d \geq 1$. Dla $n \in \mathbb{N}$ przez μ_n oznaczać będziemy rozkład X_n . Naszym celem jest wyprowadzenie zasady dużych odchyłeń dla $\{\mu_n\}_{n \in \mathbb{N}}$. Niech

$$\Lambda_n(\lambda) = \log \mathbb{E} \left[e^{\langle \lambda, X_n \rangle} \right] = \log \int e^{\langle \lambda, x \rangle} \mu_n(dx), \quad \lambda \in \mathbb{R}^d$$

oznacza funkcję generującą kumulanty μ_n . Tutaj $\langle x, y \rangle = x_1 y_1 + \dots + x_d y_d$ oznacza standardowy iloczyn skalarny wektorów $x = (x_1, \dots, x_d)$ i $y = (y_1, \dots, y_d)$ w \mathbb{R}^d . Pracować będziemy pod następującym założeniem na asymptotyczne zachowanie Λ_n .

Założenie (A) Granica funkcji generującej kumulanty

$$\Lambda(\lambda) = \lim_{n \rightarrow \infty} \Lambda_n(n\lambda)/n \in (-\infty, \infty] \quad (5.1)$$

istnieje dla każdego $\lambda \in \mathbb{R}^d$, a odwzorowanie $\lambda \mapsto \Lambda(\lambda)$ jest różniczkowalne dla każdego $\lambda \in \mathcal{D}_\Lambda = \{x \in \mathbb{R}^d : |\Lambda(x)| < \infty\}$. Dodatkowo zakładamy, że $0 \in \text{int}(\mathcal{D}_\Lambda)$.

Zauważmy, że skoro Λ_n jest wypukła dla każdego $n \in \mathbb{N}$, to Λ również jest funkcją wypukłą.

Przykład 5.1

Niech zmienne losowe $\{Y_k\}_{k \in \mathbb{N}}$ będą iid z funkcją generującą kumulanty

$$\Lambda_Y(\lambda) = \log \mathbb{E} \left[e^{\lambda Y_1} \right].$$

Wówczas dla zmiennych $X_n = (Y_1 + \dots + Y_n)/n$ mamy

$$\Lambda_n(\lambda) = n\Lambda_Y(\lambda/n).$$

Stąd

$$\Lambda(\lambda) = \lim_{n \rightarrow \infty} \Lambda_n(n\lambda)/n = \Lambda_Y(\lambda).$$

Przykład 5.2

W świetle poprzedniego przykładu możemy stwierdzić, że jeżeli $X_n = (Y_1 + \dots + Y_n)/n$, to założenie (5.1) wymusza pewną kontrolę na zależność pomiędzy zmiennymi $\{Y_k\}_{k \in \mathbb{N}}$. Istotnie, jeżeli $Y_j = Y_1$ dla każdego naturalnego j , to X_n staje się ciągiem stałym. Wówczas (5.1) zachodzi jedynie w sytuacji, gdy rozkład Y_1 jest skupiony w punkcie.

Twierdzenie 5.3 (Gärtner-Ellis)

Jeżeli spełnione jest założenie (A), to ciąg $\{\mu_n\}_{n \in \mathbb{N}}$ spełnia zasadę dużych odchyłeń z funkcją tempa

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \{ \langle \lambda, x \rangle - \Lambda(\lambda) \},$$

dla $x \in \mathbb{R}^d$.

Zanim przejdziemy do dowodu tego twierdzenia przedstawimy niezbędny z analizy wypukłej.

Definicja 5.4

Punkt $y \in \mathbb{R}^d$ jest nazywany *punktem eksponowanym* funkcji $\Lambda^*(\cdot)$ jeśli istnieje $\lambda \in \mathbb{R}^d$ taka, że dla każdego $x \neq y$

$$\langle y, \lambda \rangle - \Lambda^*(y) > \langle x, \lambda \rangle - \Lambda^*(x).$$

W takim przypadku, funkcjonal $x \mapsto \langle x, \lambda \rangle$ (lub krótko λ) jest nazywany jest nazywany *funkcjonałem eksponującą* $y \in \mathbb{R}^d$.

Lemat 5.5

Jeśli $y = \nabla \Lambda(\eta)$ dla pewnego $\eta \in \mathbb{R}^d$, to

$$\Lambda^*(y) = \langle \eta, y \rangle - \Lambda(\eta),$$

i w takim przypadku y jest punktem eksponowanym funkcji Λ^* , gdzie $\eta \in \mathbb{R}^d$ jest wektorem normalnym hiperpowierzchni eksponującej.

Dowód. Weźmy $\lambda \in \mathbb{R}^d$ i zdefiniujmy $g: [0, 1] \rightarrow \mathbb{R}$ wzorem

$$\begin{aligned} g(\alpha) &= \alpha \langle \lambda, y \rangle + (1 - \alpha) \langle \eta, y \rangle - \Lambda(\alpha \lambda + (1 - \alpha) \eta) \\ &= \alpha \langle \lambda - \eta, y \rangle - \Lambda(\eta + \alpha(\lambda - \eta)) + \langle \eta, y \rangle. \end{aligned}$$

Wówczas

$$g(0) = \langle \eta, y \rangle - \Lambda(\eta) \quad g(1) = \langle \lambda, y \rangle - \Lambda(\lambda).$$

Ponieważ $\Lambda(\cdot)$ jest wypukła, funkcja g jest wklęsła, więc dla każdego $\alpha \in [0, 1]$

$$\alpha g(1) + (1 - \alpha)g(0) \leq g(\alpha).$$

Stąd

$$g(1) - g(0) \leq \liminf_{\alpha \rightarrow 0} \frac{g(\alpha) - g(0)}{\alpha} = \langle \lambda - \eta, y - \nabla \Lambda(\eta) \rangle = 0$$

a co za tym idzie

$$\langle \lambda, y \rangle - \Lambda(\lambda) \leq \langle \eta, y \rangle - \Lambda(\eta) \leq \Lambda^*(y) \quad (5.2)$$

Ponieważ (5.2) jest prawdziwe dla każdego $\lambda \in \mathbb{R}^d$, możemy przejść do \sup_{λ} w (5.2) i stwierdzić, że

$$\langle \eta, y \rangle - \Lambda(\eta) = \Lambda^*(y).$$

Pozostaje sprawdzić, czy y jest punktem eksponowanym funkcji Λ^* . Załóżmy, że dla pewnego $x \in \mathbb{R}^d$ zachodzi

$$\Lambda(\eta) = \langle \eta, y \rangle - \Lambda^*(y) \leq \langle \eta, x \rangle - \Lambda^*(x). \quad (5.3)$$

Wówczas dla każdego $\theta \in \mathbb{R}^d$ mamy:

$$\begin{aligned}\Lambda^*(x) &= \sup_{(\eta+\theta)\in\mathbb{R}^d} \{\langle\eta+\theta, x\rangle - \Lambda(\eta+\theta)\} \\ &\geq \langle\eta+\theta, x\rangle - \Lambda(\eta+\theta)\end{aligned}$$

Stąd

$$\langle\theta, x\rangle \leq \Lambda(\eta+\theta) - \Lambda(\eta).$$

W szczególności, dla każdego $\varepsilon > 0$,

$$\langle\theta, x\rangle \leq \lim_{\varepsilon\rightarrow 0} \frac{\Lambda(\eta+\varepsilon\theta) - \Lambda(\eta)}{\varepsilon} = \langle\theta, \nabla\Lambda(\eta)\rangle \quad (5.4)$$

Ponieważ (5.4) zachodzi dla wszystkich $\theta \in \mathbb{R}^d$, rozważając $\theta = \nabla\Lambda(\eta) - x$ wnioskujemy że $x = \nabla\Lambda(\eta) = y$. Oznacza to, że (5.3) nie może zachodzić dla żadnego $x \neq y$, co dowodzi, że y jest punktem eksponowanym funkcji Λ^* . \square

Pokażemy najpierw górne oszacowanie ZWO dla $\{\mu_n\}_{n\in\mathbb{N}}$ na zbiorach zwartych.

