Wysokowymiarowy Rachunek Prawdopodobieństwa

Rafał Latała

20 listopada 2018

Poniższe notatki powstają na podstawie wykładów (i wybranych ćwiczeń) z Wysokowymiarowego Rachunku Prawdopodobieństwa, prowadzonych w semestrze jesiennym 2018/19.

Przepraszam za wszystkie nieścisłości i omyłki mogące pojawić się w tekście i jednocześnie zwracam się z prośbą do czytelników, którzy zauważyli błędy lub mają jakieś inne uwagi na temat notatek o kontakt mailowy na adres `rlatala@mimuw.edu.pl` z podaniem wersji notatek (daty) do której chcą się ustosunkować.

Dziękuję pani Marcie Strzeleckiej za uważne przejrzenie notatek.

Rafał Latała

Spis treści

1 Wstęp

W wielu problemach rachunku prawdopodobieństwa i jego zastosowań pojawiają się wielowymiarowe obiekty losowe takie jak wektory losowe, macierze losowe, procesy stochastyczne czy grafy losowe. Celem wykładu będzie przedstawienie wybranych narzędzi pozwalających badać takie obiekty. Wykład będzie dotyczył tak zwanej teorii nieasymptotycznej, tzn. nacisk będzie położony na różne szacowania, a nie na twierdzenia graniczne.

W pierwszej części wykładu omówimy pewne zagadnienia związane z teorią koncentracji miary, które pozwalają szacować odchylenia funkcji zależnej od wielu zmiennych losowych od jej wartości oczekiwanej. W drugiej pokażemy kilka metod pozwalających szacować suprema procesów stochastycznych. Omówimy też pewną liczbę bardziej konkretnych przykładów zastosowań.

Oczywiście podczas semestralnego wykładu monograficznego można omówić tylko niewielką część bogatej i ciągle rozwijającej się teorii. Dużo szerszy wybór zagadnień został przedstawiony w notatkach Ramona van Handela [6] i monografii Romana Vershynina [7], zainteresowany Czytelnik znajdzie tam też szersze zestawienie bibliografii.

2 Koncentracja miary - wprowadzenie

2.1 Funkcja koncentracji miary - definicja i przykłady.

Wiele ważnych miar probabilistycznych spełnia tzw. *fenomen koncentracji miary*. Nieformalnie rzecz biorąc polega on na tym, że większość punktów z przestrzeni leży w pobliżu zbioru wypełniającego przynajmniej połowę przestrzeni. By pojęcie to sformalizować potrzebujemy dwóch ważnych definicji.

Definicja 2.1. Niech $(\mathbb{X}, d)$ będzie przestrzenią metryczną, zaś A dowolnym podzbiorem $\mathbb{X}$. Dla $t > 0$ określamy *t-otoczenie* zbioru A wzorem

$$A_t := \{x \in X \colon d(x, A) < t\} = \bigcup_{y \in A} B(y, t),$$

gdzie $B(y, t)$ oznacza kulę otwartą w $\mathbb{X}$ o środku w y i promieniu t.

Definicja 2.2. Niech μ będzie borelowską miarą probabilistyczną na $(\mathbb{X}, d)$. *Funkcję koncentracji* miary μ definiujemy jako

$$\alpha_\mu(t) = \alpha_{(\mathbb{X},d,\mu)}(t) := \sup\Big\{1 - \mu(A_t) \colon \mu(A) \geqslant \frac{1}{2}\Big\}.$$

Na początek wykładu podamy kilka przykładów, dla których można dobrze oszacować funkcję koncentracji. Dowody podanych oszacowań przedstawimy później.

Przykład 1. Niech d oznacza odległość geodezyjną na n-wymiarowej sferze $S^n = \{x \in \mathbb{R}^{n+1} \colon |x| = 1\}$, zaś σ_n oznacza unormowaną miarę powierzchniową na S^n. Wówczas

okazuje się, że jeśli chcemy zminimalizować $\sigma_n(A_t)$ po wszystkich zbiorach ustalonej miary, ekstremalne są kule (zwane też czapeczkami), to znaczy

$$\sigma_n(A) = \sigma_n(B(x_0, r)) \;\Rightarrow\; \sigma_n(A_t) \geqslant \sigma_n(B(x_0, r)_t) = \sigma_n(B(x_0, r+t)).$$

W szczególności jeśli $\sigma_n(A) \geqslant 1/2$, to

$$\sigma_n(A_t) \geqslant \sigma_n\Big(B\Big(x_0, \frac{\pi}{2} + t\Big)\Big) \geqslant 1 - \exp\Big(-\frac{(n-1)t^2}{2}\Big).$$

Zatem $\alpha_{\sigma_n}(t) \leqslant \exp(-\frac{n-1}{2}t^2)$.

Uwaga 2.3. Zauważmy, że funkcja koncentracji σ_n szybko zbiega do 0 przy $n \to \infty$. Jedną z przyczyn tego zjawiska jest to, że miara ta nie jest dobrze unormowana. Jeśli przez $\sigma_{n,R}$ określimy rozkład jednostajny na sferze RS^n, to ponieważ jest on obrazem σ_n przy jednokładności o skali R, to

$$\alpha_{\sigma_{n,R}}(t) = \alpha_{\sigma_n}\Big(\frac{t}{R}\Big) \leqslant \exp\Big(-\frac{n-1}{2R^2}t^2\Big).$$

Zauważmy też, że

$$\int_{RS^n} x_i x_j d\sigma_{n,R}(x) = \frac{R^2}{n+1}\delta_{i,j}.$$

Zatem miara jednostajna na $\sqrt{n+1}S^n$ ma dobrą normalizację, to znaczy taką, że macierz kowariancji jest identycznością. Dla tej miary dla $n \geqslant 2$,

$$\alpha_{\sigma_{n,\sqrt{n+1}}}(t) \leqslant \exp\Big(-\frac{n-1}{2(n+1)}t^2\Big) \leqslant \exp\Big(-\frac{1}{6}t^2\Big).$$

Przykład 2. Niech γ_k oznacza kanoniczny rozkład gaussowski na $\mathbb{R}^k$, tzn. rozkład z gęstością $(2\pi)^{-k/2}\exp(-|x|^2/2)$. Wówczas ekstremalnymi zbiorami w problemie izoperymetrycznym okazują się półprzestrzenie, tzn. jeśli

$$\gamma_k(A) = \gamma_k\Big((-\infty, r] \times \mathbb{R}^{k-1}\Big) = \Phi(r),$$

to

$$\gamma_k(A_t) \geqslant \gamma_k\Big(\big((-\infty, r] \times \mathbb{R}^{k-1}\big)_t\Big) = \gamma_k\Big((-\infty, r+t] \times \mathbb{R}^{k-1}\Big) = \Phi(r+t).$$

W szczególności

$$\alpha_{\gamma_k}(t) = 1 - \Phi(t) \leqslant \frac{1}{2}e^{-t^2/2}.$$

Zauważmy, że powyższe oszacowania nie zależą od wymiaru przestrzeni.

Przykład 3. Niech ν będzie symetrycznym rozkładem wykładniczym, tzn. rozkładem na $\mathbb{R}$ z gęstością $\frac{1}{2}\exp(-|x|)$. Przez ν^k będziemy oznaczać rozkład produktowy $\nu \otimes \ldots \otimes \nu$ na

$\mathbb{R}^k$. Wyznaczenie ekstremalnych zbiorów dla problemu izoperymetrycznego związanego z tą miarą jest trudne i nieznane dla $k \neq 1$. Choć wiadomo, że ekstremalne nie są półprzestrzenie postaci $(-\infty, r] \times \mathbb{R}^{k-1}$, to są one optymalne z dokładnością do stałej, tzn.

$$\nu^k(A) = \nu((-\infty, r]) \;\Rightarrow\; \nu^k(A_t) \geqslant \nu\Big(\Big(-\infty, r + \frac{1}{2\sqrt{6}}t\Big]\Big).$$

W szczególności

$$\alpha_{\nu^k}(t) \leqslant 1 - \nu\Big(\Big(-\infty, \frac{1}{2\sqrt{6}}t\Big]\Big) = \frac{1}{2}\exp\Big(-\frac{1}{2\sqrt{6}}t\Big).$$

Zauważmy, że znowu uzyskane oszacowanie nie zależy od wymiaru przestrzeni.

Przykład 4. Niech μ będzie unormowaną miarą liczącą na kostce dyskretnej $\{0,1\}^n$ z metryką $d(x,y) = \frac{1}{n}\#\{i\colon x_i \neq y_i\}$. Tu problem izoperymetryczny daje się rozwiązać (optymalne są kule, ewentualnie z dodanymi niektórymi punktami na brzegu). W tym przypadku można pokazać, że

$$\alpha_\mu(t) \leqslant e^{-2nt^2}.$$

Krótki przegląd wyników pokazuje, że w wielu ważnych zastosowaniach można wykazać, że $\alpha_\mu(t) \leqslant C_1 \exp(-t^2/C_2)$ – mówimy wtedy, że funkcja koncentracji jest *typu gaussowskiego*. Widzieliśmy też przykład, w którym $\alpha_\mu(t) \leqslant C_1 \exp(-t/C_2)$ – mówimy wtedy o koncentracji *wykładniczej*.

2.2 Koncentracja funkcji lipschitzowskich

W wielu zastosowaniach nie interesuje nas jak zmienia się miara otoczenia zbioru, a raczej jak szybko maleją ogony funkcji określonych na przestrzeni. W tej części powiążemy ze sobą te zjawiska. Zacznijmy od definicji mediany i modułu ciągłości.

Definicja 2.4. Niech μ będzie miarą probabilistyczną na $(\mathbb{X}, d)$ oraz f będzie mierzalną funkcją z $\mathbb{X}$ w $\mathbb{R}$.
Medianą funkcji f względem miary μ nazywamy taką liczbę $M = \mathrm{Med}_\mu(f)$ dla której

$$\mu(\{x\colon f(x) \geqslant M\}) \geqslant \frac{1}{2} \quad \text{oraz} \quad \mu(\{x\colon f(x) \leqslant M\}) \geqslant \frac{1}{2}.$$

Modułem ciągłości f nazywamy funkcję

$$w_f(t) := \sup\{|f(x) - f(y)|\colon d(x,y) \leqslant t\}.$$

Fakt 2.5. *Dla dowolnej funkcji mierzalnej $F\colon \mathbb{X} \to \mathbb{R}$,*

$$\mu(\{x\colon F(x) > \mathrm{Med}_\mu(F) + w_F(t)\}) \leqslant \alpha_\mu(t)$$

oraz

$$\mu(\{x\colon |F(x) - \mathrm{Med}_\mu(F)| > w_F(t)\}) \leqslant 2\alpha_\mu(t).$$

Dowód. Niech $A := \{x\colon F(x) \leqslant \mathrm{Med}_\mu(F)\}$ wówczas $\mu(A) \geqslant 1/2$ zatem $\mu(A_t) \geqslant 1 - \alpha_\mu(t)$. Ponadto, jeśli $x \in A_t$, to istnieje $y \in A$ takie, że $d(x,y) < t$ i wówczas $F(x) \leqslant F(y) + w_F(t) \leqslant \mathrm{Med}_\mu(F) + w_F(t)$, stąd pierwsza nierówność w fakcie. Stosując ją do $-F$ i zauważając, że $\mathrm{Med}_\mu(-F) = -\mathrm{Med}_\mu(F)$ oraz $w_{-F} = w_F$ dostajemy

$$\mu(\{x\colon F(x) < \mathrm{Med}_\mu(F) - w_F(t)\}) \leqslant \alpha_\mu(t).$$

Dodając powyższą nierówność do poprzedniej otrzymamy ostatnią część faktu. □

Przypomnijmy definicję funkcji lipschitzowskiej

Definicja 2.6. Funkcję $F\colon (\mathbb{X}, d) \to \mathbb{R}$ nazywamy *lipschitzowską*, jeśli

$$\|F\|_{\mathrm{Lip}} := \sup_{x \neq y} \frac{|F(x) - F(y)|}{d(x,y)} < \infty.$$

Mówimy, że funkcja jest L-lipschitzowska jeśli $\|F\|_{\mathrm{Lip}} \leqslant L$, tzn. $|F(x) - F(y)| \leqslant L d(x,y)$ dla wszystkich $x, y \in X$.

Analogicznie można zdefiniować funkcje lipschitzowskie między przestrzeniami metrycznymi.

Fakt 2.7. *i) Jeśli F jest lipschitzowska ze stałą L, to dla $t > 0$,*

$$\mu(\{x\colon F(x) > \mathrm{Med}_\mu(F) + t\}) \leqslant \alpha_\mu(t/L)$$

oraz

$$\mu(\{x\colon |F(x) - \mathrm{Med}_\mu(F)| > t\}) \leqslant 2\alpha_\mu(t/L).$$

ii) Na odwrót, jeśli dla każdej funkcji 1-lipschitzowskiej F i ustalonego $t > 0$,

$$\mu(\{x\colon F(x) \geqslant \mathrm{Med}_\mu(F) + t\}) \leqslant \alpha,$$

to $\alpha_\mu(t) \leqslant \alpha$.

Dowód. i) Wynika z Faktu 2.5 i oczywistego szacowania $w_f(t) \leqslant tL$.

ii) Ustalmy zbiór A taki, że $\mu(A) \geqslant 1/2$ i określmy $F(x) := d(x, A)$. Wówczas F jest 1-lipschitzowska oraz $\mathrm{Med}_\mu(F) = 0$, zatem

$$\alpha \geqslant \mu(\{F \geqslant t\}) = \mu(\{x\colon d(x,A) \geqslant t\}) = 1 - \mu(A_t).$$ □

Często łatwiej i naturalniej jest wykazywać koncentrację funkcji lipschitzowskich wokół średniej a nie mediany. Kolejny fakt pokazuje jak odzyskać funkcję koncentracji w takim przypadku.

Fakt 2.8. *Załóżmy, że μ jest miarą probabilistyczną na przestrzeni metrycznej $(\mathbb{X}, d)$ oraz dla ograniczonych funkcji 1-lipschitzowskich F i $t > 0$ zachodzi*

$$\mu\Big(\Big\{x\colon F(x) > \int F d\mu + t\Big\}\Big) \leqslant \alpha(t). \tag{1}$$

Wówczas dla dowolnego zbioru borelowskiego A takiego, że $\mu(A) > 0$ zachodzi

$$1 - \mu(A_t) \leqslant \alpha(\mu(A)t).$$

W szczególności

$$\alpha_\mu(t) \leqslant \alpha\Big(\frac{t}{2}\Big).$$

Ponadto, jeśli $\lim_{t\to\infty} \alpha(t) = 0$, to dowolna funkcja 1-lipschitzowska jest całkowalna względem μ i jeśli dodatkowo α jest ciągła, to (1) zachodzi dla wszystkich funkcji 1-lipschitzowskich.

Dowód. Ustalmy zbiór borelowski A taki, że $\mu(A) > 0$ oraz liczbę $t > 0$. Zdefiniujmy $F(x) := \min\{d(x, A), t\}$, wówczas funkcja F jest ograniczona, 1-lipschitzowska i $\int F d\mu \leqslant t(1 - \mu(A))$. Stąd na mocy (1),

$$1 - \mu(A_t) = \mu(\{F \geqslant t\}) \leqslant \mu\Big(\Big\{F \geqslant \int F d\mu + \mu(A)t\Big\}\Big) \leqslant \alpha(\mu(A)t).$$

W szczególności, jeśli $\mu(A) \geqslant 1/2$, to $1 - \mu(A_t) \leqslant \alpha(t/2)$.

By udowodnić drugą część faktu, ustalmy funkcję 1-lipschitzowską F i niech $F_n := \min\{|F|, n\}$. Z (1) zastosowanej do $-F_n$ dostajemy

$$\mu\Big(\Big\{x\colon F_n(x) \leqslant \int F_n d\mu - t\Big\}\Big) \leqslant \alpha(t).$$

Wybierzmy t_0 takie, że $\alpha(t_0) < 1/2$ oraz $m := \mathrm{Med}_\mu |F|$. Wówczas $\mu(\{F_n \leqslant m\}) \geqslant 1/2$, czyli zbiory $\{F_n \leqslant m\}$ oraz $\{F_n > \int F_n d\mu - t_0\}$ mają niepuste przecięcie. Zatem $\int F_n d\mu \leqslant m + t_0$ i z twierdzenia Lebesgue'a o zbieżności monotonicznej dostajemy $\int |F| d\mu \leqslant m + t_0 < \infty$. Ostatnią część tezy dostajemy stosując (1) do $\min\{\max\{F, -n\}, n\}$ i przechodząc z $n \to \infty$. □

3 Nierówności izoperymetryczne

W tej części omówimy kilka nierówności izoperymetrycznych, pokazując różne sposoby ich dowodzenia - poprzez powiązane nierówności funkcyjne, symetryzacje czy transport miary.

3.1 Klasyczna izoperymetria

Chociaż w tym wykładzie będziemy się zajmować miarami probabilistycznymi, to przegląd nierówności izoperymetrycznych zaczniemy od klasycznego przypadku n-wymiarowej miary Lebesgue'a λ_n.

Twierdzenie 3.1. *Jeśli A jest podzbiorem borelowskim $\mathbb{R}^n$ takim, że $\lambda_n(A) = \lambda_n(B(x_0, r))$, to dla dowolnego $t > 0$,*

$$\lambda_n(A_t) \geqslant \lambda_n(B(x_0, r)_t) = \lambda_n(B(x_0, r + t)).$$

Twierdzenie 3.2 (Nierówność Prékopy-Leindlera)**.** *Jeśli $s \in [0,1]$ oraz $f, g, h \colon \mathbb{R}^n \to [0, \infty)$ spełniają warunek*

$$h(sx + (1-s)y) \geqslant f(x)^s g(y)^{1-s} \quad \text{dla } x, y \in \mathbb{R}^n, \tag{2}$$

to

$$\int_{\mathbb{R}^n} h(x)dx \geqslant \Big(\int_{\mathbb{R}^n} f(x)dx \Big)^s \Big(\int_{\mathbb{R}^n} g(x)dx \Big)^{1-s}.$$

Dowód. Najpierw wykażemy, że dla niepustych zbiorów $A, B \in \mathcal{B}(\mathbb{R}^n)$ zachodzi

$$\lambda_1(A + B) \geqslant \lambda_1(A) + \lambda_1(B).$$

Ponieważ $\lambda_1(A) = \sup\{\lambda_1(K) \colon K \subset A, K \text{ zwarty}\}$, to możemy przyjąć, że zbiory A i B są zwarte. Ponadto odpowiednio je przesuwając możemy też zakładać, że $\sup A = \inf B = 0$. Wówczas $A \cap B = \{0\}$ oraz

$$\lambda_1(A + B) \geqslant \lambda_1(A \cup B) = \lambda_1(A) + \lambda_1(B).$$

Nierówność Prékopy-Leindlera udowodnimy przez indukcję po n. Najpierw rozważmy $n = 1$. Możemy zakładać, że f, g i h są ograniczone, a z uwagi na jednorodność, że $\sup f(x) = \sup g(x) = \sup h(x) = 1$. Zauważmy, że dla $0 \leqslant r < 1$, $\{h \geqslant r\} \supset s\{f \geqslant r\} + (1-s)\{g \geqslant r\}$, więc całkując przez części dostajemy

$$\begin{aligned}
\int h(x)dx &= \int_0^1 \lambda_1(\{h \geqslant r\})dr \geqslant \int_0^1 \lambda_1(s\{f \geqslant r\} + (1-s)\{g \geqslant r\})dr \\
&\geqslant \int_0^1 \lambda_1(s\{f \geqslant r\}) + \lambda_1((1-s)\{g \geqslant r\})dr \\
&= s \int f dx + (1-s) \int g dx \geqslant \Big(\int f dx \Big)^s \Big(\int g dx \Big)^{1-s},
\end{aligned}$$

gdzie ostatnia nierówność wynika z porównywania ważonych średnich arytmetycznych i geometrycznych.

Załóżmy teraz, że $n \geqslant 2$ oraz teza twierdzenia zachodzi dla $n-1$. Niech f, g, h spełniają (2) i określmy dla $x \in \mathbb{R}$

$$F(x) = \int_{\mathbb{R}^{n-1}} f(x,z)dz, \quad G(x) = \int_{\mathbb{R}^{n-1}} g(x,z)dz \quad \text{oraz} \quad H(x) = \int_{\mathbb{R}^{n-1}} h(x,z)dz.$$

Zauważmy, że dla ustalonego $x, y \in \mathbb{R}$

$$h(sx + (1-s)y, sz_1 + (1-s)z_2) \geqslant f(x,z_1)^s g(y,z_2)^{1-s} \quad \text{dla } z_1, z_2 \in \mathbb{R}^{n-1}.$$

Zatem na mocy założenia indukcyjnego

$$H(sx + (1-s)y) \geqslant F(x)^s G(y)^{1-s}.$$

Stosując nierówność Prékopy-Leindlera w udowodnionym wcześniej przypadku $n = 1$ dostajemy

$$\begin{aligned} \int_{\mathbb{R}^n} h(x)dx = \int_{\mathbb{R}} H(x)dx &\geqslant \Big(\int_{\mathbb{R}} F(x)dx\Big)^s \Big(\int_{\mathbb{R}} G(x)dx\Big)^{1-s} \\ &= \Big(\int_{\mathbb{R}^n} f(x)dx\Big)^s \Big(\int_{\mathbb{R}^n} g(x)dx\Big)^{1-s}. \end{aligned}$$ □

Wniosek 3.3 (Nierówność Brunna-Minkowskiego)**.** *Dla dowolnych niepustych zbiorów borelowskich $A, B \subset \mathbb{R}^n$,*

$$\lambda_n(sA + (1-s)B) \geqslant \lambda_n(A)^s \lambda_n(B)^{1-s} \quad \textit{dla } s \in [0,1]$$

oraz

$$\lambda_n(A+B)^{1/n} \geqslant \lambda_n(A)^{1/n} + \lambda_n(B)^{1/n}.$$

Dowód. Pierwsza nierówność natychmiast wynika z nierówności Prékopy-Leindlera zastosowanej do funkcji $f = \mathbb{1}_A, g = \mathbb{1}_B$ oraz $h = \mathbb{1}_{sA+(1-s)B}$.

