Koncentracja Miary

Rafał Latała

22 maja 2009

Poniższe notatki powstają na podstawie (wybranych) wykładów z Koncentracji Miary, prowadzonych w semestrze wiosennym 2008/09.

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.

Dziękuje panom Kamilowi Kosińskiemu i Piotrowi Nayarowi za nadesłane komentarze.

Rafał Latała

1 Wprowadzenie

Zacznijmy od następującej definicji otoczki zbioru.

Definicja 1.1. *Niech (X, d) będzie przestrzenią metryczną, zaś A dowolnym podzbiorem 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 X o środku w y i promieniu t.

Okazuje się, że dla wielu ważnych przykładów miar probabilistycznych μ na (X, d) miara $\mu(A_t)$ szybko zbiega do 1 jeśli tylko $\mu(A) \geqslant 1/2$. Badanie tego zjawiska, zwanego *fenomenem koncentracji miary* i jego konsekwencji będzie celem tego wykładu.

Na początek wykładu podamy kilka przykładów, których dowody 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

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

Wprowadźmy kluczową definicję:

Definicja 1.2. *Niech μ będzie miarą probabilistyczną na (X, d).* Funkcją koncentracji *miary μ definiujemy jako*

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

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

Uwaga. 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

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

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

$$\alpha_{\sigma_{n,\sqrt{n}}}(t) \leqslant \exp\Big(-\frac{n-1}{2n}t^2\Big) \leqslant \exp\Big(-\frac{1}{4}t^2\Big).$$

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

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

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 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*. Widzielismy też przykład, w którym $\alpha_\mu(t) \leqslant C_1 \exp(-t/C_2)$ – mówimy wtedy o koncentracji *wykładniczej*.

1.1 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 1.3. *Niech μ będzie miarą probabilistyczną na (X,d) oraz $f\colon X \to \mathbb{R}$.*
Medianą *f względem μ nazywamy taką liczbę $M = \mathrm{Med}_\mu(f)$ dla której $\mu(\{x\colon f(x) \geqslant M\}), \mu(\{x\colon f(x) \leqslant M\}) \geqslant 1/2$.*
Modułem ciągłości *f nazywamy funkcję*

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

Fakt 1.1. *Dla dowolnej funkcji $F\colon X \to \mathbb{R}$,*

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

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. $\square$

Przypomnijmy definicję funkcji lipschitzowskiej

Definicja 1.4. *Funkcję $F\colon (X,d) \to \mathbb{R}$ 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. $|F(x) - F(y)| \leqslant Ld(x,y)$ dla wszystkich $x, y \in X$.

Analogicznie można zdefiniować funkcje lipschitzowskie między przestrzeniami metrycznymi.

Fakt 1.2. *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 1.1 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

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

$\square$

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 1.3. *Załóżmy, że μ jest miarą probabilistyczną na przestrzeni metrycznej (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 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 $t > 0$. Niech $F(x) := \min\{d(x,A), t\}$, wówczas 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 $n \to \infty$. $\square$

1.2 Obserwowalna średnica zbioru

Średnicą zbioru A w przestrzeni metrycznej (X,d) nazywamy

$$\mathrm{Diam}(A) := \sup\{d(x,y) \colon x, y \in A\}.$$

Jeśli próbujemy obserwować jakiś obiekt, to możemy przyjąć, że zbiory miary mniejszej niż pewne $\kappa > 0$ są dla nas niezauważalne (założenie to np. dla $\kappa = 10^{-10}$ wydaje się całkiem uzasadnione, jeśli miara μ jest związana ze stopniem oświetlenia obiektu). Motywuje to definicje *częściowej średnicy*

$$\mathrm{PartDiam}_\mu(X,d) = \inf\{\mathrm{Diam}(A) \colon \mu(A) \geqslant 1 - \kappa\}.$$

W praktyce również nie jesteśmy w stanie obserwować obiektów wielowymiarowych, a jedynie ich przekroje. Stosując pewne uproszczenie możemy założyć, że przestrzeń obserwujemy za pomocą funkcji 1-lipschitzowskich F, tzn. badamy rozkład $\mu \circ F^{-1}$ na $F(X)$ (np. natężenie światła). Stąd wprowadzamy

Definicja 1.5. Obserwowalną średnicą *przestrzeni* (X, d) *względem miary* μ *nazywamy wielkość*

$$\mathrm{ObsDiam}_\mu(X, d) := \sup\Big\{\mathrm{PartDiam}_{\mu\circ F^{-1}}(F(X))\colon F\colon X \to \mathbb{R}, \|F\|_{\mathrm{Lip}} \leqslant 1\Big\}.$$

Fakt 1.4.

$$\mathrm{ObsDiam}_\mu(X, d) \leqslant 2\alpha_\mu^{-1}(\kappa/2),$$

gdzie $\alpha_\mu^{-1}(\varepsilon) = \inf\{r > 0\colon \alpha_\mu(r) \leqslant \varepsilon\}$.

Dowód. Ustalmy funkcję 1-lipschitzowską F na X i niech $M = \mathrm{Med}_\mu(F)$, wówczas na podstawie Faktu 1.2,

$$\mu \circ F^{-1}([M + r, M - r]) = \mu(|F - M| \leqslant r) \geqslant 1 - 2\alpha_\mu(r).$$

Zatem

$$\mathrm{ObsDiam}_\mu(X, d) \leqslant 2 \inf\{r\colon 2\alpha_\mu(r) \leqslant \kappa\} = 2\alpha_\mu^{-1}(\kappa/2).$$

□

Przykład. Dla gaussowskiej koncentracji $\alpha_{(X,d,\mu)}(t) \leqslant C_1 \exp(-t^2/C_2)$ mamy $\mathrm{ObsDiam}_\mu(X, d) \leqslant 2\sqrt{C_2 \ln(2C_1/\kappa)}$.

W szczególności $\mathrm{ObsDiam}_{\sigma_n}(S^n)$ jest rzędu $1/\sqrt{n}$, podczas gdy nietrudno sprawdzić, że $\mathrm{Diam}(S^n) = \pi$ i $\mathrm{PartDiam}_\mu(S^n) \geqslant \pi/2$.

1.3 Transport miary

W wielu zagadnieniach będziemy używali pojęcia transportu miary.

Definicja 1.6. *Niech* μ *i* ν *będą miarami na przestrzeniach metrycznych* X *i* Y*. Powiemy, że funkcja borelowska* $\varphi\colon X \to Y$ transportuje miarę μ na miarę ν *(ew.* miara ν jest obrazem miary μ przy przekształceniu φ*) jeśli* $\nu(A) = \mu \circ \varphi^{-1}(A)$ *dla wszystkich* $A \in \mathcal{B}(Y)$.

Szczególnie wygodny jest transport lipschitzowski.

Fakt 1.5. *Jeśli* $\varphi\colon X \to Y$ *jest* L*-lipschitzowska oraz* φ *transportuje miarę* μ *na* ν*, to* $\alpha_\nu(t) \leqslant \alpha_\mu(t/L)$.

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

2 Nierówności izoperymetryczne

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

2.1 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 2.1. *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 2.2 (Nierówność Prekopy-Leindlera)**.** *Jeśli $s \in [0, 1]$ oraz $f, g, h \colon \mathbb{R}^n \to [0, \infty)$ spełniają warunek*

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

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 $A \cap B = \emptyset$ oraz

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

Nierówność Prekopy-Leindlera udowodnimy przez indukcje 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

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

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

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

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

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

Zatem na mocy założenia indukcyjnego

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

Stosując nierówność Prekopy-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 2.3 (Nierówność Brunna-Minkowskiego)**.** *Dla dowolnych niepustych zbiorów borelowskich $A, B \subset \mathbb{R}^n$,*

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

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 Prekopy-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

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

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

Dowód Twierdzenia 2.1. Niech $c_n = \lambda_n(B(0,1))$, wówczas $\lambda_n(A) = c_n r^n$ i na podstawie Wniosku 2.3,

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

□

Definicja 2.1. *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. 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 2.4. *Jeśli A jest podzbiorem borelowskim $\mathbb{R}^n$ takim, że $\lambda_n(A) = \lambda_n(B(x_0,r))$, to*

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

gdzie

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

2.2 Izoperymetria sferyczna

Twierdzenie 2.5. *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 2.6.

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

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

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

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

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

□

2.3 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

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

Twierdzenie 2.7. *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.

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

Fakt 2.8 (Lemat Poincaré)**.** *Miara $\mu_{k,n}$ zbiega słabo przy $n \to \infty$ do miary γ_k, co więcej*

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

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

Dowód Twierdzenia 2.7. 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

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

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. $\square$

Twierdzenie 2.9. *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 $\gamma_k(H) = \Phi(x)$ i H jest postaci (3), to $H_t = \{x \in \mathbb{R}^k\colon \langle x, u\rangle < r + t\}$ i $\gamma_k(H_t) = \Phi(x+t)$. $\square$

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

Wniosek 2.10. $\alpha_{\gamma_k}(t) \leqslant 1 - \Phi(t) \leqslant \frac{1}{2}\exp(-t^2/2)$.

Na podstawie Faktu 1.2 dostajemy:

Wniosek 2.11. *Jeśli* $F\colon \mathbb{R}^k \to \mathbb{R}$ *jest funkcją* L*-lipschitzowską oraz* $t \geqslant 0$ *to*

$$\gamma_k(\{x\colon F(x) \geqslant \mathrm{Med}_{\gamma_k}(F) + t\}) \leqslant 1 - \Phi(t/L) \leqslant \frac{1}{2}e^{-\frac{t^2}{2L^2}}$$

oraz

$$\gamma_k(\{x\colon |F(x) - \mathrm{Med}_{\gamma_k}(F)| \geqslant t\}) \leqslant 2(1 - \Phi(t/L)) \leqslant e^{-\frac{t^2}{2L^2}}$$

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

Wniosek 2.12. *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 f transportuje 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 1.5 i Wniosku 2.10. $\square$

Wniosek 2.13. *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 $\alpha_{\mu_{B_n}} \leqslant \frac{1}{2}\exp(-nt^2/(2C))$.*

Dowód. 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 $Tx = \frac{x}{|x|}\varphi(|x|)$. Wystarczy sprawdzić, że $\gamma_n(B(0,t)) = \mu_n(B(0,\varphi(t)))$, czyli całkując we współrzęsnych sferycznych, że dla $t \geqslant 0$,

$$\varphi(t)^n = \frac{1}{c_n}\int_0^t r^{n-1}e^{-r^2/2}dr, \tag{4}$$

gdzie

$$c_n = \int_0^\infty r^{n-1}e^{-r^2/2}dr = 2^{(n-2)/2}\Gamma(\frac{n}{2}).$$

Z (4) wynika, że

$$\varphi(t)^n \geqslant \frac{1}{c_n}e^{-t^2/2}\int_0^t r^{n-1}dr = \frac{1}{nc_n}t^n e^{-t^2/2}.$$

Różniczkując stronami (4) dostajemy $n\varphi'(t)\varphi(t)^{n-1} = t^{n-1}e^{-t^2/2}/c_n$, zatem

$$\varphi'(t) = \frac{1}{nc_n}\Big(\frac{t}{\varphi(t)}\Big)^{n-1}e^{-t^2/2} \leqslant \frac{1}{nc_n}(nc_n e^{t^2/2})^{(n-1)/n}e^{-t^2/2} \leqslant (nc_n)^{-1/n}.$$

Ze wzoru Stirlinga dostajemy

$$nc_n = 2^{n/2}\Gamma(\frac{n}{2}+1) \geqslant (\frac{n}{e})^{n/2},$$

więc

$$\|\varphi\|_{\mathrm{Lip}} = \sup_{t\geqslant 0}|\varphi'(t)| \leqslant \frac{\sqrt{e}}{\sqrt{n}}.$$

□

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.

3 Koncentracja w przestrzeniach produktowych

3.1 Transformata Laplace'a

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

Definicja 3.1. 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 3.1. *Dla dowolnej zmiennej losowej* Z,

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

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

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

to dla $t \geqslant 0$

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

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 $\int F d\mu = 0$.

3.2 Metody Martyngałowe

Twierdzenie 3.2 (Nierówność 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 3.1. $\square$

Uwaga. 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 $F\colon X \to \mathbb{R}$ i odpowiednio dobranego $(\mathcal{F}_k)$ ciągu σ-ciał podzbiorów X.

Wniosek 3.3. *Niech (X_i, d_i) będą przetrzeniami metrycznymi, $X = X_1 \times \cdots \times X_n$ z odległością l_1, to znaczy $d(x,y) = \sum_{i=1}^n d_i(x_i, y_i)$ dla $x, y \in X$ oraz niech $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ będzie produktem miar probabilistycznych μ_i na X_i. Wówczas dla dowolnej funkcji 1-lipschitzowskiej F na X*

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

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

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

Dowód. Na mocy Faktu 1.3 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

$$M_k(x) = \tilde{M}_k(x_1, \ldots, x_k) = \int_{X_{k+1}\times\ldots\times X_n} F(x) d\mu_{i+1}(x_{i+1}) \cdots d\mu_n(x_n),$$

stąd

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

i teza wynika z Twierdzenia 3.2. □

Definicja 3.2. *Mówimy, że skończona przestrzeń metryczna (X, d) ma* długość *conajwyżej l, jeśli istnieje rosnący ciąg podziałów X, $\{X\} = \mathcal{A}_0, \mathcal{A}_1, \ldots, \mathcal{A}_n = \{\{x\} \colon x \in X\}$ ($\mathcal{A}_i$ jest podpodziałem $\mathcal{A}_{i-1}$) oraz liczby $a_1, \ldots, a_n$ 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. Biorąc $\mathcal{A}_0 = \{X\}$ i $\mathcal{A}_1 = \{\{x\} \colon x \in X\}$ widzimy, że każda skończona przestrzeń metryczna ma długość nie większą niż $\mathrm{Diam}(X)$.