Fakt 5.6

Dla wszystkich zbiorów zwartych $\Gamma \subset \mathbb{R}^d$,

$$\limsup_{n\rightarrow\infty} \frac{1}{n} \log \mu_n(\Gamma) \leq - \inf_{x\in\Gamma} \Lambda^*(x). \quad (5.5)$$

Dowód. Niech $\Gamma \subset \mathbb{R}^d$ będzie zwarty. Ustalmy $\delta > 0$. Odwołując się do definicji Λ^* , dla dowolnego $q \in \Gamma$ możemy wybrać $\lambda_q \in \mathbb{R}^d$ takie, że

$$\langle\lambda_q, q\rangle - \Lambda(\lambda_q) \geq \min\{\Lambda^*(q) - \delta, 1/\delta\}. \quad (5.6)$$

Ponadto, dla każdego $q \in \Gamma$, wybieramy $r_q > 0$ dostatecznie małe, aby $r_q^2 |\lambda_q|^2 \leq \delta$. Niech $B_{r_q}(q)$ oznacza kulę euklidesową o promieniu r_q wokół q . Dla $G = B_{q, r_q}$ mamy

$$\begin{aligned}\mu_n(G) = \mathbb{P}[X_n \in G] &\leq \mathbb{P}\left[\langle\lambda, X_n\rangle \geq \inf_{g\in G} \langle\lambda, g\rangle\right] \\ &\leq \exp\left\{-\inf_{g\in G} \langle\lambda, g\rangle + \Lambda_n(\lambda)\right\}.\end{aligned}$$

Skoro dla $x \in B_{r_q}(q)$ mamy

$$\langle\lambda, x\rangle = \langle\lambda, x - q\rangle + \langle\lambda, q\rangle \geq -|\lambda|^2 r_q^2 + \langle\lambda, q\rangle,$$

to rozważając $\lambda = n\lambda_q$ mamy

$$\frac{1}{n} \log \mu_n \left(B_{r_q}(q) \right) \leq \delta - \langle \lambda_q, q \rangle + \frac{1}{n} \Lambda_n(n\lambda_q).$$

Powyżej skorzystaliśmy z warunku $r_q |2\lambda_q|^2 \leq \delta$. Z założenia (A) mamy zatem

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n \left(B_{r_q}(q) \right) \leq \delta - \langle \lambda_q, q \rangle + \Lambda(\lambda_q). \quad (5.7)$$

Przeniesiemy teraz to szacowanie na zbiór Γ . Skoro niewątpliwie $\Gamma \subset \bigcup_{q \in \Gamma} B_{r_q}(q)$ i Γ jest zwarty, to istnieje $N \in \mathbb{N}$ oraz $q_1, \dots, q_N \in \Gamma$ takie, że $\Gamma \subset \bigcup_{i=1}^N B_{r_{q_i}}(q_i)$. Z podaddytywności miary

$$\frac{1}{n} \log \mu_n(\Gamma) \leq \frac{1}{n} \log N + \max_{i \in \{1, \dots, N\}} \frac{1}{n} \log \mu_n \left(B_{r_{q_i}}(q_i) \right)$$

i stąd

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(\Gamma) &\leq \delta - \min_{i=1, \dots, N} \min \left\{ \Lambda^*(q_i) - \delta, \frac{1}{\delta} \right\} \\ &\leq \delta - \min \left\{ \Lambda^*(\Gamma) - \delta, \frac{1}{\delta} \right\} \end{aligned}$$

Przechodząc z $\delta \rightarrow 0$ dostajemy tezę. □

Fakt 5.7

Ciąg $\{\mu_n\}_{n \in \mathbb{N}}$ jest wykładniczo ciasny.

Dowód. Niech $B_M = [-M, M]^d$. Pokażemy, że

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B_M^c) = -\infty. \quad (5.8)$$

Ponieważ z założenia (A), $\Lambda(0) < \infty$, dla każdego $j = 1, \dots, d$ istnieje pewien $\epsilon > 0$ taki, że $\Lambda(\epsilon e_j) < \infty$ i $\Lambda(-\epsilon e_j) < \infty$, gdzie e_j oznacza j -ty wersor.

Niech $\mu_n^{(j)} \in \mathcal{M}_1(\mathbb{R})$ będzie j -tym rozkładem brzegowym $\mu_n \in \mathcal{M}_1(\mathbb{R}^d)$. Wówczas

$$\begin{aligned} \mu_n^j((-\infty, -M]) &= \mathbb{P}[\langle e_j, X_n \rangle < -M] \\ &\leq \mathbb{E} \left[e^{-n\epsilon(\langle e_j, X_n \rangle + M)} \right] = e^{-n\epsilon M} e^{\Lambda_n(-n\epsilon e_j)}. \end{aligned}$$

Podobnie

$$\mu_n^j([M, \infty)) \leq e^{-n\epsilon M} e^{\Lambda_n(n\epsilon e_j)}.$$

Biorąc logarytmy, dzieląc przez n , przechodząc do $\limsup_{n \rightarrow \infty}$ i następnie $\lim_{M \uparrow \infty}$ otrzymujemy

$$\lim_{M \rightarrow \infty} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n^{(j)} \left([-M, M]^C \right) = -\infty. \quad (5.9)$$

Ponieważ $B_M = [-M, M]^d$, $B_M^c = \cup_{j=1}^d \{x^{(j)} : |x^{(j)}| > M\}$, więc

$$\mu_n(B_M^c) \leq \sum_{j=1}^d \mu_n^j \left([-M, M]^C \right).$$

Przejsięcie do odpowiednich granic i zastosowanie (5.9) daje tezę. □

Dwa ostatnie fakty implikują, pełne ograniczenie górne

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(F) \leq - \inf_{x \in F} \Lambda^*(x) \quad (5.10)$$

dla dowolnego zbioru domkniętego $F \subset \mathbb{R}^d$.

Fakt 5.8

Jeśli y jest punktem eksponowanym Λ^* postaci $y = \nabla \Lambda(\eta)$ dla pewnego $\eta \in \mathcal{D}_\Lambda$, to

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) \geq -\Lambda^*(y). \quad (5.11)$$

Dowód. Rozważmy wykładniczą zamianę miary

$$\frac{d\tilde{\mu}_n}{d\mu_n}(z) = e^{n\langle \eta, z \rangle - \Lambda_n(n\eta)}. \quad (5.12)$$

Wówczas dla $\delta > 0$,

$$\mu_n(B_\delta(y)) = \int_{B_\delta(y)} e^{-n\langle \eta, z \rangle} \mathbb{E} \left[e^{n\langle \eta, X_n \rangle} \right] \tilde{\mu}_n(dz)$$

a co za tym idzie

$$\begin{aligned} \frac{1}{n} \log \mu_n(B_\delta(y)) &= \frac{1}{n} \Lambda_n(n\eta) - \langle \eta, y \rangle + \frac{1}{n} \log \int_{B_\delta(y)} e^{-n\langle \eta, z - y \rangle} \tilde{\mu}_n(dz) \\ &\geq \frac{1}{n} \Lambda_n(n\eta) - \langle \eta, y \rangle - |\eta| \delta + \frac{1}{n} \log \tilde{\mu}_n(B_\delta(y)) \end{aligned}$$

co na mocy Lematu 5.4 prowadzi z kolei do

$$\begin{aligned} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(B_{y,\delta}) &\geq -(\langle \eta, y \rangle - \Lambda(\eta)) - |\eta| \delta + \liminf_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\mu}_n(B_\delta(y)) \\ &= -\Lambda^*(y) + \lim_{\delta \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\mu}_n(B_\delta(y)). \end{aligned}$$

