Wysokowymiarowy Rachunek Prawdopodobieństwa

Rafał Latała

28 lutego 2019

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

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

Dziękuję pani Marcie Strzeleckiej i Bartłomiejowi Zawalskiemu za zgłoszone uwagi do notatek.

Rafał Latała

Spis treści

1 Wstęp

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

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

Oczywiście podczas semestralnego wykładu monograficznego można omówić tylko niewielką część bogatej i ciągle rozwijającej się teorii. Dużo szerszy wybór zagadnień został przedstawiony w notatkach Ramona van Handela [7] i monografii Romana Vershynina [8], zainteresowany Czytelnik znajdzie tam też szersze zestawienie bibliografii.

2 Koncentracja miary - wprowadzenie

2.1 Funkcja koncentracji miary - definicja i przykłady.

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

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

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

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

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

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

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

Przykład 1. Niech d oznacza odległość geodezyjną na n-wymiarowej sferze $S^n = \{x \in \mathbb{R}^{n+1} \colon |x| = 1\}$, zaś σ_n oznacza unormowaną miarę powierzchniową na S^n. Wówczas okazuje się, że jeśli chcemy zminimalizować $\sigma_n(A_t)$ po wszystkich zbiorach ustalonej miary, ekstremalne są kule (zwane też czapeczkami), to znaczy

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

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

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

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

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

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

Zauważmy też, że

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

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

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

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

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

to

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

W szczególności

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

Zauważmy, że powyższe oszacowania nie zależą od wymiaru przestrzeni.

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

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

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

W szczególności

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

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

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

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

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

2.2 Koncentracja funkcji lipschitzowskich

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

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

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

Modułem ciągłości f nazywamy funkcję

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

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

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

oraz

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

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

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

Dodając powyższą nierówność do poprzedniej otrzymamy ostatnią część faktu. □

Przypomnijmy definicję funkcji lipschitzowskiej

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

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

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

Analogicznie można zdefiniować funkcje lipschitzowskie między przestrzeniami metrycznymi.

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

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

oraz

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

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

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

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

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

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

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

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

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

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

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

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

W szczególności

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

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

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

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

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

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

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

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

3 Nierówności izoperymetryczne

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

3.1 Klasyczna izoperymetria

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

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

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

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

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

to

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

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

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

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

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

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

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

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

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

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

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

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

Zatem na mocy założenia indukcyjnego

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

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

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

□

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

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

oraz

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

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

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

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

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

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

□

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

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

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

□

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

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

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

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

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

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

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

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

gdzie

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

3.2 Izoperymetria sferyczna

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

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

Wniosek 3.9.

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

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

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

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

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

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

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

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

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

zatem

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

3.3 Izoperymetria gaussowska

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

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

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

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

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

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

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

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

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

Dowód. Proste rozumowanie pokazuje, że miara $\mu_{k,n}$ ma gęstość $g_{k,n}(x) = c_{k,n}^{-1}\tilde{g}_{k,n}(x)$, gdzie $\tilde{g}_{k,n} = (\frac{n-|x|^2}{n})^{(n-k-2)/2}\mathbb{1}_{\{|x|\leqslant\sqrt{n}\}}$ oraz $c_{k,n} = \int_{\mathbb{R}^k}\tilde{g}_{n,k}(x)dx$. Oczywiście $\lim_{n\to\infty}\tilde{g}_{k,n}(x) = \exp(-|x|^2/2)$, ponadto $|\tilde{g}_{k,n}(x)| \leqslant \exp(-(n-k-2)|x|^2/(2n)) \leqslant \exp(-|x|^2/(2n))$ dla $n \geqslant k+2$. Z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej otrzymujemy $\lim_{n\to\infty} c_{n,k} = \int_{\mathbb{R}^k}\exp(-|x|^2/2)dx$, czyli gęstość miary $\mu_{k,n}$ zbiega punktowo do gęstości miary γ_k. Teza faktu wynika z twierdzenia Scheffé'go (zob. zad.8.1.7 w [3]). □

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

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

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

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

Nietrudno zauważyć, że

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

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

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

z dowolności $r_0 < r$ wynika teza. □

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

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

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

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

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

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

Szczególnie wygodny jest transport lipschitzowski.

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

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

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

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

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

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

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

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

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

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

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

oraz

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

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

4 Metoda Martyngałowa

4.1 Transformata Laplace'a

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

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

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

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

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

Fakt 4.2. *Dla dowolnej zmiennej losowej Z,*

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

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

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

to dla $t \geqslant 0$

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

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

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

4.2 Nierówność Azumy

Poniższa nierówność to udowodnione przez Azumę uogólnienie nierówności Hoeffdinga (zob. Fakt 4.9 poniżej) na przypadek martyngałowy.