By udowodnić drugą wystarczy rozważyć przypadek, gdy A i B są zbiorami skończonej i niezerowej miary. Przyjmijmy wtedy

$$\tilde{A} = \frac{A}{s}, \quad \tilde{B} = \frac{B}{1-s} \quad \text{oraz} \quad s = \frac{\lambda_n(A)^{1/n}}{\lambda_n(A)^{1/n} + \lambda_n(B)^{1/n}}.$$

Wówczas $\lambda_n(\tilde{A}) = \lambda_n(\tilde{B}) = (\lambda_n(A)^{1/n} + \lambda_n(B)^{1/n})^n$, więc na podstawie wykazanej poprzednio nierówności

$$\lambda_n(A+B) = \lambda_n(s\tilde{A} + (1-s)\tilde{B}) \geqslant \lambda_n(\tilde{A})^s \lambda_n(\tilde{B})^{1-s} = (\lambda_n(A)^{1/n} + \lambda_n(B)^{1/n})^n.$$ □

Uwaga 3.4. Suma Minkowskiego dwu zbiorów borelowskich nie musi być zbiorem borelowskim, ale można wykazać, że jest zbiorem mierzalnym w sensie Lebesgue'a.

Dowód Twierdzenia 3.1. Niech $c_n = \lambda_n(B(0,1))$, wówczas $\lambda_n(A) = c_n r^n$ i na podstawie Wniosku 3.3,

$$\begin{aligned}\lambda_n(A_t) &= \lambda_n(A + B(0,t)) \geqslant (\lambda_n(A)^{1/n} + \lambda_n(B(0,t))^{1/n})^n \\ &= c_n(r+t)^n = \lambda_n(B(x_0, r+t)).\end{aligned}$$

□

Definicja 3.5. Dla miary μ na przestrzeni probabilistycznej (X, d) określamy *zewnętzną miarę brzegową* μ^+ wzorem

$$\mu^+(A) := \liminf_{t \to 0+} \frac{\mu(A_t) - \mu(A)}{t}.$$

Uwaga 3.6. Jeśli miara μ na $\mathbb{R}^n$ ma ciągłą gęstość $g(x)$ oraz zbiór A ma gładki brzeg, to

$$\mu^+(A) = \int_{\partial A} g(x) dH_{n-1}(x),$$

gdzie H_{n-1} oznacza $n-1$ wymiarową miarę Haussdorffa.

Równoważna różniczkowa forma klasycznej nierówności izoperymetrycznej mówi, że spośród zbiorów ustalonej objętości najmniejszą powierzchnię brzegu ma kula. Dokładniej:

Twierdzenie 3.7. *Jeśli A jest podzbiorem borelowskim $\mathbb{R}^n$ takim, że $\lambda_n(A) = \lambda_n(B(x_0, r))$, to*

$$\lambda_n^+(A) \geqslant \lambda_n^+(B(x_0, r)) = n c_n^{1/n} (\lambda_n(A))^{(n-1)/n},$$

gdzie

$$c_n = \lambda_n(B(0,1)) = \frac{\pi^{n/2}}{\Gamma(n/2+1)}.$$

3.2 Izoperymetria sferyczna

Twierdzenie 3.8. *Jeśli A jest podzbiorem borelowskim S^n takim, że $\sigma_n(A) = \sigma_n(B(x_0, r))$, to dla dowolnego $t > 0$,*

$$\sigma_n(A_t) \geqslant \sigma_n(B(x_0, r)_t) = \sigma_n(B(x_0, r+t)).$$

Wniosek 3.9.

$$\alpha_{\sigma_n}(t) \leqslant \sqrt{\frac{\pi}{8}} \exp\Big(-\frac{(n-1)}{2} t^2\Big).$$

Dowód. Dla $n = 1$ nie ma co dowodzić (bo zawsze $\alpha_\mu(t) \leqslant 1/2$). Będziemy więc zakładać, że $n \geqslant 2$. Zauważmy, że

$$\sigma_n(B(x_0, r)) = s_n^{-1} \int_0^r \sin^{n-1} t dt,$$

gdzie $s_n = \int_0^\pi \sin^{n-1} t dt$. Zatem

$$\alpha_{\sigma_n}(t) = 1 - \sigma_n(B(x_0, t + \pi/2)) = s_n^{-1} \int_{t+\pi/2}^{\pi} \sin^{n-1} u du = s_n^{-1} \int_t^{\pi/2} \cos^{n-1} u du.$$

Stosując oszacowanie $\cos u \leqslant \exp(-u^2/2)$ dla $t \in [0, \pi/2]$, dostajemy

$$\begin{aligned}\int_t^{\pi/2} \cos^{n-1} u du &\leqslant \int_t^{\pi/2} e^{-(n-1)u^2/2} du \leqslant \frac{1}{\sqrt{n-1}} \int_{t\sqrt{n-1}}^{\infty} e^{-s^2/2} ds \\ &= \frac{\sqrt{2\pi}}{\sqrt{n-1}} (1 - \Phi(t\sqrt{n-1})) \leqslant \frac{\sqrt{\pi}}{\sqrt{2(n-1)}} e^{-(n-1)t^2/2}.\end{aligned}$$

Ponadto łatwe całkowanie przez części daje, że dla $n \geqslant 3$, $s_n = \frac{n-2}{n-1} s_{n-2}$, stąd

$$\sqrt{n-1} s_n = \frac{n-2}{\sqrt{n-1}} s_{n-2} \geqslant \sqrt{n-3} s_{n-2},$$

zatem

$$\inf_{n \geqslant 2} \sqrt{n-1} s_n = \min\{s_2, \sqrt{2} s_3\} = \min\{2, \pi/\sqrt{2}\} = 2. \qquad \square$$

3.3 Izoperymetria gaussowska

Przypomnijmy, że przez γ_k oznaczamy kanoniczny rozkład gaussowski na $\mathbb{R}^k$, tzn. rozkład z gęstością $(2\pi)^{-k/2} \exp(-|x|^2/2)$.

Głównym wynikiem, który wykażemy jest to, że dla rozkładów gaussowskich optymalne dla problemu izoperymetrycznego są *półprzestrzenie afiniczne*, to znaczy zbiory postaci

$$H = \{x \in \mathbb{R}^k \colon \langle x, u \rangle < r\} \text{ dla pewnych } u \in S^{k-1} \text{ i } r \in [-\infty, \infty]. \tag{3}$$

Twierdzenie 3.10. *Niech H będzie półprzestrzenią afiniczną, a A zbiorem borelowskim w $\mathbb{R}^k$ takim, że $\gamma_k(H) = \gamma_k(A)$. Wówczas dla dowolnego $t > 0$, $\gamma_k(H_t) \leqslant \gamma_k(A_t)$*

Zanim przystąpimy do dowodu twierdzenia pokażemy, że γ_k jest granicą rzutowań rozkładów jednostajnych na $\sqrt{n} S^{n-1}$.

Niech $P = P_{k,n}$ oznacza kanoniczny rzut $\mathbb{R}^n$ na $\mathbb{R}^k$ dla $k < n$, zaś $\tilde{\sigma}_{n-1}$ oznacza unormowaną miarę powierzchniową na $\sqrt{n} S^{n-1}$. Oznaczmy przez $\mu_{k,n}$ obraz $\tilde{\sigma}_{n-1}$ przy tym rzutowaniu tzn.

$$\mu_{k,n}(A) = \tilde{\sigma}_{n-1}\Big(P_{k,n}^{-1}(A)\Big) \quad \text{dla } A \in \mathcal{B}(\mathbb{R}^k).$$

Fakt 3.11 (Lemat Poincaré)**.** *Miara $\mu_{k,n}$ zbiega słabo przy $n \to \infty$ do miary γ_k, co więcej*

$$\lim_{n\to\infty} \mu_{k,n}(A) = \gamma_k(A) \text{ dla dowolnego zbioru borelowskiego } A.$$

Dowód. Proste rozumowanie pokazuje, że miara $\mu_{k,n}$ ma gęstość $g_{k,n}(x) = c_{k,n}^{-1}\tilde{g}_{k,n}(x)$, gdzie $\tilde{g}_{k,n} = (\frac{n-|x|^2}{n})^{(n-k-2)/2}\mathbb{1}_{\{|x|\leqslant\sqrt{n}\}}$ oraz $c_{k,n} = \int_{\mathbb{R}^k} \tilde{g}_{n,k}(x)dx$. Oczywiście $\lim_{n\to\infty} \tilde{g}_{k,n}(x) = \exp(-|x|^2/2)$, ponadto $|\tilde{g}_{k,n}(x)| \leqslant \exp(-(n-k-2)|x|^2/(2n)) \leqslant \exp(-|x|^2/(2n))$ dla $n \geqslant k+2$. Z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej otrzymujemy $\lim_{n\to\infty} c_{n,k} = \int_{\mathbb{R}^k} \exp(-|x|^2/2)dx$, czyli gęstość miary $\mu_{k,n}$ zbiega punktowo do gęstości miary γ_k. Teza faktu wynika z twierdzenia Scheffé'go (zob. zad.8.1.7 w [2]). □

Dowód Twierdzenia 3.10. Ze względu na rotacyjną niezmienniczość miary γ_k możemy dla uproszczenia notacji założyć, że $H = \{x\colon x_1 < r\}$. Ustalmy dowolne $r_0 < r$ i niech $H_0 = \{x\colon x_1 < r_0\}$. Zauważmy, że $\gamma_k(H_0) < \gamma_k(A)$, zatem na podstawie Lematu Poincaré, $\mu_{k,n}(H_0) \leqslant \mu_{k,n}(A)$ dla dużych n. Ponieważ $P_{k,n}^{-1}(H_0) \cap \sqrt{n}S^{n-1}$ jest kulą w $\sqrt{n}S^{n-1}$, więc na mocy izoperymetrii sferycznej

$$\tilde{\sigma}_{n-1}\Big((P_{k,n}^{-1}(A))_t\Big) \geqslant \tilde{\sigma}_{n-1}\Big((P_{k,n}^{-1}(H_0))_t\Big).$$

Zauważmy, że przekształcenie $P_{k,n}$ jest oczywiście 1-lipschitzowskie, więc $A_t \supset P_{k,n}((P_{k,n}^{-1}(A))_t)$ i

$$\mu_{k,n}(A_t) \geqslant \mu_{k,n}(P_{k,n}((P_{k,n}^{-1}(A))_t)) \geqslant \mu_{k,n}(P_{k,n}((P_{k,n}^{-1}(H_0))_t)).$$

Nietrudno zauważyć, że

$$P_{k,n}((P_{k,n}^{-1}(H_0))_t) = \{x\colon x_1 < r_n\}$$

oraz $r_n \to r_0 + t$ przy $n \to \infty$. Stąd

$$\gamma_k(A_t) = \lim_{n\to\infty} \mu_{k,n}(A_t) \geqslant \lim_{n\to\infty} \mu_{k,n}(\{x\colon x_1 < r_n\}) = \gamma_k(\{x\colon x_1 < r_0 + t\}),$$

z dowolności $r_0 < r$ wynika teza. □

Twierdzenie 3.12. *Jeśli $\gamma_k(A) = \Phi(x)$ to $\gamma_k(A_t) \geqslant \Phi(x+t)$ oraz $\gamma_k^+(A) \geqslant I_\gamma(\gamma_k(A))$, gdzie $I_\gamma(x) := \varphi(\Phi^{-1}(x))$ oraz $\varphi(x) = \Phi'(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$.*

Dowód. Wystarczy zauważyć, że jeśli $\gamma_k(H) = \Phi(r)$ i H jest postaci (3), to $H_t = \{x \in \mathbb{R}^k\colon \langle x, u\rangle < r + t\}$ i $\gamma_k(H_t) = \Phi(r+t)$. □

Zauważając, że $\Phi(0) = 1/2$ otrzymujemy:

Wniosek 3.13. $\alpha_{\gamma_k}(t) \leqslant 1 - \Phi(t) \leqslant \frac{1}{2}\exp(-t^2/2)$.

Jak widzieliśmy już w dowodzie Twierdzenia 3.10 bardzo użyteczne jest pojęcie tzw. transportu miary.

Definicja 3.14. Niech μ i ν będą miarami na przestrzeniach mierzalnych $\mathbb{X}$ i $\mathbb{Y}$. Powiemy, że funkcja mierzalna $T\colon \mathbb{X} \to \mathbb{Y}$ *transportuje miarę* μ *na miarę* ν (ew. *miara* ν *jest obrazem miary* μ *przy przekształceniu* T) jeśli $\nu(A) = \mu(T^{-1}(A))$ dla wszystkich mierzalnych $A \subset \mathbb{Y}$.

Szczególnie wygodny jest transport lipschitzowski.

Fakt 3.15. *Jeśli* $T\colon \mathbb{X} \to \mathbb{Y}$ *jest* L*-lipschitzowska oraz* T *transportuje miarę* μ *na* ν*, to* $\alpha_\nu(t) \leqslant \alpha_\mu(t/L)$.

Dowód. Wystarczy zauważyć, że $(T^{-1}(A))_{t/L} \subset T^{-1}(A_t)$. □

Transportując w sposób lipschitzowski miarę gaussowską można uzyskać oszacowania funkcji koncentracji dla innych miar. Pokażemy dwa przykłady.

Wniosek 3.16. *Niech* $\mu_{[0,1]^n}$ *oznacza rozkład jednostajny na kostce* $[0,1]^n$*. Wówczas* $\mu_{[0,1]^n}$ *jest* $(2\pi)^{-1/2}$*-lipschitzowskim obrazem* γ_n*. W szczególności* $\alpha_{\mu_{[0,1]^n}} \leqslant \frac{1}{2}\exp(-\pi t^2)$.

Dowód. Określmy $f\colon \mathbb{R} \to (0,1)$ wzorem

$$f(x) = \mu_{[0,1]}([0, f(x)]) = \gamma_1((-\infty, x]) = \Phi(x).$$

Wówczas funkcja f transportuje miarę gaussowską γ_1 na $\mu_{[0,1]}$, to znaczy $\mu_{[0,1]} = \gamma_1 \circ f^{-1}$. Ponadto $f'(x) = (2\pi)^{-1/2}\exp(-x^2/2) \leqslant (2\pi)^{-1/2}$, czyli f jest $(2\pi)^{-1/2}$-lipschitzowska. Jeśli teraz określimy $F\colon \mathbb{R}^n \to (0,1)^n$ wzorem $F(x) = (f(x_1), \ldots, f(x_n))$, to F transportuje γ_n na μ oraz F jest $(2\pi)^{-1/2}$-lipschitzowska. Ostatnie oszacowanie w tezie wniosku jest konsekwencją Faktu 3.15 i Wniosku 3.13. □

Wniosek 3.17. *Niech* $B_n = \{x \in \mathbb{R}^n\colon |x| \leqslant 1\}$ *oznacza kulę jednostkową w* $\mathbb{R}^n$*, zaś* μ_{B_n} *będzie rozkładem jednostajnym na* B_n*. Wówczas istnieje stała* C *taka, że* μ_{B_n} *jest* $Cn^{-1/2}$*-lipschitzowskim obrazem* γ_n*. W szczególności* $\alpha_{\mu_{B_n}} \leqslant \frac{1}{2}\exp(-nt^2/(2C))$.

Ponieważ obie miary γ_n i μ_{B_n} są rotacyjnie niezmiennicze, będziemy szukać funkcji $T\colon \mathbb{R}^n \to B_n$ transportującej γ_n na μ_{B_n} postaci $Tx = \frac{x}{|x|}\varphi(|x|)$. Dalsze szczegóły pozostawiamy Czytelnikowi jako ćwiczenie.

Otwarty problem. Rozwiązać zagadnienie izoperymetryczne dla zbiorów symetrycznych, to znaczy znaleźć dla ustalonego $t > 0$, $c \in [0,1]$,

$$\inf\{\gamma_k(A_t)\colon \gamma_k(A) = c, A = -A\}$$

oraz

$$\inf\{\gamma_k^+(A)\colon \gamma_k(A) = c, A = -A\}.$$

Dość naturalna hipoteza mówi, że dla $c \geqslant 1/2$ rozwiązaniem obu problemów są zbiory postaci $[-a,a]\times\mathbb{R}^{k-1}$ zaś dla $c < 1/2$ drugi problem się optymalizuje dla $(\mathbb{R}\setminus[-a,a])\times\mathbb{R}^{k-1}$. Podobny problem można postawić dla miary σ_n, ale tam analogiczna hipoteza okazuje się być niestety fałszywa.

4 Metoda Martyngałowa

4.1 Transformata Laplace'a

Wiele dalszych szacowań będzie oparte na transformacie Laplace'a zmiennej losowej.

Definicja 4.1. *Transformatą Laplace'a* zmiennej losowej Z nazywamy funkcję

$$L_Z(\lambda) := \mathbf{E}e^{\lambda Z} \quad \lambda \in \mathbb{R}.$$

Podobnie jeśli μ jest miarą probabilistyczną na pewnej przestrzeni X oraz $F\colon X \to \mathbb{R}$, to *transformatę Laplace'a F względem μ* określamy

$$L_{F,\mu}(\lambda) := \int_X e^{\lambda F(x)} d\mu(x).$$

Fakt 4.2. *Dla dowolnej zmiennej losowej Z,*

$$\mathbf{P}(Z \geqslant t) \leqslant \inf_{\lambda \geqslant 0} e^{-\lambda t} L_Z(\lambda) \text{ dla } t \geqslant 0.$$

W szczególności, jeśli dla pewnego $a > 0$,

$$L_Z(\lambda) \leqslant \exp(a\lambda^2) \quad \lambda \in \mathbb{R},$$

to dla $t \geqslant 0$

$$\mathbf{P}(Z \geqslant t) \leqslant \exp\Big(-\frac{t^2}{4a}\Big) \text{ oraz } \mathbf{P}(|Z| \geqslant t) \leqslant 2\exp\Big(-\frac{t^2}{4a}\Big).$$

Dowód. Pierwsza część wynika z nierówności Czebyszewa, a druga z pierwszej i prostego rachunku. □

Zatem by udowodnić, że funkcja koncentracji miary μ jest gaussowska wystarczy wykazać, że $L_{F,\mu}(\lambda) \leqslant \exp(a\lambda^2)$ dla pewnego $a > 0$ i wszystkich funkcji 1-lipschitzowskich F takich, że $\int F d\mu = 0$.

4.2 Nierówność Azumy

Twierdzenie 4.3 (Nierówność Hoeffdinga-Azumy)**.** *Niech $(M_k, \mathcal{F}_k)_{k=0}^n$ będzie martyngałem o ograniczonych przyrostach takim, że $\|M_k - M_{k-1}\|_\infty \leqslant a_k$. Wówczas*

$$\mathbf{P}(M_n - M_0 \geqslant t) \leqslant \exp\Big(-\frac{t^2}{2\sum_{i=1}^n a_i^2}\Big).$$

Dowód. Określmy dla $1 \leqslant k \leqslant n$, $d_k := M_k - M_{k-1}$, wówczas $\mathbf{E}(d_k|\mathcal{F}_{k-1}) = 0$. Mamy $\frac{1-u}{2}(-x) + \frac{1+u}{2}x = ux$, więc z wypukłości $\exp(x)$,

$$e^{ux} \leqslant \frac{1-u}{2}e^{-x} + \frac{1+u}{2}e^{x} = u\sinh(x) + \cosh(x) \text{ dla } |u| \leqslant 1.$$

Stosując tę nierówność dla $u = d_k/a_k$ i $x = \lambda a_k$ dostajemy

$$\mathbf{E}(e^{\lambda d_k}|\mathcal{F}_{k-1}) \leqslant \mathbf{E}\Big(\frac{d_k}{a_k}\Big|\mathcal{F}_{k-1}\Big)\sinh(\lambda a_k) + \cosh(\lambda a_k) = \cosh(\lambda a_k).$$

Liczymy

$$\begin{aligned}\mathbf{E}e^{\lambda(M_n - M_0)} &= \mathbf{E}e^{\lambda(M_{n-1} - M_0 + d_n)} = \mathbf{E}(e^{\lambda(M_{n-1}-M_0)}\mathbf{E}(e^{\lambda d_n}|\mathcal{F}_{n-1})) \\ &\leqslant \cosh(\lambda a_n)\mathbf{E}e^{\lambda(M_{n-1}-M_0)}.\end{aligned}$$

Zatem iterując powyższą nierówność i stosując oszacowanie (wynikające np. z rozwinięcia w szereg Taylora) $\cosh(x) \leqslant \exp(x^2/2)$ dostajemy

$$L_{M_n - M_0}(\lambda) = \mathbf{E}e^{\lambda(M_n - M_0)} \leqslant \prod_{k=1}^{n}\cosh(\lambda a_k) \leqslant \exp(\frac{1}{2}\sum_{k=1}^{n} a_k^2\lambda^2).$$

Teza twierdzenia wynika z Faktu 4.2. □

Uwaga 4.4. Najczęściej będziemy mieli $\mathcal{F}_0 = \{\emptyset, \Omega\}$, wówczas M_0 jest stałe, a ponieważ martyngał ma stałą wartość oczekiwaną, to $M_0 = \mathbf{E}M_n$.

