

# Koncentracja Miary

Rafał Latała

23 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](mailto:rlatala@mimuw.edu.pl) z podaniem wersji notatek (daty) do której chcą się ustosunkować.

Dziękuję panom Kamilowi Kosińskiemu i Piotrowi Nayarowi za nadesłane komentarze.

Rafał Latała

# Spis treści

<b>1</b>	<b>Wprowadzenie</b>	<b>3</b>
1.1	Koncentracja funkcji lipschitzowskich . . . . .	5
1.2	Obserwowalna średnica zbioru . . . . .	7
1.3	Transport miary . . . . .	8
<b>2</b>	<b>Nierówności izoperymetryczne</b>	<b>8</b>
2.1	Klasyczna izoperymetria . . . . .	8
2.2	Izoperymetria sferyczna . . . . .	11
2.3	Izoperymetria gaussowska . . . . .	12
<b>3</b>	<b>Koncentracja w przestrzeniach produktowych</b>	<b>15</b>
3.1	Transformata Laplace'a . . . . .	15
3.2	Metody Martyngałowe . . . . .	15
3.3	Nierówność Poincaré . . . . .	19
3.3.1	Charakteryzacja na prostej . . . . .	24
3.4	Logarytmiczna Nierówność Sobolewa . . . . .	27
3.5	Nierówność Bobkowa . . . . .	32
3.6	Nierówności Splotu Infimum . . . . .	32
3.7	Nierówności Splotu Infimum dla Funkcji Wypukłych . . . . .	37
3.8	Nierówność Logarytmiczna Sobolewa dla Funkcji Wypukłych . . . . .	37
3.9	Aproksymacja przez otoczkę wypukłą . . . . .	38
<b>4</b>	<b>Wybrane Zastosowania</b>	<b>42</b>
4.1	Twierdzenie Dvoretzky'ego . . . . .	42
4.2	Wektory i Procesy Gaussowskie . . . . .	46
4.3	Procesy Empiryczne . . . . .	50
4.3.1	Sumy niezależnych zmiennych losowych . . . . .	50
4.3.2	Oszacowania supremów procesów empirycznych . . . . .	51

# 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

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

gdzie  $B(y, t)$  oznacza kule otwartą w  $X$  o środku w  $y$  i promieniu  $t$ .

Okazuje się, że dla wielu ważnych przykładów miar probabilistycznych  $\mu$  na  $(X, d)$  miara  $\mu(A_t)$  szybko zbiega do 1 jeśli tylko  $\mu(A) \geq 1/2$ . 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  $S^n = \{x \in \mathbb{R}^{n+1} : |x| = 1\}$ , zaś  $\sigma_n$  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 czapczkami), to znaczy

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

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

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

Wprowadźmy kluczową definicję:

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

$$\alpha_\mu(t) = \alpha_{(X, d, \mu)}(t) := \sup \left\{ 1 - \mu(A_t) : \mu(A) \geq \frac{1}{2} \right\}.$$

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

*Uwaga 1.3.* Zauważmy, że funkcja koncentracji  $\sigma_n$  szybko zbiega do 0 przy  $n \rightarrow \infty$ . 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  $\sigma_n$  przy jednokładności o skali  $R$ , to

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

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  $n \geq 2$ ,

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

**Przykład 2.** Niech  $\gamma_k$  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

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

to

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

W szczególności

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

Zauważmy, że powyższe oszacowania nie zależą od wymiaru przestrzeni.

**Przykład 3.** Niech  $\nu$  będzie symetrycznym rozkładem wykładniczym, tzn. rozkładem na  $\mathbb{R}$  z gęstością  $\frac{1}{2} \exp(-|x|)$ . Przez  $\nu^k$  będziemy oznaczać rozkład produktowy  $\nu \otimes \dots \otimes \nu$  na  $\mathbb{R}^k$ . Wyznaczenie ekstremalnych zbiorów dla problemu izoperymetrycznego związanego z tą miarą jest trudne i nieznanne 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.

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

W szczególności

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

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

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

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

Krótki przegląd wyników pokazuje, że w wielu ważnych zastosowaniach można wykazać, że  $\alpha_{\mu}(t) \leq C_1 \exp(-t^2/C_2)$  – mówimy wtedy, że funkcja koncentracji jest *typu gaussowskiego*. Widzieliśmy też przykład, w którym  $\alpha_{\mu}(t) \leq C_1 \exp(-t/C_2)$  – 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. Zaczniemy od definicji mediany i modułu ciągłości.

**Definicja 1.4.** Niech  $\mu$  będzie miarą probabilistyczną na  $(X, d)$  oraz  $f: X \rightarrow \mathbb{R}$ . Medianą funkcji  $f$  względem miary  $\mu$  nazywamy taką liczbę  $M = \text{Med}_\mu(f)$  dla której

$$\mu(\{x: f(x) \geq M\}) \geq \frac{1}{2} \quad \text{oraz} \quad \mu(\{x: f(x) \leq M\}) \geq \frac{1}{2}.$$

Modulem ciągłości  $f$  nazywamy funkcję

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

**Fakt 1.5.** Dla dowolnej funkcji  $F: X \rightarrow \mathbb{R}$ ,

$$\mu(\{x: F(x) > \text{Med}_\mu(F) + w_F(t)\}) \leq \alpha_\mu(t)$$

oraz

$$\mu(\{x: |F(x) - \text{Med}_\mu(F)| > w_F(t)\}) \leq 2\alpha_\mu(t).$$

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

$$\mu(\{x: F(x) < \text{Med}_\mu(F) - w_F(t)\}) \leq \alpha_\mu(t).$$

Dodając powyższą nierówność do poprzedniej otrzymamy ostatnią część faktu. □

Przypomnijmy definicję funkcji lipschitzowskiej

**Definicja 1.6.** Funkcję  $F: (X, d) \rightarrow \mathbb{R}$  nazywamy *lipschitzowską*, jeśli

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

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

Analogicznie można zdefiniować funkcje lipschitzowskie między przestrzeniami metrycznymi.

**Fakt 1.7.** i) Jeśli  $F$  jest lipschitzowska ze stałą  $L$ , to dla  $t > 0$ ,

$$\mu(\{x: F(x) > \text{Med}_\mu(F) + t\}) \leq \alpha_\mu(t/L)$$

oraz

$$\mu(\{x: |F(x) - \text{Med}_\mu(F)| > t\}) \leq 2\alpha_\mu(t/L).$$

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

$$\mu(\{x: F(x) \geq \text{Med}_\mu(F) + t\}) \leq \alpha,$$

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

*Dowód.* i) Wynika z Faktu 1.5 i oczywistego szacowania  $w_f(t) \leq tL$ .

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

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

□

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

**Fakt 1.8.** Załóżmy, że  $\mu$  jest miarą probabilistyczną na przestrzeni metrycznej  $(X, d)$  oraz dla ograniczonych funkcji 1-lipschitzowskich  $F$  i  $t > 0$  zachodzi

$$\mu\left(\left\{x: F(x) > \int F d\mu + t\right\}\right) \leq \alpha(t). \quad (1)$$

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

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

W szczególności

$$\alpha_\mu(t) \leq \alpha\left(\frac{t}{2}\right).$$

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

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

$$1 - \mu(A_t) = \mu(\{F \geq t\}) \leq \mu\left(\left\{F \geq \int F d\mu + \mu(A)t\right\}\right) \leq \alpha(\mu(A)t).$$

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

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

$$\mu\left(\left\{x: F_n(x) \leq \int F_n d\mu - t\right\}\right) \leq \alpha(t).$$

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

## 1.2 Obserwowalna średnica zbioru

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

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

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  $\mu$  jest związana ze stopniem oświetlenia obiektu). Motywuje to definicje *częściowej średnicy*

$$\text{PartDiam}_\mu(X, d) = \inf\{\text{Diam}(A) : \mu(A) \geq 1 - \kappa\}.$$

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 1.9.** *Obserwowalną średnicą* przestrzeni  $(X, d)$  względem miary  $\mu$  nazywamy wielkość

$$\text{ObsDiam}_\mu(X, d) := \sup\left\{\text{PartDiam}_{\mu \circ F^{-1}}(F(X)) : F : X \rightarrow \mathbb{R}, \|F\|_{\text{Lip}} \leq 1\right\}.$$

**Fakt 1.10.**

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

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

*Dowód.* Ustalmy funkcję 1-lipschitzowską  $F$  na  $X$  i niech  $M = \text{Med}_\mu(F)$ , wówczas na podstawie Faktu 1.7,

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

Zatem

$$\text{ObsDiam}_\mu(X, d) \leq 2 \inf\{r: 2\alpha_\mu(r) \leq \kappa\} = 2\alpha_\mu^{-1}(\kappa/2).$$

□

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

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

### 1.3 Transport miary

W wielu zagadnieniach będziemy używali pojęcia transportu miary.

**Definicja 1.11.** Niech  $\mu$  i  $\nu$  będą miarami na przestrzeniach metrycznych  $X$  i  $Y$ . Powiemy, że funkcja borelowska  $\varphi: X \rightarrow Y$  transportuje miarę  $\mu$  na miarę  $\nu$  (ew. miara  $\nu$  jest obrazem miary  $\mu$  przy przekształceniu  $\varphi$ ) jeśli  $\nu(A) = \mu \circ \varphi^{-1}(A)$  dla wszystkich  $A \in \mathcal{B}(Y)$ .

Szczególnie wygodny jest transport lipschitzowski.

**Fakt 1.12.** Jeśli  $\varphi: X \rightarrow Y$  jest  $L$ -lipschitzowska oraz  $\varphi$  transportuje miarę  $\mu$  na  $\nu$ , to  $\alpha_\nu(t) \leq \alpha_\mu(t/L)$ .

*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  $\lambda_n$ .