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

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

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

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

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

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

Liczymy

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

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

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

Teza twierdzenia wynika z Faktu 4.2. □

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

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

4.3 Zastosowania nierówności Azumy

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

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

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

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

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

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

stąd

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

i teza wynika z Twierdzenia 4.3. □

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

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

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

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

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

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

w szczególności

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

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

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

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

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

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

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

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

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

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

4.4 Nierówności wykładnicze dla sum niezależnych zmiennych losowych

W tej części omówimy kilka nierówności wykładniczych dla sum niezależnych zmiennych losowych, które bazują na szacowaniu transformaty Laplace'a. Dla uproszczenia notacji zdefiniujemy dla zmiennej losowej Z i $\lambda \in \mathbb{R}$,

$$\Lambda_Z(\lambda) := \ln L_Z(\lambda) = \ln \mathbf{E}e^{\lambda Z}.$$

Fakt 4.9 (Hoeffding)**.** *Jeśli X_i są niezależnymi zmiennymi losowymi takimi, że $a_i \leqslant X_i \leqslant b_i$ oraz $S = \sum_{i=1}^n X_i$, to*

$$\mathbf{P}(S \geqslant \mathbf{E}S + t) \leqslant \exp\left(-\frac{t^2}{2\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Dowód. Wystarczy zauważyć, że $M_k = \sum_{i=1}^k (X_i - \mathbf{E}X_i)$ jest martyngałem, $|M_k - M_{k-1}| = |X_k - \mathbf{E}X_k| \leqslant b_k - a_k$ i skorzystać z Twierdzenia 4.3. $\square$

Lemat 4.10. *Załóżmy, że X jest zmienną losową o średniej zero taką, że istnieją $\sigma^2, M < \infty$ spełniające warunek*

$$\mathbf{E}|X|^k \leqslant \frac{k!}{2}\sigma^2 M^{k-2} \qquad \textit{dla } k = 2, 3, \ldots.$$

Wówczas

$$\Lambda_X(\lambda) \leqslant \frac{\sigma^2 t^2}{2(1 - M|\lambda|)} \qquad \textit{dla } M|\lambda| < 1.$$

Dowód. Liczymy

$$\begin{aligned} L_X(\lambda) &= \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathbf{E}X^k \leqslant 1 + \sum_{k=2}^{\infty} \frac{|\lambda|^k}{2}\sigma^2 M^{k-2} = 1 + \frac{\lambda^2\sigma^2}{2}\sum_{k=2}^{\infty}(|\lambda| M)^{k-2} \\ &= 1 + \frac{\sigma^2\lambda^2}{2(1 - M|\lambda|)} \leqslant \exp\left(\frac{\sigma^2\lambda^2}{2(1 - M|\lambda|)}\right). \end{aligned}$$

$\square$

Twierdzenie 4.11 (Nierówność Bernsteina)**.** *Załóżmy, że X_i są niezależnymi zmiennymi losowymi o średniej zero, zaś $\sigma_i^2, M < \infty$ są takie, że*

$$\mathbf{E}|X_i|^k \leqslant \frac{k!}{2}\sigma_i^2 M^{k-2} \qquad \textit{dla } i \geqslant 1,\ k \geqslant 2. \tag{4}$$

Wówczas

$$\mathbf{E}\exp\left(\lambda \sum_{i=1}^n X_i\right) \leqslant \exp\left(\frac{\lambda^2 \sum_{i=1}^n \sigma_i^2}{2(1 - M|\lambda|)}\right) \qquad \textit{dla } M|\lambda| < 1$$

oraz dla $t > 0$,

$$\begin{aligned} \mathbf{P}\left(\sum_{i=1}^n X_i \geqslant t\right) &\leqslant \exp\left(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2 + 2Mt}\right), \\ \mathbf{P}\left(\left|\sum_{i=1}^n X_i\right| \geqslant t\right) &\leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n \sigma_i^2 + 2Mt}\right). \end{aligned}$$

Dowód. Niech $S := \sum_{i=1}^n X_i$, $\sigma^2 := \sum_{i=1}^n \sigma_i^2$, wówczas $\Lambda_S = \sum_i \Lambda_{X_i}$ i pierwsze oszacowanie wynika z Lematu 4.10. Dalej szacujemy

$$\begin{aligned} \mathbf{P}(S \geqslant t) &\leqslant \exp\left(-\sup_{\lambda > 0}(\lambda t - \Lambda_S(\lambda))\right) \leqslant \exp\left(-\sup_{0<\lambda<M^{-1}}\left(\lambda t - \frac{\lambda^2\sigma^2}{2(1 - M\lambda)}\right)\right) \\ &\leqslant \exp\left(-\frac{t^2}{2\sigma^2 + 2Mt}\right), \end{aligned}$$

gdzie ostatnią nierówność dostajemy przyjmując $\lambda = t(\sigma^2+Mt)^{-1}$. Ponieważ zmienne $-X_i$ spełniają te same założenia co X_i, więc dostajemy dla $t < 0$,

$$\mathbf{P}(S \leqslant t) = \mathbf{P}(-S \geqslant -t) \leqslant \exp\left(-\frac{t^2}{2\sigma^2 + 2Mt}\right)$$

i z tożsamości $\mathbf{P}(|S| \geqslant t) = \mathbf{P}(S \geqslant t) + \mathbf{P}(S \leqslant -t)$ wynika ostatnia część tezy. □

Uwaga 4.12. Na mocy centralnego twierdzenia granicznego oraz szacowania dystrybuanty gaussowskiej nie możemy się spodziewać lepszego oszacowania niż $\exp(-t^2/(2\sigma^2))$. Ponadto zmienne X_i o rozkładzie symetrycznym wykładniczym z parametrem 1 (tzn. zmienne z gęstością $\exp(-|x|)/2$) spełniają $\mathbf{E}|X_i|^k = k!$, czyli dla takich zmiennych zachodzą założenia Twierdzenia 4.11 z $\sigma_i^2 = 2, M = 1$. Pokazuje to, że nie możemy uzyskać szacowania lepszego niż $\exp(-t/M)$ przy $t \to \infty$.