W poniższych zastosowaniach będziemy przyjmować $M_k = \mathbf{E}_\mu(F|\mathcal{F}_k)$ dla całkowalnej funkcji $F\colon \mathbb{X} \to \mathbb{R}$ i odpowiednio dobranego $(\mathcal{F}_k)$ ciągu σ-ciał podzbiorów $\mathbb{X}$.

4.3 Zastosowania nierówności Azumy

Wniosek 4.5. *Niech $(\mathbb{X}_i, d_i)$ będą przetrzeniami metrycznymi, $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$ z odległością l_1, to znaczy $d(x,y) = \sum_{i=1}^n d_i(x_i, y_i)$ dla $x, y \in \mathbb{X}$ oraz niech $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ będzie produktem miar probabilistycznych μ_i na $\mathbb{X}_i$. Wówczas dla dowolnej funkcji 1-lipschitzowskiej F na $\mathbb{X}$*

$$\mu\Big(\Big\{x\colon F(x) \geqslant \int F d\mu + t\Big\}\Big) \leqslant \exp(-\frac{t^2}{2D^2}),$$

gdzie $D = (\sum_{i=1}^n \mathrm{Diam}(\mathbb{X}_i)^2)^{1/2}$. W szczególności

$$\alpha_\mu(t) \leqslant \exp\Big(-\frac{t^2}{8D^2}\Big).$$

Dowód. Na mocy Faktu 2.8 wystarczy wykazać pierwszą nierówność tezy. Niech $\mathcal{F}_k$ będzie σ ciałem generowanym przez pierwsze k-współrzędnych oraz $M_k := \mathbf{E}_\mu(F|\mathcal{F}_k)$. Wówczas oczywiście

$$M_k(x) = \tilde{M}_k(x_1, \dots, x_k) = \int_{\mathbb{X}_{k+1}\times\dots\times\mathbb{X}_n} F(x) d\mu_{i+1}(x_{i+1}) \cdots d\mu_n(x_n),$$

stąd

$$\begin{aligned}
&|M_k(x) - M_{k-1}(x)| = |\tilde{M}_k(x_1, \dots, x_k) - \int_{\mathbb{X}_k} \tilde{M}_k(x_1, \dots, x_k) d\mu_k(x_k)| \\
&\leqslant \sup_{y_k, z_k \in X_k} |\tilde{M}_k(x_1, \dots, x_{k-1}, y_k) - \tilde{M}_k(x_1, \dots, x_{k-1}, z_k)| \\
&\leqslant \sup_{y\in X, z_k \in \mathbb{X}_k} |F(x_1, \dots, x_{k-1}, y_k, y_{k+1}, \dots, y_n) - F(x_1, \dots, x_{k-1}, z_k, y_{k+1}, \dots, y_n)| \\
&\leqslant \sup_{y_k, z_k \in \mathbb{X}_k} d_k(y_k, z_k) \leqslant \mathrm{Diam}(\mathbb{X}_k)
\end{aligned}$$

i teza wynika z Twierdzenia 4.3. □

Przykład 1. Niech $\mathbb{X} = \{0,1\}^n$ z odległością $d(x,y) = \frac{1}{n}\#\{i\colon x_i \neq y_i\}$ i unormowaną miarą liczącą μ. Kładąc $\mathbb{X}_i = \{0,1\}$ z odległością $d_i(x,y) = \frac{1}{n} I_{\{x\neq y\}}$ widzimy, że możemy stosować poprzedni wniosek i $D = (\sum_{i=1}^n \mathrm{Diam}(\mathbb{X}_i)^2)^{1/2} = n^{-1/2}$. Zatem

$$\alpha_{(\{0,1\}^n, d, \mu)} \leqslant \exp(-\frac{nt^2}{8}).$$

Definicja 4.6. Mówimy, że skończona przestrzeń metryczna $(\mathbb{X}, d)$ ma *długość* co najwyżej l, jeśli istnieje rosnący ciąg podziałów $\mathbb{X}$, $\{\mathbb{X}\} = \mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n = \{\{x\}\colon x \in \mathbb{X}\}$ ($\mathcal{A}_i$ jest podpodziałem $\mathcal{A}_{i-1}$) oraz liczby $a_1, \dots, a_n$ spełniające $(\sum_{i=1}^n a_i^2)^{1/2} \leqslant l$ takie, że dla dowolnego $A \in \mathcal{A}_{i-1}$ oraz $B, C \in \mathcal{A}_i, B, C \subset A$ istnieje bijekcja $\Phi\colon B \to C$ dla której $d(x, \Phi(x)) \leqslant a_i$ dla $x \in B$.

Uwaga 4.7. Biorąc $\mathcal{A}_0 = \{\mathbb{X}\}$ i $\mathcal{A}_1 = \{\{x\}\colon x \in \mathbb{X}\}$ widzimy, że każda skończona przestrzeń metryczna ma długość nie większą niż $\mathrm{Diam}(\mathbb{X})$.

Twierdzenie 4.8. *Jeśli $(\mathbb{X}, d)$ jest skończoną przestrzenią metryczną o długości co najwyżej l, zaś μ unormowaną miarą liczącą na $\mathbb{X}$, to dla funkcji 1-lipschitzowskich F na $\mathbb{X}$,*

$$\mu\Big(\Big\{x\colon F(x) \geqslant \int F d\mu + t\Big\}\Big) \leqslant \exp(-\frac{t^2}{2l^2}),$$

w szczególności

$$\alpha_\mu(t) \leqslant \exp\Big(-\frac{t^2}{8l^2}\Big).$$

Dowód. Ustalmy funkcję 1-lipschitzowską F. Niech $\mathcal{F}_i$ będzie σ-ciałem generowanym przez $\mathcal{A}_i$ oraz $M_i := \mathbf{E}_\mu(F|\mathcal{F}_i)$ dla $i = 0, \ldots, n$. Wówczas

$$M_i(x) = \frac{1}{\#A} \sum_{y \in A} F(y) \text{ dla } x \in A \in \mathcal{A}_i.$$

Zatem, jeśli $A \in \mathcal{A}_{i-1}$, $B, C \in \mathcal{A}_i, B, C \subset A$ oraz $\Phi\colon B \to C$ jest bijekcją jak w Definicji 4.6, to dla $x \in B, y \in C$,

$$\begin{aligned} |M_i(x) - M_i(y)| &= \Big|\frac{1}{\#B} \sum_{z \in B} (F(z) - F(\Phi(z))\Big| \leqslant \sup_{z \in B} |F(z) - F(\Phi(z))| \\ &\leqslant \sup_{z \in B} d(z, \Phi(z)) \leqslant a_i. \end{aligned}$$

Ponieważ M_{i-1} na $A \in \mathcal{A}_{i-1}$ jest uśrednieniem M_i po $B \subset A, B \in \mathcal{A}_i$, to mamy $|M_i(x) - M_{i-1}(x)| \leqslant a_i$, czyli $\|M_i - M_{i-1}\|_\infty \leqslant a_{i-1}$. Teza wynika z Twierdzenia 4.3 oraz Faktu 2.8. □

Przykład 2. Niech Π^n będzie grupą permutacji zbioru $\{1, \ldots, n\}$ z metryką $d(\sigma, \pi) = \frac{1}{n}\#\{i\colon \sigma_i \neq \pi_i\}$, a μ unormowaną miarą liczącą na Π^n. Niech $\mathcal{A}_i$ składa się ze zbiorów postaci

$$A_{j_1,\ldots,j_i} = \{\sigma \in \Pi^n\colon \sigma(1) = j_1, \ldots, \sigma(i) = j_i\}.$$

Wówczas jeśli $B, C \in \mathcal{A}_i$ są podzbiorami pewnego $A \in \mathcal{A}_{i-1}$ to $B = A_{j_1,\ldots,j_{i-1},p}$, $C = A_{j_1,\ldots,j_{i-1},q}$ i możemy zdefiniować bijekcję Φ między B i C jako $\Phi(\sigma) = \tau_{p,q} \circ \sigma$, gdzie $\tau_{p,q}$ jest transpozycją zamieniającą p z q. Łatwo sprawdzić, że $d(\sigma, \Phi(\sigma)) \leqslant 2/n$, zatem $l = 2/\sqrt{n}$ i

$$\alpha_{(\Pi^n, d, \mu)} \leqslant \exp(-\frac{nt^2}{32}).$$

5 Nierówność Poincaré

5.1 Definicja i podstawowe własności

Definicja 5.1. Mówimy, że miara probabilistyczna μ na $(\mathbb{X}, d)$ *spełnia nierówność Poincaré* ze stałą C, jeśli dla wszystkich ograniczonych lipschitzowskich funkcji f na $\mathbb{X}$ zachodzi

$$\mathrm{Var}_\mu(f) \leqslant C \int |\nabla f|^2 d\mu, \tag{4}$$

gdzie

$$|\nabla f|(x) := \limsup_{y \to x} \frac{|f(x) - f(y)|}{d(x, y)},$$

jeśli x jest punktem skupienia $\mathbb{X}$ i $|\nabla f|(x) = 0$, jeśli x jest punktem izolowanym $\mathbb{X}$.

Uwaga 5.2. W przypadku, gdy $\mathbb{X} = \mathbb{R}^n$ ze standardową metryką euklidesową możemy użyć twierdzenia Rademachera, które mówi, że każda funkcja Lipchitzowska jest różniczkowalna prawie wszędzie i wtedy $|\nabla f|(x)$ jest dla prawie wszystkich x równy długości zwykłego gradientu f. Ponadto argument aproksymacyjny pokazuje, że by wykazać nierówność Poincaré dla miar probabilistycznych na $\mathbb{R}^n$ wystarczy sprawdzić (4) dla ograniczonych funkcji klasy $C^1(\mathbb{R}^n)$ o ograniczonych pochodnych rzędu jeden.

Uwaga 5.3. Będziemy wykorzystywali tylko dwie własności $|\nabla f|$. Mianowicie, że dla funkcji 1-lipschitzowskich $|\nabla f| \leqslant 1$ oraz, że dla dowolnej funkcji klasy $C^1(\mathbb{R})$, $|\nabla g(F)| \leqslant |g'(F)||\nabla F|$ (w szczególności $|\nabla(f+c)| = |\nabla f|$).

Uwaga 5.4. Załóżmy, że miara μ ma gęstość postaci e^{-V} na $\mathbb{R}^n$. Wówczas proste całkowanie przez części pokazuje, że

$$\int |\nabla f|^2 d\mu = \int (-\triangle f + \langle \nabla V, \nabla f \rangle) f d\mu.$$

Definiując operator $Lf := -\triangle f + \langle \nabla V, \nabla f \rangle$ widzimy, że $L1 = 0$. Nierówność Poincaré mówi, że dla funkcji f o średniej 0, czyli prostopadłych do 1, $\int fLf d\mu \geqslant C^{-1} \int f^2 d\mu$. Biorąc pod uwagę samosprzężoność L nierówność (4) jest równoważna temu, że kolejna wartość własna L to conajmniej $1/C$. Dlatego nierówność Poincaré się nazywa nierównością „luki spektralnej" (spectral gap inequality).

Czasem wygodniej w nierówności Poincaré zastąpić wariancję funkcji przez całkę kwadratu odchylenia od mediany, okazuje się, że prowadzi to do równoważnej nierówności.

Fakt 5.5. *Nierówność Poincaré jest równoważna nierówności*

$$\forall_{f \in \mathrm{Lip}(X)} \quad \mathbf{E}_\mu |f - \mathrm{Med}_\mu f|^2 \leqslant \tilde{C} \int |\nabla f|^2 d\mu.$$

Co więcej optymalne stałe w obu nierównościach spełniają $C_{\mathrm{opt}} \leqslant \tilde{C}_{\mathrm{opt}} \leqslant 3C_{\mathrm{opt}}$.

Dowód. Ponieważ

$$\mathrm{Var}_\mu(f) = \inf_{c \in \mathbb{R}} \mathbf{E}_\mu (f-c)^2 \leqslant \mathbf{E}_\mu |f - \mathrm{Med}_\mu f|^2,$$

więc oczywiście $C_{\mathrm{opt}} \leqslant \tilde{C}_{\mathrm{opt}}$.

By udowodnić przeciwne oszacowanie zauważmy, że

$$\begin{aligned} \mathrm{Var}_\mu(f) &\geqslant |\mathrm{Med}_\mu f - \mathbf{E}_\mu f|^2 \mu(\{|f - \mathbf{E}_\mu f| \geqslant |\mathrm{Med}_\mu f - \mathbf{E}_\mu f|\}) \\ &\geqslant \frac{1}{2} |\mathrm{Med}_\mu f - \mathbf{E}_\mu f|^2. \end{aligned}$$

Stąd

$$\mathbf{E}_\mu |f - \mathrm{Med}_\mu f|^2 \leqslant \mathrm{Var}_\mu(f) + |\mathrm{Med}_\mu f - \mathbf{E}_\mu f|^2 \leqslant 3\mathrm{Var}_\mu(f)$$

i otrzymujemy $\tilde{C}_{\mathrm{opt}} \leqslant 3C_{\mathrm{opt}}$. □

Fakt 5.6. *Symetryczny rozkład wykładniczy ν na $\mathbb{R}$ z gęstością $\frac{1}{2}e^{-|x|}$ spełnia nierówność Poincaré ze stałą 4.*

Dowód. Proste całkowanie przez części pokazuje, że dla funkcji $h \in C^1_{\text{ogr}}(\mathbb{R})$,

$$\int h(x)d\nu(x) = h(0) + \int \operatorname{sgn}(x)h'(x)d\nu(x).$$

Niech $f \in C^1_{\text{ogr}}(\mathbb{R})$ i $g(x) = f(x) - f(0)$ wówczas

$$\int g^2 d\nu = 2\int \operatorname{sgn}(x)g'(x)g(x)d\nu(x) \leqslant 2\Big(\int g'^2 d\nu\Big)^{1/2}\Big(\int g^2 d\nu\Big)^{1/2},$$

stąd

$$\operatorname{Var}_\nu(f) \leqslant \int g^2 d\nu \leqslant 4\int g'^2 d\nu = 4\int f'^2 d\nu.$$

□

5.2 Nierówność Poincaré a koncentracja wykładnicza

Twierdzenie 5.7. *Załóżmy, że miara μ spełnia nierówność Poincaré ze stałą C. Wówczas dla każdej funkcji 1-lipschitzowskiej F i $t > 0$*

$$\mu\Big(\Big\{F \geqslant \int F d\mu + t\Big\}\Big) \leqslant 2\exp\Big(-\frac{t}{\sqrt{C}}\Big).$$

W szczególności $\alpha_X(t) \leqslant 2\exp(-t/2\sqrt{C})$.

Dowód. Rozpatrując $F - \int F d\mu$ możemy założyć, że F ma średnią zero. Zauważmy, że dla dowolnej funkcji różniczkowalnej g mamy $|\nabla g(F)| \leqslant |g'(F)||\nabla F| \leqslant |g'(F)|$. Niech

$$M(\lambda) := M_{\mu,F}(\lambda) = \int e^{\lambda F} d\mu.$$

Stosując nierówność Poincaré do $e^{\lambda F/2}$ dostajemy

$$M(\lambda) - M\Big(\frac{\lambda}{2}\Big)^2 = \operatorname{Var}_\mu(e^{\lambda F/2}) \leqslant C\int |\nabla e^{\lambda F/2}|^2 d\mu \leqslant \frac{C\lambda^2}{4}M(\lambda).$$

Zatem dla $\lambda < 2/\sqrt{C}$ dostajemy

$$M(\lambda) \leqslant \frac{1}{1 - C\lambda^2/4}M\Big(\frac{\lambda}{2}\Big)^2.$$

Iterując tę nierówność n razy dostajemy

$$M(\lambda) \leqslant \prod_{k=0}^{n-1}\Big(\frac{1}{1 - C\lambda^2/4^{k+1}}\Big)^{2^k} M\Big(\frac{\lambda}{2^n}\Big)^{2^n}.$$

Ponieważ $M(0) = 1$ i $M'(0) = \int F d\mu = 0$, to $M(\lambda/2^n)^{2^n} \to 1$ przy $n \to \infty$ i

$$M(\lambda) \leqslant \prod_{k=0}^{\infty} \Big(\frac{1}{1 - C\lambda^2/4^{k+1}}\Big)^{2^k}.$$

Zauważmy, że

$$\prod_{k=0}^{\infty} \Big(1 - C\lambda^2 4^{-k-1}\Big)^{2^k} \geqslant 1 - C\lambda^2 \sum_{k=0}^{\infty} 2^k 4^{-k-1} = 1 - \frac{C}{2}\lambda^2.$$

W szczególności $M(1/\sqrt{C}) \leqslant 2$ i teza wynika z nierówności Czebyszewa. □

Uwaga 5.8. Nierówność Poincaré nie implikuje lepszej koncentracji niż wykładnicza. Istotnie symetryczny rozkład wykładniczy na prostej ν spełnia nierówność Poincaré ze stałą 4, a biorąc $f(x) = x$ widzimy, że dla $t > 0$,

$$\nu\left(\left\{x \in \mathbb{R}\colon f(x) \geqslant \int f d\nu + t\right\}\right) = \nu([t, \infty)) = \frac{1}{2}e^{-t}.$$

5.3 Tensoryzacja

Fakt 5.9. *Załóżmy, że μ_i są miarami probabilistycznymi na $\mathbb{X}_i$, $\mathbb{X} = \mathbb{X}_1 \times \ldots \times \mathbb{X}_n$ oraz $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$. Wówczas dla dowolnej funkcji $f \in L^2(\mathbb{X}, \mu)$*

$$\mathrm{Var}_\mu(f) \leqslant \sum_{i=1}^{n} \mathbf{E}_\mu \mathrm{Var}_{\mu_i}(f).$$

Dowód. Prosta indukcja pokazuje, że wystarczy rozpatrzeć przypadek $n = 2$. Wówczas

$$\begin{aligned}
\mathrm{Var}_\mu(f) &= \mathbf{E}_{\mu_2}\mathbf{E}_{\mu_1}(f - \mathbf{E}_\mu f)^2 = \mathbf{E}_{\mu_2}[\mathrm{Var}_{\mu_1}(f) + (\mathbf{E}_{\mu_1} f - \mathbf{E}_\mu f)^2] \\
&= \mathbf{E}_\mu \mathrm{Var}_{\mu_1}(f) + \mathbf{E}_{\mu_2}[\mathbf{E}_{\mu_1}(f - \mathbf{E}_{\mu_2} f)]^2 \\
&\leqslant \mathbf{E}_\mu \mathrm{Var}_{\mu_1}(f) + \mathbf{E}_{\mu_2}\mathbf{E}_{\mu_1}[(f - \mathbf{E}_{\mu_2} f)^2] = \mathbf{E}_\mu \mathrm{Var}_{\mu_1}(f) + \mathbf{E}_\mu \mathrm{Var}_{\mu_2}(f),
\end{aligned}$$

gdzie ostatnia nierówność wynika np. z nierówności Jensena. □

Wniosek 5.10. *Załóżmy, że miary probabilistyczne μ_i na $(\mathbb{X}_i, d_i)$ spełniają nierówność Poincaré ze stałą C_i względem gradientu $|\nabla_i|$. Wówczas miara $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ spełnia nierówność Poincaré ze stałą $C = \max_i C_i$ względem gradientu ∇f danego wzorem*

$$|\nabla f|^2 = \sum_{i=1}^{n} |\nabla_i f|^2.$$

Dowód. Z Faktu 5.9 dostajemy

$$\mathrm{Var}_\mu(f) \leqslant \sum_{i=1}^n \mathbf{E}_\mu \mathrm{Var}_{\mu_i}(f) \leqslant \sum_{i=1}^n \mathbf{E}_\mu C_i \mathbf{E}_{\mu_i} |\nabla_i f|^2 \leqslant C\mathbf{E}_\mu \sum_{i=1}^n |\nabla_i f|^2.$$

□

Wniosek 5.11. *Produktowy rozkład wykładniczy ν^n spełnia nierówność Poincaré na $\mathbb{R}^n$ ze stałą 4. W szczególności $\alpha_{\nu^n}(t) \leqslant 2\exp(-t/4)$.*

5.4 Dodatkowe własności. Charakteryzacja na prostej.

Kolejną przyjemną własnością nierówności Poincaré jest jej stabilność ze względu na zaburzenia miary μ.