**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  $t > 0$ ,

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

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

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



to

$$\int_{\mathbb{R}^n} h(x)dx \geq \left( \int_{\mathbb{R}^n} f(x)dx \right)^s \left( \int_{\mathbb{R}^n} g(x)dx \right)^{1-s}$$

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

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

Ponieważ  $\lambda_1(A) = \sup\{\lambda_1(K) : K \subset A, K \text{ zwarty}\}$ , 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

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

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  $0 \leq r < 1$ ,  $\{h \geq r\} \supset s\{f \geq r\} + (1-s)\{g \geq r\}$ , więc całkując przez części dostajemy

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

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

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

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

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

Zatem na mocy założenia indukcyjnego

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

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

$$\begin{aligned} \int_{\mathbb{R}^n} h(x)dx &= \int_{\mathbb{R}} H(x)dx \geq \left( \int_{\mathbb{R}} F(x)dx \right)^s \left( \int_{\mathbb{R}} G(x)dx \right)^{1-s} \\ &= \left( \int_{\mathbb{R}^n} f(x)dx \right)^s \left( \int_{\mathbb{R}^n} g(x)dx \right)^{1-s} \end{aligned}$$

□

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

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

oraz

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

*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

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

□

*Uwaga 2.4.* 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,

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

□

**Definicja 2.5.** Dla miary  $\mu$  na przestrzeni probabilistycznej  $(X, d)$  określamy *zewnątną miarę brzegową*  $\mu^+$  wzorem

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

*Uwaga 2.6.* Jeśli miara  $\mu$  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 Hausdorffa.

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

**Twierdzenie 2.7.** *Jeśli  $A$  jest podzbiorem borelowskim  $\mathbb{R}^n$  takim, że  $\lambda_n(A) = \lambda_n(B(x_0, r))$ , to*

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

gdzie

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

## 2.2 Izoperymetria sferyczna

**Twierdzenie 2.8.** *Jeśli  $A$  jest podzbiorem borelowskim  $S^n$  takim, że  $\sigma_n(A) = \sigma_n(B(x_0, r))$  to dla dowolnego  $t > 0$ ,*

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

**Wniosek 2.9.**

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

*Dowód.* Dla  $n = 1$  nie ma co dowodzić (bo zawsze  $\alpha_\mu(t) \leq 1/2$ ). Będziemy więc zakładać, że  $n \geq 2$ . 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  $\cos u \leq \exp(-u^2/2)$  dla  $t \in [0, \pi/2]$  dostajemy

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

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

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

zatem

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

□

### 2.3 Izoperymetria gaussowska

Przypomnijmy, że przez  $\gamma_k$  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

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

**Twierdzenie 2.10.** *Niech  $H$  będzie półprzestrzenią afiniczną, a  $A$  zbiorem borelowskim w  $\mathbb{R}^k$  takim, że  $\gamma_k(H) = \gamma_k(A)$ . Wówczas dla dowolnego  $t > 0$ ,  $\gamma_k(H_t) \leq \gamma_k(A_t)$*

Zanim przystąpimy do dowodu twierdzenia pokażemy, że  $\gamma_k$  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.

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

**Fakt 2.11** (Lemat Poincaré). *Miara  $\mu_{k,n}$  zbiega słabo przy  $n \rightarrow \infty$  do miary  $\gamma_k$ , co więcej*

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

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

*Dowód Twierdzenia 2.10.* Ze względu na rotacyjną niezmienniczość miary  $\gamma_k$  możemy dla uproszczenia notacji założyć, że  $H = \{x : x_1 < r\}$ . Ustalmy dowolne  $r_0 < r$  i niech  $H_0 = \{x : x_1 < r_0\}$ . Zauważmy, że  $\gamma_k(H_0) < \gamma_k(A)$  zatem na podstawie Lematu Poincaré  $\mu_{k,n}(H_0) \leq \mu_{k,n}(A)$  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

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

Zauważmy, że przekształcenie  $P_{k,n}$  jest oczywiście 1-lipschitzowskie, więc

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

Nietrudno zauważyć, że

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

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

$$\gamma_k(A_t) = \lim_{n \rightarrow \infty} \mu_{k,n}(A_t) \geq \lim_{n \rightarrow \infty} \mu_{k,n}(\{x: x_1 < r_n\}) = \gamma_k(\{x: x_1 < r_0 + t\}),$$

z dowolności  $r_0 < r$  wynika teza.  $\square$

**Twierdzenie 2.12.** *Jeśli  $\gamma_k(A) = \Phi(x)$  to  $\gamma_k(A_t) \geq \Phi(x+t)$  oraz  $\gamma_k^+(A) \geq I_\gamma(\gamma_k(A))$ , 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  $H_t = \{x \in \mathbb{R}^k: \langle x, u \rangle < r + t\}$  i  $\gamma_k(H_t) = \Phi(x+t)$ .  $\square$

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

**Wniosek 2.13.**  $\alpha_{\gamma_k}(t) \leq 1 - \Phi(t) \leq \frac{1}{2} \exp(-t^2/2)$ .

Na podstawie Faktu 1.7 dostajemy:

**Wniosek 2.14.** *Jeśli  $F: \mathbb{R}^k \rightarrow \mathbb{R}$  jest funkcją  $L$ -lipschitzowską oraz  $t \geq 0$  to*

$$\gamma_k(\{x: F(x) \geq \text{Med}_{\gamma_k}(F) + t\}) \leq 1 - \Phi(t/L) \leq \frac{1}{2} e^{-\frac{t^2}{2L^2}}$$

oraz

$$\gamma_k(\{x: |F(x) - \text{Med}_{\gamma_k}(F)| \geq t\}) \leq 2(1 - \Phi(t/L)) \leq e^{-\frac{t^2}{2L^2}}$$

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

**Wniosek 2.15.** *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  $\gamma_n$ . W szczególności  $\alpha_{\mu_{[0,1]^n}} \leq \frac{1}{2} \exp(-\pi t^2)$ .*

*Dowód.* Określmy  $f: \mathbb{R} \rightarrow (0, 1)$  wzorem

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

Wówczas funkcja  $f$  transportuje miarę gaussowską  $\gamma_1$  na  $\mu_{[0,1]}$ , to znaczy  $\mu_{[0,1]} = \gamma_1 \circ f^{-1}$ . Ponadto  $f'(x) = (2\pi)^{-1/2} \exp(-x^2/2) \leq (2\pi)^{-1/2}$ , czyli  $f$  jest  $(2\pi)^{-1/2}$ -lipschitzowska. Jeśli teraz określimy  $F: \mathbb{R}^n \rightarrow (0, 1)^n$  wzorem  $F(x) = (f(x_1), \dots, f(x_n))$ , to  $F$  transportuje  $\gamma_n$  na  $\mu$  oraz  $F$  jest  $(2\pi)^{-1/2}$ -lipschitzowska. Ostatnie oszacowanie w tezie wniosku jest konsekwencją Faktu 1.12 i Wniosku 2.13.  $\square$

**Wniosek 2.16.** *Niech  $B_n = \{x \in \mathbb{R}^n: |x| \leq 1\}$  oznacza kulę jednostkową w  $\mathbb{R}^n$ , zaś  $\mu_{B_n}$  będzie rozkładem jednostajnym na  $B_n$ . Wówczas istnieje stała  $C$  taka, że  $\mu_{B_n}$  jest  $Cn^{-1/2}$ -lipschitzowskim obrazem  $\gamma_n$ . W szczególności  $\alpha_{\mu_{B_n}} \leq \frac{1}{2} \exp(-nt^2/(2C))$ .*

*Dowód.* Ponieważ obie miary  $\gamma_n$  i  $\mu_{B_n}$  są rotacyjnie niezmiennicze, będziemy szukać funkcji  $T: \mathbb{R}^n \rightarrow B_n$  transportującej  $\gamma_n$  na  $\mu_{B_n}$  postaci  $Tx = \frac{x}{|x|}\varphi(|x|)$ . Wystarczy sprawdzić, że  $\gamma_n(B(0, t)) = \mu_n(B(0, \varphi(t)))$ , czyli całkując we współrzędnych sferycznych, że dla  $t \geq 0$ ,

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

gdzie

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

Z (4) wynika, że

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

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

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

Ze wzoru Stirlinga dostajemy

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

więc

$$\|\varphi\|_{\text{Lip}} = \sup_{t \geq 0} |\varphi'(t)| \leq \frac{\sqrt{e}}{\sqrt{n}}.$$

□

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

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

oraz

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

Dość naturalna hipoteza mówi, że dla  $c \geq 1/2$  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  $\sigma_n$ , 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  $\mu$  jest miarą probabilistyczną na pewnej przestrzeni  $X$  oraz  $F: X \rightarrow \mathbb{R}$ , to *transformatę Laplace'a  $F$  względem  $\mu$  określamy*

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

**Fakt 3.2.** *Dla dowolnej zmiennej losowej  $Z$ ,*

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

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

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

*to dla  $t \geq 0$*

$$\mathbf{P}(Z \geq t) \leq \exp\left(-\frac{t^2}{4a}\right) \quad \text{oraz} \quad \mathbf{P}(|Z| \geq t) \leq 2 \exp\left(-\frac{t^2}{4a}\right).$$

*Dowód.* Pierwsza część wynika z nierówności Czebyszewa, a druga z pierwszej i prostego rachunku.  $\square$

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

#### 3.2 Metody Martyngałowe

**Twierdzenie 3.3** (Nierówność Azumy). *Niech  $(M_k, \mathcal{F}_k)_{k=0}^n$  będzie martyngałem o ograniczonych przyrostach takim, że  $\|M_k - M_{k-1}\|_\infty \leq a_k$ . Wówczas*

$$\mathbf{P}(M_n - M_0 \geq t) \leq \exp\left(-\frac{t^2}{2 \sum_{i=1}^n a_i^2}\right).$$

*Dowód.* Określmy dla  $1 \leq k \leq n$ ,  $d_k := M_k - M_{k-1}$ , 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  $\exp(x)$ ,

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

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

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

Liczymy

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

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

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

Teza twierdzenia wynika z Faktu 3.2. □

*Uwaga 3.4.* Najczęściej będziemy mieli  $\mathcal{F}_0 = \{\emptyset, \Omega\}$ , wówczas  $M_0$  jest stałe, a ponieważ martyngał ma stałą wartość oczekiwaną, to  $M_0 = \mathbf{E}M_n$ .