Twierdzenie 3.4. *Jeśli (X, d) jest skończoną przestrzenią metryczną o długości co najwyżej l, zaś μ unormowaną miarą liczącą na X, to dla funkcji 1-lipschitzowskich F na 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 3.2, 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 3.2 oraz Faktu 1.3. □

Przykład 1. Niech $X = \{0,1\}^n$ z odległością $d(x,y) = \frac{1}{n}\#\{i\colon x_i \neq y_i\}$. Możemy wtedy położyć

$$\mathcal{A}_i = \{\{(x_1,\dots,x_i)\} \times \{0,1\}^{n-i}\colon x_1,\dots,x_i \in \{0,1\}\}$$

i łatwo sprawdzić, że założenia definicji są spełnione z $a_i = \frac{1}{n}$. Zatem $l = 1/\sqrt{n}$ i

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

Przykład 2. Ogólniej niech X_i będą skończonymi zbiorami, $X = X_1 \times \cdots \times X_n$, $d(x,y) = \#\{i\colon x_i \neq y_i\}$ oraz μ będzie unormowaną miarą liczącą na X. Analogicznie jak poprzednio

$$\mathcal{A}_i = \{\{(x_1,\dots,x_i)\} \times X_{i+1} \times \cdots \times X_n\colon x_j \in X_j, 1 \leqslant j \leqslant i\}$$

oraz $a_i = 1$. Zatem $l = \sqrt{n}$ i

$$\alpha_{(X,d,\mu)} \leqslant \exp(-\frac{t^2}{8n}).$$

Przykład 3. Niech Π^n będzie grupą permutacji zbioru $\{1,\dots,n\}$ z metryką $d(\sigma,\pi) = \#\{i\colon \sigma_i \neq \pi_i\}$, a μ unormowaną miarą liczącą na Π^n. Niech $\mathcal{A}_i$ składa się ze zbiorów postaci

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

Wówczas jeśli $B, C \in \mathcal{A}_i$ są podzbiorami pewnego $A \in \mathcal{A}_{i-1}$ to $B = A_{j_1,\dots,j_{i-1},p}$, $C = A_{j_1,\dots,j_{i-1},q}$ i możemy zdefiniować bijekcję Φ między B i C jako $\Phi = \tau_{p,q} \circ \sigma$, 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

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

Ostatnie twierdzenie z tej części wiąże się z miarą Haara na zwartej grupie metrycznej (G,d), to znaczy taką miarą probabilistyczną μ, że $\mu(gA) = \mu(A) = \mu(Ag)$ dla dowolnego $A \in \mathcal{B}(G)$ i $g \in G$. Zakładamy, że d jest niezmiennicza na translacje, tzn $d(hg_1, hg_2) = d(g_1, g_2) = d(g_1h, g_2h)$ dla $g_1, g_2, h \in G$. Dla podgrupy domkniętej $H \subset G$ można wprowadzić odległość na G/H wzorem

$$\rho(g_1H, g_2H) = d(g_1, g_2H) = d(g_2^{-1}g_1, H).$$

Twierdzenie 3.5. *Niech μ będzie miarą Haara na zwartej grupie metrycznej (G,d) oraz $G = G_0 \supset G_1 \supset \cdots \supset G_n = \{e\}$ będzie ciągiem domkniętych podgrup. Wówczas*

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

gdzie $l = (\sum_{i=1}^n \mathrm{Diam}(G_{i-1}/G_i)^2)^{1/2}$.

Dowód. Niech F będzie funkcją 1-lipschitzowską, $M_i = \mathbf{E}_\mu(F|\mathcal{F}_i)$ gdzie $\mathcal{F}_i$ jest σ-ciałem generowanym przez zbiory gG_i. Załóżmy, że $g_1G_i, g_2G_i \subset g_0G_{i-1}$ wówczas $g_0^{-1}g_1, g_0^{-1}g_2 \in G_{i-1}$, więc ze zwartości G_i

$$\mathrm{Diam}(G_{i-1}/G_i) \geqslant d(g_0^{-1}g_1, g_0^{-1}g_2G_i) = d(g_0^{-1}g_1, g_0^{-1}g_2h) = d(g_1, g_2h)$$

dla pewnego $h \in G_i$. Określmy przekształcenie Φ wzorem $\Phi(g) = g_2hg_1^{-1}g$, wówczas Φ zachowuje miarę μ oraz jest homeomorfizmem między g_1G i g_2G. Ponadto

$$|F(g)-F(\Phi(g))| \leqslant d(g,\Phi(g)) = d(g, g_2hg_1^{-1}g) = d(g_1, g_2h) \leqslant \mathrm{Diam}(G_{i-1}/G_i).$$

Stąd oscylacja M_i na g_0G_{i-1} jest nie większa niż $a_i = \mathrm{Diam}(G_{i-1}/G_i)$ czyli po uśrednieniu $\|M_i - M_{i-1}\|_\infty \leqslant a_i$ i możemy stosować Twierdzenie 3.2 □

3.3 Nierówność Poincaré

Definicja 3.3. *Mówimy, że miara probabilistyczna μ na (X,d)* spełnia nierówność Poincaré *ze stałą C, jeśli dla wszystkich ograniczonych lipschitzowskich funkcji f na X zachodzi*

$$\mathrm{Var}_\mu(f) \leqslant C\int |\nabla f|^2 d\mu, \tag{5}$$

gdzie

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

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

Uwaga 1. W przypadku, gdy $X = \mathbb{R}^n$ 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 standardowy argument aproksymacyjny pokazuje, że by wykazać nierówność Poincaré dla miar probabilistycznych na $\mathbb{R}^n$ wystarczy sprawdzić (5) dla ograniczonych funkcji klasy $C^1(\mathbb{R}^n)$ o ograniczonych pochodnych rzędu jeden.

Uwaga 2. 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 3. 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 (5) 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 3.6. *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 $C_{\mathrm{opt}} \leqslant \tilde{C}_{\mathrm{opt}} \leqslant (1+\sqrt{2})^2 C_{\mathrm{opt}}$.

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

$$(\mathbf{E}_\mu|f - \mathrm{Med}_\mu f|^2)^{1/2} \leqslant \mathrm{Var}_\mu(f)^{1/2} + |\mathrm{Med}_\mu f - \mathbf{E}_\mu f| \leqslant (1+\sqrt{2})\mathrm{Var}_\mu(f)^{1/2}$$

i otrzymujemy $\tilde{C}_{\mathrm{opt}} \leqslant (1+\sqrt{2})^2 C_{\mathrm{opt}}$. □

Twierdzenie 3.7. *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

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

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. $\square$

Fakt 3.8. *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 $h \in C^1_{\mathrm{ogr}}(\mathbb{R})$,

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

Niech $f \in C^1_{\mathrm{ogr}}(\mathbb{R})$ 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

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

$\square$

Fakt 3.9. *Załóżmy, że μ_i są miarami probabilistycznymi na X_i, $X = X_1 \times \ldots \times X_n$ oraz $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$. Wówczas dla dowolnej funkcji $f \in L^2(X,\mu)$*

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

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. $\square$

Wniosek 3.10. *Załóżmy, że miary probabilistyczne μ_i na (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 3.9 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 3.11. *Produktowy rozkład wykładniczy ν^n spełnia nierówność Poincaré na $\mathbb{R}^n$ ze stałą 4. W szczególności $\alpha_{\nu^n}(t) \leqslant 2\exp(-t/4)$.*

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

Fakt 3.12. *Załóżmy, że μ jest miarą probabilistyczną na 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

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

□

Fakt 3.13. *Jeśli miara ν na (Y, ρ) jest L-lipschitzowskim obrazem miary μ na (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 punktowego oszacowania $|\nabla f \circ \varphi|(x) \leqslant L|\nabla f|(\varphi(x))$. $\square$

Wniosek 3.14. *Istnieje stała uniwersalna $L_0 \leqslant 12.5$ taka, że dowolna miara symetryczna log-wklęsła μ o wariancji 1 na prostej jest L_0-lipschitzowskim obrazem symetrycznej miary wykładniczej ν. Stąd μ spełnia nierówność Poincare ze stałą $4L_0^2 \leqslant 620$.*

Dowód. Niech g będzie gęstością miary μ, wówczas $g = e^{-h}$, gdzie $h \colon \mathbb{R} \to (-\infty, \infty]$ jest funkcją parzystą wypukłą, w szczególności $g(0) = \|g\|_\infty$. Określmy

$$x_0 := \inf\{x > 0 \colon g(x) \leqslant g(0)/e\}.$$

Zauważmy, że z log-wklęsłości wynika, że $g(x) \leqslant e^{-x/x_0} g(0)$ dla $x > x_0$ oraz $g(0) \geqslant g(x) \geqslant g(0)/e$ dla $0 \leqslant x < x_0$. Stąd

$$\begin{aligned}2x_0 \frac{g(0)}{e} &\leqslant \int_{-x_0}^{x_0} g(x)dx \leqslant 1 = 2(\int_0^{x_0} + \int_{x_0}^\infty) g(x)dx \\ &\leqslant 2(x_0 g(0) + \int_{x_0}^\infty e^{-x/x_0} g(0)dx) = 2x_0 g_0 (1 + \frac{1}{e}).\end{aligned}$$

Analogicznie

$$\begin{aligned}\frac{2}{3} x_0^3 \frac{g(0)}{e} &\leqslant \int_{-x_0}^{x_0} x^2 g(x)dx \leqslant 1 = 2(\int_0^{x_0} + \int_{x_0}^\infty) x^2 g(x)dx \\ &\leqslant 2g(0)(\int_0^{x_0} x^2 dx + \int_{x_0}^\infty x^2 e^{-x/x_0} dx) = 2x_0^3 g_0 (\frac{1}{3} + \frac{5}{e}).\end{aligned}$$

Zatem

$$\frac{e}{2(e+1)} \leqslant x_0 g(0) \leqslant \frac{e}{2}, \quad \frac{3e}{2(e+15)} \leqslant x_0^3 g(0) \leqslant \frac{3e}{2},$$

skąd wnioskujemy, że

$$\sqrt{\frac{3}{e+15}} \leqslant x_0 \leqslant \sqrt{3(e+1)} \text{ oraz } \frac{e}{2\sqrt{3}(e+1)^{3/2}} \leqslant g(0) \leqslant \frac{e\sqrt{e+15}}{2\sqrt{3}}.$$

Określmy $\varphi\colon \mathbb{R} \to \mathbb{R}$ wzorem

$$\nu(-\infty, x] = \mu(-\infty, \varphi(x)].$$

Wówczas φ transportuje ν na μ, jest nieparzysta i niemalejąca, zatem

$$\|\varphi\|_{\mathrm{Lip}} = \sup_{x \geqslant 0} \varphi'(x) = \sup_{x \geqslant 0} \frac{1}{2} \frac{e^{-x}}{g(\varphi(x))}.$$

Rozpatrzmy dwie możliwości:
i) $\varphi(x) \leqslant x_0$. Wówczas $g(\varphi(x)) \geqslant g(0)/e$ i

$$\frac{1}{2} \frac{e^{-x}}{g(\varphi(x))} \leqslant \frac{e}{2g(0)} \leqslant \sqrt{3}(e+1)^{3/2}.$$

ii) $\varphi(x) > x_0$. Zauważmy, że z logarytmicznej wklęsłości g wynika, że $g(\varphi(x)+t) \leqslant g(\varphi(x))e^{-t/x_0}$ dla $t > 0$, stąd

$$\begin{aligned}\frac{1}{2}e^{-x} = \nu([x,\infty)) = \mu([\varphi(x),\infty)) &= \int_0^\infty g(\varphi(x)+t)dt \\ \leqslant \int_0^\infty g(\varphi(x))e^{-t/x_0}dt &= x_0 g(\varphi(x)).\end{aligned}$$

Zatem

$$\frac{1}{2} \frac{e^{-x}}{g(\varphi(x))} \leqslant x_0 \leqslant \sqrt{3(e+1)}.$$

Podsumowując oba przypadki otrzymujemy $\|\varphi\|_{\mathrm{Lip}} \leqslant \sqrt{3}(e+1)^{3/2} \leqslant 12.5$.
□

Uwaga. Wniosek 3.14 (z nieco gorszą stałą) jest prawdziwy bez założenia symetrii miary μ. Dowód jest bardzo podobny, choć nieco bardziej żmudny.

3.3.1 Charakteryzacja na prostej

Okazuje się, że daje się podać prostą charakteryzację miar na $\mathbb{R}$, które spełniają nierówność Poincaré. Zanim ją sformułujemy podamy ściśle związany fakt dotyczący tak zwanych ważonych nierówności Hardy'ego.

Twierdzenie 3.15 (Muckenhoupt)**.** *Załóżmy, że μ i ν są miarami na $[0,\infty)$. Wówczas istnieje stała $C < \infty$ taka, że dla każdej funkcji ograniczonej f,*

$$\int_0^\infty \Big| \int_0^x f(t)dt \Big|^2 d\mu \leqslant C \int_0^\infty |f(x)|^2 d\nu(x) \tag{6}$$

wtedy i tylko wtedy gdy

$$B := \sup_{r\geqslant 0} \mu[r,\infty) \int_0^r \frac{1}{p(x)} dx < \infty,$$

gdzie p oznacza gęstość części absolutnej ν. Co więcej, jeśli $C_{\rm opt}$ jest najmniejszą stałą C taką, że zachodzi (6), to $B \leqslant C_{\rm opt} \leqslant 4B$.

Dowód. Zamieniając f na $|f|$ i zauważając, że wyzerowanie f na nośniku części singularnej ν nie zmienia lewej strony (6) wnioskujemy, że wystarczy badać nierówność

$$\int_0^\infty \Big| \int_0^x f(t)dt\Big|^2 d\mu \leqslant C \int_0^\infty |f(x)|^2 p(x)dx \tag{7}$$

dla nieujemnych f.

Zauważmy, że dla $f \geqslant 0$ (7) implikuje

$$\mu[r,\infty)\Big(\int_0^r f(t)dt\Big)^2 \leqslant C\int_0^r |f(x)|^2 p(x)dx$$

Przyjmując $f(x) = \frac{1}{p(x)} \mathbb{1}_{\{p(x)\geqslant 1/n\}}$ i biorąc najlepszą możliwą stałą dostajemy

$$\mu[r,\infty)\int_0^r \frac{1}{p(x)} \mathbb{1}_{\{p(x)\geqslant 1/n\}} dx \leqslant C_{\rm opt},$$

skąd po przejściu z $n \to \infty$ i wzięciu supremum po r dostajemy $B \leqslant C_{\rm opt}$.