Teraz chcielibyśmy pokazać, że dla każdej $\delta > 0$,

$$\tilde{\mu}_n(B_{y,\delta}) \longrightarrow 1. \quad (5.13)$$

Jest to moralny odpowiednik słabego prawa wielkich liczb. Odwołując się do (5.10) dla zbioru domkniętego $B_{y,\delta}^c$, pokażemy, że

$$\limsup \frac{1}{n} \log \tilde{\mu}_n(B_\delta(y)^c) < 0,$$

co oczywiście wystarcza do udowodnienia (5.13). Mamy

$$\begin{aligned} \tilde{\Lambda}_n(n\lambda) &:= \log \int e^{n\langle \lambda, z \rangle} \tilde{\mu}_n(dz) \\ &= \log \int e^{n\langle \lambda, z \rangle} e^{n\langle \eta, z \rangle - \Lambda_n(n\eta)} \mu_n(dz) \\ &= \log \int e^{n\langle \lambda + \eta, z \rangle - \Lambda_n(n\eta)} \mu_n(dz) \end{aligned}$$

co implikuje

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{\Lambda}_n(n\lambda) &= \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(\lambda + \eta) - \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(n\eta) \\ &= \Lambda(\lambda + \eta) - \Lambda(\eta). \end{aligned}$$

Zatem,

$$\tilde{\Lambda}(\lambda) = \Lambda(\lambda + \eta) - \Lambda(\eta).$$

Z założeniem (A), $\tilde{\Lambda}(\lambda) < \infty$ i $\lambda \mapsto \tilde{\Lambda}(\lambda)$ jest różniczkowalna dla wszystkich $\lambda \in \eta + \mathcal{D}_\Lambda$. Innymi słowy, $\tilde{\mu}_n$ również spełnia założenie (A). Ponadto,

$$\begin{aligned} \tilde{\Lambda}^*(x) &= \sup_{\lambda} \{ \langle \lambda, x \rangle - \tilde{\Lambda}(\lambda) \} = \sup_{\lambda} \{ \langle \lambda, x \rangle - \Lambda(\lambda + \eta) + \Lambda(\eta) \} \\ &= \sup_{\lambda + \eta} \{ \langle (\lambda + \eta), x \rangle - \Lambda(\lambda + \eta) - \langle \eta, x \rangle + \Lambda(\eta) \} \\ &= \Lambda^*(x) + \Lambda(\eta) - \langle \eta, x \rangle \end{aligned}$$

Wiemy, że z założenia (A), $\Lambda^*(\cdot)$ jest dobrą i wypukłą funkcją tempa. W związku z tym $\tilde{\Lambda}^*(\cdot)$ również jest dobrą i wypukłą funkcją tempa. Ponadto, możemy skorzystać z górnego ograniczenia odchylenia (5.10) dla zbioru domkniętego $B_{y,\delta}^c$ aby wywnioskować

$$\begin{aligned} \limsup_{n \rightarrow \infty} \frac{1}{n} \log \tilde{\mu}_n(B_{y,\delta}^c) &\leq - \inf_{x \in B_{y,\delta}^c} \tilde{\Lambda}^*(x) \\ &= -\tilde{\Lambda}^*(x_0). \end{aligned}$$

dla pewnego $x_0 \notin B_{y,\delta}$. Wówczas

$$\begin{aligned}\tilde{\Lambda}^*(x_0) &= \Lambda^*(x_0) + \Lambda(\eta) - \langle \eta, x_0 \rangle \\ &\geq \Lambda^*(x_0) - \langle \eta, x_0 \rangle + \langle \eta, y \rangle - \Lambda^*(y) \\ &= \Lambda^*(x_0) - \langle \eta, x_0 \rangle - (\Lambda^*(y) - \langle \eta, y \rangle) > 0.\end{aligned}$$

Ostatnia nierówność zachodzi, ponieważ y jest punktem eksponowanym przez Λ^* , tj. $\Lambda^*(x_0) - \langle \eta, x_0 \rangle > \Lambda^*(y) - \langle \eta, y \rangle$. Ponieważ $\tilde{\Lambda}^*(x_0) > 0$, wynika stąd, że dla pewnego $c > 0$ i dostatecznie dużych n

$$\tilde{\mu}_n(B_{y,\delta}^c) \leq e^{-nc},$$

co implikuje (5.13) i tym samym (5.11), a w konsekwencji (5.14). \square

Udowodniony przez nas własnie fakt implikuje, że dla każdego otwartego $G \subset \mathbb{R}^d$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(G) \geq - \inf_{x \in G \cap \mathcal{F}} \Lambda^*(x), \quad (5.14)$$

gdzie \mathcal{F} to zbiór punktów eksponowanych przez Λ^* postaci $\nabla \Lambda(\eta)$. Pozostaje pokazać, że tych ostatnich jest dostatecznie dużo.

Dowód Twierdzenie Gärtnera-Ellisa. Na początek rozważamy

$$X_{n,\delta} X_n + \sqrt{\frac{\delta}{n}} Z,$$

gdzie $Z \sim \mathcal{N}(0, \text{Id}_d)$ jest niezależne od X_n dla dowolnego n . Naszym celem jest zastosowanie dolnego szacowania (5.14) dla $X_{n,\delta}$ dla wszystkich $\delta > 0$. Po pierwsze, musimy sprawdzić, czy rozkład $X_{n,\delta}$ spełnia Założenie (A). Faktycznie, mamy:

$$\begin{aligned}\Lambda_n^{(\delta)}(n\lambda) &:= \log \mathbb{E} \left[e^{n\langle \lambda, X_{n,\delta} \rangle} \right] = \log \mathbb{E} \left[e^{n\langle \lambda, X_n \rangle} e^{\sqrt{n\delta} \langle \lambda, Z \rangle} \right] \\ &= \log \mathbb{E} \left[e^{n\langle \lambda, X_n \rangle} \right] + \log \mathbb{E} \left[e^{\sqrt{n\delta} \langle \lambda, Z \rangle} \right]\end{aligned}$$

I stąd

$$\lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n^{(\delta)}(n\lambda) = \Lambda(\lambda) + \delta \frac{|\lambda|^2}{2} =: \Lambda_\delta(\lambda)$$

Zatem, zgodnie z Założeniem A, ta funkcja jest skończona i różniczkowalna dla wszystkich $\lambda \in \mathcal{D}_\Lambda$, co potwierdza, że rozkład $X_{n,\delta}$ spełnia Założenie A. Ponadto,

$$\Lambda_\delta^*(x) := \sup_{\lambda} \{ \langle \lambda, x \rangle - \Lambda_\delta(\lambda) \} \leq \Lambda^*(x)$$

ponieważ $\Lambda_\delta(\lambda) \geq \Lambda(\lambda)$. Korzystając z górnego oszacowania (5.10) dla zbioru domkniętego $F = \mathbb{R}^d$, otrzymujemy:

$$0 = \limsup_{n \rightarrow \infty} \frac{1}{n} \log \mu_n(\mathbb{R}^d) \leq - \inf_{x \in \mathbb{R}^d} \Lambda^*(x)$$

i $\inf_{x \in \mathbb{R}^d} \Lambda^*(x) = 0$. Zatem $\Lambda^*(y) < \infty$ dla pewnego $y \in \mathbb{R}^d$. Następnie definiujemy

$$g_\delta(\lambda) = \langle \lambda, y \rangle - \Lambda_\delta(y) = \langle \lambda, y \rangle - \Lambda(\lambda) - \delta|\lambda|^2/2.$$