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

**Wniosek 3.5.** *Niech  $(X_i, d_i)$  będą przetrzeniami metrycznymi,  $X = X_1 \times \dots \times X_n$  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  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  będzie produktem miar probabilistycznych  $\mu_i$  na  $X_i$ . Wówczas dla dowolnej funkcji 1-lipschitzowskiej  $F$  na  $X$*

$$\mu\left(\left\{x: F(x) \geq \int F d\mu + t\right\}\right) \leq \exp\left(-\frac{t^2}{2D^2}\right),$$

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

$$\alpha_\mu(t) \leq \exp\left(-\frac{t^2}{8D^2}\right).$$



*Dowód.* Na mocy Faktu 1.8 wystarczy wykazać pierwszą nierówność tezy. Niech  $\mathcal{F}_k$  będzie  $\sigma$  ciałem generowanym przez pierwsze  $k$ -współrzędnych oraz  $M_k := \mathbf{E}_\mu(F|\mathcal{F}_k)$ . Wówczas oczywiście

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

stąd

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

i teza wynika z Twierdzenia 3.3. □

**Definicja 3.6.** Mówimy, że skończona przestrzeń metryczna  $(X, d)$  ma *długość* conajwyżej  $l$ , jeśli istnieje rosnący ciąg podziałów  $X, \{X\} = \mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n = \{\{x\}: x \in X\}$  ( $\mathcal{A}_i$  jest podpodziałem  $\mathcal{A}_{i-1}$ ) oraz liczby  $a_1, \dots, a_n$  spełniające  $(\sum_{i=1}^n a_i^2)^{1/2} \leq l$  takie, że dla dowolnego  $A \in \mathcal{A}_{i-1}$  oraz  $B, C \in \mathcal{A}_i, B, C \subset A$  istnieje bijekcja  $\Phi: B \rightarrow C$  dla której  $d(x, \Phi(x)) \leq a_i$  dla  $x \in B$ .

*Uwaga 3.7.* Biorąc  $\mathcal{A}_0 = \{X\}$  i  $\mathcal{A}_1 = \{\{x\}: x \in X\}$  widzimy, że każda skończona przestrzeń metryczna ma długość nie większą niż  $\text{Diam}(X)$ .

**Twierdzenie 3.8.** *Jeśli  $(X, d)$  jest skończoną przestrzenią metryczną o długości co najwyżej  $l$ , zaś  $\mu$  unormowaną miarą liczącą na  $X$ , to dla funkcji 1-lipschitzowskich  $F$  na  $X$ ,*

$$\mu\left(\left\{x: F(x) \geq \int F d\mu + t\right\}\right) \leq \exp\left(-\frac{t^2}{2l^2}\right),$$

w szczególności

$$\alpha_\mu(t) \leq \exp\left(-\frac{t^2}{8l^2}\right).$$

*Dowód.* Ustalmy funkcję 1-lipschitzowską  $F$ . Niech  $\mathcal{F}_i$  będzie  $\sigma$ -ciałem generowanym przez  $\mathcal{A}_i$  oraz  $M_i := \mathbf{E}_\mu(F|\mathcal{F}_i)$  dla  $i = 0, \dots, n$ . 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  $\Phi: B \rightarrow C$  jest bijekcją jak w Definicji 3.6, to dla  $x \in B, y \in C$ ,

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

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  $|M_i(x) - M_{i-1}(x)| \leq a_i$ , czyli  $\|M_i - M_{i-1}\|_\infty \leq a_{i-1}$ . Teza wynika z Twierdzenia 3.3 oraz Faktu 1.8.  $\square$

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

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

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

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

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

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

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

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

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

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

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  $\Phi$  między  $B$  i  $C$  jako  $\Phi = \tau_{p,q} \circ \sigma$ , gdzie  $\tau_{p,q}$  jest transpozycją zamieniającą  $p$  z  $q$ . Łatwo sprawdzić, że  $d(\sigma, \Phi(\sigma)) \leq 2/n$ , zatem  $l = 2/\sqrt{n}$  i

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

Ostatnie twierdzenie z tej części wiąże się z miarą Haara na zwartej grupie metrycznej  $(G, d)$ , to znaczy taką miarą probabilistyczną  $\mu$ , 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 3.9.** Niech  $\mu$  będzie miarą Haara na zwartej grupie metrycznej  $(G, d)$  oraz  $G = G_0 \supset G_1 \supset \dots \supset G_n = \{e\}$  będzie ciągiem domkniętych podgrup. Wówczas

$$\alpha_{(G, d, \mu)} \leq \exp\left(-\frac{t^2}{8l^2}\right),$$

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

*Dowód.* Niech  $F$  będzie funkcją 1-lipschitzowską,  $M_i = \mathbf{E}_\mu(F|\mathcal{F}_i)$  gdzie  $\mathcal{F}_i$  jest  $\sigma$ -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  $G_i$

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

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

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

Stąd oscylacja  $M_i$  na  $g_0G_{i-1}$  jest nie większa niż  $a_i = \text{Diam}(G_{i-1}/G_i)$  czyli po uśrednieniu  $\|M_i - M_{i-1}\|_\infty \leq a_i$  i możemy stosować Twierdzenie 3.3  $\square$

### 3.3 Nierówność Poincaré

**Definicja 3.10.** Mówimy, że miara probabilistyczna  $\mu$  na  $(X, d)$  spełnia nierówność Poincaré ze stałą  $C$ , jeśli dla wszystkich ograniczonych lipschitzowskich funkcji  $f$  na  $X$  zachodzi

$$\text{Var}_\mu(f) \leq C \int |\nabla f|^2 d\mu, \quad (5)$$

gdzie

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

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

*Uwaga 3.11.* W przypadku, gdy  $X = \mathbb{R}^n$  ze standardową metryką euklidesową możemy użyć twierdzenia Rademachera, które mówi, że każda funkcja Lipschitzowska 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 3.12.* Będziemy wykorzystywali tylko dwie własności  $|\nabla f|$ . Mianowicie, że dla funkcji 1-lipschitzowskich  $|\nabla f| \leq 1$  oraz, że dla dowolnej funkcji klasy  $C^1(\mathbb{R})$ ,  $|\nabla g(F)| \leq |g'(F)||\nabla F|$  (w szczególności  $|\nabla(f+c)| = |\nabla f|$ ).

*Uwaga 3.13.* Załóżmy, że miara  $\mu$  ma gęstość postaci  $e^{-V}$  na  $\mathbb{R}^n$ . Wówczas proste całkowanie przez części pokazuje, że

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

Definiując operator  $Lf := -\Delta f + \langle \nabla V, \nabla f \rangle$  widzimy, że  $L1 = 0$ . Nierówność Poincaré mówi, że dla funkcji  $f$  o średniej 0, czyli prostopadłych do 1,  $\int f Lf d\mu \geq C^{-1} \int f^2 d\mu$ . Biorąc pod uwagę samosprężoność  $L$  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 3.14.** *Nierówność Poincaré jest równoważna nierówności*

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

Co więcej optymalne stałe w obu nierównościach spełniają  $C_{\text{opt}} \leq \tilde{C}_{\text{opt}} \leq (1 + \sqrt{2})^2 C_{\text{opt}}$ .

*Dowód.* Ponieważ

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

więc oczywiście  $C_{\text{opt}} \leq \tilde{C}_{\text{opt}}$ .

By udowodnić przeciwne oszacowanie zauważmy, że

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

Stąd

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

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

**Twierdzenie 3.15.** *Załóżmy, że miara  $\mu$  spełnia nierówność Poincaré ze stałą  $C$ . Wówczas dla każdej funkcji 1-lipschitzowskiej  $F$  i  $t > 0$*

$$\mu\left(\left\{F \geq \int F d\mu + t\right\}\right) \leq 2 \exp\left(-\frac{t}{\sqrt{C}}\right).$$

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

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

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

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

$$\text{Var}_{\mu}(e^{\lambda F/2}) = M(\lambda) - M\left(\frac{\lambda}{2}\right)^2 \leq C \int |\nabla e^{\lambda F/2}|^2 d\mu \leq \frac{C\lambda^2}{4} M(\lambda).$$

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

$$M(\lambda) \leq \frac{1}{1 - C\lambda^2/4} M\left(\frac{\lambda}{2}\right)^2.$$

Iterując tę nierówność  $n$  razy dostajemy

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

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

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

Zauważmy, że

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

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

**Fakt 3.16.** Symetryczny rozkład wykładniczy  $\nu$  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  $h \in C_{\text{ogr}}^1(\mathbb{R})$ ,

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

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

$$\int g^2 d\nu = 2 \int \text{sgn}(x) g'(x) g(x) d\nu(x) \leq 2 \left( \int g'^2 d\nu \right)^{1/2} \left( \int g^2 d\nu \right)^{1/2},$$

stąd

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

$\square$

**Fakt 3.17.** Załóżmy, że  $\mu_i$  są miarami probabilistycznymi na  $X_i$ ,  $X = X_1 \times \dots \times X_n$  oraz  $\mu = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$ . Wówczas dla dowolnej funkcji  $f \in L^2(X, \mu)$

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

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

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

gdzie ostatnia nierówność wynika np. z nierówności Jensena.  $\square$

**Wniosek 3.18.** Załóżmy, że miary probabilistyczne  $\mu_i$  na  $(X_i, d_i)$  spełniają nierówność Poincaré ze stałą  $C_i$  względem gradientu  $|\nabla_i|$ . Wówczas miara  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  spełnia nierówność Poincaré ze stałą  $C = \max_i C_i$  względem gradientu  $\nabla f$  danego wzorem

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

*Dowód.* Z Faktu 3.17 dostajemy

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

$\square$

**Wniosek 3.19.** Produktowy rozkład wykładniczy  $\nu^n$  spełnia nierówność Poincaré na  $\mathbb{R}^n$  ze stałą 4. W szczególności  $\alpha_{\nu^n}(t) \leq 2 \exp(-t/4)$ .

Kolejną przyjemną własnością nierówności Poincaré jest jej stabilność ze względu na zaburzenia miary  $\mu$ .

**Fakt 3.20.** Załóżmy, że  $\mu$  jest miarą probabilistyczną na  $X$ ,  $V$  jest ograniczoną funkcją bo-relowską oraz  $d\nu = Z^{-1} e^V d\mu$ , gdzie  $Z = \int e^V d\mu$ . Wówczas jeśli miara  $\mu$  spełnia nierówność Poincaré ze stałą  $C$  to  $\nu$  spełnia nierówność Poincaré ze stałą  $C e^{2\|V\|_\infty}$ .