Wniosek 4.13. *Załóżmy, że X_i są ograniczonymi, niezależnymi zmiennymi losowymi o średniej zero, wówczas dla $t > 0$,*

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \exp\left(-\frac{t^2}{2\sigma^2 + 2at/3}\right),$$

gdzie $\sigma^2 = \mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbf{E}X_i^2$ oraz $a = \max_i \|X_i\|_\infty$.

Dowód. Mamy dla $k \geqslant 2$,

$$\mathbf{E}|X_i|^k \leqslant a^{k-2}\mathbf{E}X_i^2 \leqslant \frac{k!}{2}\left(\frac{a}{3}\right)^{k-2}\mathbf{E}X_i^2,$$

zatem warunek (4) jest spełniony z $M = a/3$ oraz $\sigma_i = \mathbf{E}X_i^2$. □

W wielu zastosowaniach wygodniej zamiast bezpośrednio oszacowania (4) używać szacowania stałej subwykładniczej zmiennych losowych X_i.

Definicja 4.14. Mówimy, że zmienna losowa X_i jest *subwykładnicza*, jeśli $\mathbf{E}\exp(\lambda|X|) < \infty$ dla pewnego $\lambda > 0$. Dla zmiennej subwykładniczej X określamy jej *stałą subwykładniczą* wzorem

$$\|X\|_{\psi_1} := \inf\{\lambda > 0\colon\ \mathbf{E}e^{|X|/\lambda} \leqslant 2\}.$$

Wielkość $\|X\|_{\psi_1}$ to nic innego jak norma Orlicza X dla funkcji Younga $\psi_1(x) = e^x - 1$.

Lemat 4.15. *Jeśli X jest zmienną subwykładniczą, to $\|X\|_{\psi_1} < \infty$. Ponadto,*

$$\mathbf{E}|X|^k \leqslant k!\|X\|_{\psi_1}^k \quad \textit{dla } k = 1, 2, \ldots.$$

Dowód. Pierwsza część wynika stąd, że funkcja $\lambda \mapsto (\mathbf{E}\exp(\lambda|X|))^{1/\lambda}$ jest niemalejąca na $(0,\infty)$. Biorąc $t > \|X\|_{\psi_1}$ dostajemy

$$\frac{1}{k!}\mathbf{E}\left(\frac{|X|}{t}\right)^k \leqslant \mathbf{E}\exp\left(\frac{|X|}{t}\right) - 1 \leqslant 1$$

$\square$

Twierdzenie 4.16 (Nierówność Bernsteina dla zmiennych subwykładniczych)**.** *Załóżmy, że X_i są niezależnymi subwykładniczymi zmiennymi losowymi o średniej zero. Wówczas dla $t > 0$,*

$$\mathbf{P}\left(\sum_{i=1}^n X_i \geqslant t\right) \leqslant \exp\left(-\frac{t^2}{4\sum_{i=1}^n \|X_i\|_{\Psi_1}^2 + 4t\max_i \|X_i\|_{\Psi_1}}\right)$$

oraz

$$\mathbf{P}\left(\left|\sum_{i=1}^n X_i\right| \geqslant t\right) \leqslant 2\exp\left(-\frac{t^2}{4\sum_{i=1}^n \|X_i\|_{\Psi_1}^2 + 4t\max_i \|X_i\|_{\Psi_1}}\right).$$

Dowód. Lemat 4.15 implikuje, że oszacowanie (4) zachodzi z $\sigma_i^2 = 2\|X_i\|_{\psi_1}^2$ i $M = \max_i \|X\|_{\psi_1}$. Wystarczy zatem zastosować Twierdzenie 4.11. $\square$

Szacowanie podane we Wniosku 4.13 jest, z uwagi na centralne twierdzenie graniczne, bliskie optymalnego dla t małych. Jednak dla t dużych można je poprawić o czynnik logarytmiczy.

Lemat 4.17. *Załóżmy, że X jest zmienną losową o średniej zero, wariancji σ^2 oraz $\|X_i\|_\infty \leqslant a$. Wówczas*

$$\Lambda_X(\lambda) \leqslant \frac{\sigma^2}{a^2}(e^{\lambda a} - \lambda a - 1) \quad \text{dla } \lambda \geqslant 0.$$