Fakt 5.12. *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{X}$, V jest ograniczoną funkcją borelowską oraz $d\nu = Z^{-1}e^V d\mu$, gdzie $Z = \int e^V d\mu$. Wówczas jeśli miara μ spełnia nierówność Poincaré ze stałą C to ν spełnia nierówność Poincaré ze stałą $Ce^{2\|V\|_\infty}$.*

Dowód. Weźmy funkcję lipschitzowską f, odejmując stałą możemy założyć, że $\mathbf{E}_\mu f = 0$. Wówczas

$$\begin{aligned}\mathrm{Var}_\nu(f) &\leqslant \mathbf{E}_\nu f^2 = \frac{1}{Z}\int f^2 e^V d\mu \leqslant \frac{1}{Z} e^{\|V\|_\infty} \int f^2 d\mu \\ &\leqslant \frac{1}{Z} e^{\|V\|_\infty} C \int |\nabla f|^2 d\mu = Ce^{\|V\|_\infty} \int |\nabla f|^2 e^{-V} d\nu \\ &\leqslant Ce^{2\|V\|_\infty} \int |\nabla f|^2 d\nu.\end{aligned}$$

□

Fakt 5.13. *Jeśli miara ν na $(\mathbb{Y}, \rho)$ jest L-lipschitzowskim obrazem miary μ na $(\mathbb{X}, d)$ oraz μ spełnia nierówność Poincaré ze stałą C, to ν spełnia nierówność Poincaré ze stałą CL^2.*

Dowód. Niech $\nu = \mu \circ \varphi^{-1}$, gdzie $\varphi\colon X \to Y$ i $\|\varphi\|_{\mathrm{Lip}} \leqslant L$. Dla funkcji lipschitzowskich f na Y otrzymujemy

$$\begin{aligned}\mathrm{Var}_\nu(f) = \mathrm{Var}_\mu(f \circ \varphi) &\leqslant C \int |\nabla f \circ \varphi|^2 d\mu \leqslant CL^2 \int |\nabla f|^2(\varphi(x)) d\mu(x) \\ &= CL^2 \int |\nabla f|^2 d\nu,\end{aligned}$$

gdzie przedostatnia nierówność wynika z oszacowania $|\nabla f \circ \varphi|(x) \leqslant L|\nabla f|(\varphi(x))$. □

Kolejne twierdzenie (które podamy bez dowodu) charakteryzuje miary na prostej, które spełniają nierówność Poincaré.

Twierdzenie 5.14 (Muckenhaupt)**.** *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{R}$ o medianie m, zaś p oznacza gęstość jej części absolutnie ciągłej. Wówczas miara μ spełnia nierówność Poincaré ze skończoną stałą C wtedy i tylko wtedy gdy $\max\{B_+, B_-\} < \infty$, gdzie*

$$B_+ = \sup_{x>m} \mu[x,\infty) \int_m^x \frac{1}{p(y)} dy$$

$$B_- = \sup_{x<m} \mu(-\infty, x] \int_x^m \frac{1}{p(y)} dy.$$

Co więcej optymalna stała C_{opt} w nierówności Poincaré spełnia

$$\frac{1}{(1+\sqrt{2})^2} \max\{B_+, B_-\} \leqslant C_{\mathrm{opt}} \leqslant 4 \max\{B_+, B_-\}.$$

5.5 Nierówność Cheegera

W tej sekcji ν oznacza symetryczny rozkład wykładniczy na prostej z gęstością $\frac{1}{2}e^{-|x|}$. Zanim sformułujemy definicję zaczniemy od prostego faktu.

Fakt 5.15. *Niech μ będzie miarą probabilistyczną na $(\mathbb{X}, d)$. Następujące warunki są równoważne dla ustalonego $c > 0$:*
(i) $\mu^+(A) \geqslant c \min\{\mu(A), 1-\mu(A)\}$ dla dowolnego zbioru borelowskiego A,
(ii) dla dowolnego zbioru borelowskiego A i x spełniających $\mu(A) = \nu(-\infty, x]$ zachodzi $\mu(A_t) \geqslant \nu(-\infty, x+ct]$.

Dowód. (ii)⇒(i). Niech $\mu(A) = \nu(-\infty, x]$, wówczas

$$\begin{aligned}\mu^+(A) &= \liminf_{t\to 0+} \frac{\mu(A_t) - \mu(A)}{t} \geqslant \liminf_{t\to 0+} \frac{\nu((-\infty, x+ct) - \nu(-\infty, x]}{t} = \frac{1}{2}e^{-|x|} \\ &= \min\{\nu(-\infty, x], \nu(x, \infty)\} = \min\{\mu(A), 1-\mu(A)\}.\end{aligned}$$

(i)⇒(ii). Ustalmy najpierw $\delta < 1$ i niech

$$t_0 = t_0(\delta) = \inf\{t > 0 \colon\ \mu(A_t) < \nu(-\infty, x + \delta ct].$$

Załóżmy najpierw, że $t_0 < \infty$. Wówczas z monotoniczności $\mu(A_t)$ łatwo wynika, że $\mu(A_{t_0}) = \nu(-\infty, x + \delta c t_0]$, czyli

$$\begin{aligned}\mu^+(A_{t_0}) &\geqslant c \min\{\mu(A_{t_0}), 1-\mu(A_{t_0})\} = \frac{c}{2} e^{-|x+\delta c t_0|} \\ &= \frac{1}{\delta} \lim_{h\to 0+} \frac{\nu(-\infty, x+\delta c(t_0+h)] - \nu(-\infty, x+\delta c t_0]}{h}.\end{aligned}$$

Definicja dolnej i zwykłej granicy implikują, że istnieje $h_0 > 0$ takie, że dla $0 < h \leqslant h_0$,

$$\begin{aligned}\frac{\mu(A_{t_0+h}) - \mu(A)}{h} &\geqslant \frac{\mu((A_{t_0})_h) - \mu(A)}{h} \geqslant \sqrt{\delta}\frac{c}{2}e^{-|x+\delta ct_0|} \\ &\geqslant \frac{\nu(-\infty, x + \delta c(t_0 + h)] - \nu(-\infty, x + \delta ct_0]}{h}.\end{aligned}$$

Stąd $\mu(A_t) \geqslant \nu(-\infty, x + \delta ct]$ dla $t_0 \leqslant t \leqslant t_0 + h$, co przeczy definicji t_0.

Otrzymana sprzeczność pokazuje, że $t_0(\delta) = \infty$, czyli $\mu(A_t) \geqslant \nu(-\infty, x+\delta ct]$ dla $t > 0$. Przechodząc z δ do 1 otrzymujemy (ii). □

Definicja 5.16. Mówimy, że miara probabilistyczna μ na $(\mathbb{X}, d)$ *spełnia nierówność Cheegera ze stałą* $c > 0$, jeśli zachodzi jeden z warunków równoważnych Faktu 5.5.

Okazuje się, że nierówność Cheegera ma też formę funkcyjną przypominającą nierówność Poincaré.

Twierdzenie 5.17. *Miara μ spełnia nierówność Cheegera ze stałą $c > 0$ wtedy i tylko wtedy, gdy dla dowolnej funkcji lipschitzowskiej ograniczonej f zachodzi*

$$\mathbf{E}_\mu|f - \mathrm{Med}_\mu(f)| \leqslant \frac{1}{c}\int |\nabla f| d\mu$$

Do dowodu będziemy potrzebowali jednej z wersji tzw. „co-area formula".

Lemat 5.18 (Nierówność co-area)**.** *Dla dowolnej funkcji Lipschitzowskiej f na $\mathbb{X}$,*

$$\int_{\mathbb{X}} |\nabla f| d\mu \geqslant \int_{-\infty}^{\infty} \mu^+(\{f > t\}) dt.$$

Dowód. Wystarczy udowodnić nierówność dla funkcji ograniczonych. Istotnie, przyjmując $f_M = \max\{-M, \min\{f, M\}\}$, zauważamy, że $|\nabla f_M| \leqslant |\nabla f|$ i $\{f_M > t\} = \{f > t\}$ dla $|t| < M$ i przechodzimy z M do nieskończoności.

Rozpatrując zamiast f funkcję $f + c$, możemy zakładać, że f jest nieujemna. Określmy dla $t > 0$ funkcję f_t na X wzorem

$$f_t(x) := \sup\{f(y) \colon\ d(x, y) < t\}.$$

Lipschitzowskość f implikuje, że $(f_t - f)/t \leqslant M$. Łatwo sprawdzić, że $\{f_t > r\} = \{f > r\}_t$, stąd całkowanie przez części daje

$$\int_{\mathbb{X}} (f_t - f) d\mu = \int_0^{\infty} (\mu(\{f > r\}_t) - \mu(\{f > r\})) dr.$$

Mamy zatem

$$\begin{aligned}\int_{\mathbb{X}} |\nabla f| d\mu &= \int_{\mathbb{X}} \limsup_{t\to 0+} \frac{f_t - f}{t} d\mu = \limsup_{t\to 0+} \int_{\mathbb{X}} \frac{f_t - f}{t} d\mu \\ &\geqslant \liminf_{t\to 0+} \int_0^\infty \frac{\mu(\{f > r\}_t) - \mu(\{f > r\})}{t} dr \\ &\geqslant \int_0^\infty \liminf_{t\to 0+} \frac{\mu(\{f > r\}_t) - \mu(\{f > r\})}{t} dr = \int_{-\infty}^\infty \mu^+(\{f > r\}) dr,\end{aligned}$$

gdzie pierwsza i trzecia nierówność wynikają z Lematu Fatou (w pierwszej zastosowanego do funkcji nieujemnych $M - (f_t - f)/t$). □

Uwaga 5.19. Dla miar μ na $\mathbb{R}^n$ absolutnie ciągłych względem miary Lebesgue'a można udowodnić, że w nierówności co-area zachodzi równość.

Dowód Twierdzenia 5.17. „⇒". Bez straty ogólności możemy założyć, że $\mathrm{Med}_\mu(f) = 0$, wówczas $\mu\{f > t\} \leqslant 1/2$ dla $t > 0$ i $\mu\{f > t\} \geqslant 1/2$ dla $t > 0$. Nierówność co-area implikuje

$$\begin{aligned}\int |\nabla f| d\mu &\geqslant \int_{-\infty}^\infty \mu^+(\{f > t\}) dt \geqslant c \int_0^\infty \mu(\{f > t\}) dt + c \int_{-\infty}^0 (1 - \mu\{f > t\}) dt \\ &= c\mathbf{E}_\mu \max\{f, 0\} + c\mathbf{E}_\mu \max\{-f, 0\} = c\mathbf{E}_\mu |f|.\end{aligned}$$

„⇐" Udowodnimy szacowanie (i) z Faktu 5.5. Idea polega na aproksymacji 1_A przez funkcje lipschitzowskie. Jeśli $\mu(\bar{A}) > \mu(A)$, to $\mu^+(A) = \infty$ i nie ma co dowodzić, będziemy zatem zakładać, że $\mu(\bar{A}) = \mu(A)$, co jest równoważne temu, że $\mu(A_t) \to \mu(A)$ przy $t \to 0$. Dla $0 < t < 1/2$ określmy

$$f_t(x) = \frac{1}{t} \min\{\mathrm{dist}(x, A_{t^2}), t - 2t^2\}.$$

Wówczas f_t jest $1/t$-lipschitzowska, $f_t = 0$ na A_{t^2} i $f_t = 1 - 2t$ poza A_{t-t^2}, zatem $|\nabla f_t| \leqslant \frac{1}{t} I_{A_t \setminus A}$. Mamy zatem

$$\frac{\mu(A_t) - \mu(A)}{t} \geqslant \int |\nabla f_t| d\mu \geqslant c\mathbf{E}_\mu |f_t - \mathrm{Med}_\mu(f_t)|.$$

Jeśli $\mu(A) \geqslant 1/2$ to $\mathrm{Med}_\mu(f_t) = 0$ dla wszystkich t i

$$\mu^+(A) \geqslant c \liminf_{t\to 0+} \mathbf{E}_\mu |f_t| \geqslant c \liminf_{t\to 0+} (1 - 2t)(1 - \mu(A_{t-t^2})) = 1 - \mu(A).$$

Jeśli $\mu(A) < 1/2$ to $\mu(A_t) < 1/2$ dla małych t czyli $\mathrm{Med}_\mu(f_t) = 1 - 2t$ dla małych t i

$$\mu^+(A) \geqslant c \liminf_{t\to 0+} \mathbf{E}_\mu |f_t - 1 + 2t| \geqslant c \liminf_{t\to 0+} (1 - 2t)\mu(A_{t^2}) = \mu(A).$$

□

Następny fakt pokazuje, że nierówność Cheegera jest silniejsza od nierówności Poincaré.

Fakt 5.20. *Jeśli μ spełnia nierówność Cheegera ze stałą $c > 0$, to spełnia nierówność Poincaré ze stałą $4c^{-2}$.*

Dowód. Niech f będzie Lipschitzowską funkcją ograniczoną o medianie 0, zaś $g := \mathrm{sgn}(f)f^2$. Nietrudno sprawdzić, że g jest Lipschitzowska, ograniczona, ma medianę 0. Twierdzenie 5.17 implikuje

$$\mathbf{E}_\mu f^2 = \mathbf{E}_\mu |g| \leqslant \frac{1}{c}\mathbf{E}_\mu |\nabla g| = \frac{2}{c}\mathbf{E}_\mu (|f||\nabla f|) \leqslant \frac{2}{c}(\mathbf{E}_\mu |f|^2)^{1/2}(\mathbf{E}|\nabla f|^2)^{1/2}.$$

Dzieląc stronami przez $(\mathbf{E}_\mu |f|^2)^{1/2}$ dostajemy

$$\mathrm{Var}_\mu(f) \leqslant \mathbf{E}_\mu |f|^2 \leqslant \frac{4}{c^2}\mathbf{E}_\mu |\nabla f|^2.$$

□

Uwaga 5.21. Z nierówności Poincaré nie można wywnioskować nierówności Cheegera. Można pokazać, że miara z gęstością $\frac{1+\alpha}{2}|x|^\alpha I_{\{|x|\leqslant 1\}}$ dla $\alpha \in (0,1)$ spełnia nierówność Poincaré, a nie spełnia nierówności Cheegera.

Kolejne twierdzenie, pochodzące od Talagranda, rozwiązuje zagadnienie izoperymetryczne dla miary ν.

Twierdzenie 5.22. *Miara ν spełnia nierówność Cheegera ze stałą 1.*

Dowód. Dowód przeprowadzimy w kilku krokach, wykorzystując równoważności z Faktu .

Krok I. $\nu^+([a,b]) \geqslant \min\{\nu([a,b]), 1-\nu([a,b])\}$.

Rozpatrzymy trzy przypadki.

i) $a \geqslant 0$. Wówczas $\nu^+([a,b]) = e^{-a} + e^{-b} \geqslant e^{-a} - e^{-b} = \nu([a,b])$.

ii) $b \leqslant 0$. Mamy $\nu^+([a,b]) = e^a + e^b \geqslant e^b - e^a = \nu([a,b])$.

iii) $a < 0 < b$. Wtedy $\nu^+([a,b]) = e^a + e^{-b} = 1 - \nu([a,b])$.

Krok II. Jeśli A jest skończoną sumą przedziałów, to $\nu^+(A) \geqslant \min\{\nu(A), 1-\nu(A)\}$.

W rozważanym przypadku $\nu(A) = \nu(\bar{A})$ i $\nu^+(A) = \nu^+(\bar{A})$, zatem bez straty ogólności możemy zakładać, że $A = \bigcup_{i=1}^n [a_i, b_i]$ oraz $b_i < a_{i+1}$ dla $1 \leqslant i \leqslant n-1$. Niech $p_i := \nu([a_i, b_i])$. Mamy

$$\nu^+(A) = \sum_{i=1}^n \nu^+([a_i,b_i]) \geqslant \sum_{i=1}^n \min\{p_i, 1-p_i\} \geqslant \min\{\sum_i p_i, 1-\sum_i p_\} = \min\{\nu(A), 1-\nu(A)\}.$$

Pierwsza równość powyżej wynika z Kroku I, a drugą łatwo uzyskać przez rozpatrzenie dwu przypadków: $p_i \leqslant 1/2$ dla wszystkich i oraz $p_i > 1/2$ dla pewnego i.

Krok III. Jeśli A jest skończoną sumą przedziałów oraz $\nu(A) = \nu((-\infty, x])$, to $\nu(A_t) \geqslant \nu((-\infty, x+t])$.

Zauważamy, że zbiór A_t jest również skończoną sumą przedziałów, więc z Kroku II wynika, że $\nu^+(A_t) \geqslant \min\{\nu(A_t), 1-\nu(A_t)\}$. Teza Kroku III wynika z analogicznego rozumowania jak w dowodzie implikacji (i)⇒(ii) Faktu .

Krok IV. Jeśli A jest zbiorem otwartym oraz $\nu(A) = \nu((-\infty, x])$, to $\nu(A_t) \geqslant \nu((-\infty, x+t])$.

Zbiór A jest przeliczalną sumą przedziałów, więc dla $\delta > 0$ istnieje $B \subset A$, który jest skończoną sumą przedziałów i $\nu(B) \geqslant \nu((-\infty, x-\delta])$. Na mocy Kroku III $\nu(A_t) \geqslant \nu(B_t) \geqslant \nu((-\infty, x-\delta+t])$ i wystarczy przejść z δ od zera.

Krok V. Jeśli A jest dowolnym zbiorem borelowskim oraz $\nu(A) = \nu((-\infty, x])$, to $\nu(A_t) \geqslant \nu((-\infty, x+t])$.

Zauważmy, że $A_t \supset (A_\delta)_{t-\delta}$, ponadto A_δ jest zbiorem otwartym i $\nu(A_\delta) \geqslant \nu((-\infty, x])$. Korzystając z Kroku IV dostajemy $\nu(A_t) \geqslant \nu((A_\delta)_{t-\delta}) \geqslant \nu((\infty, x+t-\delta])$ i przechodzimy z δ do 0. □

Na prostej można scharakteryzować miary spełniające nierówność Cheegera.

Twierdzenie 5.23. *Niech μ będzie miarą probabilistyczną na $\mathbb{R}$, $F(x) = \mu(-\infty, x]$, zaś p będzie gęstością części absolutnie ciągłej μ. Wówczas następujące warunki są równoważne dla $c > 0$:*
i) μ spełnia nierówność Cheegera ze stałą c,
ii) μ jest $\frac{1}{c}$-lipschitzowskim obrazem ν,
iii) $\operatorname{essinf} \frac{p(x)}{\min\{F(x), 1-F(x)\}} \geqslant c$.

Szkic dowodu. Implikacja ii)⇒i) jest oczywistym wnioskiem z Twierdzenia 5.22 .

i)⇒iii). Wystarczy zauważyć, że $\mu^+((-\infty, x]) = p(x)$ dla p.w. $x \in \mathbb{R}$.

iii)⇒ii). Definiujemy $T\colon\ \mathbb{R} \mapsto \mathbb{R}$ wzorem $\nu(-\infty, x] = \mu(-\infty, Tx]$. Wówczas T transportuje ν na μ oraz

$$\nu((x,y]) = \mu((Tx, Ty]) \geqslant \int_{Tx}^{Ty} p(z)dz \geqslant c \int_{Tx}^{Ty} \min\{F(z), 1-F(z)\}dz.$$

Stąd łatwo wynika, że T jest ciągłe i $\limsup_{y \to x} \frac{Tx-Ty}{x-y} \leqslant 1/c$, czyli T jest $1/c$-Lipschitzowskie. □

6 Logarytmiczna Nierówność Sobolewa

6.1 Entropia funkcji

Definicja 6.1. Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{X}$, zaś f nieujemną funkcją mierzalną na $\mathbb{X}$. *Entropię f względem μ* definiujemy wzorem

$$\operatorname{Ent}_\mu(f) := \begin{cases} \int f \log f d\mu - \int f d\mu \log \int f d\mu & \text{jeśli } \int f \log(1+f) d\mu < \infty \\ \infty & \text{jeśli } \int f \log(1+f) d\mu = \infty. \end{cases}$$

Z wypukłości funkcji $x \log x$ na $[0, \infty)$ wynika, że $\mathrm{Ent}_\mu(f) \geqslant 0$, łatwo też zauważyć, że $\mathrm{Ent}_\mu(\lambda f) = \lambda \mathrm{Ent}_\mu(f)$ dla $\lambda \geqslant 0$.

Lemat 6.2. *Dla dowolnej funkcji nieujemnej na* $\mathbb{X}$*,*

$$\mathrm{Ent}_\mu(f) = \sup\Big\{ \int f g d\mu \colon \int e^g d\mu \leqslant 1 \Big\}. \tag{5}$$

Dowód. Z jednorodności obu stron tożsamości (5) możemy zakładać, że $\int f d\mu = 1$, wówczas $\mathrm{Ent}_\mu(f) = \int f \log f d\mu$.