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

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

$\square$

**Fakt 3.21.** *Jeśli miara  $\nu$  na  $(Y, \rho)$  jest  $L$ -lipschitzowskim obrazem miary  $\mu$  na  $(X, d)$  oraz  $\mu$  spełnia nierówność Poincaré ze stałą  $C$ , to  $\nu$  spełnia nierówność Poincaré ze stałą  $CL^2$ .*

*Dowód.* Niech  $\nu = \mu \circ \varphi^{-1}$ , gdzie  $\varphi: X \rightarrow Y$  i  $\|\varphi\|_{\text{Lip}} \leq L$ . Dla funkcji lipschitzowskich  $f$  na  $Y$  otrzymujemy

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

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

**Wniosek 3.22.** *Istnieje stała uniwersalna  $L_0 \leq 12.5$  taka, że dowolna miara symetryczna log-wkłęśła  $\mu$  o wariancji 1 na prostej jest  $L_0$ -lipschitzowskim obrazem symetrycznej miary wykładniczej  $\nu$ . Stąd  $\mu$  spełnia nierówność Poincaré ze stałą  $4L_0^2 \leq 620$ .*

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

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

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

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

Analogicznie

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

Zatem

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

skąd wnioskujemy, że

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

Określmy  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  wzorem

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

Wówczas  $\varphi$  transportuje  $\nu$  na  $\mu$ , jest nieparzysta i niemalejąca, zatem

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

Rozpatrzmy dwie możliwości:

i)  $\varphi(x) \leq x_0$ . Wówczas  $g(\varphi(x)) \geq g(0)/e$  i

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

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

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

Zatem

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

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

*Uwaga 3.23.* Wniosek 3.22 (z nieco gorszą stałą) jest prawdziwy bez założenia symetrii miary  $\mu$ . 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 3.24** (Muckenhoupt). *Załóżmy, że  $\mu$  i  $\nu$  są miarami na  $[0, \infty)$ . Wówczas istnieje stała  $C < \infty$  taka, że dla każdej funkcji ograniczonej  $f$ ,*

$$\int_0^\infty \left| \int_0^x f(t) dt \right|^2 d\mu \leq C \int_0^\infty |f(x)|^2 d\nu(x) \quad (6)$$

*wtedy i tylko wtedy gdy*

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

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



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

$$\int_0^\infty \left| \int_0^x f(t) dt \right|^2 d\mu \leq C \int_0^\infty |f(x)|^2 p(x) dx \quad (7)$$

dla nieujemnych  $f$ .

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

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

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

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

skąd po przejściu z  $n \rightarrow \infty$  i wzięciu supremum po  $r$  dostajemy  $B \leq C_{\text{opt}}$ .

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

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

wówczas na mocy nierówności Schwarz'a i twierdzenia Fubiniego

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

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)

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

Zatem

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

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

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

Wówczas

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

więc

$$\int_0^\infty \left| \int_0^x f(t) dt \right|^2 d\mu = \lim_{n \rightarrow \infty} \int_0^\infty \left| \int_0^x f(t) dt \right|^2 d\mu_n \leq 4B \int_0^\infty |f(x)|^2 p(x) dx.$$

□

**Twierdzenie 3.25.** Załóżmy, że  $\mu$  jest miarą probabilistyczną na  $\mathbb{R}$  o medianie  $m$ , zaś  $p$  oznacza gęstość jej części absolutnie ciągłej. Wówczas miara  $\mu$  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  $C_{\text{opt}}$  w nierówności Poincaré spełnia

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

*Dowód.* Z Twierdzenia 3.24 (zastosowanego do miary  $\mu$  obciętej do  $[m, \infty)$  i  $f'$  zamiast  $f$ ) wynika, że dla dowolnej funkcji Lipschitzowskiej  $f$ ,

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

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

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

Dodając stronami dostajemy

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

czyli  $C_{\text{opt}} \leq 4 \max\{B_+, B_-\}$ .

By udowodnić nierówność w drugą stronę zauważmy, że jeśli  $f$  jest ograniczoną funkcją oraz  $g(x) = 0$  dla  $x \leq m$  i  $g(x) = \int_m^x f(t) dt$  dla  $x > m$ , to  $g$  jest lipschitzowska,  $g'(x) = f(x) \mathbb{1}_{\{x \geq m\}}$  p.w. oraz  $\text{Med}_\mu(g) = 0$ . Stąd na podstawie Faktu 3.14,

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

Stąd Twierdzenie 3.24 implikuje  $B_+ \leq (1 + \sqrt{2})^2 C_{\text{opt}}$ , analogicznie pokazujemy  $B_- \leq (1 + \sqrt{2})^2 C_{\text{opt}}$ .  $\square$

### 3.4 Logarytmiczna Nierówność Sobolewa

**Definicja 3.26.** Załóżmy, że  $\mu$  jest miarą probabilistyczną na  $X$ , zaś  $f$  nieujemną funkcją mierzalną na  $X$ . *Entropię  $f$  względem  $\mu$*  definiujemy wzorem

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

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

**Definicja 3.27.** Mówimy, że miara probabilistyczna na  $(X, d)$  spełnia *logarytmiczną nierówność Sobolewa* ze stałą  $C$ , jeśli dla wszystkich ograniczonych lipschitzowskich funkcji  $f$  na  $X$  zachodzi

$$\text{Ent}_\mu(f^2) \leq 2C \int |\nabla f|^2 d\mu. \quad (8)$$

**Twierdzenie 3.28.** Załóżmy, że miara  $\mu$  spełnia logarytmiczną nierówność Sobolewa ze stałą  $C$ . Wówczas dla każdej funkcji 1-lipschitzowskiej  $F$  i  $t > 0$

$$\mu\left(\left\{F \geq \int F d\mu + t\right\}\right) \leq \exp\left(-\frac{t^2}{2C}\right).$$

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

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

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

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

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

oraz

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

Zatem (8) daje

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

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

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

oraz na podstawie (9)

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

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

**Lemat 3.29.** Dla dowolnej funkcji nieujemnej na  $X$ ,

$$\text{Ent}_\mu(f) = \sup \left\{ \int f g d\mu : \int e^g d\mu \leq 1 \right\}. \quad (10)$$

*Dowód.* Z jednorodności obu stron tożsamości (10) możemy zakładać, że  $\int f d\mu = 1$ , wówczas  $\text{Ent}_\mu(f) = \int f \log f d\mu$ .

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

$$uv \leq u \log u - u + e^v \quad \text{dla } u \geq 0, v \in \mathbb{R}. \quad (11)$$

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

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

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

Z powyższego lematu łatwo wykazać tensoryzowalność entropii:

**Fakt 3.30.** Załóżmy, że  $\mu_i$  są miarami probabilistycznymi na  $X_i$ ,  $X = X_1 \times \dots \times X_n$  oraz  $\mu = \mu_1 \otimes \mu_2 \otimes \dots \otimes \mu_n$ . Wówczas dla dowolnej nieujemnej funkcji  $f$  na  $X$

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

*Dowód.* Weźmy funkcję  $g$  na  $X$  taką, że  $\int e^g d\mu \leq 1$  oraz przyjmijmy dla  $i = 1, \dots, n$ ,

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

□

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

$$\int f g d\mu \leq \sum_{i=1}^n \int f g^i d\mu = \sum_{i=1}^n \int \left( \int f g^i d\mu_i \right) d\mu \leq \sum_{i=1}^n \int \text{Ent}_{\mu_i}(f) d\mu.$$

**Wniosek 3.31.** Załóżmy, że miary probabilistyczne  $\mu_i$  na  $(X_i, d_i)$  spełniają logarytmiczną nierówność Sobolewa ze stałą  $C_i$  względem gradientu  $|\nabla_i|$ . Wówczas miara  $\mu = \mu_1 \otimes \dots \otimes \mu_n$  spełnia logarytmiczną nierówność Sobolewa ze stałą  $C = \max_i C_i$  względem gradientu  $\nabla f$  danego wzorem

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

*Dowód.* Z Faktu 3.30 dostajemy

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

□

**Fakt 3.32.** i) Niech  $\mu_1 = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ , wówczas dla dowolnego  $f: \{-1, 1\} \rightarrow \mathbb{R}$ ,

$$\text{Ent}_{\mu_1}(f^2) \leq 2\text{Ent}_{\mu_1}|Df|^2,$$

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

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

$$\text{Ent}_{\mu_n}(f^2) \leq 2\text{Ent}_{\mu_n}|Df|^2,$$

gdzie

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

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

*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  $\alpha(t) \geq 0$ , 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

$$\alpha''(t) = \frac{1}{1 - t^2} \left( \frac{t^2}{\sqrt{1 - t^2}} - \frac{t^2}{1 + \sqrt{1 - t^2}} \right) \geq 0,$$

więc istotnie  $\alpha(t) \geq 0$ .

ii) Wynika z punktu i) i Faktu 3.30. □

**Twierdzenie 3.33.** *Miara  $\gamma_n$  spełnia logarytmiczną nierówność Sobolewa z  $C = 1$ .*

*Dowód.* Z uwagi na Fakt 3.30 wystarczy rozważyć przypadek  $n = 1$ . Niech  $f \in C_{\text{ogr}}^1(\mathbb{R})$ . Określmy  $g_n: \{-1, 1\}^n \rightarrow \mathbb{R}$  wzorem

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

Niech  $\mu_n$  i  $|Df|$  będą jak w Fakcie 3.32. Wówczas na mocy centralnego twierdzenia granicznego

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

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

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

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

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

□

**Fakt 3.34.** *Załóżmy, że  $\mu$  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  $\mu$  spełnia logarytmiczną nierówność Sobolewa ze stałą  $C$  to  $\nu$  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  $s, t$ ,  $\varphi(s+t) \geq \varphi(t) + \varphi'(t)s$ , więc

$$\varphi\left(\int f^2 d\nu\right) = \varphi\left(t + \int (f^2 - t) d\nu\right) \geq \varphi(t) + \varphi'(t) \int (f^2 - t) d\nu.$$

Zatem

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

□

Kolejny fakt dowodzimy tak samo jak dla nierówności Poincaré.