Wówczas $g_\delta(\cdot)$ jest skończona i różniczkowalna na \mathcal{D}_Λ . Pokażemy, że

$$\lim_{M \rightarrow \infty} \sup_{|\lambda| > M} g_\delta(\lambda) = -\infty. \quad (5.15)$$

Faktycznie, z nierówności Jensena

$$\begin{aligned} \Lambda(\lambda) &= \lim_n \frac{1}{n} \log \mathbb{E} \left[e^{n\langle \lambda, X_n \rangle} \right] \geq \lim_n \frac{1}{n} \mathbb{E} [n\langle \lambda, X_n \rangle] \\ &\geq \lim_n \mathbb{E} [\langle \lambda, X_n \rangle] \end{aligned}$$

Zatem kwadratowy wzrost $-\delta|\lambda|^2/2$ w λ prowadzi do (5.15). Ponadto, ponieważ $g_\delta(\lambda) \leq \Lambda^*(y) < \infty$, jej supremum musi być osiągnięte dla pewnego skończonego $\eta \in \mathbb{R}^d$. Ponieważ g_δ jest różniczkowalna, $\nabla g_\delta(\eta) = 0$. Ale

$$\nabla g_\delta(\eta) = \nabla (\langle \eta, y \rangle - \Lambda_\delta(\eta)) = y - \nabla \Lambda_\delta(\eta)$$

co implikuje $y = \nabla \Lambda_\delta(\eta)$. Zgodnie z Lematem 5.5, y jest wtedy punktem eksponowanym dla Λ_δ^* . Korzystając z dolnego oszacowania (5.14) dla $X_{n,\delta}$, mamy dla dowolnego $x \in \mathbb{R}^d$ takiego, że $\Lambda^*(x) < \infty$:

$$\begin{aligned} \liminf_n \frac{1}{n} \mathbb{P}(X_{n,\delta} \in B_{\varepsilon/2}(x)) &\geq - \inf_{y \in B_{\varepsilon/2}(x)} \Lambda_\delta^*(y) \\ &= - \inf_{y \in B_{\varepsilon/2}(x)} \Lambda_\delta^*(y) \geq -\Lambda^*(x) \end{aligned}$$

Ponieważ $Z \sim \mathcal{N}(0, \text{Id}_d)$, wiemy, że

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P} \left(\sqrt{\frac{\delta}{n}} |Z| \geq \frac{\varepsilon}{2} \right) \leq -\frac{\varepsilon^2}{8\delta}.$$

Ponieważ

$$\left\{ X_n + \sqrt{\frac{\delta}{n}} Z \in B_{\varepsilon/2}(x) \right\} \subset \{X_n \in B_\varepsilon(x)\} \cup \left\{ \sqrt{\frac{\delta}{n}} |Z| > \frac{\varepsilon}{2} \right\}$$

to

$$\mathbb{P}(X_n \in B_\varepsilon(x)) \geq \mathbb{P}\left(X_n + \sqrt{\frac{\delta}{n}}Z \in B_{\varepsilon/2}(x)\right) - \mathbb{P}\left(\sqrt{\frac{\delta}{n}}|Z| > \frac{\varepsilon}{2}\right)$$

i stąd

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_n \in B_\varepsilon(x)) \geq \max \left\{ -\Lambda^*(x), -\frac{\varepsilon^2}{8\delta} \right\} \\ \xrightarrow{\delta \rightarrow 0} -\Lambda^*(x),$$

dowodząc pożądanego dolnego ograniczenia

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{P}(X_n \in B_\varepsilon(x)) \geq -\Lambda^*(x),$$

dla dowolnego $x \in \mathbb{R}^d$ z $\Lambda^*(x) < \infty$. □

5.2 Łańcuchy Markowa

Niech $\Sigma = \{1, 2, \dots, l\}$ będzie przestrzenią stanów łańcucha Markowa $\{X_n\}_{n \in \mathbb{N}}$. Przez P oznaczamy będziemy macierz przejścia. Dla uproszczenia zakładamy będziemy, że wszystkie wyrazy P są dodatnie. (założenie to można osłabić do nierozkładalności P). Niech $f: \Sigma \rightarrow \mathbb{R}$ będzie dowolną funkcją. Rozważmy

$$S_n = \sum_{k=1}^n f(X_k).$$

Przez μ_n oznaczamy będziemy rozkład S_n/n . Pokażemy, że ciąg $\{\mu_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyłeń.

Twierdzenie 5.9 (Perron-Frobenius)

Dla macierzy kwadratowej $B = (b_{ij})_{n \times n}$ z dodatnimi wyrazami istnieje rzeczywista i dodatnia wartość własna λ (zwana wartością własną Perrona-Frobeniusa), która spełnia następujące właściwości:

1. λ jest rzeczywista i dodatnia.
2. Dla każdej wartości własnej $\lambda' \neq \lambda$, $|\lambda'| < \lambda$.
3. λ ma krotność 1.
4. λ ma wektor własny z wszystkimi dodatnimi wyrazami.
5. Dla każdego $1 \leq i \leq n$ i każdego wektora Φ z dodatnimi wyrazami:

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \left(\sum_{j=1}^n B^m(i, j) \Phi_j \right) = \lim_{m \rightarrow \infty} \frac{1}{m} \log \left(\sum_{j=1}^n B^m(j, i) \Phi_j \right) = \log \lambda$$

Dowód. Niech

$$C = \left\{ \varphi \in \mathbb{R}^n : \forall i, \varphi(i) \geq 0, \sum_{i=1}^n \varphi(i) = 1 \right\}.$$

Rozważmy $f : C \rightarrow C$ zadane jako $f(\varphi) = B\varphi / \|B\varphi\|$. Funkcja f jest ciągła. Zgodnie z twierdzeniem Brouwer'a o punkcie stałym, istnieje $\varphi \in C$ takie, że $f(\varphi) = \varphi$. Zauważmy, że istnieje $j_0 > 0$ takie, że $\varphi(j_0) > 0$. Niech $\rho = \|B\varphi\| > 0$. Wtedy $B\varphi = \rho\varphi$. Dla każdego i :

$$\varphi(i) = \frac{B\varphi(i)}{\rho} = \frac{1}{\rho} \sum_{j=1}^n B_{ij} \varphi_j \geq \frac{1}{\rho} B_{ij_0} \varphi_{j_0} > 0$$

Zatem wszystkie wyrazy φ są dodatnie. Znaleźliśmy wartość własną ρ , która jest rzeczywista, dodatnia i posiada wektor własny z dodatnimi współrzędnymi.

Zdefiniujmy nową macierz $A = (a_{ij})_{n \times n}$ jako $a_{ij} = b_{ij} \cdot \varphi_j / \varphi_i$. Niech λ będzie wartością własną macierzy B z wektorem własnym α . Niech β będzie wektorem, gdzie $\beta_i = \alpha_i / \varphi_i$. Wówczas:

$$(A\beta)_i = \sum_{j=1}^n a_{ij} \beta_j = \sum_{j=1}^n b_{ij} \cdot \frac{\varphi_j}{\varphi_i} \cdot \alpha_j = \frac{1}{\varphi_i} \cdot \sum_{j=1}^n b_{ij} \alpha_j = \frac{1}{\varphi_i} (B\alpha)_i = \rho \alpha_i = \rho \beta_i$$

czyli β jest wektorem własnym macierzy A z wartością własną λ . Jeżeli rozważymy $\lambda = \rho$, to stwierdzimy, że $(1, \dots, 1)^T$ jest wektorem własnym macierzy A z wartością własną ρ . Oznacza to, że dla każdego i : $\sum_{j=1}^n a_{ij} = \rho$.