By udowodnić, że $C_{\rm opt} \leqslant 4B$ wystarczy wykazać, że (7) zachodzi z $C = 4B$. Załóżmy wpierw, że miara μ ma gęstość g i przyjmijmy

$$h(x) := \Big(\int_0^x \frac{1}{p(t)}dt\Big)^{1/2}$$

wówczas na mocy nierówności Schwarza i twierdzenia Fubiniego

$$\begin{aligned}
\int_0^\infty \Big|\int_0^x f(t)dt\Big|^2 d\mu &= \int_0^\infty \Big|\int_0^x f(t)dt\Big|^2 g(x)dx \\
&\leqslant \int_0^\infty g(x)\Big[\int_0^x f^2(t)p(t)h(t)dt \int_0^x \frac{1}{p(u)h(u)}du\Big]dx \\
&= \int_0^\infty f^2(t)p(t)h(t)\Big[\int_t^\infty g(x)\int_0^x \frac{1}{p(u)h(u)}dudx\Big]dt
\end{aligned}$$

Zauważmy, że

$$\int_0^x \frac{1}{p(u)h(u)}du = \int_0^x 2h'(u)du = 2h(x),$$

więc na mocy definicji B (użytej dwa razy)

$$\begin{aligned}
&\int_t^\infty g(x)\int_0^x \frac{1}{g(u)h(u)}dudx = 2\int_t^\infty g(x)\Big(\int_0^x \frac{1}{p(u)}du\Big)^{1/2}dx\\
&\leqslant 2\sqrt{B}\int_t^\infty g(x)\Big(\int_x^\infty g(u)du\Big)^{-1/2}dx = 4\sqrt{B}\int_t^\infty \frac{d}{dx}\Big(-\int_x^\infty g(u)du\Big)^{1/2}dx\\
&= 4\sqrt{B}\Big(\int_t^\infty g(x)dx\Big)^{1/2} \leqslant 4B\Big(\int_t^\infty \frac{1}{p(x)}dx\Big)^{-1/2} = 4Bh^{-1}(t).
\end{aligned}$$

Zatem

$$\int_0^\infty \Big|\int_0^x f(t)dt\Big|^2 d\mu \leqslant \int_0^\infty f^2(t)p(t)h(t)4Bh^{-1}(t) = 4B\int_0^\infty f^2(t)p(t)dt.$$

Jeśli miara μ nie ma gęstości, to znajdujemy miary μ_n z gęstościami zbieżne słabo do μ takie, że $\mu_n[x,\infty) \leqslant \mu[x,\infty)$ Można np przyjąć

$$\mu_n(A) = n\int_0^{1/n} \mu(A+t)dt$$

Wówczas

$$\sup_{r\geqslant 0}\mu_n[r,\infty)\int_0^r \frac{1}{p(x)}dx \leqslant \sup_{r\geqslant 0}\mu[r,\infty)\int_0^r \frac{1}{p(x)}dx = B,$$

więc

$$\int_0^\infty \Big|\int_0^x f(t)dt\Big|^2 d\mu = \lim_{n\to\infty}\int_0^\infty \Big|\int_0^x f(t)dt\Big|^2 d\mu_n \leqslant 4B\int_0^\infty |f(x)|^2p(x)dx.$$

□

Twierdzenie 3.16. *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 $C_{\rm opt}$ w nierówności Poincaré spełnia

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

Dowód. Z Twierdzenia 3.15 (zastosowanego do miary μ obciętej do $[m,\infty)$ i f' zamiast f) wynika, że dla dowolnej funkcji Lipschitzowskiej f,

$$\int_m^\infty (f(x)-f(m))^2 d\mu(x) = \int_m^\infty \Big(\int_m^x f'(t)dt\Big)^2 d\mu(x) \leqslant 4B_+ \int_m^\infty f'^2(x)d\mu(x).$$

Analogicznie rozpatrując miarę μ na odcinku $(-\infty, m)$,

$$\int_{-\infty}^m (f(x)-f(m))^2 d\mu(x) \leqslant 4B_- \int_m^\infty f'^2(x)d\mu(x).$$

Dodając stronami dostajemy

$$\mathrm{Var}_\mu(f) \leqslant \int (f(x)-f(m))^2 d\mu(x) \leqslant 4\max\{B_+, B_-\} \int f'^2(x)d\mu(x),$$

czyli $C_{\mathrm{opt}} \leqslant 4\max\{B_+, B_-\}$.

By udowodnić nierówność w drugą stronę zauważmy, że jeśli f jest ograniczoną funkcją oraz $g(x) = 0$ dla $x \leqslant m$ i $g(x) = \int_m^x f(t)dt$ dla $x > m$, to g jest lipschitzowska, $g'(x) = f(x)\mathbb{1}_{\{x \geqslant m\}}$ p.w. oraz $\mathrm{Med}_\mu(g) = 0$. Stąd na podstawie Faktu 3.6,

$$\begin{aligned}\int_m^\infty \Big|\int_m^x f(t)dt\Big|^2 d\mu &= \int |g(x) - \mathrm{Med}_\mu g|^2 d\mu(x) \\ &\leqslant (1+\sqrt{2})^2 C_{\mathrm{opt}} \int_0^\infty g'^2(x)d\mu(x) \\ &= (1+\sqrt{2})^2 C_{\mathrm{opt}} \int_m^\infty |f(x)|^2 d\mu(x).\end{aligned}$$

Stąd Twierdzenie 3.15 implikuje $B_+ \leqslant (1+\sqrt{2})^2 C_{\mathrm{opt}}$, analogicznie pokazujemy $B_- \leqslant (1+\sqrt{2})^2 C_{\mathrm{opt}}$. □

3.4 Logarytmiczna Nierówność Sobolewa

Definicja 3.4. *Załóżmy, że μ jest miarą probabilistyczną na X, zaś f nieujemną funkcją mierzalną na X.* Entropię f względem μ *definiujemy wzorem*

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

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 $\lambda \geqslant 0$.

Definicja 3.5. *Mówimy, że miara probabilistyczna na* (X, d) spełnia logarytmiczną nierówność Sobolewa *ze stałą* C*, jeśli dla wszystkich ograniczonych lipschitzowskich funkcji* f *na* X *zachodzi*

$$\mathrm{Ent}_\mu(f^2) \leqslant 2C \int |\nabla f|^2 d\mu. \tag{8}$$

Twierdzenie 3.17. *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 (8) daje

$$\lambda M'(\lambda) - M(\lambda) \log M(\lambda) \leqslant C \frac{\lambda^2}{2} M(\lambda). \tag{9}$$

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 (9)

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

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

Lemat 3.18. *Dla dowolnej funkcji nieujemnej na X,*

$$\mathrm{Ent}_\mu(f) = \sup\Big\{ \int fg d\mu \colon \int e^g d\mu \leqslant 1 \Big\}. \tag{10}$$

Dowód. Z jednorodności obu stron (10) 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

$$uv \leqslant u \log u - u + e^v \text{ dla } u \geqslant 0, v \in \mathbb{R}. \tag{11}$$

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

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

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

Z powyższego lematu łatwo wykazać tensoryzowalność entropii:

Fakt 3.19. *Załóżmy, że μ_i są miarami probabilistycznymi na X_i, $X = X_1 \times \ldots \times X_n$ oraz $\mu = \mu_1 \otimes \mu_2 \otimes \cdots \otimes \mu_n$. Wówczas dla dowolnej nieujemnej funkcji f na X*

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

Dowód. Weźmy funkcję g na 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

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

Wniosek 3.20. *Załóżmy, że miary probabilistyczne μ_i na (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 3.19 dostajemy

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

□

Fakt 3.21. *i) Niech $\mu_1 = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$, wówczas dla dowolnego $f\colon \{-1,1\} \to \mathbb{R}$,*

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

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 $f\colon \{-1,1\}^n \to \mathbb{R}$,

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

gdzie $|Df|^2(x) = \frac{1}{4}\sum_{i=1}^n (f(x) - f(s_i(x)))^2$ oraz $s_i((x_1,\ldots,x_n)) = (x_1,\ldots,x_{i-1},-x_i,x_{i+1},\ldots,x_n)$.

Dowód. i) Z uwagi na jednorodność możemy zakładać, że $\mathbf{E}_{\mu_1} f^2 = 1$, wówczas istnieje $t \in [-1,1]$ takie, że $f(1) = 1+t$ oraz $f(-1) = 1-t$ 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 3.19. □

Twierdzenie 3.22. *Miara γ_n spełnia logarytmiczną nierówność Sobolewa z $C = 1$.*

Dowód. Z uwagi na Fakt 3.19 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

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

Niech μ_n i $|Df|$ będą jak w Fakcie 3.21. 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 $T_n(x) = n^{-1/2}(x_1 + \ldots + x_n)$

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

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 3.23. *Załóżmy, że μ jest miarą probabilistyczną na 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

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

□

Kolejny fakt dowodzimy tak samo jak dla nierówności Poincaré.

Fakt 3.24. *Jeśli miara ν na (Y, ρ) jest L-lipschitzowskim obrazem miary μ na (X, d) oraz μ spełnia logarytmiczną nierówność Sobolewa ze stałą C, to ν spełnia logarytmiczną nierówność Sobolewa ze stałą CL^2.*

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

Fakt 3.25. *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 3.26. *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

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

3.5 Nierówność Bobkowa

Nierówność logarytmiczna 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 3.6. *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*

$$I\Big(\int f d\mu\Big) \leqslant \int \sqrt{I(f)^2 + C|\nabla f|^2} d\mu. \tag{12}$$

Fakt 3.27. *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 3.28. *Jeśli miara probabilistyczna μ na $\mathbb{R}^n$ spełnia nierówność Bobkowa na ze stałą C, to*

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

oraz

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

Twierdzenie 3.29. *Kanoniczna miara gaussowska γ_n spełnia nierówność Bobkowa z $C = 1$.*

3.6 Nierówności Splotu Infimum

Zacznijmy od zaproponowanej przez Maureya definicji.

Definicja 3.7. *Splotem infimum dwu funkcji f i g określonych na $\mathbb{R}^n$ nazywamy funkcję $f\Box g$ daną wzorem*

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

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

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

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

Pierwszą użyteczną cechą własności (τ) jest jej tensoryzowalność.

Fakt 3.30. *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 g na $\mathbb{R}^{n_2}$ jako

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

Dodatkowo wprowadźmy oznaczenie $f^y(x) = f(x,y)$. Wówczas dla dowolnych $y, \tilde{y}$

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

na mocy własności (τ) dla (μ_1, φ_1). Stąd

$$\int e^{f\Box\varphi(x,y)} d\mu_1(x) \leqslant e^{g\Box\varphi_2(y)}$$

i korzystając z Twierdzenia Fubiniego i (τ) dla (μ_1, φ_2).

$$\begin{aligned}\int e^{f\Box\varphi} d\mu_1 \otimes \mu_2 \leqslant \int e^{g\Box\varphi_2(y)} d\mu_2(y) \leqslant \Big(\int e^{-g(y)} d\mu_2(y)\Big)^{-1} \\ = (\int e^{-f} d\mu_1 \otimes \mu_2)^{-1}.\end{aligned}$$

□

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

Fakt 3.31. *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

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

Zatem

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

□

By sformułować związki nierówności splotu infimum z koncentracją określmy zbiór

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

Zacznijmy od prostego faktu

Fakt 3.32. *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. 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 3.33. *Załóżmy, że para* (μ, φ) *ma własność* (τ)*. Wówczas dla dowolnego zbioru borelowskiego* A *i* $t > 0$*,*

$$\mu(A + B_\varphi(t)) \geqslant \frac{e^t \mu(A)}{(e^t - 1)\mu(A) + 1}. \tag{13}$$

W szczególności

$$\mu(A + B_\varphi(t)) > \min\{e^{t/2} \mu(A), 1/2\} \tag{14}$$

oraz

$$\mu(A) \geqslant \frac{1}{2} \ \Rightarrow \ 1 - \mu(A + B_\varphi(t)) < e^{-t/2}(1 - \mu(A)). \tag{15}$$

Ponadto

$$\mu(A) = \nu(-\infty, x] \ \Rightarrow \ \mu(A + B_\varphi(t)) \geqslant \nu(-\infty, x + t/2]. \tag{16}$$

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

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

skąd bezpośredni rachunek prowadzi do (13).

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 (14). 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 (15).

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 (13) implikuje (16).