**Fakt 3.35.** *Jeśli miara  $\nu$  na  $(Y, \rho)$  jest  $L$ -lipschitzowskim obrazem miary  $\mu$  na  $(X, d)$  oraz  $\mu$  spełnia logarytmiczną nierówność Sobolewa ze stałą  $C$ , to  $\nu$  spełnia logarytmiczną nierówność Sobolewa ze stałą  $CL^2$ .*

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

**Fakt 3.36.** *Jeśli miara probabilistyczna  $\mu$  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 3.37.** *Załóżmy, że  $\mu$  jest miarą probabilistyczną na  $\mathbb{R}$  o medianie  $m$ , zaś  $p$  oznacza gęstość jej części absolutnie ciągłej. Wówczas miara  $\mu$  spełnia logarytmiczną nierówność Sobolewa ze skończoną stałą  $C$  wtedy i tylko wtedy gdy  $\max\{B_+, B_-\} < \infty$ , gdzie*

$$\begin{aligned} B_+ &= \sup_{x > m} \mu[x, \infty) \ln\left(\frac{1}{\mu[x, \infty)}\right) \int_m^x \frac{1}{p(y)} dy \\ B_- &= \sup_{x < m} \mu(-\infty, x] \ln\left(\frac{1}{\mu(-\infty, x]}\right) \int_x^m \frac{1}{p(y)} dy. \end{aligned}$$

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

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

### 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 funkcję 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 3.38.** Mówimy, że miara probabilistyczna  $\mu$  na  $\mathbb{R}^n$  spełnia nierówność Bobkowa ze stałą  $C$ , jeśli dla wszystkich  $f \in C_{\text{ogr}}^1(\mathbb{R}^n)$  o wartościach w przedziale  $[0, 1]$  zachodzi

$$I\left(\int f d\mu\right) \leq \int \sqrt{I(f)^2 + C|\nabla f|^2} d\mu. \quad (12)$$

**Fakt 3.39.** Jeśli miary  $\mu_i$  spełniają nierówność Bobkowa ze stałymi  $C_i$ , to miara  $\mu_1 \otimes \dots \otimes \mu_n$  spełnia nierówność Bobkowa ze stałą  $\max_i C_i$ .

**Twierdzenie 3.40.** Jeśli miara probabilistyczna  $\mu$  na  $\mathbb{R}^n$  spełnia nierówność Bobkowa na ze stałą  $C$ , to

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

oraz

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

**Twierdzenie 3.41.** Kanoniczna miara gaussowska  $\gamma_n$  spełnia nierówność Bobkowa z  $C = 1$ .

### 3.6 Nierówności Splotu Infimum

Zacznijmy od zaproponowanej przez Maureya definicji.

**Definicja 3.42.** Splotem infimum dwu funkcji  $f$  i  $g$  określonych na  $\mathbb{R}^n$  nazywamy funkcję  $f \square g$  daną wzorem

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

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

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

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



Pierwszą użyteczną cechą własności  $(\tau)$  jest jej tensoryzowalność.

**Fakt 3.43.** *Jeśli pary  $(\mu_i, \varphi_i)$  mają własność  $(\tau)$ ,  $\mu = \mu_1 \otimes \cdots \otimes \mu_n$  oraz*

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

to również para  $(\mu, \varphi)$  ma własność  $(\tau)$ .

*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

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

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

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

na mocy własności  $(\tau)$  dla  $(\mu_1, \varphi_1)$ . Stąd

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

i korzystając z Twierdzenia Fubiniiego i  $(\tau)$  dla  $(\mu_1, \varphi_2)$ .

$$\begin{aligned} \int e^{f \square \varphi} d\mu_1 \otimes \mu_2 &\leq \int e^{g \square \varphi_2(y)} d\mu_2(y) \leq \left( \int e^{-g(y)} d\mu_2(y) \right)^{-1} \\ &= \left( \int e^{-f} d\mu_1 \otimes \mu_2 \right)^{-1}. \end{aligned}$$

□

Następny fakt pokazuje w jaki sposób można transportować  $(\tau)$ .

**Fakt 3.44.** *Załóżmy, że  $\mu$  jest miarą probabilistyczną na  $\mathbb{R}^n$ , zaś  $\varphi$  funkcją kosztu na  $\mathbb{R}^n$  taką, że  $(\mu, \varphi)$  spełnia własność  $(\tau)$ . Jeśli  $T: \mathbb{R}^n \rightarrow \mathbb{R}^m$  oraz funkcja  $\psi$  na  $\mathbb{R}^m$  spełnia  $\psi(Tx - Ty) \leq \varphi(x - y)$  dla wszystkich  $x, y$ , to para  $(\mu \circ T^{-1}, \psi)$  ma własność  $(\tau)$ .*

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

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

Zatem

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

□

By sformułować związki nierówności spłotu infimum z koncentracją określmy zbiór

$$B_\varphi(t) = \{x: \varphi(x) \leq t\}.$$

Zacznijmy od prostego faktu

**Fakt 3.45.** *Jeśli  $(\mu, \varphi)$  ma własność  $(\tau)$  to dla dowolnego zbioru borelowskiego  $A$  takiego, że  $\mu(A) > 0$  mamy*

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

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

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

□

*Uwaga 3.46.* Funkcja  $f$  w poprzednim dowodzie nie była oczywiście ograniczona, ale łatwo ominąć ten problem stosując nierówność  $(\tau)$  do  $f_n = n \mathbb{1}_{\mathbb{R}^n \setminus A}$  dla  $n \geq t$ .

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

**Fakt 3.47.** *Załóżmy, że para  $(\mu, \varphi)$  ma własność  $(\tau)$ . Wówczas dla dowolnego zbioru borelowskiego  $A$  i  $t > 0$ ,*

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

*W szczególności*

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

*oraz*

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

*Ponadto*

$$\mu(A) = \nu(-\infty, x] \Rightarrow \mu(A + B_\varphi(t)) \geq \nu(-\infty, x + t/2]. \quad (16)$$

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

Zatem własność  $(\tau)$  daje

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

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  $p \leq e^{-t/2}/2$ ,

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

skąd otrzymujemy (14). Ponadto dla  $p \geq 1/2$ ,

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

i dostajemy (15).

Niech  $F(x) = \nu(-\infty, x]$  i  $g_t(p) = F(F^{-1}(p) + t)$ . Poprzednie rachunki pokazują, że dla  $t, p > 0$ ,  $f_t(p) \geq g_{t/2}(p)$ , jeśli  $F^{-1}(p) + t/2 \leq 0$  lub  $F^{-1}(p) \geq 0$ . Ponieważ  $g_{t+s} = g_t \circ g_s$  i  $f_{t+s} = f_t \circ f_s$ , otrzymujemy  $f_t(p) \geq g_{t/2}(p)$  dla wszystkich  $t, p > 0$ , zatem (13) implikuje (16). □

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

**Fakt 3.48.** Para  $(\nu_+, \varphi_0)$  ma własność  $(\tau)$ , gdzie

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

**Lemat 3.49.** Dla wszystkich  $x \in \mathbb{R}$  mamy  $2|\varphi_0'(x)| \leq 1$  oraz

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

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

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

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

*Dowód Faktu 3.48.* Ustalmy funkcję ograniczoną  $f$ , przyjmijmy  $g := f \square \varphi_0$  i niech

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

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

$$\int_0^{x(t)} e^{-f(x)-x} dx = t I_0 \text{ oraz } \int_0^{y(t)} e^{g(x)-x} dx = t I_1.$$

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  $g$ ,  $g(y(t)) \leq f(x(t)) + \varphi_0(y(t) - x(t))$ , więc

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

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

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

Pisząc dla uproszczenia  $x$  i  $y$  zamiast  $x(t)$  i  $y(t)$  stosując poprzednie oszacowanie  $y'(t)$  oraz nierówność między średnią arytmetyczną i geometryczną dostajemy (wykorzystując parzystość  $\varphi_0$ )

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

Zatem na mocy Lematu 3.49,  $(-e^{-z(t)})' = e^{-z(t)} z'(t) \geq \sqrt{I_0 I_1}$ , co po odcałkowaniu daje  $\sqrt{I_0 I_1} \leq 1$ .  $\square$

*Uwaga 3.50.* 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 3.51.** *Miara  $\nu$  spełnia nierówność infimum z funkcją kosztu  $\varphi_1$  postaci*

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

*Dowód.* Z wypukłości funkcji  $\varphi_0$  łatwo wynika, że  $\varphi_1 = \varphi_0 \square \varphi_0$ . Ponieważ miara  $\nu_-$  jest symetrycznym odbiciem  $\nu_+$  a funkcja  $\varphi_0$  jest symetryczna, to  $(\nu_-, \varphi_0)$  ma własność  $(\tau)$ , więc  $(\nu_+ \otimes \nu_-, \varphi_0(x) + \varphi_0(y))$  też ma  $(\tau)$ . Miara  $\nu$  jest splotem miar  $\nu_+$  i  $\nu_-$ , czyli obrazem  $\nu_+ \otimes \nu_-$  przy przekształceniu  $T(x, y) = x + y$ . Teza wynika z Faktu 3.44  $\square$

Wiemy, że miara  $\nu$  a zatem i miara produktowa  $\nu^n$  spełniają nierówność Poincaré, więc jeśli  $\nu^n(A) \geq \frac{1}{2}$ , to  $\nu^n(A + tB_2^n) \geq 1 - e^{-t/C}$  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  $1 \leq p < \infty$

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

**Twierdzenie 3.52.** *Dla dowolnego zbioru borelowskiego  $A$  w  $\mathbb{R}^n$  takiego, że  $\nu^n(A) > 0$  mamy dla  $t \geq 0$ ,*

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

*Ponadto*

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

*Dowód.* Para  $(\nu^n, \varphi_n)$  ma własność  $(\tau)$ , gdzie  $\varphi_n(x_1, \dots, x_n) = \varphi_1(x_1) + \dots + \varphi_1(x_n)$ . Łatwo sprawdzić, że

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

Teza wynika zatem z Faktów 3.45 i 3.47. □

### 3.7 Nierówności Splotu Infimum dla Funkcji Wypukłych

### 3.8 Nierówność Logarytmiczna Sobolewa dla Funkcji Wypukłych

**Fakt 3.53.** *Załóżmy, że  $F: \mathbb{R}^n \rightarrow \mathbb{R}$  jest lipschitzowska oraz wypukła po każdej współrzędnej. Wówczas dla dowolnej produktowej miary probabilistycznej  $\mu$  na  $\mathbb{R}^n$ ,*