Niech λ będzie wartością własną z wektorem własnym β (dla A) i niech j_0 będzie takie, że $|\beta_{j_0}| \geq |\beta_j|$ dla każdego j . Wówczas

$$|\lambda| = \left| \frac{(A\beta)_{j_0}}{\beta_{j_0}} \right| = \frac{1}{|\beta_{j_0}|} \sum_{j=1}^n a_{j_0j} \beta_j \leq \left(\sum_{j=1}^n a_{j_0j} \right) \frac{|\beta_j|}{|\beta_{j_0}|} \leq \sum_{j=1}^n a_{j_0j} = \rho. \quad (5.16)$$

Musimy pokazać, że jeśli λ nie jest rzeczywistą wartością dodatnią, to $|\lambda| < \rho$. Załóżmy b.z.o., że $\beta_{j_0} > 0$. Jeśli λ nie jest rzeczywistą wartością dodatnią, to istnieje j_1 takie, że $\beta_{j_1} \neq 0$, ale nie jest dodatnie. Wtedy jednak

$$|a_{j_0j_0} \beta_{j_0} + a_{j_0j_1} \beta_{j_1}| < |a_{j_0j_0} \beta_{j_0}| + |a_{j_0j_1} \beta_{j_1}|,$$

więc nierówność w(5.16) jest ostra i $|\lambda| < \rho$.

Teraz pokażemy, że ρ ma krotność 1. Niech

$$D = \left\{ \beta \in \mathbb{C}^d : \sum_{i=1}^n |\beta_i| = 1, \sum_{i=1}^n \beta_i = 0 \right\}.$$

Wtedy każde $\gamma \in \mathbb{C}^d$ ma prawie jednoznaczne przedstawienie jako $a \cdot (1, \dots, 1)^T + b \cdot \theta$, gdzie $\theta \in D$. Dla każdego $\theta \in D$ zachodzi $\|A\mathcal{V}\|_1 < \rho \|\mathcal{V}\|_1$. Ze zwartości $\sup_{\theta \in D} \|A\theta\|_1 =: \kappa < \rho$. Załóżmy, że α jest innym wektorem własnym z wartością własną ρ . Dla każdej stałej a , $a\alpha - (1, \dots, 1)^T$ jest wektorem własnym z wartością własną ρ , ale istnieje a takie, że $a\alpha - (1, \dots, 1)^T \in D$. Sprzeczność.

Niech Φ ma dodatnie współrzędne. Wtedy $\Phi = a\varphi + u$, gdzie u składa się z innych wektorów własnych. Wówczas $B^m\Phi = aB^m\varphi + B^m u = a\lambda^m\varphi + B^m u$ i $\|B^m u\| \leq c \cdot \kappa^m$, więc

$$\sum_{j=1}^n B^m(i, j)\Phi_j = (B^m\Phi)_i = a\rho^m i + o(\rho^m)$$

i stąd

$$\lim_{m \rightarrow \infty} \frac{1}{m} \log \left(\sum_{j=1}^n B^m(i, j)\Phi_j \right) = \log \rho.$$

□

Zamiast dodatnich współrzędnych B możemy założyć ierozkładalność i aperiodyczność, tj. istnieje m takie, że dla każdego i, j , $B^m(i, j) > 0$. Przypomnijmy, że $\Sigma = \{1, \dots, l\}$. Niech P będzie macierzą przejść, tj. dla każdego i : $\sum_{j=1}^l P_{ij} = 1$ i dla każdego i, j : $P_{ij} \geq 0$. Załóżmy, że P jest nierozkładalne i aperiodyczne. Dla ustalonego $\theta \in \Sigma$, niech (X_n) będzie łańcuchem Markowa takim, że $X_0 = \theta$ i $\mathbb{P}[X_{n+1} = j | X_n = i] = P_{ij}$. Niech $f : \Sigma \rightarrow \mathbb{R}^d$ i niech $Y_n = f(X_n)$, $Z_n = \frac{1}{n} \sum_{k=1}^n Y_k$.

Twierdzenie 5.10

Ciąg rozkładów $\{Z_n\}_{n \in \mathbb{N}}$ spełnia zasadę wielkich odchyień.

Dowód. Dla każdego λ rozważmy

$$\Lambda_n(\lambda) := \log E \left[e^{\langle \lambda, Z_n \rangle} \right].$$

Aby uzasadnić zasadę wielkich odchyień, musimy pokazać, że

$$\Lambda(\lambda) = \lim_{n \rightarrow \infty} \frac{1}{n} \Lambda_n(n\lambda)$$

istnieje i jest różniczkowalna. Udowodnimy jedynie istnienie:

$$n\Lambda_n(n\lambda) = \frac{1}{n} \log \mathbb{E} \left[e^{\langle n\lambda, Z_n \rangle} \right] = \frac{1}{n} \log \mathbb{E} \left[\exp \left\{ \left\langle \lambda, \sum_{k=1}^n Y_k \right\rangle \right\} \right].$$

Z własności Markowa wartość oczekiwana pod logarytmem rozpisuje się jako

$$\begin{aligned} \mathbb{E} \left[\exp \left\{ \left\langle \lambda, \sum_{k=1}^n Y_k \right\rangle \right\} \right] &= \sum_{x_1, \dots, x_n \in \Sigma} \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] \prod_{k=1}^n e^{\langle \lambda, f(x_k) \rangle} \\ &= \sum_{x_1, \dots, x_n \in \Sigma} P(\theta, x_1) \cdots P(x_{n-1}, x_n) \prod_{k=1}^n e^{\langle \lambda, f(x_k) \rangle} \end{aligned}$$

Dla każdego $\lambda \in \mathbb{Z}^d$, zdefiniujmy macierz P^λ wzorem

$$P^\lambda(i, j) := P(i, j) e^{\langle \lambda, f(j) \rangle}.$$

P^λ jest nieujemna, nieredukowalna i aperiodyczna. Mamy

$$\begin{aligned} \mathbb{E} \left[\exp \left\{ \left\langle \lambda, \sum_{k=1}^n Y_k \right\rangle \right\} \right] &= \sum_{x_1, \dots, x_n \in \Sigma} P^\lambda(\theta, x_1) \cdots P^\lambda(x_{n-1}, x_n) \\ &= \sum_{j \in \Sigma} (P^\lambda)^n(\theta, j) \end{aligned}$$

Podsumowując

$$n\Lambda_n(n\lambda) = \frac{1}{n} \log \mathbb{E} \left[\exp \left\{ \left\langle \lambda, \sum_{k=1}^n Y_k \right\rangle \right\} \right] = \frac{1}{n} \log \sum_{j \in \Sigma} (P^\lambda)^n(\theta, j)$$

Zbieżność powyższego wyniku z Twierdzenia Perrona-Frobeniusa □

Dyfuzje

Streszczenie Zaczniemy

Istnieje wiele sytuacji, w których rozwiązania równań różniczkowych cząstkowych są wyrażane jako całki po przestrzeniach funkcyjnych. Najprostszym przykładem nietrywialnym jest wzór Feynmana-Kaca, który wyraża rozwiązanie równanie

$$\partial_t u = \frac{1}{2} \Delta u + v(x)u,$$

przy warunku początkowym $u(0, x) = f(x)$ jako całkę po przestrzeni funkcyjnej

$$u(t, x) = \mathbb{E} \left[\exp \left(\int_0^t v(B_s^x) ds \right) \cdot f(B_t^x) \right]$$