□

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 3.34. *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 3.35. *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 3.34. Ustalmy funkcję ograniczoną f, przyjmijmy $g := f\Box\varphi_0$ 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 3.35, $(-e^{-z(t)})' = e^{-z(t)}z'(t) \geqslant \sqrt{I_0 I_1}$, co po odcałkowaniu daje $\sqrt{I_0 I_1} \leqslant 1$. □

Uwaga. 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 3.36. *Miara ν spełnia nierówność infimum z funkcją kosztu φ_1 postaci*

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

Dowód. Z wypukłości funkcji φ_0 łatwo wynika, że $\varphi_1 = \varphi_0 \Box \varphi_0$. 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 3.31 $\Box$

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 3.37. *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 3.32 i 3.33. $\Box$

3.7 Nierówności Splotu Infimum dla Funkcji Wypukłych

3.8 Nierówność Logarytmiczna Sobolewa dla Funkcji Wypukłych

Fakt 3.38. *Załóżmy, że $F \colon \mathbb{R}^n \to \mathbb{R}$ jest lipschitzowska oraz wypukła po każdej współrzędnej. Wówczas dla dowolnej produktowej miary probabilistycznej μ na $\mathbb{R}^n$,*

$$\mathrm{Ent}_\mu(e^f) \leqslant \int\int \sum_{i=1}^n (x_i - y_i)^2 (\partial_i f)^2(x) e^{f(x)} d\mu(x) d\mu(y).$$

Twierdzenie 3.39. *Załóżmy, że* $F\colon [0,1]^n \to \mathbb{R}$ *jest L-lipschitzowska oraz wypukła po każdej współrzędnej. Wówczas dla dowolnej produktowej miary probabilistycznej* μ *na* $[0,1]^n$,

$$\mu\Big(\Big\{F \geqslant \int F d\mu + t\Big\}\Big) \leqslant \exp\Big(-\frac{t^2}{4L^2}\Big) \text{ dla } t \geqslant 0.$$

Dowód. Bez straty ogólności (zastępując F przez $aF+b$) możemy zakładać, że $L=1$ oraz $\int F = 0$. Wówczas $|\nabla F| \leqslant 1$ p.w., więc Fakt 3.38 implikuje, że dla $\lambda > 0$,

$$\begin{aligned}\mathrm{Ent}_\mu(e^{\lambda F}) &\leqslant \lambda^2 \int\int \sum_{i=1}^n (x_i - y_i)^2 (\partial_i F)^2(x) e^{\lambda F(x)} d\mu(x) d\mu(y) \\ &\leqslant \lambda^2 \int\int |\nabla F|^2(x) e^{\lambda F(x)} d\mu(x) d\mu(y) \leqslant \lambda^2 \int e^{\lambda F} d\mu.\end{aligned}$$

Kładąc $M(\lambda) = \int e^{\lambda F}$ dostajemy dla $\lambda > 0$,

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

Czyli funkcja $H(\lambda) = \frac{1}{\lambda} \log M(\lambda)$ spełnia $H'(\lambda) \leqslant 1$, a ponieważ $H(0+) = \int F d\mu = 0$, więc $H(\lambda) \leqslant \lambda$, czyli $M(\lambda) \leqslant \exp(\lambda^2)$. Stąd Fakt 3.1 implikuje tezę. □

Uwaga. Jeśli F jest wypukła, to $-F$ jest wklęsła, a dla funkcji wklęsłych nie da się łatwo zmodyfikować dowodu Faktu 3.38. Dlatego podane metody dają tylko oszacowanie górnego odchylenia od średniej funkcji wypukłych.

3.9 Aproksymacja przez otoczkę wypukłą

W tej części będziemy zakładać, że przestrzeń X ma strukturę produktową, tzn. $X = X_1 \times X_2 \times \cdots \times X_n$. Określmy

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

Z Wniosku 3.3 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 $A \subset X$ i $x \in X$ określmy

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

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

Dla $A \subset X$ i $x \in X$ 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

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

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 $\mathcal{D}^c_A(x)$.

Fakt 3.40. *Dla dowolnego $A \subset X$ i $x \in X$,*

$$d(0, V_A(x)) = \inf_{y \in V_A(x)} |y| = \mathcal{D}^c_A(x).$$

Dowód. i) $\mathcal{D}^c_A(x) \leqslant d(0, V_A(x))$. Niech $z \in V_A(x)$ takie, że $|z| = d(0, V_A(x))$. 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 $\mathcal{D}^c_A(x) \leqslant |z| = d(0, V_A(x))$.

ii) $\mathcal{D}^c_A(x) \geqslant d(0, V_A(x))$. Ustalmy $z \in V_A(x)$ taki, że $|z| = d(0, V_A(x))$. 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

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

$\square$

Twierdzenie 3.41. *Załóżmy, że $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ jest probabilistyczną miarą produktową na $X = X_1 \times \cdots \times X_n$. Wówczas dla dowolnego niepustego, mierzalnego zbioru A w 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

$$\tilde{X} = X_1 \times \cdots \times X_{n-1}, \quad \tilde{\mu} = \mu_1 \otimes \cdots \otimes \mu_{n-1}$$

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

$$B = \{\tilde{x} \colon x = (\tilde{x}, x_n) \in A\} \text{ oraz } A(x_n) = \{\tilde{x} \colon x = (\tilde{x}, x_n) \in A\} \text{ dla } x_n \in X_n.$$

Zauważmy, że jeśli $s \in U_{A(x_n)}$, 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 $\tilde{X}$) dostajemy dla dowolnego $\theta \in [0,1]$,

$$\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}_A^c(\tilde{x}, x_n)^2}{4}\Big) d\tilde{\mu}(\tilde{x}) \leqslant e^{(1-\theta)^2/4}\Big(\frac{1}{\tilde{\mu}(A(x_n))}\Big)^{\theta}\Big(\frac{1}{\tilde{\mu}(B)}\Big)^{1-\theta}. \tag{17}$$

Zauważmy teraz, że

$$\inf_{\theta\in[0,1]} e^{(1-\theta)^2/4} u^{-\theta} \leqslant 2 - u \text{ dla } u \in [0,1]. \tag{18}$$

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 (17) oraz (18) 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

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

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

Przykład. Niech $X = \{0,1\}^n$ oraz $\mu = \mu_p^n$, gdzie $\mu_p = p\delta_1 + (1-p)\delta_0$. Załóżmy, że zbiór $A \subset \{0,1\}^n$ jest *monotonicznie dziedziczny*, w sensie

$$x \in A, y \in \{0,1\}^n, y \leqslant x \ \Rightarrow\ y \in A.$$

Niech dla $x \in X$

$$N(x) := \#\{1 \leqslant i \leqslant n \colon x_i = 1\},$$

wówczas

$$d_H(x, A) \leqslant \mathcal{D}_A^c(x)\sqrt{N(x)},$$

gdzie d_H oznacza metrykę Hamminga. Istotnie, przyjmijmy $a = N(x)^{-1/2}(\mathbb{1}_{\{x_i=1\}})_i$ i weźmy $y \in A$ takie, że

$$d_a(x,y) = \frac{1}{\sqrt{N(x)}}\sum_{x_i=1} \mathbb{1}_{\{y_i \neq x_i\}} \leqslant \mathcal{D}_A^c(x).$$

Z uwagi na monotoniczną dziedziczność A możemy przyjąć, że $y_i = 0$ dla $x_i = 0$, zatem

$$d_H(x,y) = \sum_{i=1}^{n} \mathbb{1}_{\{x_i \neq y_i\}} = \sum_{x_i=1} \mathbb{1}_{\{y_i \neq x_i\}} \leqslant \sqrt{N(x)}\mathcal{D}_A^c(x).$$

Stąd dla $s > 0$,

$$\begin{aligned}\mu_p^n(\{d_H(x,A) \geqslant r\}) &\leqslant \mu_p^n(\{\mathcal{D}_A^c(x) \geqslant rs^{-1/2}\}) + \mu_p^n(\{N(x) > s\}) \\ &\leqslant \frac{1}{\mu_p^n(A)} e^{-\frac{r^2}{4s}} + \mu_p^n(\{N(x) > s\}).\end{aligned}$$

Można sprawdzić, że drugi czynnik jest mały dla $s = n\alpha$ z $\alpha > p$.

Twierdzenie 3.41 prowadzi do koncentracji pewnej klasy funkcji lipschitzowskich w odpowiednim sensie . Mianowicie zachodzi

Wniosek 3.42. *Załóżmy, że funkcja $F\colon X \to \mathbb{R}$ spełnia warunek*

$$\forall_{x\in X} \exists_{a=a(x)\in S^{n-1}} \forall_{y \in X}\ F(x) \leqslant F(y) + d_a(x,y). \tag{19}$$

Wówczas dla dowolnej probabilistycznej miary produktowej μ na X,

$$\mu(\{|F - \mathrm{Med}_\mu(F)| \geqslant t\}) \leqslant 4e^{-t^2/4} \ \textit{dla}\ t > 0.$$

Dowód. Dla $m \in \mathbb{R}$ połóżmy $A = \{F \leqslant m\}$, zauważmy, że warunek (19) implikuje, że dla dowolnego $x \in X$,

$$F(x) \leqslant m + d_{a(x)}(x,A) \leqslant m + \mathcal{D}_A^c(x),$$

stąd

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

Zatem dla dowolnego m,

$$\mu(\{F \leqslant m\})\mu(\{F \geqslant m+t\}) \leqslant e^{-t^2/4}.$$

Biorąc $m = \mathrm{Med}_\mu(F)$ i $m = \mathrm{Med}_\mu(F) - t$ dostajemy tezę. $\square$

4 Wybrane Zastosowania

4.1 Twierdzenie Dvoretzky'ego

W tej części wykorzystamy izoperymetrię gaussowską do dowodu twierdzenia Dvoretzky'ego, które mówi, że jeśli wymiar przestrzeni unormowanej jest odpowiednio wysoki, to da się w nią włożyć prawie izometrycznie k-wymiarową przestrzeń euklidesową.

Musimy najpierw wyjaśnić co to znaczy prawie izometryczne włożenie. Ponieważ wszystkie przestrzenie euklidesowe są izometryczne, będziemy mówić o włożeniach l_2^k.

Definicja 4.1. *Powiemy, że przestrzeń l_2^k się wkłada ze stałą C w przestrzeń unormowaną E, jeśli istnieją wektory $v_1, \ldots, v_k \in E$ takie, że*

$$|t| \leqslant \Big\| \sum_{i=1}^{k} t_i v_i \Big\| \leqslant C|t| \ \textit{dla wszystkich } t \in \mathbb{R}^k.$$

Uwaga 4.1 Ostatnią definicję można sformułować równoważnie (biorąc $u_i = v_i/\sqrt{C}$) jako istnienie $u_1, \ldots, u_k \in E$ takich, że

$$\frac{1}{\sqrt{C}}|t| \leqslant \Big\| \sum_{i=1}^{k} t_i u_i \Big\| \leqslant \sqrt{C}|t| \text{ dla wszystkich } t \in S^{k-1}.$$

Definicja 4.2. *Dla wektora gaussowskiego w przestrzeni unormowanej E postaci $X = \sum_{i=1}^{m} v_i g_i$, gdzie $g_1, \ldots, g_n$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$ określmy*

$$\sigma(X) := \sup\big\{(\mathbf{E}\varphi(X)^2)^{1/2} \colon \varphi \in E^*, \|\varphi\| \leqslant 1\big\} \ \textit{oraz } d(X) := \Big(\frac{\mathbf{E}\|X\|}{\sigma(X)}\Big)^2.$$

Dla przestrzeni unormowanej E określamy

$$d(E) = \sup\Big\{d(X) \colon X = \sum_{i=1}^{m} v_i g_i, v_i \in E\Big\}.$$

Fakt 4.1. *Jeśli $X = \sum_{i=1}^{m} v_i g_i$ jest wektorem gaussowskim w E, to*

$$\sigma(X) = \sup_{|t|=1} \Big\| \sum_{i=1}^{m} t_i v_i \Big\|.$$

Dowód. Mamy

$$\sup_{|t|=1}\Big\|\sum_{i=1}^m t_i v_i\Big\| = \sup_{|t|=1}\sup_{\|\varphi\|\leqslant 1}\varphi\Big(\sum_{i=1}^m t_i v_i\Big) = \sup_{\|\varphi\|\leqslant 1}\sup_{|t|=1}\sum_{i=1}^m t_i\varphi(v_i)$$
$$= \sup_{\|\varphi\|\leqslant 1}\Big(\sum_{i=1}^m \varphi^2(v_i)\Big)^{1/2} = \sup_{\|\varphi\|\leqslant 1}(\mathbf{E}\varphi(X)^2)^{1/2} = \sigma(X).$$

□

Jesteśmy teraz gotowi, by sformułować twierdzenie Dvoretzky'ego w wersji pochodzącej od Milmanna.

Twierdzenie 4.2. *Dla dowolnego $\varepsilon > 0$ istnieje stała $c(\varepsilon) > 0$ taka, że dla dowolnej przestrzeni unormowanej E oraz $k \leqslant c(\varepsilon)d(E)$, l_2^k się wkłada ze stałą $1+\varepsilon$ w E.*

Zanim przejdziemy do dowodu tego twierdzenia wprowadzimy jedną ważną definicję i udowodnimy kilka lematów.

Definicja 4.3. *Zbiór $A \subset X$ nazywamy ε-siecią przestrzeni metrycznej (X,d), jeśli dla dowolnego $x \in X$ istnieje $y \in A$ takie, że $d(x,y) \leqslant \varepsilon$.*

Dalej na przestrzeni S^{n-1} będziemy rozważać metrykę dziedziczoną z przestrzeni euklidesowej $\mathbb{R}^n$.