Dowód. Liczymy

$$\mathbf{E}e^{\lambda X} = 1 + \lambda\mathbf{E}X + \sum_{k=2}^\infty \frac{\lambda^k \mathbf{E}X^k}{k!} \leqslant 1 + \sigma^2\sum_{k=2}^\infty \frac{\lambda^k a^{k-2}}{k!} = 1 + \frac{\sigma^2}{a^2}(e^{\lambda a} - \lambda a - 1)$$

i teza wynika natychmiast z nierówności $\ln(1+x) \leqslant x$. $\square$

Twierdzenie 4.18 (nierówność Bennetta)**.** *Załóżmy, że X_i są ograniczonymi niezależnymi zmiennymi losowymi o średniej zero, $\sigma^2 = \mathrm{Var}(\sum_{i=1}^n X_i) = \sum_{i=1}^n \mathbf{E}X_i^2$ oraz $a \geqslant \max_i \|X_i\|_\infty$. Wówczas dla $\lambda > 0$,*

$$\mathbf{E}\exp\left(\lambda\sum_{i=1}^n X_i\right) \leqslant \exp\left(\frac{\sigma^2}{a^2}(e^{\lambda a} - \lambda a - 1)\right)$$

oraz dla $t > 0$,

$$\mathbf{P}\left(\sum_{i=1}^{n} X_i \geqslant t\right) \leqslant \exp\left(-\frac{\sigma^2}{a^2} h\left(\frac{ta}{\sigma^2}\right)\right) \leqslant \exp\left(-\frac{t}{2a}\ln\left(1+\frac{ta}{\sigma^2}\right)\right),$$

gdzie

$$h(x) := (1+x)\ln(1+x) - x.$$

Dowód. Pierwsza część wynika natychmiast z Lematu 4.17. By pokazać drugą zauważamy, że dla $S = \sum_{i=1}^n X_i$,

$$\mathbf{P}(S \geqslant t) \leqslant \exp\left(-\sup_{\lambda>0}(\lambda t - \Lambda_S(\lambda))\right) \leqslant \exp\left(-\sup_{\lambda>0}\left(\lambda t - \frac{\sigma^2}{a^2}(e^{\lambda a} - \lambda a - 1)\right)\right).$$

Prosty rachunek pokazuje, że powyższe supremum jest osiągane w punkcie

$$\lambda = \frac{1}{a}\ln\Big(1 + \frac{at}{\sigma^2}\Big)$$

i wynosi

$$\frac{\sigma^2}{a^2}\Big[\Big(1+\frac{ta}{\sigma^2}\Big)\ln\Big(1+\frac{ta}{\sigma^2}\Big) - \frac{ta}{\sigma^2}\Big] = \frac{\sigma^2}{a^2}h\Big(\frac{ta}{\sigma^2}\Big) \geqslant \frac{t}{2a}\ln\Big(1+\frac{ta}{\sigma^2}\Big),$$

gdzie ostatnie oszacowanie wynika z poniższego lematu. □

Lemat 4.19. *Dla dowolnego $x \geqslant 0$,*

$$(1+x)\ln(1+x) - x \geqslant \frac{x}{2}\ln(1+x).$$

Dowód. Niech $f(x) = (1+x)\ln(1+x) - x - (x/2)\ln(1+x) = (1+x/2)\ln(1+x) - x$. Liczymy $f'(x) = (\ln(1+x) - x(1+x)^{-1})/2$, $f''(x) = x(1+x)^{-2}$, zatem $f(0) = f'(0) = 0$ oraz $f''(x) \geqslant 0$ dla $x \geqslant 0$. □

Uwaga 4.20. Jeśli $\mathbf{P}(Y_{n,i} = 1) = 1 - \mathbf{P}(Y_{n,i} = 0) = 1/n$ oraz $Y_{n,i}$ są niezależne, to rozkład $\sum_{i=1}^n Y_{i,n}$ zbiega do rozkładu Poissona z parametrem 1. Biorąc $X_{n,i} = Y_{n,i} - 1/n$ mamy $\sum_{i=1}^n \mathbf{E}X_{i,n}^2 \leqslant 1$ oraz $\max_i \|X_{i,n}\|_\infty \leqslant 1$. To pokazuje, że przy założeniach Twierdzenia 4.18 nie można uzyskać przy $t \to \infty$ oszacowania lepszego rzędu niż $t \ln t$.

Uwaga 4.21. Nierówność Bennetta ma swoją wersję martyngałową. Mianowicie, dla martyngału $(M_k, \mathcal{F}_k)_{k=0}^n$ spełniającego warunki

$$\max_k \|M_k - M_{k-1}\|_\infty \leqslant a$$

i

$$\sum_{k=1}^n \|\mathbf{E}((M_k - M_{k-1})^2 | \mathcal{F}_{k-1})\|_\infty \leqslant \sigma^2,$$

zachodzi nierówność

$$\mathbf{P}(M_n - M_0 \geqslant t) \leqslant \exp\left(-\frac{\sigma^2}{a^2}h\left(\frac{ta}{\sigma^2}\right)\right) \leqslant \exp\left(-\frac{t}{2a}\ln\left(1+\frac{ta}{\sigma^2}\right)\right).$$

5 Nierówność Poincaré

5.1 Definicja i podstawowe własności

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

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

gdzie

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

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

Uwaga 5.2. W przypadku, gdy $\mathbb{X} = \mathbb{R}^n$ ze standardową metryką euklidesową możemy użyć twierdzenia Rademachera, które mówi, że każda funkcja Lipchitzowska jest różniczkowalna prawie wszędzie i wtedy $|\nabla f|(x)$ jest dla prawie wszystkich x równy długości zwykłego gradientu f. Ponadto argument aproksymacyjny pokazuje, że by wykazać nierówność Poincaré dla miar probabilistycznych na $\mathbb{R}^n$ wystarczy sprawdzić (5) dla ograniczonych funkcji klasy $C^1(\mathbb{R}^n)$ o ograniczonych pochodnych rzędu jeden.