Nietrudno sprawdzić, że dla $u > 0$, $\sup_{v \in \mathbb{R}}(uv - e^v) = u \log u - u$, zatem

$$uv \leqslant u \log u - u + e^v \text{ dla } u \geqslant 0, v \in \mathbb{R}. \tag{6}$$

Zatem biorąc g takie, że $\int e^g d\mu \leqslant 1$ dostajemy

$$\int f g d\mu \leqslant \int (f \log f - f + e^g) d\mu = \mathrm{Ent}_\mu(f) - 1 + \int e^g d\mu \leqslant \mathrm{Ent}_\mu(f).$$

By udowodnić nierówność w przeciwną stronę wystarczy przyjąć $g = \log f$. □

Z powyższego lematu łatwo wykazać tensoryzowalność entropii:

Fakt 6.3. *Załóżmy, że* μ_i *są miarami probabilistycznymi na* $\mathbb{X}_i$*,* $\mathbb{X} = \mathbb{X}_1 \times \ldots \times \mathbb{X}_n$ *oraz* $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$*. Wówczas dla dowolnej nieujemnej funkcji* f *na* $\mathbb{X}$ *zachodzi*

$$\mathrm{Ent}_\mu(f) \leqslant \sum_{i=1}^n \mathbf{E}_\mu \mathrm{Ent}_{\mu_i}(f).$$

Dowód. Weźmy funkcję g na $\mathbb{X}$ taką, że $\int e^g d\mu \leqslant 1$ oraz przyjmijmy dla $i = 1, \ldots, n$,

$$g^i(x_1, \ldots, x_n) := \log\Big(\frac{\int e^{g(x_1, \ldots, x_n)} d\mu_1(x_1) \cdots d\mu_{i-1}(x_{i-1})}{\int e^{g(x_1, \ldots, x_n)} d\mu_1(x_1) \cdots d\mu_i(x_i)} \Big).$$

□

Wówczas $g \leqslant \sum_{i=1}^n g^i$ oraz $\int e^{g^i} d\mu_i \leqslant 1$, stąd

$$\int f g d\mu \leqslant \sum_{i=1}^n \int f g^i d\mu = \sum_{i=1}^n \int \Big(\int f g^i d\mu_i \Big) d\mu \leqslant \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f) d\mu.$$

6.2 LNS - definicja, tensoryzowalność, związek z koncentracją

Definicja 6.4. Mówimy, że miara probabilistyczna na $(\mathbb{X}, d)$ *spełnia logarytmiczną nierówność Sobolewa* ze stałą C, jeśli dla wszystkich ograniczonych lipschitzowskich funkcji f na $\mathbb{X}$ zachodzi

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C \int |\nabla f|^2 d\mu. \tag{7}$$

Fakt 6.5. *Załóżmy, że miary probabilistyczne μ_i na $(\mathbb{X}_i, d_i)$ spełniają logarytmiczną nierówność Sobolewa ze stałą C_i względem gradientu $|\nabla_i|$. Wówczas miara $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ spełnia logarytmiczną nierówność Sobolewa ze stałą $C = \max_i C_i$ względem gradientu ∇f danego wzorem*

$$|\nabla f|^2 = \sum_{i=1}^n |\nabla_i f|^2.$$

Dowód. Z Faktu 6.3 dostajemy

$$\mathrm{Ent}_\mu(f^2) \leqslant \sum_{i=1}^n \mathbf{E}_\mu \mathrm{Ent}_{\mu_i}(f^2) \leqslant \sum_{i=1}^n \mathbf{E}_\mu 2C_i \mathbf{E}_{\mu_i} |\nabla_i f|^2 \leqslant 2C \mathbf{E}_\mu \sum_{i=1}^n |\nabla_i f|^2.$$

□

Twierdzenie 6.6. *Załóżmy, że miara μ spełnia logarytmiczną nierówność Sobolewa ze stałą C. Wówczas dla każdej funkcji 1-lipschitzowskiej F i $t > 0$,*

$$\mu\Big(\Big\{F \geqslant \int F d\mu + t\Big\}\Big) \leqslant \exp\Big(-\frac{t^2}{2C}\Big).$$

W szczególności $\alpha_X(t) \leqslant \exp(-t^2/8C)$.

Dowód. Ustalmy ograniczoną funkcję 1-Lipschitzowską F taką, że $\int F d\mu = 0$. Wystarczy, że pokażemy iż dla $\lambda \geqslant 0$

$$M(\lambda) := M_{F,\lambda} = \int e^{\lambda F} d\mu \leqslant e^{C\lambda^2/2}.$$

Zastosujmy logarytmiczną nierówność Sobolewa do $f^2 := e^{\lambda F}$. Wówczas

$$\mathrm{Ent}_\mu(f^2) = \lambda \mathbf{E}_\mu F e^{\lambda F} - \mathbf{E}_\mu e^{\lambda F} \log \mathbf{E}_\mu e^{\lambda F} = \lambda M'(\lambda) - M(\lambda) \log M(\lambda)$$

oraz

$$\int |\nabla f|^2 d\mu = \frac{\lambda^2}{4} \int |\nabla F|^2 e^{\lambda F} \leqslant \frac{\lambda^2}{4} M(\lambda).$$

Zatem (7) daje

$$\lambda M'(\lambda) - M(\lambda) \log M(\lambda) \leqslant C \frac{\lambda^2}{2} M(\lambda). \tag{8}$$

Określmy $H(\lambda) := \frac{1}{\lambda}\log M(\lambda)$ dla $\lambda > 0$. Wówczas

$$\lim_{\lambda\to 0} H(\lambda) = \frac{M'(0)}{M(0)} = \int F d\mu = 0$$

oraz na podstawie (8)

$$H'(\lambda) = -\frac{1}{\lambda^2}\log M(\lambda) + \frac{1}{\lambda}\frac{M'(\lambda)}{M(\lambda)} \leqslant \frac{C}{2}.$$

Zatem $H(\lambda) \leqslant C\lambda/2$ czyli $M(\lambda) \leqslant \exp(C\lambda^2/2)$. □

6.3 LNS dla miary gaussowskiej

Fakt 6.7. *i) Niech $\mu_1 = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, wówczas dla dowolnego $f\colon \{-1,1\} \to \mathbb{R}$,*

$$\mathrm{Ent}_{\mu_1}(f^2) \leqslant 2\mathbf{E}_{\mu_1}|Df|^2,$$

gdzie $Df(x) = \frac{1}{2}(f(x) - f(-x))$.

ii) Niech $\mu_n = \mu_1 \otimes \cdots \otimes \mu_1$ będzie rozkładem jednostajnym na $\{-1,1\}^n$, wówczas dla dowolnego $f\colon \{-1,1\}^n \to \mathbb{R}$,

$$\mathrm{Ent}_{\mu_n}(f^2) \leqslant 2\mathbf{E}_{\mu_n}|Df|^2,$$

gdzie

$$|Df|^2(x) = \frac{1}{4}\sum_{i=1}^{n}(f(x) - f(s_i(x)))^2,$$

oraz $s_i((x_1,\ldots,x_n)) = (x_1,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)$ dla $1 \leqslant i \leqslant n$.

Dowód. i) Z uwagi na jednorodność możemy zakładać, że $\mathbf{E}_{\mu_1} f^2 = 1$, wówczas istnieje $t \in [-1,1]$ takie, że $f(1) = \sqrt{1+t}$ oraz $f(-1) = \sqrt{1-t}$ i nierówność z punktu i) ma postać $\alpha(t) \geqslant 0$, gdzie

$$\alpha(t) := 1 - \sqrt{1-t^2} - \frac{1+t}{2}\log(1+t) - \frac{1-t}{2}\log(1-t).$$

Nietrudno sprawdzić, że $\alpha(0) = \alpha'(0) = 0$ oraz

$$\alpha''(t) = \frac{1}{1-t^2}\Big(\frac{t^2}{\sqrt{1-t^2}} - \frac{t^2}{1+\sqrt{1-t^2}}\Big) \geqslant 0,$$

więc istotnie $\alpha(t) \geqslant 0$.

ii) Wynika z punktu i) i Faktu 6.3. □

Twierdzenie 6.8. *Miara γ_n spełnia logarytmiczną nierówność Sobolewa z $C = 1$.*

Dowód. Z uwagi na Fakt 6.3 wystarczy rozważyć przypadek $n = 1$. Niech $f \in C^1_{\mathrm{ogr}}(\mathbb{R})$. Określmy $g_n \colon \{-1,1\}^n \to \mathbb{R}$ wzorem

$$g_n(x) := f\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right).$$

Niech μ_n i $|Df|$ będą jak w Fakcie 6.7. Wówczas na mocy centralnego twierdzenia granicznego

$$\mathrm{Ent}_{\mu_n}(g_n^2) = \int g_n^2 \log g_n^2 d\mu_n - \int g_n^2 d\mu_n \log \int g_n^2 d\mu_n \to \mathrm{Ent}_{\gamma_1}(f^2).$$

Ponadto kładąc $T_n(x) = n^{-1/2}(x_1 + \ldots + x_n)$

$$|Dg_n|(x)^2 = \frac{1}{4}\sum_{i=1}^n \left(f(T_n(x)) - f\left(T_n(x) - 2\frac{x_i}{\sqrt{n}}\right)\right)^2 = f'(T_n(x))^2 + r_n$$

gdzie r_n zbiega do zera jednostajnie względem $|T_n(x)|$. Zatem

$$\lim_{n\to\infty} \mathbf{E}_{\mu_n} |Dg_n|(x)^2 = \lim_{n\to\infty} \mathbf{E}_{\mu_n} f'(T_n(x))^2 = \mathbf{E}_{\gamma_1} f'(x)^2.$$

□

Fakt 6.9. *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{X}$, V jest ograniczoną funkcją borelowską oraz $d\nu = Z^{-1}e^V d\mu$, gdzie $Z = \int e^V d\mu$. Wówczas jeśli miara μ spełnia logaryticzną nierówność Sobolewa ze stałą C to ν spełnia logarytmiczną nierówność Sobolewa ze stałą $Ce^{4\|V\|_\infty}$.*

Dowód. Funkcja $\varphi(u) = u \log u$ jest wypukła na $[0,\infty)$ stąd dla dowolnych s, t, $\varphi(s+t) \geqslant \varphi(t) + \varphi'(t)s$, więc

$$\varphi\Big(\int f^2 d\nu\Big) = \varphi\Big(t + \int (f^2 - t) d\nu\Big) \geqslant \varphi(t) + \varphi'(t)\int (f^2 - t) d\nu.$$

Zatem

$$\begin{aligned}
\mathrm{Ent}_\nu(f^2) &= \inf_{t\in\mathbb{R}} \int \left[\varphi(f^2) - \varphi(t) - \varphi'(t)(f^2 - t)\right] d\nu \\
&\leqslant \frac{1}{Z} e^{\|V\|_\infty} \inf_{t\in\mathbb{R}} \int \left[\varphi(f^2) - \varphi(t) - \varphi'(t)(f^2 - t)\right] Z e^{-V} d\nu \\
&= \frac{1}{Z} e^{\|V\|_\infty} \mathrm{Ent}_\mu(f^2) \leqslant \frac{2C}{Z} e^{\|V\|_\infty} \int |\nabla f|^2 d\mu \\
&\leqslant 2Ce^{2\|V\|_\infty} \int |\nabla f|^2 d\nu.
\end{aligned}$$

□

Kolejny fakt dowodzimy tak samo jak dla nierówności Poincaré.

Fakt 6.10. *Jeśli miara ν na $(\mathbb{Y},\rho)$ jest L-lipschitzowskim obrazem miary μ na $(\mathbb{X},d)$ oraz μ spełnia logarytmiczną nierówność Sobolewa ze stałą C, to ν spełnia logarytmiczną nierówność Sobolewa ze stałą CL^2.*

Stosując logarytmiczną nierówność Sobolewa do funkcji $f = 1 + \varepsilon g$ dowodzimy

Fakt 6.11. *Jeśli miara probabilistyczna μ spełnia logarytmiczną nierówność Sobolewa ze stałą C, to spełnia również nierówność Poincaré ze stałą C.*

Opierając się na twierdzeniu Muckenhoupta da się wyprowadzić kryterium równoważne nierówności logarytmicznej Sobolewa dla miar na prostej.

Twierdzenie 6.12. *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{R}$ o medianie m, zaś p oznacza gęstość jej części absolutnie ciągłej. Wówczas miara μ spełnia logarytmiczną nierówność Sobolewa ze skończoną stałą C wtedy i tylko wtedy gdy $\max\{B_+, B_-\} < \infty$, gdzie*

$$B_+ = \sup_{x>m} \mu[x,\infty) \ln\Big(\frac{1}{\mu[x,\infty)}\Big) \int_m^x \frac{1}{p(y)} dy$$

$$B_- = \sup_{x<m} \mu(-\infty,x] \ln\Big(\frac{1}{\mu(-\infty,x]}\Big) \int_x^m \frac{1}{p(y)} dy.$$

Co więcej optymalna stała C_{opt} w nierówności Poincaré spełnia

$$\frac{1}{150}(B_+ + B_-) \leqslant C_{\mathrm{opt}} \leqslant 468(B_+ + B_-).$$

6.4 Nierówność Bobkowa

Logarytmiczna nierówność Sobolewa implikuje koncentrację gaussowską, ale nie implikuje gaussowskiej izoperymetrii. Okazuje się, że jest silniejsza nierówność, która implikuje gaussowską izoperymetrię, a jednocześnie ma szereg równie dobrych własności jak nierówność Poincaré czy logarytmiczna nierówność Sobolewa.

Przedstawione poniżej rozumowania można podobnie jak w poprzednich sekcjach prowadzić w większej ogólności, jednak by uniknąć szczegółów technicznych ograniczymy się do miar na $\mathbb{R}^n$ i funkcji gładkich.

W tej części przez I będziemy oznaczać gaussowską funckję izoperymetryczną, tzn $I(x) = \varphi(\Phi^{-1}(x))$, gdzie $\varphi = (2\pi)^{-1/2}\exp(-|x|^2/2)$. Dodatkowo określamy $I(0) = I(1) = 0$.

Definicja 6.13. Mówimy, że miara probabilistyczna μ na $\mathbb{R}^n$ *spełnia nierówność Bobkowa* ze stałą C, jeśli dla wszystkich $f \in C^1_{\mathrm{ogr}}(\mathbb{R}^n)$ o wartościach w przedziale $[0,1]$ zachodzi

$$I\Big(\int f d\mu\Big) \leqslant \int \sqrt{I(f)^2 + C^2|\nabla f|^2} d\mu. \tag{9}$$

Fakt 6.14. *Jeśli miary μ_i spełniają nierówność Bobkowa ze stałymi C_i, to miara $\mu_1 \otimes \cdots \otimes \mu_n$ spełnia nierówność Bobkowa ze stałą* $\max_i C_i$.

Twierdzenie 6.15. *Jeśli miara probabilistyczna μ na $\mathbb{R}^n$ spełnia nierówność Bobkowa na ze stałą C, to*

$$\mu^+(A) \geqslant \frac{1}{C} I(\mu(A)) \qquad \textit{dla } A \in \mathcal{B}(\mathbb{R}^n)$$

oraz

$$\mu(A_t) \geqslant \Phi(\Phi^{-1}(\mu(A)) + t/C) \qquad \textit{dla } A \in \mathcal{B}(\mathbb{R}^n),\ t > 0.$$

Twierdzenie 6.16. *Kanoniczna miara gaussowska γ_n spełnia nierówność Bobkowa z $C = 1$.*

7 Nierówności Splotu Infimum

7.1 Własność (τ) Maureya

Zacznijmy od zaproponowanej przez Maureya definicji.

Definicja 7.1. Splotem infimum dwu funkcji f i g określonych na $\mathbb{R}^n$ nazywamy funkcję $f \square g$ daną wzorem

$$f \square g(x) := \inf\{f(y) + g(x-y) \colon y \in \mathbb{R}^n\}.$$

Niech μ będzie miarą probabilistyczną na $\mathbb{R}^n$ oraz $\varphi \colon \mathbb{R}^n \to [0, \infty]$. Mówimy, że para (μ, φ) ma własność (τ) bądź, że miara μ spełnia nierówność splotu infimum z funkcją kosztu φ jeśli

$$\int e^{f \square \varphi} d\mu \int e^{-f} d\mu \leqslant 1$$

dla dowolnej ograniczonej mierzalnej funkcji f na $\mathbb{R}^n$.

Pierwszą użyteczną cechą własności (τ) jest jej tensoryzowalność.

Fakt 7.2. *Jeśli pary (μ_i, φ_i) mają własność (τ), $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ oraz*

$$\varphi(x_1, \ldots, x_n) = \varphi_1(x_1) + \ldots + \varphi_n(x_n),$$

to również para (μ, φ) ma własność (τ).

Dowód. Prosty argument indukcyjny pokazuje, że wystarczy udowodnić tezę dla $n = 2$. Niech $f = f(x, y)$ będzie ograniczoną funkcją na $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$, określmy $f^y(x) = f(x, y)$ oraz zdefiniujmy g na $\mathbb{R}^{n_2}$ jako

$$g(y) := \ln\Big(\int e^{f^y \square \varphi_1(x)} d\mu_1(x) \Big).$$

Własność (τ) dla (μ_1, φ_1) implikuje, że $g(y) \leqslant -\ln(\int e^{-f^y} d\mu_1)$, zatem

$$\int e^{-g} d\mu_2 \geqslant \int e^{-f} d\mu_1 \otimes \mu_2.$$

Ponadto dla dowolnych $y, \tilde{y}$

$$\int e^{f \square \varphi(x,y)} d\mu_1(x) \leqslant \int e^{f^{\tilde{y}} \square \varphi_1(x) + \varphi_2(y-\tilde{y})} d\mu_1(x) = e^{g(\tilde{y}) + \varphi_2(y-\tilde{y})},$$

więc $g \square \varphi_2(y) \geqslant \ln(\int e^{f \square \varphi(x,y)} d\mu_1(x))$ i

$$\int e^{g \square \varphi_2} d\mu_2 \geqslant \int e^{f \square \varphi} d\mu_1 \otimes \mu_2.$$

Teza wynika z powyższych nierówności i własności τ dla (μ_2, φ_2). □

Następny fakt pokazuje w jaki sposób można transportować (τ).

Fakt 7.3. *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{R}^n$, zaś φ funkcją kosztu na $\mathbb{R}^n$ taką, że (μ, φ) spełnia własność (τ). Jeśli $T\colon \mathbb{R}^n \to \mathbb{R}^m$ oraz funkcja ψ na $\mathbb{R}^m$ spełnia $\psi(Tx - Ty) \leqslant \varphi(x - y)$ dla wszystkich x, y, to para $(\mu \circ T^{-1}, \psi)$ ma własność (τ).*

Dowód. Niech f będzie ograniczoną funkcją na $\mathbb{R}^m$. Zauważmy, że

$$f \circ T \square \varphi(x) = \inf_y (f(Ty) + \varphi(x-y)) \geqslant \inf_y (f(Ty) + \psi(Tx - Ty)) \geqslant f \square \psi(Tx).$$

Zatem

$$\begin{aligned} \int e^{f \square \psi} d\mu \circ T^{-1} &= \int e^{f \square \psi(Tx)} d\mu(x) \leqslant \int e^{f \circ T \square \varphi(x)} d\mu(x) \leqslant \Big(\int e^{-f \circ T} d\mu \Big)^{-1} \\ &= \Big(\int e^{-f} d\mu \circ T^{-1} \Big)^{-1}. \end{aligned}$$

□

7.2 Splot infimum a koncentracja

By sformułować związki nierówności splotu infimum z koncentracją określmy zbiór

$$B_\varphi(t) = \{x\colon \varphi(x) \leqslant t\}.$$

Zacznijmy od prostego faktu

Fakt 7.4. *Jeśli (μ, φ) ma własność (τ) to dla dowolnego zbioru borelowskiego A takiego, że $\mu(A) > 0$ mamy*

$$1 - \mu(A + B_\varphi(t)) \leqslant \frac{1}{\mu(A)} e^{-t}.$$

Dowód. Zastosujmy własność (τ) do funkcji $f = 0$ na zbiorze A i $f = \infty$ poza zbiorem A. Zauważmy, że $f\Box\varphi \geqslant t$ poza zbiorem $A + B_\varphi(t)$, zatem

$$1 \geqslant \int e^{f\Box\varphi} d\mu \int e^{-f} d\mu \geqslant e^t(1 - \mu(A + B_\varphi(t))\mu(A).$$

□

Uwaga 7.5. Funkcja f w poprzednim dowodzie nie była oczywiście ograniczona, ale łatwo ominąć ten problem stosując nierówność (τ) do $f_n = n\mathbb{1}_{\mathbb{R}^n \setminus A}$ dla $n \geqslant t$.

Poprzedni Fakt daje dobre oszacowanie tylko dla dużych wartości t. Nieco modyfikując jego dowód da się uzyskać też nierówności koncentracyjne dla małych t.

Fakt 7.6. *Załóżmy, że para (μ, φ) ma własność (τ). Wówczas dla dowolnego zbioru borelowskiego A i $t > 0$,*

$$\mu(A + B_\varphi(t)) \geqslant \frac{e^t \mu(A)}{(e^t - 1)\mu(A) + 1}. \tag{10}$$

W szczególności

$$\mu(A + B_\varphi(t)) > \min\{e^{t/2}\mu(A), 1/2\} \tag{11}$$

oraz

$$\mu(A) \geqslant \frac{1}{2} \;\Rightarrow\; 1 - \mu(A + B_\varphi(t)) < e^{-t/2}(1 - \mu(A)). \tag{12}$$

Ponadto

$$\mu(A) = \nu(-\infty, x] \;\Rightarrow\; \mu(A + B_\varphi(t)) \geqslant \nu(-\infty, x + t/2]. \tag{13}$$

Dowód. Niech $f(x) = t\mathbb{1}_{\mathbb{R}^n \setminus A}$. Wówczas f jest nieujemna, więc $f\Box\varphi$ też jest nieujemna (rozpatrujemy tylko nieujemne funkcje kosztu). Dla $x \notin A + B_\varphi(t)$ mamy $f\Box\varphi(x) \geqslant t$.