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

**Twierdzenie 3.54.** *Załóżmy, że  $F: [0, 1]^n \rightarrow \mathbb{R}$  jest  $L$ -lipschitzowska oraz wypukła po każdej współrzędnej. Wówczas dla dowolnej produktowej miary probabilistycznej  $\mu$  na  $[0, 1]^n$ ,*

$$\mu\left(\left\{F \geq \int F d\mu + t\right\}\right) \leq \exp\left(-\frac{t^2}{4L^2}\right) \text{ dla } t \geq 0.$$

*Dowód.* Bez straty ogólności (zastępując  $F$  przez  $aF + b$ ) możemy zakładać, że  $L = 1$  oraz  $\int F = 0$ . Wówczas  $|\nabla F| \leq 1$  p.w., więc Fakt 3.53 implikuje, że dla  $\lambda > 0$ ,

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

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

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

Czyli funkcja  $H(\lambda) = \frac{1}{\lambda} \log M(\lambda)$  spełnia  $H'(\lambda) \leq 1$ , a ponieważ  $H(0+) = \int F d\mu = 0$ , więc  $H(\lambda) \leq \lambda$ , czyli  $M(\lambda) \leq \exp(\lambda^2)$ . Stąd Fakt 3.2 implikuje tezę.  $\square$

*Uwaga 3.55.* Jeśli  $F$  jest wypukła, to  $-F$  jest wklęsła, a dla funkcji wklęsłych nie da się łatwo zmodyfikować dowodu Faktu 3.53. 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 3.5 wynika, że dla  $|a| = 1$ ,  $\alpha_{\mu, X, d_a}(t) \leq \exp(-|t|^2/8)$ , 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  $\mathcal{D}_A^c(x)$  można zdefiniować w równoważny, nieco bardziej abstrakcyjny sposób.

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

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

oraz

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

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

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

**Fakt 3.56.** Dla dowolnego  $A \subset X$  i  $x \in X$ ,

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

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

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

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

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

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

ii)  $\mathcal{D}_A^c(x) \geq d(0, V_A(x))$ . 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

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

Biorąc  $\theta \rightarrow 0+$  dostajemy  $\langle z, s - z \rangle \geq 0$ , czyli

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

Stąd

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

□

**Twierdzenie 3.57.** 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  $X$ ,

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

W szczególności dla  $t > 0$ ,

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

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

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

Założmy, że  $n \geq 2$  i teza zachodzi dla  $n - 1$ . Dla uproszczenia notacji przyjmijmy

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

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

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

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  $|x|^2$ ,

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

czyli z dowolności wyboru  $t$  i  $s$ ,

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

Odcałkowując i korzystając z nierówności Höldera dostajemy

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

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

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

Zauważmy teraz, że

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

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

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

Nierówności (17) oraz (18) z  $u = \tilde{\mu}(A(x_n))/\tilde{\mu}(B)$  implikują

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



Zatem

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

gdź  $v(2-v) \leq 1$  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

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

Niech dla  $x \in X$

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

wówczas

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

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

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

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

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

Stąd dla  $s > 0$ ,

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

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

Twierdzenie 3.57 prowadzi do koncentracji pewnej klasy funkcji lipschitzowskich w odpowiednim sensie. Mianowicie zachodzi

**Wniosek 3.58.** *Założmy, że funkcja  $F: X \rightarrow \mathbb{R}$  spełnia warunki*

$$\forall x \in X \exists a = a(x) \in S^{n-1} \forall y \in X \quad F(x) \leq F(y) + d_a(x, y). \quad (19)$$

Wówczas dla dowolnej probabilistycznej miary produktowej  $\mu$  na  $X$ ,

$$\mu(\{|F - \text{Med}_\mu(F)| \geq t\}) \leq 4e^{-t^2/4} \quad \text{dla } t > 0.$$

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

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

stąd

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

Zatem dla dowolnego  $m$ ,

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

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

## 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. Wszystkie przestrzenie euklidesowe są izometryczne, więc wystarczy określić włożenie  $l_2^k$ .

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

$$|t| \leq \left\| \sum_{i=1}^k t_i v_i \right\| \leq C|t| \text{ dla wszystkich } t \in \mathbb{R}^k.$$

*Uwaga 4.2.* Ostatnią definicję można sformułować równoważnie (biorąc  $u_i = v_i/\sqrt{C}$ ) jako istnienie  $u_1, \dots, u_k \in E$  takich, że

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

**Definicja 4.3.** Dla wektora gaussowskiego w przestrzeni unormowanej  $E$  postaci  $X = \sum_{i=1}^m v_i g_i$ , gdzie  $g_1, \dots, g_n$  są niezależnymi zmiennymi  $\mathcal{N}(0, 1)$  określmy

$$\sigma(X) := \sup \{(\mathbf{E}\varphi(X)^2)^{1/2} : \varphi \in E^*, \|\varphi\| \leq 1\} \text{ oraz } d(X) := \left( \frac{\mathbf{E}\|X\|}{\sigma(X)} \right)^2.$$

Dla przestrzeni unormowanej  $E$  określamy

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

**Fakt 4.4.** Jeśli  $X = \sum_{i=1}^m v_i g_i$  jest wektorem gaussowskim w  $E$ , to

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

*Dowód.* Mamy

$$\begin{aligned} \sup_{|t|=1} \left\| \sum_{i=1}^m t_i v_i \right\| &= \sup_{|t|=1} \sup_{\|\varphi\| \leq 1} \varphi \left( \sum_{i=1}^m t_i v_i \right) = \sup_{\|\varphi\| \leq 1} \sup_{|t|=1} \sum_{i=1}^m t_i \varphi(v_i) \\ &= \sup_{\|\varphi\| \leq 1} \left( \sum_{i=1}^m \varphi^2(v_i) \right)^{1/2} = \sup_{\|\varphi\| \leq 1} (\mathbf{E} \varphi(X)^2)^{1/2} = \sigma(X). \end{aligned}$$

□

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

**Twierdzenie 4.5.** Dla dowolnego  $\varepsilon > 0$  istnieje stała  $c(\varepsilon) > 0$  taka, że dla dowolnej przestrzeni unormowanej  $E$  oraz  $k \leq c(\varepsilon)d(E)$ ,  $l_2^k$  się wkłada ze stałą  $1 + \varepsilon$  w  $E$ .

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

**Definicja 4.6.** Zbiór  $A \subset X$  nazywamy  $\varepsilon$ -sieciami przestrzeni metrycznej  $(X, d)$ , jeśli dla dowolnego  $x \in X$  istnieje  $y \in A$  takie, że  $d(x, y) \leq \varepsilon$ .

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

**Lemat 4.7.** Dla  $0 < \varepsilon \leq 1$  istnieje  $\varepsilon$ -sieć  $N$  w  $S^{n-1}$  taka, że  $\#N \leq \left(\frac{3}{\varepsilon}\right)^n$ .

*Dowód.* Niech  $N$  będzie maksymalnym podzbiorem  $S^{n-1}$  którego dowolne dwa elementy są odległe o więcej niż  $\varepsilon$ . Wówczas  $N$  jest  $\varepsilon$ -sieciami 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

$$\begin{aligned} c_n \left(1 + \frac{\varepsilon}{2}\right)^n &= \lambda_n \left( B \left( 0, 1 + \frac{\varepsilon}{2} \right) \right) \geq \lambda_n \left( \bigcup_{x \in N} B \left( x, \frac{\varepsilon}{2} \right) \right) \\ &= \sum_{x \in N} \lambda_n \left( B \left( x, \frac{\varepsilon}{2} \right) \right) = \#N c_n \left( \frac{\varepsilon}{2} \right)^n, \end{aligned}$$

zatem  $\#N \leq \left(\frac{2}{\varepsilon} + 1\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n$ .

□

**Lemat 4.8.** Niech  $N$  będzie  $\varepsilon$ -siecią w  $S^{n-1}$  dla pewnego  $\varepsilon \in [0, 1]$ . Wówczas dla dowolnych wektorów  $v_1, \dots, v_n$  w  $E$ ,

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

oraz

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

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

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

□

*Dowód Twierdzenia 4.5.* Bez straty ogólności możemy przyjąć, że  $\varepsilon < 1$ , dobierzmy  $\delta > 0$  takie, że  $(1+\delta)/(1-\delta) \leq \sqrt{1+\varepsilon}$  oraz  $1-\delta-\delta(1+\delta)/(1-\delta) \geq 1/\sqrt{1+\varepsilon}$  (np. można przyjąć  $\delta = \varepsilon/20$ ). Niech  $N$  będzie  $\delta$ -siecią w  $S^{k-1}$  mocy nie większej niż  $(\frac{3}{\delta})^k$ . Na podstawie Lematu 4.8 i Uwagi 4.2 wystarczy wykazać, że istnieją wektory  $u_1, \dots, u_k$  w  $E$  takie, że

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

Niech  $X = \sum_{i=1}^m v_i g_i$  spełnia  $d(X) \geq d(E)/2$ , dla uproszczenia notacji przyjmijmy  $d = d(E)$ ,  $\sigma = \sigma(X)$  oraz  $M = \mathbf{E}\|X\|$ . Niech  $F: \mathbb{R}^m \rightarrow \mathbb{R}$  będzie dane wzorem  $F(x) = \left\| \sum_{i=1}^m v_i x_i \right\|$ . Wówczas na mocy Faktu 4.4 funkcja  $F$  jest  $\sigma$ -Lipschitzowska, ponadto  $X \sim F(G)$ , gdzie  $G \sim \gamma_m$ . Stąd

$$\mathbf{P}\left(\left|\frac{\|X\|}{M} - 1\right| \geq \delta\right) = \gamma_m\left(\left|F(G) - \int F d\gamma_m\right| \geq \delta M\right) \leq 2 \exp\left(-\frac{\delta^2 M^2}{2\sigma^2}\right),$$

gdzie ostatnia nierówność jest konsekwencją Twierdzeń 3.28 i 3.33. Niech  $X_1, \dots, X_k$  będą niezależnymi kopiami zmiennej  $X$ . Wówczas, dla dowolnego  $t \in S^{n-1}$ , zmienne  $\sum_{i=1}^k t_i X_i$  ma ten sam rozkład co  $X$ . Zatem

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

czyli

$$\mathbf{P}\left(\exists t \in N \left\|\sum_{i=1}^k t_i \frac{X_i}{M}\right\| - 1 \geq \delta\right) \leq 2\#N e^{-\delta^2 d/4} \leq 2\left(\frac{3}{\delta}\right)^k e^{-\delta^2 d/4} < 1,$$

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

By móc stosować Twierdzenie 4.5 potrzebujemy oszacowań z dołu  $d(E)$ . Zaczniemy od geometrycznego lematu.