Lemat 4.3. *Dla $0 < \varepsilon \leqslant 1$ istnieje ε-sieć N w S^{n-1} taka, że $\#N \leqslant (\frac{3}{\varepsilon})^n$.*

Dowód. Niech N będzie maksymalnym podzbiorem S^{n-1} którego dowolne dwa elementy są odległe o więcej niż ε. Wówczas N jest ε-siecią oraz kule (w $\mathbb{R}^n$) $(B(x,\varepsilon/2))_{x\in N}$ są rozłączne i zawierają się w kuli $B(0,1+\varepsilon/2)$. Stąd przyjmując $c_n = \lambda_n(B(0,1))$ dostajemy

$$c_n\Big(1+\frac{\varepsilon}{2}\Big)^n = \lambda_n\Big(B\Big(0,1+\frac{\varepsilon}{2}\Big)\Big) \geqslant \lambda_n\Big(\bigcup_{x\in N} B\Big(x,\frac{\varepsilon}{2}\Big)\Big)$$
$$= \sum_{x\in N}\lambda_n\Big(B\Big(x,\frac{\varepsilon}{2}\Big)\Big) = \#N c_n\Big(\frac{\varepsilon}{2}\Big)^n,$$

zatem $\#N \leqslant (\frac{2}{\varepsilon}+1)^n \leqslant (\frac{3}{\varepsilon})^n$. □

Lemat 4.4. *Niech N będzie ε-siecią w S^{n-1} dla pewnego $\varepsilon \in [0,1]$. Wówczas dla dowolnych wektorów $v_1,\ldots,v_n$ w E,*

$$\sup_{|t|=1}\Big\|\sum_{i=1}^n v_i t_i\Big\| \leqslant \frac{1}{1-\varepsilon}\sup_{s\in N}\Big\|\sum_{i=1}^n v_i s_i\Big\|$$

oraz

$$\inf_{|t|=1}\Big\|\sum_{i=1}^n v_i t_i\Big\| \geqslant \inf_{s\in N}\Big\|\sum_{i=1}^n v_i s_i\Big\| - \frac{\varepsilon}{1-\varepsilon}\sup_{s\in N}\Big\|\sum_{i=1}^n v_i s_i\Big\|.$$

Dowód. Z definicji ε-sieci łatwo wynika, że dowolny wektor $t \in S^{n-1}$ można zapisać w postaci $t = s + au$, $s \in N$, $u \in S^{n-1}$, $a \leqslant \varepsilon$, stąd

$$\Big|\Big\|\sum_{i=1}^n v_i t_i\Big\| - \Big\|\sum_{i=1}^n v_i s_i\Big\|\Big| \leqslant a\Big\|\sum_{i=1}^n v_i u_i\Big\| \leqslant \varepsilon \sup_{u\in S^{n-1}}\Big\|\sum_{i=1}^n v_i u_i\Big\|.$$

□

Dowód Twierdzenia 4.2. Bez straty ogólności możemy przyjąć, że $\varepsilon < 1$, dobierzmy $\delta > 0$ takie, że $(1+\delta)/(1-\delta) \leqslant \sqrt{1+\varepsilon}$ oraz $1-\delta-\delta(1+\delta)/(1-\delta) \geqslant 1/\sqrt{1+\varepsilon}$ (np. można przyjąć $\delta = \varepsilon/20$). Niech N będzie δ-siecią w S^{k-1} mocy nie większej niż $(\frac{3}{\delta})^k$. Na podstawie Lematu 4.4 i Uwagi 4.1 wystarczy wykazać, że istnieją wektory $u_1, \ldots, u_k$ w E takie, że

$$1-\delta \leqslant \Big\|\sum_{i=1}^k t_i u_i\Big\| \leqslant 1+\delta \text{ dla } t \in N.$$

Niech $X = \sum_{i=1}^m v_i g_i$ spełnia $d(X) \geqslant d(E)/2$, dla uproszczenia notacji przyjmijmy $d = d(E)$, $\sigma = \sigma(X)$ oraz $M = \mathbf{E}\|X\|$. Niech $F\colon \mathbb{R}^m \to \mathbb{R}$ będzie dane wzorem $F(x) = \|\sum_{i=1}^m v_i x_i\|$. Wówczas na mocy Faktu 4.1 funkcja F jest σ-Lipschitzowska, ponadto $X \sim F(G)$, gdzie $G \sim \gamma_m$. Stąd

$$\mathbf{P}\Big(\Big|\frac{\|X\|}{M} - 1\Big| \geqslant \delta\Big) = \gamma_m\Big(\Big|F(G) - \int F d\gamma_m\Big| \geqslant \delta M\Big) \leqslant 2\exp\Big(-\frac{\delta^2 M^2}{2\sigma^2}\Big),$$

gdzie ostatnia nierówność jest konsekwencją Twierdzeń 3.17 i 3.22. Niech $X_1, \ldots, X_k$ będą niezależnymi kopiami zmiennej X. Wówczas, dla dowolnego $t \in S^{n-1}$, zmienne $\sum_{i=1}^m t_i X_i$ ma ten sam rozkład co X. Zatem

$$\mathbf{P}\Big(\Big|\Big\|\sum_{i=1}^k t_i \frac{X_i}{M}\Big\| - 1\Big| \geqslant \delta\Big) = \mathbf{P}\Big(\Big|\frac{\|X\|}{M} - 1\Big| \geqslant \delta\Big) \leqslant 2e^{-\delta^2 d(X)/2} \leqslant 2e^{-\delta^2 d/4},$$

czyli

$$\mathbf{P}\Big(\exists_{t\in N}\Big|\Big\|\sum_{i=1}^k t_i \frac{X_i}{M}\Big\| - 1\Big| \geqslant \delta\Big) \leqslant 2\#N e^{-\delta^2 d/4} \leqslant 2\Big(\frac{3}{\delta}\Big)^k e^{-\delta^2 d/4} < 1,$$

jeśli tylko $k \leqslant c(\delta)d$, czyli można za u_i przyjąć $X_i(\omega)/M$ dla pewnego ω. □

By móc stosować Twierdzenie 4.2 potrzebujemy oszacowań z dołu $d(E)$. Zacznijmy od geometrycznego lematu.

Lemat 4.5 (Dvoretzky-Rogers)**.** *Niech E będzie n-wymiarową przestrzenią unormowaną. Wówczas istnieją wektory $v_1, \ldots, v_n \in E$ takie, że $\|\sum_{i=1}^n t_i v_i\| \leqslant |t|$ dla $t \in \mathbb{R}^n$ oraz $\|v_i\| \geqslant \frac{n-i+1}{n}$.*

Dowód. Przypomnijmy, że dla $T \colon l_2^n \to E$ określamy

$$\|T\| = \sup\{\|Tx\| \colon |x| \leqslant 1\}.$$

Zbiór operatorów o normie nie większej niż jeden można traktować jako zwarty podzbiór $\mathbb{R}^{n^2}$, w szczególności istnieje operator T taki, że $\|T\| \leqslant 1$ oraz

$$\det(T) = \max\{\det(S) \colon\ S \colon l_2^n \to E, \|S\| \leqslant 1\}.$$

Wówczas dla dowolnego $\varepsilon > 0$,

$$\det(T + \varepsilon S) \leqslant \det(T)\|T + \varepsilon S\|^n.$$

Mamy jednak

$$\det(T + \varepsilon S) = \det(T)\det(I + \varepsilon T^{-1}S) = \det(T)(1 + \varepsilon \mathrm{tr}(T^{-1}S) + o(\varepsilon)),$$

więc

$$1 + \varepsilon \mathrm{tr}(T^{-1}S) \leqslant \|T + \varepsilon S\|^n + o(\varepsilon) \leqslant (1 + \varepsilon\|S\|)^n + o(\varepsilon) \leqslant 1 + \varepsilon n\|S\| + o(\varepsilon).$$

Wykazaliśmy w ten sposób, że $\mathrm{tr}(T^{-1}S) \leqslant n\|S\|$ dla dowolnego $S \colon l_2^n \to E$. Niech $S = TP$, gdzie P oznacza rzut ortogonalny na podrzestrzeń V. Wówczas

$$\|TP\| \geqslant \mathrm{tr}(T^{-1}TP) = \mathrm{tr}(P) = \dim V.$$

Niech $y_1 \in \mathbb{R}^n$ będzie takie, że $|y_1| = 1$ i $\|Ty_1\| = \|T\|$. Jeśli wybraliśmy już $y_1, \ldots, y_k$, to kładziemy $V_k = \{y_1, \ldots, y_k\}^{\perp}$ i wybieramy $y_{k+1} \in V_k$ takie, że $|y_{k+1}| = 1$ oraz $\|Ty_{k+1}\| = \|TP_k\| \geqslant (n-k)/n$, gdzie P_k oznacza rzut ortogonalny na V_k. Wystarczy, że położymy $v_k = Ty_k$. $\square$

Lemat 4.6. *Niech ε_i będą niezależnymi symetrycznymi zmiennymi losowymi przyjmującymi wartości ± 1. Wówczas dla dowolnych wektorów $u_i \in E$, $\mathbf{E}\|\sum_{i=1}^n u_i\varepsilon_i\| \geqslant \max_{1\leqslant i\leqslant n}\|u_i\|$.*

Dowód. Mamy dla dowolnego $1 \leqslant i \leqslant n$,

$$\begin{aligned}\mathbf{E}\Big\|\sum_{k=1}^{n} u_k\varepsilon_k\Big\| &= \mathbf{E}\frac{1}{2}\Big(\Big\|u_i\varepsilon_i + \sum_{k\neq i}^{n} u_k\varepsilon_k\Big\| + \Big\|u_i\varepsilon_i - \sum_{k\neq i}^{n} u_k\varepsilon_k\Big\|\Big)\\ &\geqslant \mathbf{E}\|u_i\varepsilon_i\| = \|u_i\|.\end{aligned}$$

□

Wniosek 4.7. *Dla dowolnej przestrzeni unormowanej E wymiaru n, $d(E) \geqslant \frac{1}{C}\log n$.*

Dowód. Niech v_i będą takie jak w Lemacie 4.5 oraz $X = \sum_{i=1}^{n} v_i g_i$. Wówczas $\sigma(X) = 1$. Niech (ε_i) będzie ciągiem niezależnych symetrycznymi zmiennych losowych przyjmujących wartości ± 1, niezależnym od (g_i). Wówczas na mocy symetrii g_i,

$$\mathbf{E}\|X\| = \mathbf{E}\Big\|\sum_{i=1}^{n} \varepsilon_i g_i\Big\| \geqslant \mathbf{E}\max_{i\leqslant n}\|v_i g_i\| \geqslant \frac{1}{2}\mathbf{E}\max_{i\leqslant\lceil n/2\rceil}|g_i| \geqslant c\sqrt{\log n}.$$

□

Wkładanie l_2^k w przestrzeń E można wysłowić w bardziej geometryczny sposób. Mianowicie obrazem kuli w l_2^k przy liniowym, nieosobliwym przekształceniu jest k-wymiarowa elipsoida. To, że przekształcenie liniowe jest prawie izometrią znaczy, że odpowiednie k-wymiarowe przekroje kuli jednostkowej w E są bliskie tej elipsoidzie. Istnieje też wzajemna odpowiedniość między kulami jednostkowymi w n-wymiarowych przestrzeniach unormowanych a n-wymiarowymi symetrycznymi ciałami wypukłymi (tzn. zwartymi symetrycznymi zbiorami w $\mathbb{R}^n$ o niepustym wnetrzu). Stąd otrzymujemy równoważne sformułowanie twierdzenia Dvoretzky'ego.

Twierdzenie 4.8. *Istnieje stała $c(\varepsilon)$ taka, że dla dowolnego n-wymiarowego symetrycznego ciała wypukłego i $k \leqslant c(\varepsilon)\log n$ istnieje k wymiarowa podprzestrzeń V i elipsoida $\mathcal{E}$ w V taka, że*

$$\mathcal{E} \subset K \cap V \subset (1+\varepsilon)\mathcal{E}.$$

4.2 Wektory i Procesy Gaussowskie

Procesy i zmienne gaussowskie odgrywają kluczową rolę w rachunku prawdopodobieństwa i statystyce matematycznej. W poprzednim paragrafie widzieliśmy też, że wektory gaussowskie pełnią ważną rolę w zastosowaniach geometrycznych. Udowodniliśmy tam też pierwszą wersję koncentracji, która jest kluczowa w rozmaitych zastosowaniach, w tej części podamy jej dokładne sformułowanie.

Zacznijmy od przypomnienia definicji.

Definicja 4.4. *Proces $(G_t)_{t\in T}$ nazywamy* procesem gaussowskim*, jeśli dla dowolnych $t_1,\ldots,t_n\in T$ wektor losowy $(G_{t_1},\ldots,G_{t_n})$ ma rozkład gaussowski. Proces nazywamy* scentrowanym*, jeśli $\mathbf{E}G_t=0$ dla $t\in T$.*

By uniknąć problemów związanych z mierzalnością będziemy zakładać, że zbiór T jest przeliczalny. Alternatywnie można zakładać ośrodkowość procesu.

Twierdzenie 4.9. *Załóżmy, że $(G_t)_{t\in T}$ jest procesem gaussowskim, indeksowanym przez przeliczalny zbiór indeksów T, takim, że $\sup_{t\in T}G_t<\infty$ prawie na pewno. Wówczas $\mathbf{E}\sup_{t\in T}G_t<\infty$ oraz dla $u>0$,*

$$\mathbf{P}\Big(\sup_{t\in T}G_t-\mathbf{E}\sup_{t\in T}G_t\geqslant u\Big)\leqslant\exp\Big(-\frac{u^2}{2\sigma^2}\Big)\tag{20}$$

i

$$\mathbf{P}\Big(\sup_{t\in T}G_t-\mathbf{E}\sup_{t\in T}G_t\leqslant -u\Big)\leqslant\exp\Big(-\frac{u^2}{2\sigma^2}\Big),\tag{21}$$

gdzie

$$\sigma:=\sup_{t\in T}(\mathrm{Var}(G_t))^{1/2}.$$

Dowód. **Krok I.** $T=\{t_1,\ldots,t_n\}$ jest zbiorem skończonym. Wówczas istnieje macierz $A=(a_{ij})_{1\leqslant i\leqslant n,1\leqslant j\leqslant k}$ oraz wektor $m\in\mathbb{R}^n$ takie, że

$$(G_{t_1},\ldots,G_{t_n})\sim m+AX,\quad X\sim\gamma_k.$$

Określmy $F\colon\mathbb{R}^k\to\mathbb{R}$ wzorem

$$F(x)=\max_{1\leqslant i\leqslant n}\Big\{m_i+\sum_{j=1}^k a_{ij}x_j\Big\},$$

wówczas

$$\|F\|_{\mathrm{Lip}}=\max_i\Big(\sum_{j=1}^k a_{ij}^2\Big)^{1/2}=\max_i\mathrm{Var}(G_{t_i})^{1/2}=\sigma.$$

Stąd

$$\mathbf{P}\Big(\sup_{t\in T} G_t - \mathbf{E}\sup_{t\in T} G_t \geqslant u\Big) = \gamma_k\Big(\Big\{F - \int F d\gamma_k \geqslant u\Big\}\Big) \leqslant e^{-u^2/2\sigma^2},$$

gdzie ostatnia nierówność wynika z Twierdzeń 3.17 i 3.22. Korzystając z tego, że funkcja $-F$ też jest σ-lipschitzowska wyprowadzamy (21) w przypadku skończonego T.