Zatem własność (τ) daje

$$\begin{aligned} 1 &\geqslant \int e^{f\Box\varphi(x)} d\mu(x) \int e^{-f(x)} d\mu(x) \\ &\geqslant \Big[\mu(A + B_\varphi(t)) + e^t(1 - \mu(A + B_\varphi(t)))\Big]\,[\mu(A) + e^{-t}(1 - \mu(A))], \end{aligned}$$

skąd bezpośredni rachunek prowadzi do (10).

Niech $f_t(p) := e^t p/((e^t - 1)p + 1)$, zauważmy, że f_t is rosnąca względem p oraz dla $p \leqslant e^{-t/2}/2$,

$$(e^t - 1)p + 1 \leqslant e^{t/2} + 1 - \frac{1}{2}(e^{t/2} + e^{-t/2}) < e^{t/2},$$

skąd otrzymujemy (11). Ponadto dla $p \geqslant 1/2$,

$$1 - f_t(p) = \frac{1 - p}{(e^t - 1)p + 1} \leqslant \frac{1 - p}{(e^t + 1)/2} < e^{-t/2}(1 - p)$$

i dostajemy (12).

Niech $F(x) = \nu(-\infty, x]$ i $g_t(p) = F(F^{-1}(p) + t)$. Poprzednie rachunki pokazują, że dla $t, p > 0$, $f_t(p) \geqslant g_{t/2}(p)$, jeśli $F^{-1}(p) + t/2 \leqslant 0$ lub $F^{-1}(p) \geqslant 0$. Ponieważ $g_{t+s} = g_t \circ g_s$ i $f_{t+s} = f_t \circ f_s$, otrzymujemy $f_t(p) \geqslant g_{t/2}(p)$ dla wszystkich $t, p > 0$, zatem (10) implikuje (13).

□

7.3 Dwupoziomowa koncentracja dla rozkładu wykładniczego

Niech jak do tej pory ν oznacza miarę na $\mathbb{R}$ z gęstością $\frac{1}{2}e^{-|x|}$, zaś ν_+, ν_- miary z gęstościami odpowiednio $e^{-x}\mathbb{1}_{[0,\infty)}$ i $e^{x}\mathbb{1}_{(-\infty,0]}$.

Fakt 7.7. *Para (ν_+, φ_0) ma własność (τ), gdzie*

$$\varphi_0(x) = \begin{cases} \frac{1}{18}x^2 & \text{dla } |x| \leqslant 2 \\ \frac{2}{9}(|x| - 1) & \text{dla } |x| > 2. \end{cases}$$

Lemat 7.8. *Dla wszystkich $x \in \mathbb{R}$ mamy $2|\varphi_0'(x)| \leqslant 1$ oraz*

$$(1 - 4\varphi_0'(x)^2)e^{\varphi_0(x)} \geqslant 1.$$

Dowód. Pierwszą nierówność otrzymujemy przez łatwe sprawdzenie. By udowodnić drugą, z uwagi na symetrię φ_0, wystarczy rozpatrywać przypadek $x \geqslant 0$. Ponadto $\varphi_0'(x)$ jest stałe dla $x \geqslant 2$ a φ_0 rosnące na tym przedziale, więc możemy zakładać, że $0 \leqslant x \leqslant 2$. Wówczas nierówność po podstawieniu $y = x^2/18$ ma postać

$$e^{-y} \leqslant 1 - \frac{8}{9}y, \quad 0 \leqslant y \leqslant \frac{2}{9}.$$

Funkcja e^{-y} jest wypukła, więc wystarczy sprawdzić tylko $y = 0$ i $y = 2/9$. □

Dowód Faktu 7.7. Ustalmy funkcję ograniczoną f, przyjmijmy $g := f \square \varphi_0$ i niech

$$I_0 := \int_0^\infty e^{-f(x)-x}dx, \quad I_1 := \int_0^\infty e^{g(x)-x}dx.$$

Musimy pokazać, że $I_0 I_1 \leqslant 1$. Dla $t \in (0, 1)$ zdefiniujmy $x(t)$ i $y(t)$ wzorami

$$\int_0^{x(t)} e^{-f(x)-x}dx = tI_0 \text{ oraz } \int_0^{y(t)} e^{g(x)-x}dx = tI_1.$$

Wówczas

$$x'(t) = I_0 e^{f(x(t))+x(t)}, \quad y'(t) = I_1 e^{-g(y(t))+y(t)}.$$

Na mocy definicji g, $g(y(t)) \leqslant f(x(t)) + \varphi_0(y(t) - x(t))$, więc

$$y'(t) \geqslant I_1 e^{-f(x(t))-\varphi_0(y(t)-x(t))+y(t)}.$$

Niech $z(t) = \frac{1}{2}(x(t) + y(t)) - \varphi_0(x(t) - y(t))$, wówczas

$$z'(t) = \Big(\frac{1}{2} - \varphi_0'(x(t) - y(t))\Big)x'(t) + \Big(\frac{1}{2} + \varphi_0'(x(t) - y(t))\Big)y'(t).$$

Pisząc dla uproszczenia x i y zamiast $x(t)$ i $y(t)$ stosując poprzednie oszacowanie $y'(t)$ oraz nierówność między średnia arytmetyczną i geometryczną dostajemy (wykorzystując parzystość φ_0)

$$\begin{aligned} z'(t) &\geqslant \frac{1}{2}(1 - 2\varphi_0'(x-y))I_0 e^{x+f(x)} + \frac{1}{2}(1 + 2\varphi_0'(x-y))I_1 e^{-\varphi_0(x-y)+y-f(x)} \\ &\geqslant \sqrt{1 - 4\varphi_0'(x-y)^2}\sqrt{I_0 I_1}\, e^{\frac{1}{2}(x+y)-\frac{1}{2}\varphi_0(x-y)} \\ &= \sqrt{I_0 I_1}\, e^{z(t)}\sqrt{1 - 4\varphi_0'(x-y)^2}\, e^{\frac{1}{2}\varphi_0(x-y)}. \end{aligned}$$

Zatem na mocy Lematu 7.8, $(-e^{-z(t)})' = e^{-z(t)}z'(t) \geqslant \sqrt{I_0 I_1}$, co po odcałkowaniu daje $\sqrt{I_0 I_1} \leqslant 1$. □

Uwaga 7.9. Funkcja g jest ciągła, więc y jest różniczkowalna. Funkcja f nie musi być ciągła więc x nie musi być różniczkowalna. Jednak z ograniczoności f łatwo wywnioskować lokalną Lipschitzowskość x (stąd też z), a zatem różniczkowalność x prawie wszędzie. Funkcja $e^{-z(t)}$ jest zatem lokalnie lipschitzowska, czyli jest całką swojej pochodnej, która istnieje p.w..

Wniosek 7.10. *Miara ν spełnia nierówność infimum z funkcją kosztu φ_1 postaci*

$$\varphi_1(t) = 2\varphi_0(\frac{t}{2}) = \begin{cases} \frac{1}{36}t^2 & dla\ |t| \leqslant 4 \\ \frac{2}{9}(|t| - 2) & dla\ |t| > 4. \end{cases}$$

Dowód. Z wypukłości funkcji φ_0 łatwo wynika, że $\varphi_1 = \varphi_0 \square \varphi_0$. Ponieważ miara ν_- jest symetrycznym odbiciem ν_+ a funkcja φ_0 jest symetryczna, to (ν_-, φ_0) ma własność (τ), więc $(\nu_+ \otimes \nu_-, \varphi_0(x) + \varphi_0(y))$ też ma (τ). Miara ν jest splotem miar ν_+ i ν_-, czyli obrazem $\nu_+ \otimes \nu_-$ przy przekształceniu $T(x, y) = x + y$. Teza wynika z Faktu 7.3 □

Wiemy, że miara ν a zatem i miara produktowa ν^n spełniają nierówność Poincaré, więc jeśli $\nu^n(A) \geqslant \frac{1}{2}$, to $\nu^n(A + tB_2^n) \geqslant 1 - e^{-t/C}$ dla pewnej stałej absolutnej C. Okazuje się, że można tę nierówność wzmocnić.

Zanim sformułujemy twierdzenie (które pierwszy z gorszymi stałymi udowodnił Talagrand) wprowadźmy następujące oznaczenie kuli jednostkowej w l_p^n dla $1 \leqslant p < \infty$

$$B_p^n := \{x \in \mathbb{R}^n \colon \sum_{i=1}^n |x_i|^p \leqslant 1\}.$$

Twierdzenie 7.11. *Dla dowolnego zbioru borelowskiego A w $\mathbb{R}^n$ takiego, że $\nu^n(A) > 0$ mamy dla $t \geqslant 0$,*

$$1 - \nu^n(A + 6\sqrt{t}B_2^n + 9tB_1^n) \leqslant \frac{1}{\nu^n(A)} e^{-t}.$$

Ponadto

$$\nu^n(A) = \nu(-\infty, x] \ \Rightarrow \nu^n(A + 6\sqrt{2t}B_2^n + 18tB_1^n) \geqslant \nu(-\infty, x + t].$$

Dowód. Para (ν^n, φ_n) ma własność (τ), gdzie $\varphi_n(x_1, \ldots, x_n) = \varphi_1(x_1) + \ldots + \varphi_1(x_n)$. Łatwo sprawdzić, że

$$B_{\varphi_n}(t) \subset 6\sqrt{t}B_2^n + 9tB_1^n.$$

Teza wynika zatem z Faktów 7.4 i 7.6. □

7.4 Wypukła własność (τ)

Definicja 7.12. Niech μ będzie miarą probabilistyczną na $\mathbb{R}^n$ oraz $\varphi\colon \mathbb{R}^n \to [0, \infty]$ wypukła. Mówimy, że para (μ, φ) ma *wypukłą własność* (τ) bądź, że miara μ *spełnia nierówność splotu infimum z funkcją kosztu* φ jeśli

$$\int e^{f \square \varphi} d\mu \int e^{-f} d\mu \leqslant 1$$

dla dowolnej wypukłej funkcji f na $\mathbb{R}^n$.

Wypukła nierówność (τ) się tensoryzuje podobnie jak zwykła nierówność (τ).

Fakt 7.13. *Jeśli pary (μ_i, φ_i) mają wypukłą własność (τ), $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ oraz*

$$\varphi(x_1, \ldots, x_n) = \varphi_1(x_1) + \ldots + \varphi_n(x_n),$$

to również para (μ, φ) ma wypukłą własność (τ).

Dowód. Dowód przebiega podobnie do dowodu Faktu 7.2. Stosując taką jak w tamtym dowodzie notację, wystarczy zauważyć, że funkcja $y \mapsto f^y \square \varphi_1$ jest wypukła (wykorzystujemy tu zarówno wypukłość f jak i φ_1), i wywnioskować z nierówności Höldera wypukłość g. □

Tak samo jak Faktu 7.4 dowodzimy, że wypukła nierówność (τ) implikuje koncentrację dla zbiorów wypukłych.

Fakt 7.14. *Jeśli (μ, φ) ma wypukłą własność (τ) to dla dowolnego wypukłegp zbioru borelowskiego A takiego, że $\mu(A) > 0$ mamy*

$$1 - \mu(A + B_\varphi(t)) \leqslant \frac{1}{\mu(A)} e^{-t}.$$

Lemat 7.15 (Maurey)**.** *Załóżmy, że μ jest miarą probabilistyczną na $\mathbb{R}^n$ skupioną na zbiorze o średnicy nie większej niż Δ. Wówczas para $(\mu, \frac{|x|^2}{4\Delta^2})$ ma wypukłą własność τ.*

Dowód. Załóżmy, że μ jest skupiona na zbiorze A i $\operatorname{diam}(A) \leqslant \Delta$. Niech f będzie wypukłą funkcją na $\mathbb{R}^n$, $\varphi(x) = \frac{1}{4\Delta^2}|x|^2$ oraz $g := f \square \varphi$. Ewentualnie odejmując od f stałą możemy zakładać, że $\inf_A f = 0$. Ustalmy $\varepsilon > 0$ i wybierzmy $a \in A$ taki, że $f(a) \leqslant \varepsilon$. Wówczas dla $x \in A$ i $\lambda \in [0,1]$ mamy

$$\begin{aligned} g(x) &\leqslant f(\lambda a + (1-\lambda)x) + \varphi(\lambda(x-a)) \leqslant \lambda f(a) + (1-\lambda)f(x) + \frac{\lambda^2|x-a|^2}{4\Delta^2} \\ &\leqslant \lambda\varepsilon + (1-\lambda)f(x) + \frac{1}{4}\lambda^2. \end{aligned}$$

Z dowolności $\varepsilon > 0$ dostajemy

$$g(x) \leqslant \inf_{\lambda\in[0,1]} (1-\lambda)f(x) + \frac{1}{4}\lambda^2 = k(f(x)) \quad \text{dla } x \in A,$$

gdzie $k(x) = u - u^2$ dla $u \in [0, 1/2]$ i $k(x) = 1/4$ dla $x \geqslant 1/2$.

Pokażemy, że $e^{k(u)} \leqslant 2 - e^{-u}$. Wystarczy tę nierówność oczywiście pokazać dla $u \in [0, 1/2]$, ale wtedy

$$\frac{1}{2}(e^{k(u)} + e^{-u}) = e^{-u^2/2}\cosh(u - u^2/2) \leqslant e^{-u^2/2}\cosh(u) \leqslant 1.$$

Mamy zatem

$$\int e^g d\mu \leqslant \int e^{k(f)} d\mu \leqslant 2 - \int e^{-f} d\mu \leqslant \left(\int e^{-f} d\mu\right)^{-1}.$$

□

Twierdzenie 7.16. *Jeśli μ jest rozkładem jednostajnym na $\{a,b\}^n$ (lub ogólniej dowolnym rozkładem produktowym o nośniku w $[a,b]^n$), zaś A jest wypukłym podzbiorem $[a,b]^n$, to*

$$\int \exp\left(\frac{1}{4(b-a)^2}\operatorname{dist}(x,A)^2\right) d\mu \leqslant \frac{1}{\mu(A)}.$$

W szczególności

$$1 - \mu(A_t) \leqslant \frac{1}{\mu(A)} \exp\left(-\frac{t^2}{4(b-a)^2}\right) \qquad \text{dla } t > 0.$$

Dowód. Na mocy Lematu 7.15 i tensoryzacji wiemy, że μ spełnia wypukłą własność splotu infimum z funkcją kosztu $\varphi(x) = \frac{1}{4(b-a)^2}|x|^2$. Stosujemy własność (τ) do funkcji $f = 0$ na A i $f = \infty$ poza A i dostajemy

$$1 \geqslant \int e^{-f} d\mu \int e^{f\square\varphi} d\mu = \mu(A) \int \exp\left(\frac{1}{4(b-a)^2}\operatorname{dist}(x,A)^2\right) d\mu.$$

□

Uwaga 7.17. Twierdzenie powyższe jest nieprawdziwe bez założenia wypukłości A. Weźmy bowiem za μ_n rozkład jednostajny na $\{-1,1\}^n$, oraz

$$A = \left\{ x \in \{-1,1\}^n \colon \ \sum_{i=1}^n x_i \leqslant 0 \right\}.$$

Wówczas $\mu_n(A) \geqslant 1/2$ oraz korzystając z tego, że $|a-b| \leqslant \frac{1}{4}|a-b|^2$ dla $a, b = \pm 1$ dostajemy

$$A_t \cap \{-1,1\}^n \subset \left\{ x \in \{-1,1\}^n \colon \ \sum_{i=1}^n x_i \leqslant \frac{t^2}{4} \right\}.$$

Na mocy centralnego twierdzenia granicznego $\limsup_n \mu_n(A_{tn^{1/4}}) = \Phi(t^2/4) < 1$.

8 Nierówności transportowe

8.1 Koszt optymalnego transportu

By zdefiniować koszt transportu miar będziemy potrzebowali kilku definicji.

Definicja 8.1. Przez $\mathcal{P}(\mathbb{X})$ będziemy oznaczać rodzinę miar probabilistycznych na przestrzeni mierzalnej $\mathbb{X}$. Dla $\mu, \nu \in \mathcal{P}(\mathbb{X})$ przez $\Pi(\mu,\nu)$ będziemy oznaczali zbiór wszystkich miar probabilistycznych π na $\mathbb{X} \times \mathbb{X}$ takich, że μ i ν są miarami brzegowymi π, czyli $\pi(A \times \mathbb{X}) = \mu(A)$ i $\pi(\mathbb{X} \times A) = \nu(A)$ dla dowolnego zbioru mierzalnego $A \subset \mathbb{X}$.

Uwaga 8.2. Zbiór $\Pi(\mu,\nu)$ jest niepusty, gdyż zawiera miarę produktową $\mu \otimes \nu$. Zauważmy też, że jeśli T transportuje μ na ν oraz X ma rozkład μ, to rozkład zmiennej (X, TX) należy do $\Pi(\mu,\nu)$.

Definicja 8.3. Załóżmy, że $c\colon \mathbb{X} \times \mathbb{X} \to [0,\infty]$ jest funkcją mierzalną. Dla $\mu, \nu \in \mathcal{P}(\mathbb{X})$ definiujemy *optymalny koszt transportu miary μ na ν z funkcją kosztu c* wzorem

$$T_c(\mu,\nu) := \inf \left\{ \int_{\mathbb{X}\times\mathbb{X}} c(x,y) d\pi(x,y) \colon \pi \in \Pi(\mu,\nu) \right\}.$$

W przypadku, gdy $(\mathbb{X}, d)$ jest przestrzenią metryczną, a $c(x,y) = d^p(x,y)$ będziemy pisać T_p zamiast T_c. Określamy też *odległość Wassersteina* miar $\mu, \nu \in \mathcal{P}(\mathbb{X})$ jako

$$W_p(\mu,\nu) := T_p(\mu,\nu)^{1/p} = \inf \left\{ \left(\int_{\mathbb{X}\times\mathbb{X}} d^p(x,y) d\pi(x,y) \right)^{1/p} : \pi \in \Pi(\mu,\nu) \right\}, \quad 1 \leqslant p < \infty,$$

$$W_p(\mu,\nu) := T_p(\mu,\nu) = \inf \left\{ \int_{\mathbb{X}\times\mathbb{X}} d^p(x,y) d\pi(x,y) \colon \pi \in \Pi(\mu,\nu) \right\}, \quad p \in (0,1].$$

Uwaga 8.4. Można udowodnić, że jeśli $\mathbb{X}$ jest przestrzenią polską, to W_p jest metryką na przestrzeni miar probabilistycznych μ na $\mathbb{X}$ takich, że $\int_{\mathbb{X}} d(x,x_0)^p d\mu(x) < \infty$ dla pewnego (równoważnie każdego) $x_0 \in \mathbb{X}$.

Uwaga 8.5. Równoważnie możemy zdefiniować

$$T_c(\mu,\nu) = \inf\{\mathbf{E}c(X,Y)\colon\ X\sim\mu, Y\sim\nu\}.$$

Uwaga 8.6. Zauważmy, że

$$T_c(\mu,\nu) \leqslant \inf\{\mathbf{E}_\mu c(x,Tx)\colon\ T \text{ transportuje } \mu \text{ na } \nu\}.$$

W wielu przypadkach można udowodnić, że w powyższej nierówności zachodzi równość, ale nie jest tak zawsze – np. gdy μ ma atomy, a ν jest bezatomowa, to nie istnieje transport μ na ν.

Definicja 8.7. Jeśli $(\mathbb{X},d)$ jest przestrzenią metryczną, to określamy odległość Monge'a-Kantorowicza miar $\mu,\nu\in\mathcal{P}(\mathbb{X})$ wzorem

$$W_1^{\mathrm{Lip}}(\mu,\nu) := \sup\left\{\left|\int f d\mu - \int f d\nu\right| : f\colon \mathbb{X}\to\mathbb{R} \text{ 1-Lipschitzowska, ograniczona}\right\}.$$

Fakt 8.8. *Dla dowolnej przestrzeni metrycznej $(\mathbb{X},d)$ zachodzi*

$$W_1^{\mathrm{Lip}}(\mu,\nu) \leqslant W_1(\mu,\nu) \quad \textit{dla } \mu,\nu\in\mathcal{P}(\mathbb{X}).$$

Dowód. Zauważmy, że dla dowolnego $\pi\in\Pi(x,y)$ i f 1-Lipschitzowskiego mamy

$$\left|\int f d\mu - \int f d\nu\right| = \left|\int (f(x)-f(y)) d\pi(x,y)\right| \leqslant \int |f(x)-f(y)| d\pi(x,y) \leqslant \int d(x,y) d\pi(x,y).$$

Biorąc supremum po f i infimum po π dostajemy tezę. □

Przy dodatkowym założeniu ośrodkowości odległości W_1^{Lip} i W_1 się pokrywają.