Tu B^x jest d -wymiarowym ruchem Browna zapoczątkowanym w punkcie x . Kolejnym przykładem jest rozwiązanie zagadnienia Dirichleta

$$L_1 u = \frac{1}{2} \Delta u + \langle \nabla u, b \rangle = 0$$

w ograniczonym otwartym zbiorze G w \mathbb{R}^d z gładkim brzegiem ∂G i warunkiem brzegowym Dirichleta $u = f$ na ∂G . Tutaj $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ jest gładkim polem wektorowym. Rozwiązanie zagadnienie można przedstawić za pomocą dyfuzji $X^x = (X_t^x)_{t \geq 0}$ spełniającej stochastyczne równanie różniczkowe

$$dX_t^x = dB_t + b(X_t^x)dt$$

z warunkiem początkowym $X_0^x = x$. W tym przypadku proces X^x można zdefiniować jako rozwiązanie

$$X_t^x = x + \int_0^t b(X_s^x) ds + B_t.$$

Dla dowolnej funkcji $v \in C^2(\mathbb{R}^d)$ proces

$$M_t = v(X_t^x) - \int_0^t \langle \nabla v(X_s^x), b(X_s^x) \rangle + \frac{1}{2} \Delta v(X_s^x) ds$$

jest martyngałem (zwanym niekiedy martyngałem Dynkina) o średniej $v(x)$ (zadanie). Jeżeli rozważymy czas zatrzymania

$$\tau_x = \inf\{s \geq 0 : X_s^x \notin G\},$$

można pokazać, że $\mathbb{E}[\tau_x] < \infty$ i skorzystać z twierdzenia Dooba o zatrzymaniu aby otrzymać

$$\mathbb{E}[M_\tau] = v(x).$$

Jeżeli $v = u$ jest rozwiązaniem wspomnianego zagadnienia Dirichleta, to powyższe redukuje się do

$$\mathbb{E}[u(X_{\tau_x}^x)] = u(x).$$

Jeżeli dodamy do tego fakt, że $X_{\tau_x}^x \in \partial G$, to powyższe daje probabilistyczną reprezentację rozwiązania jako

$$\mathbb{E}[f(X_{\tau_x}^x)] = u(x).$$

które można wyrazić jako:

$$u(x) = \mathbb{E}^*[f(x(t))]$$

Chcemy zbadać jaki wpływ na zachowanie rozwiązania ma wygaszanie części Gaussowskiej w rozwiązaniu. Dla $\epsilon \in [0, 1]$ rozważmy

$$L_\epsilon = \frac{\epsilon}{2} \Delta u + \langle \nabla u, b \rangle.$$

Niech $u = u_\epsilon$ będzie rozwiązaniem zagadnienia

$$L_\epsilon u = 0$$

w G z warunkiem brzegowym $u = f$ na ∂G . Chcemy badać jak zachowuje się u_ϵ przy $\epsilon \rightarrow 0$. W naturalny sposób zwracamy się do stowarzyszonego procesu dyfuzji

$$X_t^{x,\epsilon} = x + \int_0^t b(X_s^{x,\epsilon}) ds + \sqrt{\epsilon} B_t.$$

Widzimy, że przypadek graniczny $\epsilon = 0$ odpowiada po prostu rozwiązaniu równania zwyczajnego

$$\frac{dF(t)}{dt} = b(F(t)), \quad F(0) = x.$$

Skoro ostatecznie interesuje nas problem wyjścia $X^{x,\epsilon}$ z G widzimy, że kluczowe jest zachowanie portretów fazowych powyższego równania. Jeżeli F wychodzi z G przez powiedzmy punkt $y_0 \in \partial G$, to łatwo zgadnąć, że $u_\epsilon(x) \rightarrow f(y_0)$. Sytuacja, w której F nie wychodzi z G jest problematyczna.

6.1 Dyfuzje

Niech $B = (B_t)_{t \geq 0}$ będzie d -wymiarowym ruchem Browna. Dla $x \in \mathbb{R}^d$ i $\epsilon > 0$ rozważmy proces $X^{x,\epsilon}$ zadany przez

$$X_t^{x,\epsilon} = x + \int_0^t b(X_s^{x,\epsilon}) ds + \sqrt{\epsilon} B_t,$$

gdzie $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ jest funkcją spełniającą warunek Lipschitza. Chcemy opisać duże odchylenia dla $X^{x,\epsilon}$ przy $\epsilon \rightarrow 0$. Przez $\mathbf{P}_{x,\epsilon}$ będzie rozkładem $X^{x,\epsilon}$ na $C(\mathbb{R}^d)$.

Ustalmy $x \in \mathbb{R}^d$. Niech $\kappa, g: [0, +\infty) \rightarrow \mathbb{R}^d$ będą związane

$$\kappa(t) = x + \int_0^t b(\kappa(s)) ds + g(t).$$

Funkcja uwikłana $x \mapsto g$ jest ciągła. Chcemy zbadać ciągłość odwzorowania odwrotnego. Rozważmy w tym celu $\eta, h: \rightarrow \mathbb{R}^d$ uwikłane tym samym równaniem

$$\eta(t) = x + \int_0^t b(\eta(s)) ds + h(t).$$

Wówczas

$$\kappa(t) - \eta(t) = \int_0^t b(\kappa(s)) - b(\eta(s)) ds + g(t) - h(t).$$

Położmy

$$\Delta(t) = \sup_{s \leq t} \|\kappa(s) - \eta(s)\|, \quad \alpha(t) = \sup_{s \leq t} \|g(s) - h(s)\|.$$

Dla $s \leq t$ mamy

$$\begin{aligned} \|\kappa(s) - \eta(s)\| &\leq A \int_0^s \|\kappa(u) - \eta(u)\| du + \sup_{u \leq s} \|g(u) - h(u)\| \\ &\leq A \int_0^t \|\kappa(u) - \eta(u)\| du + \sup_{u \leq t} \|g(u) - h(u)\| \end{aligned}$$

Biorąc supremum po $s \leq t$ otrzymujemy

$$\Delta(t) \leq A \int_0^t \Delta(u) du + \alpha(t) \tag{6.1}$$

Mnożąc stronami przez czynnik całkujący e^{-At} dostajemy

$$\int_0^t \Delta(s) ds \leq e^{At} + \int_0^t \alpha(s) e^{-As} ds \leq e^{At} \alpha(t) \frac{1 - e^{-At}}{A}.$$

Wstawiając to oszacowanie do (6.1) otrzymujemy

$$\Delta(t) \leq e^{At} \sup_{s \leq t} \|g(s) - h(s)\|.$$

Powyższe rozumowanie jest po prostu dowodem Lematu Gronwalla. Wynika stąd, że $g \mapsto x$ jest ciągłe $C_0([0, t], \mathbb{R}^d) \rightarrow C([0, T], \mathbb{R}^d)$, gdzie dla $x \in \mathbb{R}^d$,

$$C_x([0, t], \mathbb{R}^d) = \{f: [0, T] \rightarrow \mathbb{R}^d : f \text{ ciągłe}, f(0) = x\}.$$

Twierdzenie 6.1

Ustalmy $T > 0$ oraz $x \in \mathbb{R}^d$. Miary $\{\mathbf{P}_{x,\epsilon}\}_{\epsilon>0}$ spełniają zasadę wielkich odchyień z prędkością $1/\epsilon$ i funkcją tempa

$$I_b(f) = \begin{cases} \frac{1}{2} \int_0^T \|f'(t) - b(f(t))\|^2 dt & f \in \mathcal{H} \\ +\infty & f \notin \mathcal{H} \end{cases}$$

gdzie

$$\mathcal{H} = \left\{ f(t) = \int_0^t g(s) ds : g \in L^2([0, t], \mathbb{R}^d) \right\}$$