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

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

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

Definiując operator $Lf := -\triangle f + \langle \nabla V, \nabla f \rangle$ widzimy, że $L1 = 0$. Nierówność Poincaré mówi, że dla funkcji f o średniej 0, czyli prostopadłych do 1, $\int fLfd\mu \geqslant C^{-1} \int f^2 d\mu$. Biorąc pod uwagę samosprzężoność L nierówność (5) jest równoważna temu, że kolejna wartość własna L to conajmniej $1/C$. Dlatego nierówność Poincaré się nazywa nierównością „luki spektralnej" (spectral gap inequality).

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

Fakt 5.5. *Nierówność Poincaré jest równoważna nierówności*

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

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

Dowód. Ponieważ

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

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

By udowodnić przeciwne oszacowanie zauważmy, że

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

Stąd

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

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

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

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

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

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

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

stąd

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

□

5.2 Nierówność Poincaré a koncentracja wykładnicza

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

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

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

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

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

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

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

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

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

Iterując tę nierówność n razy dostajemy

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

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

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

Zauważmy, że

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

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

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

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

5.3 Tensoryzacja

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

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

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

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

gdzie ostatnia nierówność wynika np. z nierówności Jensena. □

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

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

Dowód. Z Faktu 5.9 dostajemy

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

□

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

5.4 Dodatkowe własności. Charakteryzacja na prostej.

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

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

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

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

□

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

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

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

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

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

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

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

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

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

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

5.5 Nierówność Cheegera

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

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

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

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

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

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

Załóżmy najpierw, że $t_0 < \infty$. Wówczas z monotoniczności $\mu(A_t)$ łatwo wynika, że $\mu(A_{t_0}) = \nu(-\infty, x+\delta ct_0]$, czyli

$$\begin{aligned}\mu^+(A_{t_0}) &\geqslant c\min\{\mu(A_{t_0}), 1-\mu(A_{t_0})\} = \frac{c}{2}e^{-|x+\delta ct_0|} \\ &= \frac{1}{\delta}\lim_{h\to 0+} \frac{\nu(-\infty, x+\delta c(t_0+h)] - \nu(-\infty, x+\delta ct_0]}{h}.\end{aligned}$$

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

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

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

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

Definicja 5.16. Mówimy, że miara probabilistyczna μ na $(\mathbb{X}, d)$ *spełnia nierówność Cheegera ze stałą* $c > 0$, jeśli zachodzi jeden z warunków równoważnych Faktu 5.15.

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

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

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

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

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

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

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

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

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

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

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

Mamy zatem

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

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

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

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

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

„$\Leftarrow$" Udowodnimy szacowanie (i) z Faktu 5.15. Idea polega na aproksymacji 1_A przez funkcje lipschitzowskie. Jeśli $\mu(\bar{A}) > \mu(A)$, to $\mu^+(A) = \infty$ i nie ma co dowodzić, będziemy zatem zakładać, że $\mu(\bar{A}) = \mu(A)$, co jest równoważne temu, że $\mu(A_t) \to \mu(A)$ przy $t \to 0$. Dla $0 < t < 1/2$ określmy

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

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

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

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

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

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

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

□

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

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

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

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

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

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

□

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

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

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

Dowód. Dowód przeprowadzimy w kilku krokach, wykorzystując równoważności z Faktu 5.15.

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

Rozpatrzymy trzy przypadki.
i) $a \geqslant 0$. Wówczas $\nu^+([a,b]) = e^{-a}+e^{-b} \geqslant e^{-a}-e^{-b} = \nu([a,b])$.
ii) $b \leqslant 0$. Mamy $\nu^+([a,b]) = e^{a}+e^{b} \geqslant e^{b}-e^{a} = \nu([a,b])$.
iii) $a < 0 < b$. Wtedy $\nu^+([a,b]) = e^{a}+e^{-b} = 1-\nu([a,b])$.

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

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

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

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

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

Zauważamy, że zbiór A_t jest również skończoną sumą przedziałów, więc z Kroku II wynika, że $\nu^+(A_t) \geqslant \min\{\nu(A_t), 1-\nu(A_t)\}$. Teza Kroku III wynika z analogicznego rozumowania jak w dowodzie implikacji (i)$\Rightarrow$(ii) Faktu 5.15.

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

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

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

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

Na prostej można scharakteryzować miary spełniające nierówność Cheegera.