Twierdzenie 8.9 (Dualność Monge'a-Kantorowicza-Rubinsteina)**.** *Załóżmy, że $(\mathbb{X},d)$ jest ośrodkową przestrzenią metryczną. Wówczas*

$$W_1(\mu,\nu) = W_1^{\mathrm{Lip}}(\mu,\nu) \quad \textit{dla } \mu,\nu\in\mathcal{P}(\mathbb{X}).$$

8.2 Względna entropia

Definicja 8.10. Niech μ, ν będą dwiema miarami probabilistycznymi na $\mathbb{X}$. Określamy *entropię miary ν względem miary μ* wzorem

$$H(\nu|\mu) := \begin{cases} \mathrm{Ent}_\mu \frac{d\nu}{d\mu} = \mathbf{E}_\nu \log(\frac{d\nu}{d\mu}), & \text{jeśli } \nu\ll\mu, \\ +\infty & \text{w przeciwnym przypadku.} \end{cases}$$

Lemat 8.11 (Zasada wariacyjna Gibbsa)**.** *Dla dowolnej ograniczonej z góry funkcji mierzalnej f,*

$$\log \mathbf{E}_\mu e^f = \sup_\nu\{\mathbf{E}_\nu f - H(\nu|\mu)\}$$

Dowód. Określmy miarę $\tilde{\mu}$ wzorem

$$d\tilde{\mu} = \frac{e^f}{\mathbf{E}_\mu e^f} d\mu.$$

Wówczas dla dowolnej miary probabilistycznej $\nu \ll \mu$,

$$\begin{aligned}\mathbf{E}_\nu f - H(\nu|\mu) &= \mathbf{E}_\nu f - \mathbf{E}_\nu \log(\frac{d\nu}{d\mu}) = \mathbf{E}_\nu f - \mathbf{E}_\nu \log(\frac{d\tilde{\mu}}{d\mu}) - \mathbf{E}_\nu \log(\frac{d\nu}{d\tilde{\mu}}) \\ &= \log(\mathbf{E}_\mu e^f) - H(\nu|\tilde{\mu}).\end{aligned}$$

Wystarczy zauważyć, że $H(\nu|\tilde{\mu}) \geqslant 0$ i $H(\tilde{\mu}|\tilde{\mu}) = 0$. □

Twierdzenie 8.12 (Bobkow-Goetze)**.** *Niech μ będzie miarą probabilistyczną na przestrzeni metrycznej $(\mathbb{X}, d)$ i $\alpha > 0$. Wówczas n.w.s.r.*
i) $W_1^{\mathrm{Lip}}(\nu, \mu) \leqslant \sqrt{2\alpha H(\nu|\mu)}$ dla dowolnej miary probabilistycznej ν,
ii) dla dowolnej funkcji 1-Lipschitzowskiej ograniczonej f,

$$\mathbf{E}_\mu e^{\lambda(f - \mathbf{E}_\mu f)} \leqslant e^{\alpha\lambda^2/2} \text{ dla } \lambda \in \mathbb{R}.$$

Dowód. Zamieniając f na $-f$ widzimy, że ii) wystarczy dowodzić dla $\lambda \geqslant 0$. Zasada wariacyjna Gibbsa pokazuje, że warunek ii) jest równoważny

$$\begin{aligned}0 &\geqslant \sup_{\lambda \geqslant 0} \sup_f \sup_\nu \left\{ \lambda(\mathbf{E}_\nu f - \mathbf{E}_\mu f) - H(\nu|\mu) - \frac{\alpha\lambda^2}{2} \right\} \\ &= \sup_\nu \sup_{\lambda \geqslant 0} \sup_f \left\{ \lambda(\mathbf{E}_\nu f - \mathbf{E}_\mu f) - H(\nu|\mu) - \frac{\alpha\lambda^2}{2} \right\} \\ &= \sup_\nu \sup_{\lambda \geqslant 0} \left\{ \lambda W_1^{\mathrm{Lip}}(\mu, \nu) - H(\nu|\mu) - \frac{\alpha\lambda^2}{2} \right\} = \sup_\nu \left\{ \frac{W_1^{\mathrm{Lip}}(\mu, \nu)^2}{2\alpha} - H(\nu|\mu) \right\},\end{aligned}$$

co jest oczywiście równoważne warunkowi i). □

Uwaga 8.13. Logarytmiczna nierówność Sobolewa ze stałą C implikuje zachodzenie warunku ii) z $\alpha = C$ (zob. dowód Twierdzenia 6.6). W szczególności miara gaussowska γ_n spełnia warunki twierdzenia Bobkowa-Goetzego z $\alpha = 1$.

8.3 Tensoryzacja nierówności transportowych

Definicja 8.14. Powiemy, że miara probabilistyczna μ na X *spełnia nierówność T_p ze stałą α*, jeśli

$$T_p(\nu, \mu) \leqslant (2\alpha H(\nu|\mu))^{p/2} \quad \text{dla } \nu \in \mathcal{P}(\mathbb{X}).$$

Uwaga 8.15. Dla $p > q$ i $\mu, \nu \in \mathcal{P}(\mathbb{X})$ zachodzi $T_p(\mu, \nu)^{1/p} \geqslant T_q(\mu, \nu)^{1/q}$, zatem nierówność T_p pociąga za sobą nierówność T_q dla $q < p$.

Naturalne jest pytanie czy nierówności T_p się tensoryzują. Wykorzystamy do tego ogólne twierdzenie.

Twierdzenie 8.16 (Marton)**.** *Załóżmy, że funkcja* $\varphi\colon [0,\infty) \to [0,\infty)$ *jest wypukła oraz dla* $i = 1,\dots,n$, c_i *są nieujemnymi mierzalnymi funkcjami na* $\mathbb{X}_i \times \mathbb{X}_i$, *zaś* $\mu_i \in \mathcal{P}(\mathbb{X}_i)$ *spełniają warunek*

$$\inf_{\pi\in\Pi(\mu_i,\nu)} \varphi(\mathbf{E}_\pi c_i(x,y)) \leqslant H(\nu|\mu_i) \quad \textit{dla wszystkich } \nu \in \mathcal{P}(\mathbb{X}_i).$$

Wówczas dla wszystkich miar probabilistycznych ν *na* $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$ *zachodzi*

$$\inf_{\pi\in\Pi(\mu_1\otimes\cdots\otimes\mu_n,\nu)} \sum_{i=1}^n \varphi(\mathbf{E}_\pi c_i(x_i,y_i)) \leqslant H(\nu|\mu_1\otimes\cdots\otimes\mu_n).$$

Do dowodu twierdzenia 8.16 przydatny będzie lemat o dekompozycji miary. Nie podamy jego dowodu, gdyż wykorzystamy go tylko dla miar z gęstością jak w Przykładzie 2 poniżej, ale ogólne sformułowanie przydaje się, gdy np. chcemy dowieść, że odległość Wassersteina jest metryką.

Twierdzenie 8.17. *Załóżmy, że* $\mathbb{X}$ *i* $\mathbb{Y}$ *są przestrzeniami polskimi oraz* $\pi \in \mathcal{P}(\mathbb{X}\times\mathbb{Y})$. *Niech* π_1 *będzie rozkładem brzegowym* π. *Wówczas istnieje rodzina miar probabilistycznych* $(\pi_{2,x})_{x\in\mathbb{X}}$ *taka, że*
i) dla dowolnego zbioru borelowskiego $A \subset \mathbb{X}\times\mathbb{Y}$ *przekształcenie* $x \to (\delta_x \otimes \pi_{2,x})(A)$ *jest mierzalne,*
ii) $\pi = \int_{\mathbb{X}} \delta_x \otimes \pi_{2,x} d\pi_1(x)$.

Przykład 1. Jeśli rozkład brzegowy π_1 jest miarą dyskretną $\sum_i p_i\delta_{x_i}$, to możemy przyjąć

$$\pi_{2,x}(B) = \begin{cases} \frac{\pi(\{x\}\times B)}{\pi_1(\{x\})} & \text{jeśli } \pi_1(\{x\}) > 0 \\ 0 & \text{jeśli } \pi_1(\{x\}) = 0. \end{cases}$$

Przykład 2. Jeśli π ma gęstość g względem pewnej miary produktowej $\mu_1\otimes\mu_2$, to definiujemy $d\pi_{2,x} = g_{2,x}d\mu_2$, gdzie

$$g_{2,x}(y) = \begin{cases} \frac{g(x,y)}{\int_{\mathbb{X}} g(x,y)d\mu_2(y)} & \text{jeśli } \int_{\mathbb{X}} g(x,y)d\mu_2(y) > 0 \\ 0 & \text{jeśli } \int_{\mathbb{X}} g(x,y)d\mu_2(y) = 0. \end{cases}$$

Dowód twierdzenia 8.16. Twierdzenie udowodnimy przez indukcję po n. Dla $n = 1$ teza jest oczywista. Załóżmy zatem, że $n \geqslant 2$ i teza indukcyjna zachodzi dla $n-1$, pokażemy, że jest też prawdą dla n. Dla uproszczenia notacji przyjmijmy

$$\tilde{\mathbb{X}} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_{n-1}, \quad \tilde{\mu} = \mu_1 \otimes \cdots \otimes \mu_{n-1},$$

ponadto dla $x \in \mathbb{X}$ będziemy pisać $x = (\tilde{x}, x_n)$, gdzie $\tilde{x} \in \tilde{\mathbb{X}}$.

Ustalmy miarę probabilistyczną ν na $\mathbb{X} = \tilde{\mathbb{X}} \times \mathbb{X}_n$ taką, że $H(\nu|\tilde{\mu} \otimes \mu_n) < \infty$. Wówczas jak wiemy z lematu o dekompozycji (zob. Przykład 2 powyżej)

$$\nu = \int_{\tilde{\mathbb{X}}} \delta_{\tilde{x}} \otimes \nu_{\tilde{x}} d\tilde{\nu}(\tilde{x}),$$

gdzie $\tilde{\nu}$ oznacza brzegowy rozkład ν na $\tilde{\mathbb{X}}$. Łatwo sprawdzić (zob. Przykład 2), że

$$H(\nu|\tilde{\mu} \otimes \mu_n) = H(\tilde{\nu}|\tilde{\mu}) + \int_{\tilde{\mathbb{X}}} H(\nu_{\tilde{x}}|\mu_n) d\tilde{\nu}(\tilde{x}).$$

Ustalmy $\varepsilon > 0$. Założenie indukcyjne implikuje, że istnieje miara probabilistyczna $\tilde{\pi} \in \Pi(\tilde{\mu}, \tilde{\nu})$ taka, że

$$\sum_{i=1}^{n-1} \varphi(\mathbf{E}_{\tilde{\pi}} c_i(x_i, y_i)) \leqslant H(\tilde{\nu}|\tilde{\mu}) + \varepsilon.$$

Z założenia twierdzenia wynika natomiast, że dla $\tilde{x} \in \tilde{\mathbb{X}}$ istnieje miara $\pi_{\tilde{x}} \in \Pi(\mu_n, \nu_{\tilde{x}})$ dla której

$$\varphi(\mathbf{E}_{\pi_{\tilde{x}}} c_n(x, y)) \leqslant H(\nu_{\tilde{x}}|\mu_n) + \varepsilon.$$

Określmy π jako miarę na $\mathbb{X} \times \mathbb{X}$, które możemy w naturalny sposób utożsamiać z $\tilde{\mathbb{X}} \times \tilde{\mathbb{X}} \times \mathbb{X}_n \times \mathbb{X}_n$, wzorem

$$\pi := \int_{\tilde{\mathbb{X}} \times \tilde{\mathbb{X}}} \delta_{\tilde{x}, \tilde{y}} \otimes \pi_{\tilde{y}} d\pi(\tilde{x}, \tilde{y}).$$

Wówczas $\pi \in \Pi(\mu_1 \otimes \cdots \otimes \mu_n, \nu)$,

$$\sum_{i=1}^{n-1} \varphi(\mathbf{E}_{\pi} c_i(x_i, y_i)) = \sum_{i=1}^{n-1} \varphi(\mathbf{E}_{\tilde{\pi}} c_i(x_i, y_i)) \leqslant H(\tilde{\nu}|\tilde{\mu}) + \varepsilon$$

i z wypukłości φ

$$\begin{aligned} \varphi(\mathbf{E}_{\pi} c_n(x_n, y_n)) &= \varphi(\mathbf{E}_{\tilde{\pi}} \mathbf{E}_{\pi_{\tilde{y}}} c_n(x_n, y_n)) \leqslant \mathbf{E}_{\tilde{\pi}} \varphi(\mathbf{E}_{\pi_{\tilde{y}}} c_n(x_n, y_n)) \ \leqslant \mathbf{E}_{\tilde{\pi}} H(\nu_{\tilde{y}}|\mu_n) + \varepsilon \\ &= \int_{\tilde{\mathbb{X}}} H(\nu_{\tilde{y}}|\mu_n) d\tilde{\nu}(\tilde{y}) + \varepsilon. \end{aligned}$$

Zatem

$$\sum_{i=1}^{n} \varphi(\mathbf{E}_{\pi} c_i(x_i, y_i)) \leqslant H(\nu|\mu) + 2\varepsilon$$

i z dowolności $\varepsilon > 0$ otrzymujemy dowód kroku indukcyjnego. □

Wniosek 8.18. *Załóżmy, że miary probabilistyczne μ_i na (X_i, d_i) spełniają nierówność T_1 ze stałymi α_i, $1 \leqslant i \leqslant n$. Na $\mathbb{X} = \mathbb{X}_1 \times \ldots \times \mathbb{X}_n$ określmy ważoną l_1-metrykę $d_c(x, y) := \sum_{i=1}^n c_i d_i(x_i, y_i)$. Wówczas miara $\mu_1 \otimes \cdots \otimes \mu_n$ spełnia nierówność T_1 na $(\mathbb{X}, d_c)$ ze stałą $(\sum_{i=1}^n c_i^2)^{1/2} \max_i \alpha_i$.*

Dowód. Niech ν i ρ będą miarami probabilistycznymi na $(\mathbb{X}, d_c)$. Wówczas

$$T_1(\nu,\rho) = \inf_{\pi\in\Pi(\nu,\rho)} \sum_{i=1}^n c_i \mathbf{E}_\pi d_i(x_i,y_i) \leqslant \sqrt{\sum_{i=1}^n c_i^2} \inf_{\pi\in\Pi(\nu,\rho)} \left(\sum_{i=1}^n (\mathbf{E}_\pi d_i(x_i,y_i))^2\right)^{1/2}.$$

Teza wniosku wynika teraz łatwo z Twierdzenia 8.16 z $c_i(x_i,y_i) = d_i(x_i,y_i)$. $\varphi(x) = (x/2\alpha)^2$, $\alpha := \max_i \alpha_i$. □

Innym wnioskiem z Twierdzenia Marton jest tensoryzowalność nierówności T_2 względem metryki l_2.

Wniosek 8.19. *Załóżmy, że miary probabilistyczne μ_i na $(\mathbb{X}_i, d_i)$ spełniają nierówność T_2 ze stałymi α_i. Na $\mathbb{X} = \mathbb{X}_1 \times \ldots \times \mathbb{X}_n$ określmy l_2-metrykę $d(x,y) := (\sum_{i=1}^n d_i(x_i,y_i)^2)^{1/2}$. Wówczas miara $\mu_1 \otimes \cdots \otimes \mu_n$ spełnia nierówność T_2 ze stałą $\max_i \alpha_i$ na $(\mathbb{X}, d)$.*

Dowód. Stosujemy Twierdzenie 8.16 z $\varphi(x) := \frac{1}{2\alpha}x$, $\alpha = \max_i \alpha_i$ oraz $c_i(x_i,y_i) := d_i^2(x_i,y_i)$. □

8.4 Nierówność T_2 Talagranda a bezwymiarowa koncentracja

Wniosek 8.20. *Załóżmy, że miara μ spełnia nierówność T_2 ze stałą α na przestrzeni metrycznej $(\mathbb{X}, d)$. Wówczas dla dowolnej funkcji 1-Lipschitzowskiej na $\mathbb{X}^n$ z l_2-metryką $d_n(x,y) := (\sum_{i=1}^n d(x_i,y_i)^2)^{1/2}$ zachodzi*

$$\mu^n(\{x \in \mathbb{X}^n \colon f(x) - \mathbf{E}_{\mu^n} f \geqslant t\}) \leqslant e^{-t^2/2\alpha}.$$

W szczególności $\alpha_{\mu^n}(t) \leqslant \exp(-t^2/8\alpha)$.

Dowód. Z Wniosku 8.19 wynika, że μ^n spełnia nierówność T_2 ze stałą α, zatem dla dowolnej miary probabilistycznej ν na X^n zachodzi

$$W_1^{\mathrm{Lip}}(\mu^n,\nu) \leqslant W_1(\mu^n,\nu) \leqslant W_2(\mu^n,\nu) \leqslant \sqrt{2\alpha H(\nu|\mu^n)}$$

i teza łatwo wynika z Twierdzenia 8.12. □

Okazuje się, że nierówność T_2 jest równoważna bezwymiarowej koncentracji.

Twierdzenie 8.21 (Gozlan)**.** *Załóżmy, że μ jest miarą probablistyczną na ośrodkowej przestrzeni polskiej $(\mathbb{X}, d)$, zaś d_n są l_2-metrykami na $\mathbb{X}^n$. Wówczas następujące warunki są równoważne:*
i) μ spełnia nierówność T_2 na $(\mathbb{X}, d)$ ze stałą α:

$$W_2(\nu,\mu) \leqslant \sqrt{2\alpha H(\nu|\mu)} \quad \text{dla każdego } \nu \in (\mathbb{X}),$$

ii) dla każdego n miara μ^n spełnia nierówność T_1 na $(\mathbb{X}^n, d_n)$ ze stałą α:

$$W_1(\nu, \mu^n) \leqslant \sqrt{2\alpha H(\nu|\mu^n)} \quad \text{dla każdego } \nu \in \mathcal{P}(\mathbb{X}^n),$$

iii) istnieje stała C taka, że dla każdego n i każdej funckji 1-Lipschitzowskiej f na $(\mathbb{X}^n, d_n)$,

$$\mu^n(\{x \in \mathbb{X}^n\colon\ f(x) - \mathbf{E}_{\mu^n} f \geqslant t\}) \leqslant Ce^{-t^2/2\alpha}.$$

Dowód $i) \Rightarrow ii) \Rightarrow iii)$ dowodzimy jak we Wniosku 8.20. By udowodnić najbardziej zaskakującą implikację $iii) \Rightarrow i)$ wykorzystamy twierdzenie o wielkich odchyleniach Sanowa.

Twierdzenie 8.22 (Sanow)**.** *Niech $X_1, X_2, \ldots$ będą niezależnymi zmiennymi losowymi o wartościach w przestrzeni polskiej $\mathbb{X}$ i jednakowym rozkładzie μ. Wówczas dla dowolnego zbioru otwartego G w przestrzeni miar probabilistycznych na $\mathbb{X}$ z topologią słabej zbieżności zachodzi*

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in G\right) \geqslant -\inf_{\nu\in G} H(\nu|\mu). \tag{14}$$

Uwaga 8.23. Twierdzenie 8.22 to tak naprawdę tylko połowa twierdzenia Sanowa dotycząca szacowania wielkich odchyleń dla miar empirycznych z dołu. Druga część mówi, że dla dowolnego zbioru zwartego F w przestrzeni miar probabilistycznych na $\mathbb{X}$ z topologią słabej zbieżności mamy

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in F\right) \leqslant -\inf_{\nu\in F} H(\nu|\mu).$$

Dowód Twierdzenia 8.22. Ustalmy $\nu \in U$ takie, że $H(\nu|\mu) < \infty$ (jeśli takie ν nie istnieje, to infimum po lewej stronie (14) jest równe $+\infty$ i nierówność jest oczywista). Niech $g = \frac{d\nu}{d\mu}$ oraz $Y_1, Y_2, \ldots$ będą niezależnymi zmiennymi o rozkładzie ν. Wówczas $g(Y_i) > 0$ p.n. oraz dla dowolnej funkcji mierzalnej f na $\mathbb{X}^n$,

$$\mathbf{E} f(Y_1, \ldots, Y_n) = \mathbf{E}(f(X_1, \ldots, X_n) \prod_{k=1}^{n} g(X_k)).$$

Mamy

$$\begin{aligned}
\mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in G\right) &\geqslant \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in G,\ \prod_{k=1}^{n} g(X_k) > 0\right) \\
&= \mathbf{E}\left(\mathbb{1}_{\{\frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k} \in G\}} \prod_{k=1}^{n} g(Y_k)^{-1}\right) \\
&\geqslant e^{-n(\mathbf{E}_\nu \log g + \varepsilon)} \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k} \in G, \frac{1}{n}\sum_{k=1}^{n} \log g(Y_k) \leqslant \mathbf{E}_\nu \log g + \varepsilon\right).
\end{aligned}$$

Mocne prawo wielkich liczb implikuje, że z prawdopodobieństwem 1 przy $n \to \infty$ zachodzi $\frac{1}{n}\sum_{k=1}^{n} \log g(Y_k) \to \mathbf{E}_\nu \log g$ oraz $\frac{1}{n}\sum_{k=1}^{n} \delta_{Y_k} \to \nu$ słabo. Stąd z otwartości G otrzymujemy, że

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in G\right) \geqslant -\mathbf{E}_\nu \log g - \varepsilon = -H(\nu|\mu) - \varepsilon.$$

Przechodząc z ε do 0 i biorąc supremum prawej strony ostatniej nierówności po $\nu \in G$ dostajemy tezę. □

Zanim udowodnimy twierdzenie Gozlana, wykażemy kilka faktów dotyczących metryki Wassersteina. We wszystkich trzech faktach zakładamy, że μ jest rozkładem probabilistycznym na przestrzeni polskiej $\mathbb{X}$ oraz $1 \leqslant p < \infty$.