Dowód. Niech \mathbf{Q}_ϵ będzie rozkładem $\sqrt{\epsilon}B$ na $C_0([0, T], \mathbb{R}^d)$. Wówczas $\{\mathbf{Q}_\epsilon\}_{\epsilon>0}$ spełniają zasadę wielkich odchyień z prędkością $1/\epsilon$ i funkcją tempa

$$I(f) = \begin{cases} \frac{1}{2} \int_0^T \|f'(t)\|^2 dt & f \in \mathcal{H} \\ +\infty & f \notin \mathcal{H} \end{cases}$$

Niech teraz $\phi_x(g) = \kappa$ będzie rozwiązaniem

$$\kappa(t) = x + \int_0^t b(\kappa(s)) ds + g(t).$$

Z dyskusji poprzedzającej dowód twierdzenia wiemy, że ϕ_x jest funkcją ciągłą. Skoro $\mathbf{P}_{x\epsilon} = \mathbf{Q}_\epsilon \circ \phi_x^{-1}$, to z zasady kontrakcji $\{\mathbf{P}_{x,\epsilon}\}_\epsilon$ spełniają zasadę wielkich odchyień z funkcją tempa

$$I_b(f) = \inf\{I(g) : \phi_x(g) = f\}.$$

Jeżeli $\phi_x(g) = f$, dla $g \in \mathcal{H}$, to

$$f'(t) = b(f(t)) + g'(t)$$

a co za tym idzie

$$I(g) = \frac{1}{2} \int_0^T \|g'(s)\|^2 ds = \frac{1}{2} \int_0^T \|f'(t) - b(f(t))\|^2 dt.$$

□

6.2 Problem wyjścia

Założenie (V) Funkcja b jest postaci $b(x) = -\nabla V(x)$, gdzie V jest gładką funkcją taką, że

- $\inf_{\partial G} V$ jest przyjęte w dokładniej jednym punkcie $y_0 \in \partial G$
- Istnieje $x_0 \in G$ takie, że wszystkie rozwiązania $u: [0, +\infty) \rightarrow \mathbb{R}^d$ ($Du(t) = b(u(t))$) z warunkiem początkowym $u(0) \in G$ pozostają w G oraz $u(t) \rightarrow x_0$ gdy $t \rightarrow \infty$.

Lemat 6.2

Przy założeniu (V), dla dowolnego $x \in G$,

$$\inf_{T>0} \inf_h \int_0^T \|h'(s) + \nabla V(h(s))\|^2 ds = 4(V(x) - V(x_0)),$$

gdzie infimum jest wzięte po wszystkich $h: [0, T] \rightarrow G$ takich, że $h(0) = x$ i $h(T) = x_0$.

Dowód. Dla $T > 0$ rozważmy zagadnienie

$$x'(t) = -\nabla V(x(t)), \quad x(0) = x.$$

Odwracając kierunek otrzymujemy ścieżkę $h(t) = x(T - t)$ otrzymujemy sieżkę między $h(0) = x(T)$ a $h(T) = x$. Skoro

$$h'(t) = \nabla V(h(t))$$

to

$$\begin{aligned} \int_0^T \|h'(s) + \nabla V(h(s))\|^2 ds &= \\ &= \int_0^T \|h'(s) - \nabla V(h(s))\|^2 ds + 4 \int_0^T \langle \nabla V(h(s)), h'(t) \rangle ds \\ &= 4(V(h(T)) - V(h(0))) = 4(V(x) - V(x(T))). \end{aligned}$$

Skoro $x(T) \rightarrow x_0$ przy $T \rightarrow \infty$, to

$$\inf_{T>0} \inf_h \int_0^T \|h'(s) + \nabla V(h(s))\|^2 ds \leq 4(V(x) - V(x_0)),$$

Z drugiej strony dla dowolnego h takiego, że $h(T) = x$ i $h(0) = x_0$,

$$\begin{aligned} 4(V(x) - V(x_0)) &= 4 \int_0^T \langle \nabla V(h(t)), h'(t) \rangle dt \\ &= \int_0^T \|h'(s) + \nabla V(h(s))\|^2 ds - \int_0^T \|h'(s) - \nabla V(h(s))\|^2 ds \\ &\leq \int_0^T \|h'(s) + \nabla V(h(s))\|^2 ds. \end{aligned}$$

□

Lemat 6.3

Niech U będzie dowolnym otoczeniem x_0 . Połóżmy

$$\Lambda(U, T) = \inf_f \int_0^T \|f'(s) + \nabla V(f(s))\|^2 ds,$$

gdzie infimum jest wzięte po wszystkich $f: [0, T] \rightarrow G$ takich, że $f(t) \in G \setminus U$. Wówczas

$$\liminf_{T \rightarrow \infty} \Lambda(U, T) = \infty.$$

Powyższy lemat mówi, że ścieżki X^x , które przez długi czas pozostają oddalone od x_0 są mało prawdopodobne.

Dowód. Załóżmy nie wprost, że istnieje $M < \infty$ takie, że

$$\liminf_{T \rightarrow \infty} \Lambda(U, T) \leq M.$$

Oznacza to, że dla każdego $T > 0$ istnieje f_T takie, że f_T jest odcięte od x_0 i

$$\int_0^T \|f'_T(t) + \nabla V(f(t))\|^2 dt \leq 2M.$$

Skoro ∇V jest ciągła, to możemy bez zmniejszania ogólności założyć, że

$$\int_0^T \|f'_T(t)\|^2 dt \leq 2M$$

potencjalnie zwiększając M . Dla każdego $C > 0$ istnieje zatem t_C takie, że

$$\int_{t_C}^{t_C+C} \|f'(t)\|^2 + \|f'_T(t) + \nabla V(f(t))\|^2 dt \leq 4MC/T.$$

Przy ustalonym $C > 0$ funkcje $f_T(t - t_C)$ są jednakowo ciągłe. Wybierając podciąg zbieżny dostajemy $f: [0, C] \rightarrow \mathbb{R}$ rozwiązujące $f'(t) = -\nabla V(f(t))$. Z jedyności rozwiązania te muszą być zgodne dla różnych $C > 0$. Możemy więc przejść z C do nieskończoności i dostać rozwiązanie $f'(t) = -\nabla V(f(t))$, które jest odcięte od x_0 . \square

6.3 Własność Markowa

Dla uproszczenia notacji połóżmy $X = X^{0,1}$, czyli rozwiązanie zagadnienia

$$X_t = \int_0^t b(X_s) ds + B_t.$$

Wszelkie własności probabilistyczne X przenoszą się łatwo na $X^{x,\epsilon}$.

Definicja 6.4

Niech $(\mathcal{F}_t)_{t \geq 0}$ będzie ustaloną filtracją. Powiemy, że proces $Y = (Y_t)_{t \geq 0}$ o trajektoriach ciągłych jest procesem Markowa, jeżeli dla dowolnej funkcji $\Phi: C([0, +\infty) \rightarrow [0, +\infty))$ mamy i dowolnego $s > 0$ mamy

$$\mathbb{E} [\Phi((Y_{t+s})_{t \geq 0}) | \mathcal{F}_s] = \mathbb{E} [\Phi((Y_{t+s})_{t \geq 0}) | X_s].$$

Przykład 6.5

Ruch Browna ma własność Markowa. Przypomnijmy, że dla dowolnego $s > 0$ proces $B_t^{(s)} = B_{t+s} - B_s$ jest niezależny od σ -ciała $\mathcal{F}_s^B = \sigma(B_r : r \leq s)$. Stąd dla dowolnej funkcji $\psi: C([0, +\infty)) \rightarrow [0, +\infty)$

$$\mathbb{E} [\psi((B_{t+s})_{t \geq s}) | \mathcal{F}_s^B] = \mathbb{E} [\psi(B^{(s)} + B_s) | \mathcal{F}_s^B] = \Psi(B_s),$$

gdzie

$$\Psi(x) = \mathbb{E} [\psi(B + x)].$$

W szczególności

$$\begin{aligned} \mathbb{E} [\psi((B_{t+s})_{t \geq s}) | B_s] &= \mathbb{E} \left[\mathbb{E} [\psi((B_{t+s})_{t \geq s}) | \mathcal{F}_s^B] | B_s \right] = \\ &= \mathbb{E} [\Psi(B_s) | B_s] = \Psi(B_s) = \mathbb{E} [\psi((B_{t+s})_{t \geq s}) | \mathcal{F}_s^B]. \end{aligned}$$

Fakt 6.6

Proces X ma własność Markowa.