Twierdzenie 5.23. *Niech μ będzie miarą probabilistyczną na $\mathbb{R}$, $F(x) = \mu(-\infty,x]$, zaś p będzie gęstością części absolutnie ciągłej μ. Wówczas następujące warunki są równoważne dla $c > 0$:*

i) μ *spełnia nierówność Cheegera ze stałą* c*,*
ii) μ *jest* $\frac{1}{c}$*-lipschitzowskim obrazem* ν*,*
iii) $\operatorname{essinf} \frac{p(x)}{\min\{F(x), 1-F(x)\}} \geqslant c$*.*

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

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

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

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

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

6 Logarytmiczna Nierówność Sobolewa

6.1 Entropia funkcji

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

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

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

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

$$\mathrm{Ent}_\mu(f) = \sup\Big\{\int fg d\mu\colon\ \int e^g d\mu \leqslant 1\Big\}. \tag{6}$$

Dowód. Z jednorodności obu stron tożsamości (6) możemy zakładać, że $\int f d\mu = 1$, wówczas $\mathrm{Ent}_\mu(f) = \int f\log f d\mu$.

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

$$uv \leqslant u\log u - u + e^v \text{ dla } u \geqslant 0, v\in\mathbb{R}. \tag{7}$$

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

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

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

Z powyższego lematu łatwo wykazać tensoryzowalność entropii:

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

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

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

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

□

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

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

6.2 LNS - definicja, tensoryzowalność, związek z koncentracją

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

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

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

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

Dowód. Z Faktu 6.3 dostajemy

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

□

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

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

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

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

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

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

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

oraz

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

Zatem (8) daje

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

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

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

oraz na podstawie (9)

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

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

6.3 LNS dla miary gaussowskiej

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

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

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

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

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

gdzie

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

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

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

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

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

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

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

ii) Wynika z punktu i) i Faktu 6.3. $\square$

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

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

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

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

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

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

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

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

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

$\square$

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

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

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

Zatem

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

□

Kolejny fakt dowodzimy tak samo jak dla nierówności Poincaré.

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

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

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

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

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

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

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

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

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

6.4 Nierówność Bobkowa

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

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

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

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

$$I\Big(\int f d\mu\Big) \leqslant \int \sqrt{I(f)^2 + C^2|\nabla f|^2} d\mu. \tag{10}$$

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

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

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

oraz

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

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

6.5 Wektory i Procesy Gaussowskie

Procesy i wektory gaussowskie odgrywają kluczową rolę w rachunku prawdopodobieństwa i statystyce matematyczne, jak również w wielu zastosowaniach.

Zacznijmy od przypomnienia definicji.

Definicja 6.17. Proces $(G_t)_{t\in T}$ nazywamy *procesem gaussowskim*, jeśli dla dowolnych $t_1, \ldots, t_n \in T$ wektor losowy $(G_{t_1}, \ldots, G_{t_n})$ ma rozkład gaussowski. Proces nazywamy *scentrowanym*, jeśli $\mathbf{E}G_t = 0$ dla $t \in T$.

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

Twierdzenie 6.18. *Załóżmy, że $(G_t)_{t\in T}$ jest procesem gaussowskim, indeksowanym przez przeliczalny zbiór T, takim, że $Z := \sup_{t\in T} G_t < \infty$ prawie na pewno. Wówczas $\mathbf{E}Z < \infty$,*

$$\mathbf{E}e^{\lambda(Z-\mathbf{E}Z)} \leqslant e^{-\lambda^2\sigma^2/2} \qquad \textit{dla } \lambda \in \mathbb{R}, \tag{11}$$

$$\mathbf{P}(Z-\mathbf{E}Z \geqslant u) \leqslant \exp\Big(-\frac{u^2}{2\sigma^2}\Big) \quad \textit{i} \quad \mathbf{P}(Z-\mathbf{E}Z \leqslant -u) \leqslant \exp\Big(-\frac{u^2}{2\sigma^2}\Big) \quad \textit{dla } u>0, \tag{12}$$

gdzie

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

Dowód. Nierówność (12) wynika z (11), udowodnimy zatem tę pierwszą.

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

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

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

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

wówczas Z ma ten sam rozkład co $F(X)$ i

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

Stąd

$$\mathbf{E}e^{\lambda(Z-\mathbf{E}Z)} = \mathbf{E}e^{\lambda(F(Z)-\mathbf{E}F(Z))} \leqslant e^{-\lambda^2\sigma^2/2},$$

gdzie ostatnia nierówność wynika z Twierdzeń 6.6 i 6.8.

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

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

Niech M spełnia $\mathbf{P}(Z \geqslant M) \leqslant 1/4$. Z Kroku I dostajemy

$$\mathbf{P}\Big(Z_n - \mathbf{E}Z_n \leqslant -\sigma_n\Big) \leqslant e^{-1/2},$$

zatem

$$\mathbf{P}\Big(Z_n > \mathbf{E}Z_n - \sigma_n\Big) \geqslant 1 - e^{-1/2} > \frac{1}{4} \geqslant \mathbf{P}\Big(Z_n \geqslant M\Big),$$

stąd $M \geqslant \mathbf{E}Z_n - \sigma_n$. Mamy więc z twierdzenia Lebesgue'a o zbieżności monotonicznej,

$$\mathbf{E}Z = \lim_n \mathbf{E}Z_n \leqslant \sup_n \{M + \sigma_n\} \leqslant M + \sigma < \infty.$$

Stosując oszacowanie z Kroku I otrzymujemy

$$\mathbf{E}e^{\lambda(Z-\mathbf{E}Z)} = \lim_n \mathbf{E}e^{\lambda(Z_n - \mathbf{E}Z_n)} \leqslant \lim_n e^{\lambda^2\sigma_n^2/2} = e^{-u^2\sigma^2/2},$$

gdzie pierwsza równość wynika z tego, że $\mathbf{E}Z_n \to \mathbf{E}Z$ oraz twierdzenia Lebesgue'a o zbieżności monotonicznej dla $\lambda \geqslant 0$, bądź zmajoryzowanej dla $\lambda < 0$. □

Uwaga 6.19. Łącząc nierówności (12) dostajemy

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

Zauważmy też, że $|G_t| = \max\{G_t, -G_t\}$, więc w Twierdzeniu 6.18 i powyższej nierówności można zastąpić G_t przez $|G_t|$.