Fakt 8.24. *Funkcja $\nu \mapsto W_p(\nu, \mu)$ jest półciągła z dołu na $\mathcal{P}(\mathbb{X})$, tzn. jeśli ν_n zbiega słabo do μ, to*

$$\liminf_{n\to\infty} W_p(\nu_n, \mu) \geqslant W_p(\nu, \mu).$$

Dowód. Niech $\pi_n \in \Pi(\nu_n, \mu)$ będą takie, że

$$W_p(\nu_n, \nu) \geqslant (\mathbf{E}_{\pi_n} d(x,y)^p)^{1/p} - \frac{1}{n}.$$

Pokażemy najpierw, że ciąg (π_n) jest ciasny w $\mathcal{P}(\mathbb{X}\times\mathbb{X})$. Dla $\varepsilon > 0$ z ciasności ciągu (ν_n) możemy znaleźć zbiór zwarty $K_1 \subset \mathbb{X}$ taki, że $\nu_n(K_1) \geqslant 1 - \varepsilon/2$ dla wszystkich n. Istnieje też zbiór zwarty $K_2 \subset \mathbb{X}$ taki, że $\mu(K_2) \geqslant 1 - \varepsilon/2$. Ponieważ ν_n i μ to rozkłady brzegowe π_n, więc $1 - \pi_n(K_1 \times K_2) \leqslant 1 - \nu_n(K_1) + 1 - \mu(K_2) \leqslant \varepsilon$.

Używając ciasności (π_n) możemy wybrać podciąg ciąg n_k taki, że $\pi_{n_k} \to \pi$ słabo w $\mathcal{P}(\mathbb{X}\times\mathbb{X})$ i $\liminf_n W_p(\nu_n, \mu) = \lim_k W_p(\nu_{n_k}, \mu)$. Łatwo sprawdzamy, że $\pi \in \Pi(\nu, \mu)$ oraz dla dowolnego $a < \infty$,

$$\begin{aligned}\liminf_n W_p(\nu_n, \mu) &\geqslant \liminf_k (\mathbf{E}_{\pi_{n_k}} d(x,y)^p)^{1/p} \geqslant \liminf_k (\mathbf{E}_{\pi_{n_k}} \min\{a, d^p(x,y)\})^{1/p} \\ &= (\mathbf{E}_\pi \min\{a, d^p(x,y)\})^{1/p}.\end{aligned}$$

Z dowolności $a > 0$ mamy

$$W_p(\nu, \mu) \leqslant (\mathbf{E}_\pi d^p(x,y))^{1/p} \leqslant \liminf_n W_p(\nu_n, \mu).$$

□

Fakt 8.25. *Funkcja $g_n(x_1, \ldots, x_n) := W_p(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, \mu)$ jest $n^{-1/p}$-lipschitzowska na $\mathbb{X}^n$ z l_p-metryką $d_p(x,y) := (\sum_{k=1}^{n} d(x_k, y_k)^p)^{1/p}$.*

Dowód. Zauważmy wpierw, że każde $\pi \in \Pi(\frac{1}{n}\sum_{k=1}^n \delta_{x_k}, \mu)$ jest postaci $\pi = \frac{1}{n}\sum_{k=1}^n \delta_{x_k} \otimes \mu_k$ dla $\mu_1, \dots, \mu_n \in \mathcal{P}(\mathbb{X})$ takich, że $\frac{1}{n}\sum_{k=1}^n \mu_k = \mu$. Stąd dla $x, y \in \mathbb{X}^n$ mamy

$$\begin{aligned}
&W_p\left(\frac{1}{n}\sum_{k=1}^n \delta_{x_k}, \mu\right) - W_p\left(\frac{1}{n}\sum_{k=1}^n \delta_{y_k}, \mu\right) \\
&= \inf_{\frac{1}{n}\sum_{k=1}^n \mu_k = \mu} \left(\frac{1}{n}\sum_{k=1}^n \int d(x_i, z)^p d\mu_i(z)\right)^{1/p} - \inf_{\frac{1}{n}\sum_{k=1}^n \mu_k = \mu} \left(\frac{1}{n}\sum_{k=1}^n \int d(y_i, z)^p d\mu_i(z)\right)^{1/p} \\
&\leqslant \sup_{\frac{1}{n}\sum_{k=1}^n \mu_k = \mu} \left[\left(\frac{1}{n}\sum_{k=1}^n \int d(x_i, z)^p d\mu_i(z)\right)^{1/p} - \left(\frac{1}{n}\sum_{k=1}^n \int d(y_i, z)^p d\mu_i(z)\right)^{1/p}\right] \\
&\leqslant \sup_{\frac{1}{n}\sum_{k=1}^n \mu_k = \mu} \left(\frac{1}{n}\sum_{k=1}^n \int |d(x_i, z) - d(y_i, z)|^p d\mu_i(z)\right)^{1/p} \leqslant n^{-1/p}\left(\sum_{k=1}^n d(x_k, y_k)^p\right)^{1/p}.
\end{aligned}$$

□

Fakt 8.26. *Jeśli $X_1, X_2, \dots$ są niezależne o rozkładzie μ, oraz $\mathbf{E}_\mu d(x, x_0)^{p+\varepsilon} < \infty$ dla pewnego $x_0 \in X$ i $\varepsilon > 0$, to $\mathbf{E}W_p(\frac{1}{n}\sum_{k=1}^n \delta_{X_k}, \mu)^p = 0$.*

Dowód. Ustalmy $a > 0$. Mamy

$$\begin{aligned}
W_p(\nu, \mu)^p &= \inf_{\pi \in \Pi(\nu,\mu)} \left(\mathbf{E}_\pi d^p(x, y) \mathbb{1}_{\{d(x,y) \leqslant a\}} + \mathbf{E}_\pi d^p(x, y) \mathbb{1}_{\{d(x,y) > a\}}\right) \\
&\leqslant \inf_{\pi \in \Pi(\nu,\mu)} \left(a^{p-1} \mathbf{E}_\pi \min\{d(x, y), a\} + \frac{\mathbf{E}_\pi (d(x_0, x) + d(x_0, y))^{p+\varepsilon}}{a^\varepsilon}\right). \\
&\leqslant a^{p-1} \inf_{\pi \in \Pi(\nu,\mu)} \mathbf{E}_\pi \min\{d(x, y), a\} + 2^{p+\varepsilon} \frac{\mathbf{E}_\nu d(x_0, x)^{p+\varepsilon} + \mathbf{E}_\mu d(x_0, y)^{p+\varepsilon}}{a^\varepsilon}.
\end{aligned}$$

Z twierdzenia Skorochoda wynika, że jeśli $\nu_n \to \mu$ słabo, to istnieją zmienne losowe $Y_n \sim \nu_n$ i $Y \sim \mu$ takie, że $Y_n \to Y$ p.n, w konsekwencji z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej dostajemy

$$\nu_n \to \mu \text{ słabo w } \mathcal{P}(\mathbb{X}) \quad \Rightarrow \quad \inf_{\pi \in \Pi(\nu_n, \mu)} \mathbf{E}_\pi \min\{d(x, y), a\} \to 0.$$

Ponieważ $\frac{1}{n}\sum_{k=1}^n \delta_{X_k} \to \mu$ słabo z prawdopodobieństwem 1 przy $n \to \infty$, więc ponownie używając twierdzenia Lebesgue'a o zbieżności zmajoryzowanej dostajemy

$$\mathbf{E} \inf_{\pi \in \Pi(\frac{1}{n}\sum_{k=1}^n \delta_{X_k}, \mu)} \mathbf{E}_\pi \min\{d(x, y), a\} \to 0.$$

Stąd dla dowolnego $a > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \mathbf{E}W_p\left(\frac{1}{n}\sum_{k=1}^n \delta_{X_k}, \mu\right)^p &\leqslant 2^{p+\varepsilon} a^{-\varepsilon} \left(\mathbf{E}\frac{1}{n}\sum_{k=1}^n d(x_0, X_k)^{p+\varepsilon} + \mathbf{E}_\mu d(x_0, y)^{p+\varepsilon}\right) \\
&\leqslant 2^{p+\varepsilon+1} a^{-\varepsilon} \mathbf{E}_\mu d(x_0, y)^{p+\varepsilon}
\end{aligned}$$

i biorąc $a \to \infty$ dostajemy tezę.

□

Dowód Twierdzenia 8.21. i)⇒ii). Stosujemy Wniosek 8.19 i to, że $W_1 \leqslant W_2$.
ii)⇒iii). Wystarczy wykorzystać to, że $W_1^{\mathrm{Lip}} \leqslant W_1$ i Twierdzenie 8.12.
iii)⇒i). Określmy

$$g_n(x_1, \ldots, x_n) := W_2\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{x_k}, \mu\right).$$

Fakt 8.24 implikuje, że zbiór

$$G_t := \{\nu \in \mathcal{P}(\mathbb{X}) \colon\ W_2(\nu, \mu) > t\}$$

jest otwarty. Zatem z twierdzenia Sanowa

$$-\inf_{\nu \in G_t} H(\nu|\mu) \leqslant \liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(g_n(X_1, \ldots, X_n) > t\right).$$

Z założenia iii) i $n^{-1/2}$-lipschitzowskości g_n (Fakt 8.25) dostajemy

$$\mathbf{P}\left(g_n(X_1, \ldots, X_n) > t\right) \leqslant C \exp\left(-\frac{n}{2\alpha}(t - \mathbf{E}g_n(X_1, \ldots, X_n))_+^2\right).$$

Stąd

$$-\inf_{\nu \in G_t} H(\nu|\mu) \leqslant -\limsup_{n\to\infty} \frac{(t - \mathbf{E}g_n(X_1, \ldots, X_n))_+^2}{2\alpha} = -\frac{t^2}{2\alpha},$$

gdzie ostatnia nierówność wynika z Faktu 8.26 (warunek iii) z $n = 1$ i 1-Lipschitzowskość metryki implikują, że $\mathbf{E}_\mu d(x_0, x)^p < \infty$ dla dowolnego $p < \infty$). Otrzymana nierówność jest równoważna

$$\sqrt{2\alpha H(\nu|\mu)} \geqslant t, \quad \text{jeśli } W_2(\mu, \nu) > t,$$

skąd łatwo wynika nierówność T_2. □

9 Aproksymacja przez otoczkę wypukłą

9.1 Definicje

W tej części będziemy zakładać, że przestrzeń $\mathbb{X}$ ma strukturę produktową, tzn. $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$. Określmy metrykę na $\mathbb{X}$ wzorem

$$d_a(x, y) = \sum_{i=1}^{n} a_i \mathbb{1}_{\{x_i \neq y_i\}}.$$

Z Wniosku 4.5 wynika, że dla $|a| = 1$, $\alpha_{\mu,X,d_a}(t) \leqslant \exp(-|t|^2/8)$, jednak poszerzenie zbioru w każdej z metryk d_a wygląda nieco inaczej. Celem tego rozdziału jest uzyskanie jednostajnej wersji tego wyniku.

Dla $A \subset \mathbb{X}$ i $x \in \mathbb{X}$ określmy

$$\mathcal{D}_A^c(x) := \sup_{|a|=1} d_a(x, A).$$

Okazuje się, że $\mathcal{D}_A^c(x)$ można zdefiniować w równoważny, nieco bardziej abstrakcyjny sposób.

Dla $A \subset \mathbb{X}$ i $x \in \mathbb{X}$ określmy

$$U_A(x) := \{(\mathbb{1}_{\{x_i \neq y_i\}})_{1 \leqslant i \leqslant n} \colon y \in A\} \subset \{0,1\}^n$$

oraz

$$V_A(x) := \operatorname{conv}\{U_A(x)\} \subset [0,1]^n.$$

Łatwo zauważyć, że $V_A(x)$ jest domkniętym wielościanem wypukłym. Ponadto $0 \in V_A(x)$ wtedy i tylko wtedy gdy $0 \in A$.

Kolejny fakt łączy $V_A(x)$ i $\mathcal{D}_A^c(x)$.

Fakt 9.1. *Dla dowolnego $A \subset \mathbb{X}$ i $x \in \mathbb{X}$,*

$$\operatorname{dist}(0, V_A(x)) = \inf_{y \in V_A(x)} |y| = \mathcal{D}_A^c(x).$$

Dowód. i) $\mathcal{D}_A^c(x) \leqslant \operatorname{dist}(0, V_A(x))$. Niech $z \in V_A(x)$ takie, że $|z| = \operatorname{dist}(0, V_A(x))$. Ustalmy $a \in S^{n-1}$, wtedy

$$\inf_{s \in U_A(x)} \langle a, s\rangle = \inf_{y \in V_A(x)} \langle a, y\rangle \leqslant \langle a, z\rangle \leqslant |z|.$$

Zatem istnieje $y \in A$ takie, że $s = (\mathbb{1}_{\{x_i \neq y_i\}})_i \in U_A(x)$ spełnia $\langle a, s\rangle \leqslant |z|$. Stąd

$$d_a(x, A) \leqslant d_a(x, y) = \sum_{i=1}^n a_i \mathbb{1}_{\{x_i \neq y_i\}} = \langle a, s\rangle \leqslant |z|,$$

czyli $\mathcal{D}_A^c(x) \leqslant |z| = \operatorname{dist}(0, V_A(x))$.

ii) $\mathcal{D}_A^c(x) \geqslant \operatorname{dist}(0, V_A(x))$. Ustalmy $z \in V_A(x)$ taki, że $|z| = \operatorname{dist}(0, V_A(x))$. Jeśli $z = 0$, to nierówność jest oczywista, w przeciwnym przypadku niech $a := z/|z|$. Zauważmy, że dla dowolnego $s \in V_A(x)$ i $\theta \in [0,1]$, $\theta s + (1-\theta)z \in V_A(x)$, zatem

$$|z|^2 \leqslant |\theta s + (1-\theta)z|^2 = |z + \theta(s-z)|^2 = |z|^2 + 2\theta\langle z, s-z\rangle + \theta^2|s|^2.$$

Biorąc $\theta \to 0+$ dostajemy $\langle z, s-z\rangle \geqslant 0$, czyli

$$\langle a, s\rangle = \frac{1}{|z|}\langle z, s\rangle \geqslant \frac{1}{|z|}\langle z, z\rangle = |z|.$$

Stąd

$$\mathcal{D}_A^c(x) \geqslant d_a(x, A) = \inf_{s \in U_A(x)} \langle a, s\rangle \geqslant |z| = \operatorname{dist}(0, V_A(x)).$$

□

9.2 Twierdzenie Talagranda

Twierdzenie 9.2. *Załóżmy, że $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ jest produktową miarą probabilistyczną na $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_n$. Wówczas dla dowolnego niepustego, mierzalnego zbioru A w $\mathbb{X}$,*

$$\int \exp\Big(\frac{(\mathcal{D}_A^c)^2}{4}\Big) d\mu \leqslant \frac{1}{\mu(A)}.$$

W szczególności dla $t > 0$,

$$\mu(\{\mathcal{D}_A^c \geqslant t\}) \leqslant \frac{1}{\mu(A)} e^{-t^2/4}.$$

Dowód. Przeprowadzimy indukcję po n. Dla $n = 1$, mamy $\mathcal{D}_A^c(x) = \mathbb{1}_{X\setminus A}(x)$, więc

$$\int \exp\Big(\frac{(\mathcal{D}_A^c)^2}{4}\Big) d\mu = e^{1/4}(1-\mu(A)) + \mu(A) \leqslant 2(1-\mu(A)) + \mu(A) \leqslant \frac{1}{\mu(A)}.$$

Załóżmy, że $n \geqslant 2$ i teza zachodzi dla $n-1$. Dla uproszczenia notacji przyjmiemy

$$\tilde{\mathbb{X}} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_{n-1}, \quad \tilde{\mu} = \mu_1 \otimes \cdots \otimes \mu_{n-1}$$

oraz dla $x \in \mathbb{X}$ będziemy pisać $x = (\tilde{x}, x_n)$, gdzie $\tilde{x} \in \tilde{\mathbb{X}}$. Ustalmy $A \subset \mathbb{X} = \tilde{\mathbb{X}} \times \mathbb{X}_n$ i przyjmijmy

$$B = \{\tilde{x} \colon \exists y \in \mathbb{X}_n \ x = (\tilde{x}, y) \in A\} \quad \text{oraz} \quad A(y) = \{\tilde{x} \colon x = (\tilde{x}, y) \in A\} \text{ dla } y \in \mathbb{X}_n.$$

Zauważmy, że jeśli $s \in U_{A(x_n)}(x)$, to $(s,0) \in U_A(x)$, a jeśli $t \in U_B(x)$, to $(t,1)$ lub $(t,0)$ należą do $U_A(x)$. Zatem jeśli wybierzemy $s \in V_{A(x_n)}(x)$ oraz $t \in V_B(x)$, to $(s,0) \in U_A(x)$ oraz $(t,b) \in V_A(x)$ dla pewnego $b \in [0,1]$, czyli z wypukłości zbioru $V_A(x)$, $(\theta s + (1-\theta)t, (1-\theta)b) \in V_A(x)$. Stąd z wypukłości funkcji $|x|^2$,

$$\mathcal{D}_A^c(x)^2 \leqslant |\theta s + (1-\theta)t|^2 + |(1-\theta)b|^2 \leqslant \theta|s|^2 + (1-\theta)|t|^2 + (1-\theta)^2,$$

czyli z dowolności wyboru t i s,

$$\mathcal{D}_A^c(x)^2 \leqslant \theta \mathcal{D}_{A(x_n)}^c(\tilde{x})^2 + (1-\theta)\mathcal{D}_B^c(\tilde{x})^2 + (1-\theta)^2.$$

Odcałkowywując i korzystając z nierówności Höldera dostajemy

$$\begin{aligned}
&\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_A^c(\tilde{x}, x_n)^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \\
&\leqslant e^{(1-\theta)^2/4} \Big(\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_{A(x_n)}^c(\tilde{x})^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \Big)^{\theta} \Big(\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_B^c(\tilde{x})^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \Big)^{1-\theta}.
\end{aligned}$$

Zatem na mocy założenia indukcyjnego (zastosowanego do zbiorów $A(x_n)$ i B w $\tilde{\mathbb{X}}$) dostajemy dla dowolnego $\theta \in [0,1]$,

$$\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_A^c(\tilde{x},x_n)^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \leqslant e^{(1-\theta)^2/4}\Big(\frac{1}{\tilde{\mu}(A(x_n))}\Big)^{\theta}\Big(\frac{1}{\tilde{\mu}(B)}\Big)^{1-\theta}. \tag{15}$$

Zauważmy teraz, że

$$\inf_{\theta\in[0,1]} e^{(1-\theta)^2/4}u^{-\theta} \leqslant 2-u \quad \text{ dla } u\in[0,1]. \tag{16}$$

Istotnie dla $u \geqslant e^{-1/2}$ możemy przyjąć $\theta = 1+2\log u$ i po zlogarytmowaniu pozostaje sprawdzić, że $f(u) := \log(2-u)+\log(u)+\log^2(u) \geqslant 0$. Prosty rachunek pokazuje, że dla $u\in[0,1]$, $(uf')' = -2(u-2)^{-2}+2u^{-1} \geqslant 0$, czyli $uf'(u) \leqslant f'(1) = 0$, więc $f(u) \geqslant f(1) = 0$.

Dla $u \leqslant e^{-1/2}$ kładziemy $\theta = 0$ i sprawdzamy (numerycznie lub korzystając z poprzedniego rozumowania dla $u = e^{-1/2}$), że $e^{1/4} \leqslant 2-e^{-1/2} \leqslant 2-u$.

Nierówności (15) oraz (16) z $u = \tilde{\mu}(A(x_n))/\tilde{\mu}(B)$ implikują

$$\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_A^c(\tilde{x},x_n)^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \leqslant \frac{1}{\tilde{\mu}(B)}\Big(2-\frac{\tilde{\mu}(A(x_n))}{\tilde{\mu}(B)}\Big).$$

Zatem

$$\begin{aligned}\int_X \exp\Big(\frac{\mathcal{D}_A^c(\tilde{x},x_n)^2}{4}\Big) d\mu(x) &\leqslant \int_{X_n} \frac{1}{\tilde{\mu}(B)}\Big(2-\frac{\tilde{\mu}(A(x_n))}{\tilde{\mu}(B)}\Big) d\mu_n(x_n)\\ &= \frac{1}{\tilde{\mu}(B)}\Big(2-\frac{\mu(A)}{\tilde{\mu}(B)}\Big) \leqslant \frac{1}{\mu(A)},\end{aligned}$$

gdyż $v(2-v) \leqslant 1$ dla $v\in[0,1]$. □

Literatura

[1] S. Boucheron, G. Lugosi, P. Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013.

[2] J. Jakubowski, R. Sztencel, *Wstęp do teorii prawdopodobieństwa*, wyd II, Script, Warszawa 2001.

[3] M. Ledoux, *The concentration of measure phenomenon*, American Mathematical Society, Providence 2001.

[4] M. Ledoux, M. Talagrand, *Probability in Banach spaces*, Springer-Verlag, Berlin, 1991.

[5] W. Rudin, *Analiza rzeczywista i zespolona*, wyd. II, PWN, Warszawa 1998.

[6] R. van Handel, *Probability in High Dimension*, APC 550 Lecture Notes, Princeton University, https://web.math.princeton.edu/ rvan/APC550.pdf.

[7] R. Vershynin, *High-Dimensional Probability*, Cambridge University Press, Cambridge 2018.