**Lemat 4.9** (Dvoretzky-Rogers). *Niech  $E$  będzie  $n$ -wymiarową przestrzenią unormowaną. Wówczas istnieją wektory  $v_1, \dots, v_n \in E$  takie, że  $\|\sum_{i=1}^n t_i v_i\| \leq |t|$  dla  $t \in \mathbb{R}^n$  oraz  $\|v_i\| \geq \frac{n-i+1}{n}$ .*

*Dowód.* Przypomnijmy, że dla  $T: l_2^n \rightarrow E$  określamy

$$\|T\| = \sup\{\|Tx\|: |x| \leq 1\}.$$

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

$$\det(T) = \max\{\det(S): S: l_2^n \rightarrow E, \|S\| \leq 1\}.$$

Wówczas dla dowolnego  $\varepsilon > 0$ ,

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

Mamy jednak

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

więc

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

Wykazaliśmy w ten sposób, że  $\operatorname{tr}(T^{-1}S) \leq n \|S\|$  dla dowolnego  $S: l_2^n \rightarrow E$ . Niech  $S = TP$ , gdzie  $P$  oznacza rzut ortogonalny na podprzestrzeń  $V$ . Wówczas

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

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

**Lemat 4.10.** *Niech  $\varepsilon_i$  będą niezależnymi symetrycznymi zmiennymi losowymi przyjmującymi wartości  $\pm 1$ . Wówczas dla dowolnych wektorów  $u_1, \dots, u_n \in E$  zachodzi*

$$\mathbf{E} \left\| \sum_{i=1}^n u_i \varepsilon_i \right\| \geq \max_{1 \leq i \leq n} \|u_i\|.$$

*Dowód.* Mamy dla dowolnego  $1 \leq i \leq n$ ,

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

□

**Wniosek 4.11.** *Dla dowolnej przestrzeni unormowanej  $E$  wymiaru  $n$ ,  $d(E) \geq \frac{1}{C} \log n$ .*

*Dowód.* Niech  $v_i$  będą takie jak w Lemacie 4.9 oraz  $X = \sum_{i=1}^n v_i g_i$ . Wówczas  $\sigma(X) = 1$ . Niech  $(\varepsilon_i)$  będzie ciągiem niezależnych symetrycznymi zmiennych losowych przyjmujących wartości  $\pm 1$ , niezależnym od  $(g_i)$ . Wówczas na mocy symetrii  $g_i$ ,

$$\mathbf{E} \|X\| = \mathbf{E} \left\| \sum_{i=1}^n \varepsilon_i g_i \right\| \geq \mathbf{E} \max_{i \leq n} \|v_i g_i\| \geq \frac{1}{2} \mathbf{E} \max_{i \leq \lceil n/2 \rceil} |g_i| \geq c \sqrt{\log n}.$$

□

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 wnętrzu). Stąd otrzymujemy równoważne sformułowanie twierdzenia Dvoretzky'ego.

**Twierdzenie 4.12.** *Istnieje stała  $c(\varepsilon)$  taka, że dla dowolnego  $n$ -wymiarowego symetrycznego ciała wypukłego i  $k \leq c(\varepsilon) \log n$  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 4.13.** Proces  $(G_t)_{t \in T}$  nazywamy *procesem gaussowskim*, jeśli dla dowolnych  $t_1, \dots, t_n \in T$  wektor losowy  $(G_{t_1}, \dots, G_{t_n})$  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 4.14.** *Załóżmy, że  $(G_t)_{t \in T}$  jest procesem gaussowskim, indeksowanym przez przeliczalny zbiór indeksów  $T$ , takim, że  $\mathbf{P}(\sup_{t \in T} G_t < \infty) = 1$ . Wówczas  $\mathbf{E} \sup_{t \in T} G_t < \infty$  oraz dla  $u > 0$ ,*

$$\mathbf{P}\left(\sup_{t \in T} G_t - \mathbf{E} \sup_{t \in T} G_t \geq u\right) \leq \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad (20)$$

i

$$\mathbf{P}\left(\sup_{t \in T} G_t - \mathbf{E} \sup_{t \in T} G_t \leq -u\right) \leq \exp\left(-\frac{u^2}{2\sigma^2}\right), \quad (21)$$

gdzie

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

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

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

Określmy  $F: \mathbb{R}^k \rightarrow \mathbb{R}$  wzorem

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

wówczas

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

Stąd

$$\mathbf{P}\left(\sup_{t \in T} G_t - \mathbf{E} \sup_{t \in T} G_t \geq u\right) = \gamma_k\left(\left\{F - \int F d\gamma_k \geq u\right\}\right) \leq e^{-u^2/2\sigma^2},$$

gdzie ostatnia nierówność wynika z Twierdzeń 3.28 i 3.33. Korzystając z tego, że funkcja  $-F$  też jest  $\sigma$ -lipschitzowska wyprowadzamy (21) w przypadku skończonego  $T$ .

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

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

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

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

zatem

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

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

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

Ponadto z Kroku I wynika

$$\mathbf{P}\left(\max_{1 \leq i \leq n} G_{t_i} \geq M_n + u\right) \leq e^{-u^2/2\sigma_n^2} \leq e^{-u^2/2\sigma^2}$$

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

*Uwaga 4.15.* Łącząc (20) i (21) dostajemy

$$\mathbf{P}\left(\left|\sup_{t \in T} G_t - \mathbf{E} \sup_{t \in T} G_t\right| \geq u\right) \leq 2 \exp\left(-\frac{u^2}{2\sigma^2}\right) \text{ dla } u > 0.$$

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 4.16.* Korzystając z izoperymetrii gaussowskiej (Wniosek 2.14) zamiast nierówności logarytmicznej Sobolewa możemy udowodnić, że przy oznaczeniach Twierdzenia 4.14 dla  $u > 0$ ,

$$\mathbf{P}\left(\sup_{t \in T} G_t - \text{Med}\left(\sup_{t \in T} G_t\right) \geq u\right) \leq \Phi\left(\frac{u}{\sigma}\right) \leq \frac{1}{2} \exp\left(-\frac{u^2}{2\sigma^2}\right)$$

oraz

$$\mathbf{P}\left(\left|\sup_{t \in T} G_t - \text{Med}\left(\sup_{t \in T} G_t\right)\right| \geq u\right) \leq 2\Phi\left(\frac{u}{\sigma}\right) \leq \exp\left(-\frac{u^2}{2\sigma^2}\right).$$

**Wniosek 4.17.** *Przy założeniach i oznaczeniach Twierdzenia 4.14,*

$$\lim_{t \rightarrow \infty} \frac{1}{u^2} \log \mathbf{P}\left(\sup_{t \in T} G_t \geq u\right) = -\frac{1}{2\sigma^2}.$$

*Ponadto,*

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

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



*Dowód.* Z Twierdzenia 4.14

$$\frac{1}{u^2} \log \mathbf{P} \left( \sup_{t \in T} G_t \geq u \right) \leq -\frac{1}{2\sigma^2}.$$

Z drugiej strony,

$$\begin{aligned} \liminf_{t \rightarrow \infty} \frac{1}{u^2} \log \mathbf{P} \left( \sup_{t \in T} G_t \geq u \right) &\geq \sup_{t \in T} \liminf_{t \rightarrow \infty} \frac{1}{u^2} \log \mathbf{P}(G_t \geq u) \\ &= \sup_{t \in T} -\frac{1}{2\text{Var}(G_t)} = -\frac{1}{2\sigma^2}. \end{aligned}$$

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  $0 \leq \alpha < 1/2\sigma_t^2$

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

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

**Definicja 4.18.** 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 funkcyjałów o normie jeden.

**Twierdzenie 4.19.** *Załóżmy, że  $X$  jest wektorem gaussowskim w ośrodkowej przestrzeni Banacha. Wówczas  $\mathbf{E}\|X\| < \infty$  oraz dla  $u > 0$ ,*

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

gdzie

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

*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 4.14 dla procesu gaussowskiego  $(\varphi(X))_{\varphi \in D}$ .  $\square$

**Wniosek 4.20.** *Przy oznaczeniach Twierdzenia 4.19 dla  $p \geq 1$ ,*

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

gdzie  $C$  jest pewną stałą uniwersalną.

Uwaga 4.21. Jak nietrudno zauważyć

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

Wynika też stąd, że

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

### 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  $X_1, X_2, \dots, X_n$  są niezależne oraz  $S = \sum_{i=1}^n X_i$ .

**Twierdzenie 4.22** (Nierówność Hoeffdinga). *Dla dowolnego  $t > 0$*

$$\mathbf{P}(S - \mathbf{E}S \geq t) \leq \exp\left(-\frac{t^2}{2D^2}\right),$$

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  $\text{Var}(S)$ . W przeciwnym przypadku dobrze jest stosować precyzyjniejsze nierówności.