Uwaga 6.20. Korzystając z izoperymetrii gaussowskiej (Wniosek 3.13) zamiast nierówności logarytmicznej Sobolewa możemy udowodnić, że przy oznaczeniach Twierdzenia 6.18 dla $u > 0$,

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

oraz

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

Wniosek 6.21. *Przy założeniach i oznaczeniach Twierdzenia 6.18,*

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

Ponadto,

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

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

Dowód. Z Twierdzenia 6.18

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

Z drugiej strony,

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

Druga część tezy dla $\alpha<\frac{1}{2\sigma^2}$ wynika natychmiast z (12) (dla $|G_t|$). Ponadto, jeśli $G_t\sim\mathcal{N}(a_t,\sigma_t^2)\sim a_t+\sigma_t g$, to dla $0\leqslant\alpha<1/2\sigma_t^2$

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

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

Definicja 6.22. Wektor losowy X w ośrodkowej przestrzeni Banacha F nazywamy *gaussowskim*, jeśli dla dowolnego $\varphi\in F^*$, $\varphi(X)$ ma rozkład gaussowski.

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

Twierdzenie 6.23. *Załóżmy, że X jest wektorem gaussowskim w ośrodkowej przestrzeni Banacha. Wówczas $\mathbf{E}\|X\|<\infty$,*

$$\mathbf{E}e^{\lambda(\|X\|-\mathbf{E}\|X\|)}\leqslant e^{\lambda^2\sigma^2/2}\quad \textit{dla } \lambda\in\mathbb{R},$$

$$\mathbf{P}(\|X\|-\mathbf{E}\|X\|\geqslant u)\leqslant e^{-\frac{u^2}{2\sigma^2}}\quad \textit{oraz}\quad \mathbf{P}(\|X\|-\mathbf{E}\|X\|\leqslant -u)\leqslant e^{-\frac{u^2}{2\sigma^2}}\qquad \textit{dla } u\geqslant 0,$$

gdzie

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

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 6.18 dla procesu gaussowskiego $(\varphi(X))_{\varphi\in D}$. □

Wniosek 6.24. *Przy oznaczeniach Twierdzenia 6.23 dla $p\geqslant 1$,*

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

gdzie C,C' są pewnymi stałymi uniwersalnymi.

Dowód. Pierwsza nierówność wynika łatwo z Twierdzenia 6.23 i całkowania przez części. By udowodnić drugą wystarczy zauważyć, że dla dowolnego funkcjonału $\varphi \in F^*$ o normie 1 zachodzi

$$(\mathrm{Var}(\varphi(X))^{1/2} = \sqrt{\frac{\pi}{2}}\mathbf{E}|\varphi(X) - \mathbf{E}\varphi(X)\rangle| \leqslant \sqrt{\frac{\pi}{2}}\mathbf{E}\|X - \mathbf{E}X\| \leqslant \sqrt{2\pi}\mathbf{E}\|X\|.$$

□

Uwaga 6.25. Jak nietrudno zauważyć

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

Stąd dla $p \geqslant 1$,

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

7 Nierówności Splotu Infimum

7.1 Własność (τ) Maureya

Zacznijmy od zaproponowanej przez Maureya definicji.

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

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

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

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

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

Pierwszą użyteczną cechą własności (τ) jest jej tensoryzowalność.

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

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

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

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

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

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

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

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

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

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

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

Teza wynika z powyższych nierówności i własności τ dla (μ_2, φ_2). $\Box$

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

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

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

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

Zatem

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

$\Box$

7.2 Splot infimum a koncentracja

By sformułować związki nierówności splotu infimum z koncentracją określmy zbiór

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

Zacznijmy od prostego faktu

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

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

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

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

$\square$

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

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

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

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

W szczególności

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

oraz

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

Ponadto

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

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

Zatem własność (τ) daje

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

skąd bezpośredni rachunek prowadzi do (13).

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

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

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

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

i dostajemy (15).

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

□

7.3 Dwupoziomowa koncentracja dla rozkładu wykładniczego

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

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

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

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

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

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

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

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

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

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

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

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

Wówczas

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Ponadto

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

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

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

Teza wynika zatem z Faktów 7.4 i 7.6. □

7.4 Wypukła własność (τ)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Z dowolności $\varepsilon > 0$ dostajemy

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

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

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

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

Mamy zatem

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

□

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

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

W szczególności

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

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

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

□

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

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

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

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

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

8 Nierówności transportowe

8.1 Koszt optymalnego transportu

By zdefiniować koszt transportu miar będziemy potrzebowali kilku definicji.