Krok II. $T = \{t_1, t_2, \ldots\}$ jest nieskończone. Połóżmy $T_n := \{t_1, \ldots, t_n\}$,

$$M_n := \mathbf{E}\max_{1\leqslant i\leqslant n} G_{t_i} \text{ oraz } \sigma_n := \max_{1\leqslant i\leqslant n} \mathrm{Var}(G_{t_i})^{1/2}.$$

Niech A spełnia $\mathbf{P}(\sup_{t\in T} G_t \geqslant A) \leqslant 1/4$. Z Kroku I dostajemy

$$\mathbf{P}\Big(\max_{1\leqslant i\leqslant n} G_{t_i} - M_n \leqslant -\sigma_n\Big) \leqslant e^{-1/2},$$

zatem

$$\mathbf{P}\Big(\max_{1\leqslant i\leqslant n} G_{t_i} > M_n - \sigma_n\Big) \geqslant 1 - e^{-1/2} > \frac{1}{4} \geqslant \mathbf{P}\Big(\sup_{t\in T} G_t \geqslant A\Big),$$

stąd $A \geqslant M_n - \sigma_n$. Mamy więc

$$\mathbf{E}\sup_{t\in T} G_t = \sup_n M_n \leqslant \sup_n\{A + \sigma_n\} \leqslant A + \sigma < \infty.$$

Ponadto z Kroku I wynika

$$\mathbf{P}\Big(\max_{1\leqslant i\leqslant n} G_{t_i} \geqslant M_n + u\Big) \leqslant e^{-u^2/2\sigma_n^2} \leqslant e^{-u^2/2\sigma^2}$$

i przechodząc z $n \to \infty$ otrzymujemy (20). Podobnie wyprowadzamy (21). $\square$

Uwaga 4.2 Łącząc (20) i (21) dostajemy

$$\mathbf{P}\Big(\Big|\sup_{t\in T} G_t - \mathbf{E}\sup_{t\in T} G_t\Big| \geqslant u\Big) \leqslant 2\exp\Big(-\frac{u^2}{2\sigma^2}\Big) \text{ dla } u > 0.$$

Zauważmy też, że $|G_t| = \max\{G_t, -G_t\}$, więc w (20), (21) i powyższej nierówności można zastąpić G_t przez $|G_t|$.

Uwaga 4.3 Korzystając z izoperymetrii gaussowskiej (Wniosek 2.11) zamiast nierówności logarytmicznej Sobolewa możemy udowodnić, że przy oznaczeniach Twierdzenia 4.9 dla $u > 0$,

$$\mathbf{P}\Big(\sup_{t\in T} G_t - \mathrm{Med}\Big(\sup_{t\in T} G_t\Big) \geqslant u\Big) \leqslant \Phi\Big(\frac{u}{\sigma}\Big) \leqslant \frac{1}{2}\exp\Big(-\frac{u^2}{2\sigma^2}\Big)$$

oraz

$$\mathbf{P}\Big(\Big|\sup_{t\in T} G_t - \mathrm{Med}\Big(\sup_{t\in T} G_t\Big)\Big| \geqslant u\Big) \leqslant 2\Phi\Big(\frac{u}{\sigma}\Big) \leqslant \exp\Big(-\frac{u^2}{2\sigma^2}\Big).$$

Wniosek 4.10. *Przy założeniach i oznaczeniach Twierdzenia 4.9,*

$$\lim_{t\to\infty} \frac{1}{u^2} \log \mathbf{P}\Big(\sup_{t\in T} G_t \geqslant u\Big) = -\frac{1}{2\sigma^2}.$$

Ponadto,

$$\mathbf{E}\exp\Big(\alpha \sup_{t\in T} G_t^2\Big) < \infty$$

wtedy i tylko wtedy gdy $\alpha < \frac{1}{2\sigma^2}$.

Dowód. Z Twierdzenia 4.9

$$\frac{1}{u^2} \log \mathbf{P}\Big(\sup_{t\in T} G_t \geqslant u\Big) \leqslant -\frac{1}{2\sigma^2}.$$

Z drugiej strony,

$$\begin{aligned}\liminf_{t\to\infty} \frac{1}{u^2} \log \mathbf{P}\Big(\sup_{t\in T} G_t \geqslant u\Big) &\geqslant \sup_{t\in T} \liminf_{t\to\infty} \frac{1}{u^2} \log \mathbf{P}(G_t \geqslant u) \\ &= \sup_{t\in T} -\frac{1}{2\mathrm{Var}(G_t)} = -\frac{1}{2\sigma^2}.\end{aligned}$$

Druga część tezy dla $\alpha < \frac{1}{2\sigma^2}$ wynika natychmiast z (20) (dla $|G_t|$). Ponadto, jeśli $G_t \sim \mathcal{N}(a_t, \sigma_t^2) \sim a_t + \sigma_t g$, to dla $0 \leqslant \alpha < 1/2\sigma_t^2$

$$\mathbf{E}e^{\alpha G_t^2} \geqslant \mathbf{E}e^{\alpha\sigma_t^2 g^2} \mathbb{1}_{g\geqslant 0} = \frac{1}{2}\mathbf{E}e^{\alpha\sigma_t^2 g^2} = \frac{1}{2}\sqrt{\frac{1}{1-2\alpha\sigma_t^2}},$$

więc $\mathbf{E}\exp(\alpha \sup_{t\in T} G_t^2) \geqslant \sup_{t\in T} \mathbf{E}\exp(\alpha G_t^2) = \infty$ dla $\alpha \geqslant \frac{1}{2\sigma^2}$. □

Definicja 4.5. *Wektor losowy X w ośrodkowej przestrzeni Banacha F nazywamy* gaussowskim, *jeśli dla dowolnego $\varphi \in F^*$, $\varphi(X)$ ma rozkład gaussowski.*

Założenie o ośrodkowości F ma charakter techniczny, służy uniknięciu problemów z mierzalnością (w nieośrodkowej przestrzeni Banacha suma dwóch wektorów losowych nie musi być mierzalna). Alternatywnie można zakładać, że norma w F jest wybijana przez przeliczalny ciąg funkcjonałów o normie jeden.

Twierdzenie 4.11. *Załóżmy, że X jest wektorem gaussowskim w ośrodkowej przestrzeni Banacha. Wówczas $\mathbf{E}\|X\| < \infty$ oraz dla $u > 0$,*

$$\mathbf{P}(\|X\| - \mathbf{E}\|X\| \geqslant u) \leqslant e^{-\frac{u^2}{2\sigma^2}} \text{ oraz } \mathbf{P}(\|X\| - \mathbf{E}\|X\| \leqslant -u) \leqslant e^{-\frac{u^2}{2\sigma^2}},$$

gdzie

$$\sigma := \{\mathrm{Var}(\varphi(X))^{1/2} \colon \varphi \in F^*, \|\varphi\| \leqslant 1\}.$$

Dowód. Wystarczy zauważyć, że istnieje przeliczalny podzbiór D kuli jednostkowej w F^* taki, że $\|x\| = \sup_{\varphi \in D} \varphi(X)$ i skorzystać z Twierdzenia 4.9 dla procesu gaussowskiego $(\varphi(X))_{\varphi \in D}$. □

Wniosek 4.12. *Przy oznaczeniach Twierdzenia 4.11 dla $p \geqslant 1$,*

$$(\mathbf{E}\|X\|^p)^{1/p} \leqslant \mathbf{E}\|X\| + C\sqrt{p}\sigma,$$

gdzie C jest pewną stałą uniwersalną.

Uwaga 4.4 Jak nietrudno zauważyć

$$(\mathbf{E}\|X\|^p)^{1/p} \geqslant \sup_{\|\varphi\| \leqslant 1} (\mathbf{E}|\varphi(X)|^p)^{1/p} \geqslant \frac{1}{C}\sqrt{p}\sigma.$$

Wynika też stąd, że

$$(\mathbf{E}\|X\|^p)^{1/p} \leqslant \mathbf{E}\|X\| + \tilde{C} \sup_{\|\varphi\| \leqslant 1} (\mathbf{E}|\varphi(X)|^p)^{1/p}.$$

4.3 Procesy Empiryczne

4.3.1 Sumy niezależnych zmiennych losowych

Zacznijmy od przypomnienia trzech nierówności dla sum niezależnych, rzeczywistych, ograniczonych zmiennych losowych.

Zakładamy, że zmienne $X_1, X_2, \ldots, X_n$ są niezależne oraz $S = \sum_{i=1}^n X_i$.

Twierdzenie 4.13 (Nierówność Hoeffdinga)**.** *Dla dowolnego $t > 0$*

$$\mathbf{P}(S - \mathbf{E}S \geqslant t) \leqslant \exp\Big(-\frac{t^2}{2D^2}\Big),$$

gdzie $D = (\sum_{i=1}^n \|X_i - \mathbf{E}X_i\|_\infty^2)^{1/2}$.

Nierówność Hoeffdinga daje dobre oszacowanie tylko, gdy D^2 jest zbliżona do $\mathrm{Var}(S)$. W przeciwnym przypadku dobrze jest stosować precyzyjniejsze nierówności.

Twierdzenie 4.14 (Nierówność Bernsteina)**.** *Załóżmy, że* $|X_i - \mathbf{E}X_i| \leqslant C$ *dla* $1 \leqslant i \leqslant n$*. Wówczas dla* $t \geqslant 0$*,*

$$\mathbf{P}(S - \mathbf{E}S \geqslant t) \leqslant \exp\Big(-\frac{t^2}{2\mathrm{Var}(S) + \frac{2}{3}Ct}\Big).$$

Nierówność Bernsteina dla t małych daje ogon gaussowski S, a dla t dużych ogon wykładniczy. Okazuje się, że można ją polepszyć by dla dużych t otrzymać ogon poissonowski.

Twierdzenie 4.15 (Nierówność Bennetta)**.** *Załóżmy, że* $|X_i - \mathbf{E}X_i| \leqslant C$ *dla* $1 \leqslant i \leqslant n$*. Wówczas dla* $t \geqslant 0$*,*

$$\mathbf{P}(S-\mathbf{E}S \geqslant t) \leqslant \exp\Big(-\frac{\mathrm{Var}(S)}{C^2}h\Big(\frac{Ct}{\mathrm{Var}(S)}\Big)\Big) \leqslant \exp\Big(-\frac{t}{2C}\ln\Big(1+\frac{tC}{\mathrm{Var}(S)}\Big)\Big),$$

gdzie $h(x) := (1+x)\ln(1+x) - x$ *dla* $x \geqslant 0$*.*

4.3.2 Oszacowania supremów procesów empirycznych

W tej części będziemy zakładać, że X_i są niezależnymi zmiennymi losowymi o wartościach w pewnej przestrzeni mierzalnej $(V, \mathcal{V})$, zaś $\mathcal{F}$ pewną przeliczalną rodziną funkcji mierzalnych na V. Proces postaci

$$\Big(\sum_{i=1}^n f(X_i)\Big)_{f\in\mathcal{F}}$$

nazywamy *procesem empirycznym.* Będzie nas interesowało poszukiwanie oszacowań supremum takiego procesu, to znaczy zmiennej

$$Z = \sup_{f\in\mathcal{F}} \sum_{i=1}^n f(X_i).$$

Na pierwszy rzut oka powyższa definicja nie wydaje się naturalna. Poniższe dwa przykłady powinny pomóc w zrozumieniu motywacji za nią stojących.

Przykład 1. Sumy niezależnych wektorów losowych. Załóżmy, że X_i są niezależnymi wektorami losowymi o wartościach w pewnej ośrodkowej przestrzeni Banacha F i niech $S = \sum_{i=1}^n X_i$. Wybierzmy taki przeliczalny zbiór funcjonałów D w kuli jednostkowej F^*, by $\|x\| = \sup_{\varphi \in D} \varphi(x)$ dla wszystkich $x \in F$. Wówczas

$$\|S\| = \Big\| \sum_{i=1}^n X_i \Big\| = \sup_{\varphi \in D} \sum_{i=1}^n \varphi(X_i).$$

Przykład 2. Rozkłady empiryczne. Niech μ bedzie pewnym (ustalonym, ale nieznanym) rozkładem probabilistycznym na $(V, \mathcal{V})$, zaś $X_1, X_2, \ldots$ niezależnymi zmiennymi losowymi o rozkładzie μ. Wówczas μ możemy estymować przez *rozkład empiryczny*

$$\mu_n := \frac{1}{n} \sum_{i=1}^n \delta_{X_i}.$$

Naturalne pytanie brzmi: czy μ_n zbiega do μ i jak szybko? Przykładowo, twierdzenie Gliwienki-Cantelli mówi, że jeśli $V = \mathbb{R}$, to dystrybuanta μ_n zbiega do dystrybuanty μ jednostajnie prawie na pewno. By badać szybkość zbieżności ustala się pewną klasę funkcji $\mathcal{F}$ na V i bada się

$$\sup_{f \in \mathcal{F}} \Big| \int f d\mu_n - \int f d\mu \Big| = \sup_{f \in \mathcal{F}} \Big| \sum_{i=1}^n \frac{1}{n} (f(X_i) - \int f d\mu) \Big|.$$

Dowody twierdzeń, które omówimy w tym paragrafie będą opartę o metodę entropijną, którą już wcześniej stosowaliśmy przy badaniu nierówności logarytmicznej Sobolewa.

Twierdzenie 4.16. *Załóżmy, że $0 \leqslant f \leqslant 1$ dla $f \in \mathcal{F}$. Wówczas dla dowolnego $\lambda \geqslant 0$,*

$$\mathbf{E} e^{\lambda Z} \leqslant e^{(e^\lambda - 1)\mathbf{E}Z}.$$

W szczególności dla $t \geqslant 0$,

$$\mathbf{P}(Z - \mathbf{E}Z \geqslant t) \leqslant \exp\Big(- \mathbf{E}Z h\Big(\frac{t}{\mathbf{E}Z}\Big)\Big) \leqslant \exp\Big(- \frac{t}{2} \log\Big(1 + \frac{t}{\mathbf{E}Z}\Big)\Big),$$

gdzie $h(x) = (1 + x)\log(1 + x) - x$ dla $x \geqslant 0$.