**Twierdzenie 4.23** (Nierówność Bernsteina). *Załóżmy, że  $|X_i - \mathbf{E}X_i| \leq C$  dla  $1 \leq i \leq n$ . Wówczas dla  $t \geq 0$ ,*

$$\mathbf{P}(S - \mathbf{E}S \geq t) \leq \exp\left(-\frac{t^2}{2\text{Var}(S) + \frac{2}{3}Ct}\right).$$

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 4.24** (Nierówność Bennetta). *Załóżmy, że  $|X_i - \mathbf{E}X_i| \leq C$  dla  $1 \leq i \leq n$ . Wówczas dla  $t \geq 0$ ,*

$$\mathbf{P}(S - \mathbf{E}S \geq t) \leq \exp\left(-\frac{\text{Var}(S)}{C^2} h\left(\frac{Ct}{\text{Var}(S)}\right)\right) \leq \exp\left(-\frac{t}{2C} \ln\left(1 + \frac{tC}{\text{Var}(S)}\right)\right),$$

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

### 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

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

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 funkcjonałów  $D$  w kuli jednostkowej  $F^*$ , by  $\|x\| = \sup_{\varphi \in D} \varphi(x)$  dla wszystkich  $x \in F$ . Wówczas

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

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

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

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

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

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

**Twierdzenie 4.25.** Załóżmy, że  $0 \leq f \leq 1$  dla  $f \in \mathcal{F}$ . Wówczas dla dowolnego  $\lambda \geq 0$ ,

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

W szczególności dla  $t \geq 0$ ,

$$\mathbf{P}(Z - \mathbf{E}Z \geq t) \leq \exp\left(-\mathbf{E}Zh\left(\frac{t}{\mathbf{E}Z}\right)\right) \leq \exp\left(-\frac{t}{2} \log\left(1 + \frac{t}{\mathbf{E}Z}\right)\right),$$

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

**Lemat 4.26.** Niech  $\Phi(u) = e^{-u} + u - 1$ , wówczas dla  $\lambda \geq 0$ ,

$$\text{Ent}_{\mathbf{P}}(e^{\lambda Z}) \leq \Phi(\lambda)\mathbf{E}Ze^{\lambda Z}.$$

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

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

Bez straty ogólności możemy przyjąć, że  $\#\mathcal{F} = N < \infty$ , niech  $\mu_i$  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

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

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

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

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

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

wtedy z nierówności (22),

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

Określmy mierzalne rozbitcie  $E^n$  na zbiory  $A_k$ ,  $1 \leq k \leq N$  spełniające

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

Wówczas

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

Z wypukłości  $\Phi$ ,  $\Phi(\lambda u) \leq u\Phi(\lambda)$  dla  $u \in [0, 1]$ ,  $\lambda \geq 0$ , zatem

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

W efekcie,

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

□

*Dowód Twierdzenia 4.25.* Bez straty ogólności możemy założyć, że  $Z$  nie jest zdegenerowane. Niech  $M(\lambda) = \mathbf{E}e^{\lambda Z}$ , wówczas na mocy Lematu 4.26 dostajemy dla  $\lambda > 0$ ,

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

co po uwzględnieniu definicji  $\Phi$  daje

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

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

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

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

$$(\log J(\lambda) - \log(e^\lambda - 1))' \leq 0,$$

zatem

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

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

$$\begin{aligned} \mathbf{P}(Z - \mathbf{E}Z \geq t) &\leq \inf_{\lambda \geq 0} e^{-\lambda t + \mathbf{E}Z} \mathbf{E}e^{\lambda Z} \leq \exp\left(-\lambda t + (e^\lambda - \lambda - 1)\mathbf{E}Z\right) \\ &= \exp\left(-\mathbf{E}Z h\left(\frac{t}{\mathbf{E}Z}\right)\right). \end{aligned}$$

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

**Twierdzenie 4.27.** *Załóżmy, że  $|f| \leq C$  dla wszystkich  $f \in \mathcal{F}$ . Wówczas dla  $t \geq 0$ ,*

$$\mathbf{P}\left(|Z - \mathbf{E}Z| \geq t\right) \leq 2 \exp\left(-\frac{1}{10} \min\left\{\frac{t}{C}, \frac{t^2}{4\mathbf{E}\Sigma^2}\right\}\right),$$

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

**Lemat 4.28.** *Dla dowolnej funkcji  $F$*

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

*Dowód.* Na mocy nierówności Jensena

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

Zauważmy, że dla dowolnych liczb  $a, b$

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

Stąd

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

□

**Lemat 4.29.** *Załóżmy, że  $|f| \leq 1$  dla wszystkich  $f \in \mathcal{F}$ , wówczas dla dowolnego  $\lambda \in \mathbb{R}$ ,*

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

*Dowód.* Bez straty ogólności możemy przyjąć, że  $\#\mathcal{F} = N < \infty$ . Jak w dowodzie Lematu 4.26 niech  $\mu_i$  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

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

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

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

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

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

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 4.28,

$$\begin{aligned} & \int \text{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x) \\ & \leq \iiint ((\lambda(Z_i(x_i) - Z_i(y_i)))_+)^2 e^{\lambda Z_i(x_i)} d\mu_i(x_i) d\mu_i(y_i) d\mu(x). \end{aligned}$$

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

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

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

$$\begin{aligned} & \int \text{Ent}_{\mu_i}(e^{\lambda Z(x)}) d\mu(x) \\ & \leq \lambda^2 e^{2|\lambda|} \iiint ((Z_i(x_i) - Z_i(y_i))_+)^2 e^{\lambda Z_i(x_i)} d\mu_i(x_i) d\mu_i(y_i) d\mu(x). \end{aligned}$$

Zauważmy, że

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

więc

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

Zatem

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

□

*Dowód Twierdzenia 4.27.* Z uwagi na jednorodność możemy założyć, że  $C = 1$ . Przyjmijmy  $\tilde{Z} := Z - \mathbf{E}Z$  oraz  $\tilde{M}(\lambda) := \mathbf{E}e^{\lambda \tilde{Z}}$ . Ustalmy  $\lambda_0 > 0$  takie, że  $c_0 := 2e^{2\lambda_0}$  spełnia  $c_0 \lambda_0^2 < 1$ . Lemat 4.29 daje dla  $|\lambda| \leq \lambda_0$ ,

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

Oszacujmy ostatni składnik, mamy

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

gdzie ostatnia nierówność wynika z nierówności Younga (11)  $uv \leq u \log u - u + e^v$  dla  $u = e^{\lambda \tilde{Z}}$ ,  $v = \Sigma^2 - (e-1)\mathbf{E}\Sigma^2$ . Zauważmy, że  $0 \leq f^2 \leq 1$  dla  $f \in \mathcal{F}$ , więc do  $\Sigma^2$  możemy stosować Twierdzenie 4.25, aby dostać  $\mathbf{E}e^{\Sigma^2 - (e-1)\mathbf{E}\Sigma^2} \leq 1$ . Z nierówności Jensena,  $\mathbf{E}e^{\lambda \tilde{Z}} \geq e^{\lambda \mathbf{E}\tilde{Z}} = 1$ , zatem

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

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

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

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

$$H'(\lambda) \leq c_0 \left( e \mathbf{E}\Sigma^2 + \lambda \frac{\tilde{M}'(\lambda)}{\tilde{M}(\lambda)} \right). \quad (23)$$



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

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

Podobnie

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

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

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

Stąd

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

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

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

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

$$\mathbf{P}(Z - \mathbf{E}Z \geq t) \leq \exp\left(-\min\left\{\frac{\lambda_0 t}{2}, \frac{t^2}{4\kappa_0 \mathbf{E}\Sigma^2}\right\}\right).$$

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

**Twierdzenie 4.30.** *Załóżmy, że  $|f| \leq C$  dla wszystkich  $f \in \mathcal{F}$ . Wówczas dla  $t \geq 0$ ,*

$$\mathbf{P}(|Z - \mathbf{E}Z| \geq t) \leq 3 \exp\left(-\frac{t}{KC} \log\left(1 + \frac{Ct}{\mathbf{E}\Sigma^2}\right)\right),$$

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

*Dowód.* Bez straty ogólności możemy założyć, że  $C = 1$ . Nasz dowód będzie się opierał na Twierdzeniu 4.27 i odpowiednim obcinaniu. Dla  $\rho > 0$  określmy

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

oraz

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

Wówczas dla  $t > 0$ ,

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

Stosując Twierdzenie 4.27 do  $Z_\rho$  dostajemy

$$\mathbf{P}(|Z_\rho - \mathbf{E}Z_\rho| \geq t) \leq 2 \exp\left(-\frac{1}{10} \min\left\{\frac{t}{\rho}, \frac{t^2}{4\mathbf{E}\Sigma^2}\right\}\right).$$

By oszacować  $W_\rho$  założmy, że  $t \geq \mathbf{E}W_\rho$  i zastosujmy Twierdzenie 4.25,

$$\mathbf{P}(W_\rho + \mathbf{E}W_\rho \geq 3t) \leq \mathbf{P}(W_\rho - \mathbf{E}W_\rho \geq t) \leq \exp\left(-\frac{t}{2} \log\left(1 + \frac{t}{\mathbf{E}W_\rho}\right)\right).$$

Wybierzmy teraz

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

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

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

Mamy więc dla tak dobranego  $\rho$ ,

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

Zauważmy, że

$$\min\left\{u, \frac{u}{4}\right\} \geq \frac{1}{16} \log(1 + 4u) \text{ dla } u \geq 0,$$

więc

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

Ponadto

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

Zatem

$$\mathbf{P}(|Z - \mathbf{E}Z| \geq 4t) \leq 3 \exp\left(-\frac{t}{160} \log\left(1 + \frac{4t}{\mathbf{E}\Sigma^2}\right)\right)$$

i zamieniając  $4t$  na  $t$  dostajemy tezę. □

*Uwaga 4.31.* Rozpatrując rodzinę funkcji  $\mathcal{F} \cup -\mathcal{F}$  widzimy, że wszystkie udowodnione w tej części oszacowania pozostają też prawdziwe, gdy  $Z$  zamienimy na

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

By móc sensownie stosować Twierdzenia 4.27 i 4.30 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 4.32.** *Załóżmy, że  $|f| \leq C$  oraz  $\mathbf{E}f(X_i) = 0$  dla  $f \in \mathcal{F}$ , wówczas*

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

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

**Fakt 4.33.** *Dla dowolnej przeliczalnej rodziny funkcji ograniczonych  $\mathcal{F}$ ,*

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

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

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

*Dowód.* Niech  $(X'_i)$  będzie kopią ciągu  $(X_i)$  niezależną od  $(X_i)$  oraz  $(\varepsilon_i)$ . Wówczas z nierówności trójkąta oraz nierówności Jensena,

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

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

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

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

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

stąd

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

□

**Twierdzenie 4.34** (Zasada kontrakcji). *Niech  $\varphi_i: \mathbb{R} \rightarrow \mathbb{R}$  dla  $1 \leq i \leq n$  będą funkcjami 1-lipschitzowskimi oraz  $\varphi_i(0) = 0$ . Wówczas dla dowolnego ograniczonego  $T \subset \mathbb{R}^n$  oraz funkcji wypukłej niemalejącej  $F: [0, \infty) \rightarrow [0, \infty)$ ,*

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

*Dowód Twierdzenia 4.32.* Zauważmy, że na podstawie pierwszej części Faktu 4.33,

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

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

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

gdzie ostatnia nierówność wynika z drugiej części Faktu 4.33. □

## 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.