Dowód. Dla $g \in C([0, +\infty))$ przez $\phi(g) = y$ oznaczmy rozwiązanie równania całkowego

$$y(t) = \int_0^t b(y(s)) ds + g(t).$$

Wtedy $\phi: C([0, +\infty) \rightarrow C([0, +\infty))$ jest ciągła. Dodatkowo

$$X = \phi(B)$$

Jeżeli rozważymy teraz proces przesunięty $X_t^{[s]} = X_{t+s}$ to dla $s > 0$,

$$X_t^{[s]} = \int_0^s b(X_r^{[s]}) dr + B^{(s)} + X_s.$$

Czyli

$$X^{[s]} = \phi(B^{(s)} + x).$$

Dla dowolnej $\psi: C([0, +\infty) \rightarrow [0, +\infty))$ mamy

$$\mathbb{E}[\psi(X^{[s]})\mathcal{F}_s^B] = \mathbb{E}[\psi \circ \phi(B^{(s)} + X_s) | \mathcal{F}_s^B] = \Psi(X_s)$$

gdzie

$$\Psi(x) = \mathbb{E}[\psi \circ \phi(B^{(s)} + x)]$$

Własność Markowa wynika z powyższego dokładnie tak samo, jak miało to miejsce dla ruchu Browna. \square

6.4 Problem wyjścia 2

Przypomnijmy, że dla $G \subset \mathbb{R}^d$ otwartego, ograniczonego, spójnego i o gładkim brzegu rozważamy rozwiązanie $u = u_\epsilon$ zagadnienia

$$\begin{aligned} 0 &= \epsilon \Delta u / 2 + \langle b | \nabla u \rangle && \text{w } G \\ u &= f && \text{na } \partial G. \end{aligned}$$

Twierdzenie 6.7

Przy założeniu (V),

$$\lim_{\epsilon \rightarrow 0^+} u_\epsilon(x) = f(y_0)$$

dla każdego $x \in G$.

Rozważane zagadnienie ma reprezentację probabilistyczną

$$u_\epsilon(x) = \mathbb{E}[f(X_{\tau^{x,\epsilon}}^{\epsilon,x})],$$

gdzie

$$\tau^{x,\epsilon} = \inf\{t \geq 0 : X^{x,\epsilon} \notin G\}.$$

Dowód. Niech N będzie dowolnym otoczeniem y_0 w ∂G . Powiedzmy, że

$$\inf_{y \in \partial G \setminus N} V(y) = V(y_0) + \theta$$

dla pewnej $\theta > 0$. Niech teraz U_2 będzie takim otoczeniem x_0 w G , że $V(x) \leq V(x_0) + \theta/10$ dla $x \in U_2$. Niech $U_1 \subseteq \text{cl}(U_1) \subseteq U_2$ będzie kolejnym otoczeniem x_0 . Dla $g \in C([0, +\infty))$ niech

$$\tau = \tau(g) = \inf\{t \geq 0 : g(t) \notin G\}.$$

Wówczas

$$u_\epsilon(x) = \int f(g(\tau(g))) \mathbf{P}_{x,\epsilon}(dg).$$

Pokażemy, że dla każdego $x \in G$,

$$\lim_{\epsilon \rightarrow 0} \mathbf{P}_{x,\epsilon}[g : g(\tau) \notin N] = 0.$$

Rozważmy ciąg czasów zatrzymania $\tau_j = \tau_j(g)$,

$$\begin{aligned}\tau_1 &= \inf\{t \geq 0 : g(t) \in cl(U_1)\} \wedge \tau \\ \tau_2 &= \inf\{t \geq \tau_1 : g(t) \notin U_2\} \\ \tau_{2k+1} &= \inf\{t \geq \tau_{2k} : g(t) \in cl(U_1)\} \wedge \tau \\ \tau_{2k+2} &= \inf\{t \geq \tau_{2k+1} : g(t) \notin U_2\}\end{aligned}$$

Korzystając z mocnej własności Markowa dla X wystarczy pokazać, że

$$\lim_{\epsilon \rightarrow 0} \frac{\sup_{x \in \partial U_2} \mathbf{P}_{x,\epsilon}[\tau_1 > \tau, g(\tau) \notin N]}{\inf_{x \in \partial U_2} \mathbf{P}_{x,\epsilon}[\tau_1 > \tau, g(\tau) \in N]} = 0.$$

Dla licznika mamy

$$a(x, \epsilon) = \mathbf{P}_{x,\epsilon}[\tau_1 = \tau, g(\tau) \notin N] \mathbf{P}_{x,\epsilon}[T > \tau_1 = \tau, g(\tau) \notin N] + \mathbf{P}_{x,\epsilon}[T < \tau_1]$$

Dla drugiego składnika, na mocy udowodnionego wcześniej lematu

$$\lim_{T \rightarrow \infty} \limsup_{\epsilon \rightarrow 0} \epsilon \log \mathbf{P}_{x,\epsilon}[T < \tau_1] = -\infty.$$

Dla pierwszego mamy

$$\begin{aligned}\lim_{T \rightarrow \infty} \limsup_{\epsilon \rightarrow 0} \epsilon \log \sup_{x \in \partial U_2} \mathbf{P}_{x,\epsilon}[T > \tau_1 = \tau, g(\tau) \notin N] \\ - 2 \inf_{y \in N^c} \inf_{x \in \partial U_2} (V(y) - V(x)) \leq -3\theta/2 - 2(V(y_0) - V(x_0)).\end{aligned}$$

Stąd

$$\limsup_{\epsilon \rightarrow 0} \sup_{x \in \partial U_2} a(x, \epsilon) \leq -3\theta/2 - 2(V(y_0) - V(x_0)).$$

Mianownik szacujemy z dołu przez

$$\begin{aligned}\liminf_{\epsilon \rightarrow 0} \inf_{y \in \partial U_2} \log \mathbf{P}_{x,\epsilon}[\tau_1 = \tau, g(\tau) \in N] \\ \geq -2 \sup_{x \in \partial U_2} (V(y_0) - V(x)) \geq -2(V(y_0) - V(x_0)) - \theta/5.\end{aligned}$$

□

Uwaga 6.8

Założenie, że $b = -\nabla V$ nie jest istotne. Można założyć, że rozwiązania $g' = b(g)$ zaczęte w G w nim pozostają i zbiegają do jedyne punktu stacjonarnego x_0 oraz założyć, że

$$V(x) = \frac{1}{4} \inf_{T > 0} \inf_{h: h(0)=0, h(T)=x_0} \int_0^T \|h'(t) - b(h(t))\|^2 dt$$

przyjmuje z $\inf_{\partial G}$ przyjętym dokładnie w jednym punkcie y_0 .

Literatura

Cra38. Harald Cramér. Sur un nouveau theoreme-limite de la theorie des probabilités. *Scientifiques et Industrielles*, 736:5–23, 1938.