Lemat 4.17. *Niech $\Phi(u) = e^{-u} + u - 1$, wówczas dla $\lambda \geqslant 0$,*

$$\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) \leqslant \Phi(\lambda)\mathbf{E}Z e^{\lambda Z}.$$

Dowód. Stosując nierówność $\log x \leqslant x - 1$ dla $t/x = \int e^f d\mu$, a następnie podstawienie $t = e^u$ dostajemy dla dowolnej miary probabilistycznej μ,

$$\begin{aligned}\mathrm{Ent}_\mu(e^f) &= \int e^f(f - \log \int e^f) d\mu = \inf_{t>0} \int [fe^f - (\log t + 1)e^f + t] d\mu \\ &= \inf_{u\in\mathbb{R}} \int \Phi(f-u)e^f d\mu. \end{aligned} \tag{22}$$

Bez straty ogólności możemy przyjąć, że $\#\mathcal{F} = N < \infty$, niech μ_i oznacza rozkład $(f(X_i))_{f\in\mathcal{F}}$ na $E = [0,1]^N$ oraz $\mu = \mu_1 \otimes \cdots \otimes \mu_n$. Określmy też

$$Z(x) = \max_{1\leqslant k\leqslant N} \sum_{i=1}^n x_i^k \text{ dla } x = (x_1, \ldots, x_n) \in E^n.$$

Wówczas rozkład Z względem $\mathbf{P}$ jest taki sam jak $Z(x)$ względem μ, czyli

$$\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) = \mathrm{Ent}_\mu(e^{\lambda Z(x)}) \leqslant \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x).$$

Dla $x = (x_1, \ldots, x_n) \in E^n$ określmy

$$y_i(x) = (x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_n) \in E^n, \quad 1 \leqslant i \leqslant n,$$

wtedy z nierówności (22),

$$\int \mathrm{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x) \leqslant \int \Phi(\lambda(Z(x) - Z(y_i(x))))e^{\lambda Z(x)} d\mu.$$

Określmy mierzalne rozbicie E^n na zbiory A_k, $1 \leqslant k \leqslant N$ spełniające

$$A_k \subset \Big\{x \in E^n \colon Z(x) = \sum_{i=1}^n x_i^k\Big\}.$$

Wówczas

$$0 \leqslant Z(x) - Z(y_i(x)) \leqslant \sum_{k=1}^N \mathbb{1}_{A_k}(x) x_i^k.$$

Z wypukłości Φ, $\Phi(\lambda u) \leqslant u\Phi(\lambda)$ dla $u \in [0,1]$, $\lambda \geqslant 0$, zatem

$$\Phi(\lambda(Z(x) - Z(y_i(x)))) \leqslant \sum_{k=1}^N \mathbb{1}_{A_k}(x) x_i^k \Phi(\lambda).$$

W efekcie,

$$\begin{aligned}
\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) &\leqslant \sum_{i=1}^{n}\int\sum_{k=1}^{N}\mathbb{1}_{A_k}(x)x_i^k\Phi(\lambda)e^{\lambda Z(x)}d\mu \\
&= \Phi(\lambda)\int\sum_{k=1}^{N}\mathbb{1}_{A_k}(x)\sum_{i=1}^{n}x_i^k e^{\lambda Z(x)}d\mu = \Phi(\lambda)\int Z(x)e^{\lambda Z(x)}d\mu \\
&= \Phi(\lambda)\mathbf{E}Ze^{\lambda Z}.
\end{aligned}$$

□

Dowód Twierdzenia 4.16. Bez straty ogólności możemy założyć, że Z nie jest zdegenerowane. Niech $M(\lambda) = \mathbf{E}e^{\lambda Z}$, wówczas na mocy Lematu 4.17 dostajemy dla $\lambda > 0$,

$$\lambda M'(\lambda) - M(\lambda)\log M(\lambda) = \mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) \leqslant \Phi(\lambda)\mathbf{E}Ze^{\lambda Z} = \Phi(\lambda)M'(\lambda),$$

co po uwzględnieniu definicji Φ daje

$$(1 - e^{-\lambda})M'(\lambda) \leqslant M(\lambda)\log M(\lambda).$$

Określmy $J(\lambda) = \log M(\lambda)$, wówczas

$$J'(\lambda) \leqslant \frac{e^{\lambda}}{e^{\lambda}-1}J(\lambda).$$

Zauważmy, że wobec $Z \geqslant 0$, $J(\lambda) \geqslant 0$ i nierówność jest ostra dla $\lambda > 0$. Poprzednia nierówność daje dla $\lambda > 0$,

$$(\log J(\lambda) - \log(e^{\lambda} - 1))' \leqslant 0,$$

zatem

$$\frac{J(\lambda)}{e^{\lambda}-1} \leqslant \lim_{\lambda\to 0+}\frac{J(\lambda)}{e^{\lambda}-1} = \mathbf{E}Z.$$

Stąd $M(\lambda) \leqslant \exp((e^{\lambda} - 1)\mathbf{E}Z)$, drugą część twierdzenia otrzymujemy w standardowy sposób z nierówności Czebyszewa:

$$\begin{aligned}
\mathbf{P}(Z - \mathbf{E}Z \geqslant t) &\leqslant \inf_{\lambda\geqslant 0} e^{-\lambda t + \mathbf{E}Z}\mathbf{E}e^{\lambda Z} \leqslant \exp\Big(-\lambda t + (e^{\lambda} - \lambda - 1)\mathbf{E}Z\Big) \\
&= \exp\Big(-\mathbf{E}Zh\Big(\frac{t}{\mathbf{E}Z}\Big)\Big).
\end{aligned}$$

Ostatnia nierówność w twierdzeniu wynika z szacowania $h(x) \geqslant \frac{x}{2}\log(1 + x)$. □

Twierdzenie 4.18. *Załóżmy, że $|f| \leqslant C$ dla wszystkich $f \in \mathcal{F}$. Wówczas dla $t \geqslant 0$,*

$$\mathbf{P}\Big(|Z - \mathbf{E}Z| \geqslant t\Big) \leqslant 2\exp\Big(-\frac{1}{10}\min\Big\{\frac{t}{C}, \frac{t^2}{4\mathbf{E}\Sigma^2}\Big\}\Big),$$

gdzie $\Sigma^2 = \sup_{f\in\mathcal{F}} \sum_{i=1}^n f^2(X_i)$.

Lemat 4.19. *Dla dowolnej funkcji F*

$$\mathrm{Ent}_\mu(e^F) \leqslant \iint ((F(x) - F(y))_+)^2 e^{F(x)} d\mu(x) d\mu(y).$$

Dowód. Na mocy nierówności Jensena

$$\begin{aligned}\mathrm{Ent}_\mu(e^F) &= \int Fe^F d\mu - \int e^F d\mu \log \int e^F d\mu \leqslant \int Fe^F d\mu - \int e^F d\mu \int F d\mu \\ &= \frac{1}{2}\iint (F(x) - F(y))(e^{F(x)} - e^{F(y)}) d\mu(x) d\mu(y).\end{aligned}$$

Zauważmy, że dla dowolnych liczb a, b

$$(a - b)(e^a - e^b) \leqslant (a - b)^2 e^{\max\{a,b\}} = ((a - b)_+)^2 e^a + ((b - a)_+)^2 e^b.$$

Stąd

$$\begin{aligned}&\frac{1}{2}\iint (F(x) - F(y))(e^{F(x)} - e^{F(y)}) d\mu(x) d\mu(y) \\ &\leqslant \frac{1}{2}\iint \big[((F(x) - F(y))_+)^2 e^{F(x)} + ((F(y) - F(x))_+)^2 e^{F(y)}\big] d\mu(x) d\mu(y) \\ &= \iint ((F(x) - F(y))_+)^2 e^{F(x)} d\mu(x) d\mu(y).\end{aligned}$$

□

Lemat 4.20. *Załóżmy, że $|f| \leqslant 1$ dla wszystkich $f \in \mathcal{F}$, wówczas dla dowolnego $\lambda \in \mathbb{R}$,*

$$\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) \leqslant 2\lambda^2 e^{2|\lambda|}(\mathbf{E}\Sigma^2 \mathbf{E}e^{\lambda Z} + \mathbf{E}(\Sigma^2 e^{\lambda Z})).$$

Dowód. Bez straty ogólności możemy przyjąć, że $\#\mathcal{F} = N < \infty$. Jak w dowodzie Lematu 4.17 niech μ_i oznacza rozkład $(f(X_i))_{f\in\mathcal{F}}$ na $E = [-1, 1]^N$ oraz $\mu = \mu_1 \otimes \cdots \otimes \mu_n$. Określmy też

$$Z(x) = \max_{1\leqslant k\leqslant N} \sum_{i=1}^n x_i^k \text{ dla } x = (x_1, \ldots, x_n) \in E^n.$$

Wówczas rozkład Z względem $\mathbf{P}$ jest taki sam jak $Z(x)$ względem μ, czyli

$$\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z}) = \mathrm{Ent}_{\mu}(e^{\lambda Z(x)}) \leqslant \sum_{i=1}^{n} \int \mathrm{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x).$$

Określmy mierzalne rozbicie E^n na zbiory A_k, $1 \leqslant k \leqslant N$ spełniające

$$A_k \subset \Big\{ x \in E^n \colon Z(x) = \sum_{i=1}^{n} x_i^k \Big\}.$$

Dla ustalonych $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) \in E^{n-1}$ niech

$$Z_i(z) = Z_i(x_1, \dots, x_{i-1}, z, x_{i+1}, \dots, x_n) \text{ dla } z \in E,$$

wówczas na podstawie Lematu 4.19,

$$\int \mathrm{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x)$$
$$\leqslant \iiint ((\lambda(Z_i(x_i) - Z_i(y_i)))_+)^2 e^{\lambda Z_i(x_i)} d\mu_i(x_i) d\mu_i(y_i) d\mu(x).$$

Dla $\lambda \geqslant 0$, $((\lambda(Z_i(x_i) - Z_i(y_i)))_+)^2 = \lambda^2((Z_i(x_i) - Z_i(y_i))_+)^2$, zaś dla $\lambda < 0$,

$$\begin{aligned} ((\lambda(Z_i(x_i) - Z_i(y_i)))_+)^2 e^{\lambda Z_i(x_i)} &= \lambda^2((Z_i(y_i) - Z_i(x_i))_+)^2 e^{\lambda Z_i(x_i)} \\ &\leqslant \lambda^2 e^{2|\lambda|}((Z_i(y_i) - Z_i(x_i))_+)^2 e^{\lambda Z_i(y_i)} \end{aligned}$$

gdyż $|Z_i(x_i) - Z_i(y_i)| \leqslant 2$ wobec $|f| \leqslant 1$. Stąd

$$\int \mathrm{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x)$$
$$\leqslant \lambda^2 e^{2|\lambda|} \iiint ((Z_i(x_i) - Z_i(y_i))_+)^2 e^{\lambda Z_i(x_i)} d\mu_i(x_i) d\mu_i(y_i) d\mu(x).$$

Zauważmy, że

$$Z_i(x_i) - Z_i(y_i) \leqslant \sum_{k=1}^{N} \mathbb{1}_{A_k}(x)(x_i^k - y_i^k),$$

więc

$$((Z_i(x_i) - Z_i(y_i))_+)^2 \leqslant 2 \sum_{k=1}^{N} \mathbb{1}_{A_k}(x)(|x_i^k|^2 + |y_i^k|^2).$$

Zatem

$$
\begin{aligned}
&\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z})\\
&\leqslant 2\lambda^2 e^{2|\lambda|}\sum_{i=1}^{n}\iiint\sum_{k=1}^{N}\mathbb{1}_{A_k}(x)(|x_i^k|^2+|y_i^k|^2)e^{\lambda Z_i(x_i)}d\mu_i(x_i)d\mu_i(y_i)d\mu(x)\\
&=2\lambda^2 e^{2|\lambda|}\iint\sum_{k=1}^{N}\mathbb{1}_{A_k}(x)\sum_{i=1}^{n}(|x_i^k|^2+|y_i^k|^2)e^{\lambda Z(x)}d\mu(x)d\mu(y)\\
&\leqslant 2\lambda^2 e^{2|\lambda|}\iint(\max_{1\leqslant k\leqslant N}|x_i^k|^2+\max_{1\leqslant k\leqslant N}|x_i^k|^2)e^{\lambda Z(x)}d\mu(x)d\mu(y)\\
&=(\mathbf{E}\Sigma^2\mathbf{E}e^{\lambda Z}+\mathbf{E}(\Sigma^2 e^{\lambda Z})).
\end{aligned}
$$

□

Dowód Twierdzenia 4.18. Z uwagi na jednorodność możemy założyć, że $C = 1$. Przyjmijmy $\tilde{Z} := Z - \mathbf{E}Z$ oraz $\widetilde{M}(\lambda) := \mathbf{E}e^{\lambda\tilde{Z}}$. Ustalmy $\lambda_0 > 0$ takie, że $c_0 := 2e^{2\lambda_0}$ spełnia $c_0\lambda_0^2 < 1$. Lemat 4.20 daje dla $|\lambda| \leqslant \lambda_0$,

$$\lambda\widetilde{M}'(\lambda)-\widetilde{M}(\lambda)\log\widetilde{M}(\lambda)=e^{-\lambda\tilde{Z}}\mathrm{Ent}_{\mathbf{P}}(e^{\lambda Z})\leqslant c_0\lambda^2\big(\mathbf{E}\Sigma^2\widetilde{M}(\lambda)+\mathbf{E}(\Sigma^2e^{\lambda\tilde{Z}})\big).$$

Oszacujmy ostatni składnik, mamy

$$
\begin{aligned}
\mathbf{E}(\Sigma^2 e^{\lambda\tilde{Z}}) &= (e-1)\mathbf{E}\Sigma^2\mathbf{E}e^{\lambda\tilde{Z}}+\mathbf{E}([\Sigma^2-(e-1)\mathbf{E}\Sigma^2]e^{\lambda\tilde{Z}})\\
&\leqslant (e-1)\mathbf{E}\Sigma^2\mathbf{E}e^{\lambda\tilde{Z}}+\lambda\mathbf{E}(\tilde{Z}e^{\lambda\tilde{Z}})-\mathbf{E}e^{\lambda\tilde{Z}}+\mathbf{E}e^{\Sigma^2-(e-1)\mathbf{E}\Sigma^2},
\end{aligned}
$$

gdzie ostatnia nierówność wynika z nierówności Younga (11) $uv \leqslant u\log u - u + e^v$ dla $u = e^{\lambda\tilde{Z}}$, $v = \Sigma^2 - (e-1)\mathbf{E}\Sigma^2$. Zauważmy, że $0 \leqslant f^2 \leqslant 1$ dla $f \in \mathcal{F}$, więc do Σ^2 możemy stosować Twierdzenie 4.16, aby dostać $\mathbf{E}e^{\Sigma^2-(e-1)\mathbf{E}\Sigma^2} \leqslant 1$. Z nierówności Jensena, $\mathbf{E}e^{\lambda\tilde{Z}} \geqslant e^{\lambda\mathbf{E}\tilde{Z}} = 1$, zatem

$$\mathbf{E}(\Sigma^2 e^{\lambda\tilde{Z}}) \leqslant (e-1)\mathbf{E}\Sigma^2\widetilde{M}(\lambda)+\lambda\widetilde{M}'(\lambda).$$

Na podstawie wcześniejszych oszacowań dostajemy dla $|\lambda| \leqslant \lambda_0$,

$$\lambda\widetilde{M}'(\lambda)-\widetilde{M}(\lambda)\log\widetilde{M}(\lambda)\leqslant c_0\lambda^2(e\mathbf{E}\Sigma^2\widetilde{M}(\lambda)+\lambda\widetilde{M}'(\lambda)).$$

Określmy $H(\lambda) = \frac{1}{\lambda}\log\widetilde{M}(\lambda)$ dla $\lambda \neq 0$ i $H(0) = 0$, wówczas

$$H'(\lambda)\leqslant c_0\Big(e\mathbf{E}\Sigma^2+\lambda\frac{\widetilde{M}'(\lambda)}{\widetilde{M}(\lambda)}\Big). \tag{23}$$

Zauważmy, że $\widetilde{M}''(\lambda) = \mathbf{E}\tilde{Z}^2 e^{\lambda\tilde{Z}} \geqslant 0$ oraz $\widetilde{M}'(\lambda) = \mathbf{E}\tilde{Z} = 0$, więc dla $\lambda \geqslant 0$, $\widetilde{M}'(\lambda) \geqslant 0$ oraz

$$\int_0^\lambda s\frac{\widetilde{M}'(s)}{\widetilde{M}(s)}ds \leqslant \lambda\int_0^\lambda \frac{\widetilde{M}'(s)}{\widetilde{M}(s)}ds = \lambda\log\widetilde{M}(\lambda).$$

Podobnie

$$\int_{-\lambda}^0 s\frac{\widetilde{M}'(s)}{\widetilde{M}(s)}ds \leqslant -\lambda\log\widetilde{M}(\lambda).$$

Odcałkowanie (23) daje zatem dla $|\lambda| \leqslant \lambda_0$,

$$\log\widetilde{M}(\lambda) \leqslant c_0(e\lambda^2\mathbf{E}\Sigma^2 + \lambda^2\log\widetilde{M}(\lambda)).$$

Stąd

$$\widetilde{M}(\lambda) \leqslant e^{\kappa_0\lambda^2\mathbf{E}\Sigma^2}$$

dla $\kappa_0 := ec_0(1 - c_0\lambda_0^2)^{-1}$. Nierówność Czebyszewa implikuje dla $t \geqslant 0$,

$$\mathbf{P}(Z - \mathbf{E}Z \geqslant t) \leqslant \inf_{0\leqslant\lambda\leqslant\lambda_0} e^{-\lambda t + \kappa_0\lambda^2\mathbf{E}\Sigma^2}.$$

Wybierając $\lambda = t/2\kappa_0\mathbf{E}\Sigma^2$ dla $t \leqslant 2\kappa_0\lambda_0\mathbf{E}\Sigma^2$ i $\lambda = \lambda_0$ dla $t \geqslant 2\kappa_0\lambda_0\mathbf{E}\Sigma^2$ dostajemy

$$\mathbf{P}(Z - \mathbf{E}Z \geqslant t) \leqslant \exp\Big(-\min\Big\{\frac{\lambda_0 t}{2}, \frac{t^2}{4\kappa_0\mathbf{E}\Sigma^2}\Big\}\Big).$$

Podobny argument pracuje dla $\mathbf{P}(Z - \mathbf{E}Z \leqslant -t)$. Wybierając $\lambda_0 = 1/5$ i sprawdzając, że wtedy $\kappa_0 \leqslant 10$ dostajemy tezę. □

Twierdzenie 4.21. *Załóżmy, że $|f| \leqslant C$ dla wszystkich $f \in \mathcal{F}$. Wówczas dla $t \geqslant 0$,*

$$\mathbf{P}\Big(|Z - \mathbf{E}Z| \geqslant t\Big) \leqslant 3\exp\Big(-\frac{t}{KC}\log\Big(1 + \frac{Ct}{\mathbf{E}\Sigma^2}\Big)\Big),$$

gdzie $\Sigma^2 = \sup_{f\in\mathcal{F}}\sum_{i=1}^n f^2(X_i)$, a $K \leqslant 640$ jest stałą numeryczną.

Dowód. Bez straty ogólności możemy założyć, że $C = 1$. Nasz dowód będzie się opierał na Twierdzeniu 4.18 i odpowiednim obcinaniu. Dla $\rho > 0$ określmy

$$\mathcal{F}_\rho := \{f\mathbb{1}_{\{|f|\leqslant\rho\}} : f \in \mathcal{F}\}, \quad Z_\rho := \sup_{f\in\mathcal{F}_\rho}\sum_{i=1}^n f(X_i)$$

oraz

$$W_\rho := \sup_{f\in\mathcal{F}} \sum_{i=1}^n |f(X_i)|\mathbb{1}_{\{|f(X_i)|\geqslant\rho\}}.$$

Wówczas dla $t > 0$,

$$\mathbf{P}(|Z - \mathbf{E}Z| \geqslant 4t) \leqslant \mathbf{P}(|Z_\rho - \mathbf{E}Z_\rho| \geqslant t) + \mathbf{P}(W_\rho + \mathbf{E}W_\rho \geqslant 3t).$$

Stosując Twierdzenie 4.18 do Z_ρ dostajemy

$$\mathbf{P}(|Z_\rho - \mathbf{E}Z_\rho| \geqslant t) \leqslant 2\exp\Big(-\frac{1}{10}\min\Big\{\frac{t}{\rho}, \frac{t^2}{4\mathbf{E}\Sigma^2}\Big\}\Big).$$

By oszacować W_ρ załóżmy, że $t \geqslant \mathbf{E}W_\rho$ i zastosujmy Twierdzenie 4.16,

$$\mathbf{P}(W_\rho + \mathbf{E}W_\rho \geqslant 3t) \leqslant \mathbf{P}(W_\rho - \mathbf{E}W_\rho \geqslant t) \leqslant \exp\Big(-\frac{t}{2}\log\Big(1 + \frac{t}{\mathbf{E}W_\rho}\Big)\Big).$$

Wybierzmy teraz

$$\rho = \rho(t) := \min\Big\{1, \sqrt{\frac{\mathbf{E}\Sigma^2}{t}}\Big\},$$

wówczas albo $\rho = 1$ i wtedy $W_\rho = 0$ albo $\rho < 1$, co wobec $W_\rho \leqslant \Sigma^2/\rho$ daje

$$\mathbf{E}W_\rho \leqslant \frac{\mathbf{E}\Sigma^2}{\rho} = \sqrt{t\mathbf{E}\Sigma^2} \leqslant t.$$

Mamy więc dla tak dobranego ρ,

$$\mathbf{P}(|Z-\mathbf{E}Z| \geqslant 4t) \leqslant 2\exp\Big(-\frac{1}{10}\min\Big\{\frac{t}{\rho}, \frac{t^2}{4\mathbf{E}\Sigma^2}\Big\}\Big) + \exp\Big(-\frac{t}{2}\log\Big(1+\frac{t}{\mathbf{E}W_\rho}\Big)\Big).$$

Zauważmy, że

$$\min\Big\{u, \frac{u}{4}\Big\} \geqslant \frac{1}{16}\log(1 + 4u) \text{ dla } u \geqslant 0,$$

więc

$$\min\Big\{\frac{t}{\rho}, \frac{t^2}{4\mathbf{E}\Sigma^2}\Big\} \geqslant t\min\Big\{\sqrt{\frac{t}{\mathbf{E}\Sigma^2}}, \frac{t}{4\mathbf{E}\Sigma^2}\Big\} \geqslant \frac{t}{16}\log\Big(1 + \frac{4t}{\mathbf{E}\Sigma^2}\Big).$$

Ponadto

$$\log\Big(1 + \frac{t}{\mathbf{E}W_\rho}\Big) \geqslant \log\Big(1 + \sqrt{\frac{t}{\mathbf{E}\Sigma^2}}\Big) \geqslant \frac{1}{4}\log\Big(1 + \frac{4t}{\mathbf{E}\Sigma^2}\Big).$$

Zatem

$$\mathbf{P}(|Z - \mathbf{E}Z| \geqslant 4t) \leqslant 3\exp\Big(-\frac{t}{160}\log\Big(1 + \frac{4t}{\mathbf{E}\Sigma^2}\Big)\Big)$$

i zamieniając $4t$ na t dostajemy tezę. $\square$

Uwaga 4.5 Rozpatrując rodzinę funkcji $\mathcal{F} \cup -\mathcal{F}$ widzimy, że wszystkie udowodnione w tej części oszacowania pozostają też prawdziwe, gdy Z zamienimy na

$$\bar{Z} = \Big|\sup_{f \in \mathcal{F}} \sum_{i=1}^{n} f(X_i)\Big|$$

By móc sensownie stosować Twierdzenia 4.18 i 4.21 potrzebujemy oszacowań $\mathbf{E}\Sigma^2$. Zdefiniujmy słabą wariancję empiryczną Z jako

$$\sigma^2 := \sup_{f \in \mathcal{F}} \mathbf{E} \sum_{i=1}^{n} f^2(X_i).$$

Twierdzenie 4.22. *Załóżmy, że* $|f| \leqslant C$ *oraz* $\mathbf{E}f(X_i) = 0$ *dla* $f \in \mathcal{F}$, *wówczas*

$$\mathbf{E}\Sigma^2 \leqslant \sigma^2 + 16C\mathbf{E}\bar{Z}.$$

Dowód będzie oparty na symetryzacji i zasadzie kontrakcji. By sformułować tę pierwszą wprowadźmy niezależny od pozostałych zmiennych ciąg $(\varepsilon_i)_{1 \leqslant i \leqslant n}$ niezależnych symetrycznych zmiennych losowych przyjmujących wartości ± 1.

Fakt 4.23. *Dla dowolnej przeliczalnej rodziny funkcji ograniczonych* $\mathcal{F}$,

$$\mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} f(X_i)\Big| \leqslant \sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} \mathbf{E}f(X_i)\Big| + 2\mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} \varepsilon_i f(X_i)\Big|.$$

Ponadto, jeśli $\mathbf{E}f(X_i) = 0$ *dla* $f \in \mathcal{F}$, *to*

$$\mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} \varepsilon_i f(X_i)\Big| \leqslant 2\mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} f(X_i)\Big|.$$

Dowód. Niech (X_i') będzie kopią ciągui (X_i) niezależną od (X_i) oraz (ε_i). Wówczas z nierówności trójkąta oraz nierówności Jensena,

$$\begin{aligned}\mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} f(X_i)\Big| &\leqslant \sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} \mathbf{E}f(X_i)\Big| + \mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} (f(X_i) - \mathbf{E}f(X_i))\Big| \\ &\leqslant \sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} \mathbf{E}f(X_i)\Big| + \mathbf{E}\sup_{f \in \mathcal{F}} \Big|\sum_{i=1}^{n} (f(X_i) - f(X_i'))\Big|.\end{aligned}$$

Z symetrii rodzina zmiennych $(f(X_i) - f(X_i'))_{f\in\mathcal{F}, i\leqslant n}$ ma ten sam rozkład co $(\varepsilon_i(f(X_i) - f(X_i')))_{f\in\mathcal{F}, i\leqslant n}$, zatem

$$\begin{aligned}
\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n (f(X_i) - f(X_i'))\Big| &\leqslant \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i(f(X_i) - f(X_i'))\Big| \\
&\leqslant \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big| + \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f(X_i')\Big| \\
&= 2\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big|.
\end{aligned}$$

By wykazać drugie oszacowanie zauważmy, że na mocy nierówności Jensena dla $A \subset \{1,\ldots,n\}$,

$$\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in A} f(X_i)\Big| = \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\in A} f(X_i) + \sum_{i\notin A}\mathbf{E}f(X_i)\Big| \leqslant \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n f(X_i)\Big|,$$

stąd

$$\begin{aligned}
\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big| &\leqslant \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\colon \varepsilon_i=1} \varepsilon_i f(X_i)\Big| + \mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i\colon \varepsilon_i=-1} \varepsilon_i f(X_i)\Big| \\
&\leqslant 2\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n f(X_i)\Big|.
\end{aligned}$$

□

Twierdzenie 4.24 (Zasada kontrakcji)**.** *Niech $\varphi_i\colon \mathbb{R}\to\mathbb{R}$ dla $1\leqslant i\leqslant n$ będą funkcjami 1-lipschitzowskimi oraz $\varphi_i(0) = 0$. Wówczas dla dowolnego ograniczonego $T\subset\mathbb{R}^n$ oraz funkcji wypukłej niemalejącej $F\colon [0,\infty)\to [0,\infty)$,*

$$\mathbf{E}F\Big(\frac{1}{2}\sup_{t\in T}\Big|\sum_{i=1}^n \varepsilon_i\varphi_i(t_i)\Big|\Big) \leqslant \mathbf{E}F\Big(\sup_{t\in T}\Big|\sum_{i=1}^n \varepsilon_i t_i\Big|\Big).$$

Dowód Twierdzenia 4.22. Zauważmy, że na podstawie pierwszej części Faktu 4.23,

$$\mathbf{E}\Sigma^2 \leqslant \sigma^2 + 2\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f(X_i)\Big|.$$

Funkcja $t\to\frac{1}{2C}t^2$ jest 1-lipschitzowska na $[-C,C]$ więc na mocy zasady kontrakcji (zastosowanej warunkowo względem X_i),

$$\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f^2(X_i)\Big| \leqslant 4C\mathbf{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^n \varepsilon_i f^2(X_i)\Big| \leqslant 8C\mathbf{E}\bar{Z},$$

gdzie ostatnia nierówność wynika z drugiej części Faktu 4.23. □

Literatura

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

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

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