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

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

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

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

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

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

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

Uwaga 8.5. Równoważnie możemy zdefiniować

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

Uwaga 8.6. Zauważmy, że

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

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

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

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

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

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

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

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

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

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

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

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

8.2 Względna entropia

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

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

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

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

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

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

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

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

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

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

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

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

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

co jest oczywiście równoważne warunkowi i). □

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

8.3 Tensoryzacja nierówności transportowych

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

i z wypukłości φ

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

Zatem

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

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

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

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

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

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

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

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

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

8.4 Nierówność T_2 Talagranda a bezwymiarowa koncentracja

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

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

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

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

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

i teza łatwo wynika z Twierdzenia 8.12. □

Okazuje się, że nierówność T_2 jest równoważna bezwymiarowej koncentracji.

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

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

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

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

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

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

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

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

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbf{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{X_k} \in G\right) \geqslant -\inf_{\nu\in G} H(\nu|\mu). \tag{17}$$

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

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

Dowód Twierdzenia 8.22. Ustalmy $\nu \in U$ takie, że $H(\nu|\mu) < \infty$ (jeśli takie ν nie istnieje, to infimum po lewej stronie (17) jest równe $+\infty$ i nierówność jest oczywista). Niech $g = \frac{d\nu}{d\mu}$ oraz $Y_1, Y_2, \ldots$ będą niezależnymi zmiennymi o rozkładzie ν. Wówczas $g(Y_i) > 0$ p.n. oraz dla dowolnej funkcji mierzalnej f na $\mathbb{X}^n$,

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

Mamy

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

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

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

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

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

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

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

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

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

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

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

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

Z dowolności $a > 0$ mamy

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

$\square$

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

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

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

$\square$

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

Dowód. Ustalmy $a > 0$. Mamy

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

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

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

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

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

Stąd dla dowolnego $a > 0$,

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

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

□

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

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

Fakt 8.24 implikuje, że zbiór

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

jest otwarty. Zatem z twierdzenia Sanowa

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

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

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

Stąd

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

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

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

skąd łatwo wynika nierówność T_2.

□

9 Aproksymacja przez otoczkę wypukłą

9.1 Definicje

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

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

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

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

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

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

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

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

oraz

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Stąd

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

□

9.2 Twierdzenie Talagranda

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

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

W szczególności dla $t > 0$,

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

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

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

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

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

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

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

Zauważmy, że jeśli $s \in U_{A(x_n)}(x)$, to $(s, 0) \in U_A(x)$, a jeśli $t \in U_B(x)$, to $(t, 1)$ lub $(t, 0)$ należą do $U_A(x)$. Zatem jeśli wybierzemy $s \in V_{A(x_n)}(x)$ oraz $t \in V_B(x)$, to $(s, 0) \in U_A(x)$

oraz $(t,b) \in V_A(x)$ dla pewnego $b \in [0,1]$, czyli z wypukłości zbioru $V_A(x)$, $(\theta s + (1-\theta)t, (1-\theta)b) \in V_A(x)$. Stąd z wypukłości funkcji $|x|^2$,

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

czyli z dowolności wyboru t i s,

$$\mathcal{D}^c_A(x)^2 \leqslant \theta\mathcal{D}^c_{A(x_n)}(\tilde{x})^2 + (1-\theta)\mathcal{D}^c_B(\tilde{x})^2 + (1-\theta)^2.$$

Odcałkowywując i korzystając z nierówności Höldera dostajemy

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

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

$$\int_{\tilde{X}} \exp\Big(\frac{\mathcal{D}^c_A(\tilde{x},x_n)^2}{4}\Big)d\tilde{\mu}(\tilde{x}) \leqslant e^{(1-\theta)^2/4}\Big(\frac{1}{\tilde{\mu}(A(x_n))}\Big)^{\theta}\Big(\frac{1}{\tilde{\mu}(B)}\Big)^{1-\theta}. \tag{18}$$

Zauważmy teraz, że

$$\inf_{\theta\in[0,1]} e^{(1-\theta)^2/4}u^{-\theta} \leqslant 2-u \quad \text{dla } u \in [0,1]. \tag{19}$$

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

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

Nierówności (18) oraz (19) z $u = \tilde{\mu}(A(x_n))/\tilde{\mu}(B)$ implikują

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

Zatem

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

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

9.3 Wybrane zastosowania

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

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

Niech dla $x \in \mathbb{X}$

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

wówczas

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

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

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

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

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

Stąd dla $s > 0$,

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

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

Twierdzenie 9.2 prowadzi do koncentracji pewnej klasy funkcji lipschitzowskich w odpowiednim sensie . Mianowicie zachodzi

Wniosek 9.3. *Załóżmy, że funkcja $F \colon \mathbb{X} \to \mathbb{R}$ spełnia warunek*

$$\forall_{x \in \mathbb{X}} \exists_{a=a(x)} \forall_{y \in \mathbb{X}} \; F(x) \leqslant F(y) + d_a(x,y) = F(y) + \sum_{i=1}^n a_i(x) \mathbb{1}_{\{x_i \neq y_i\}}. \tag{20}$$

Wówczas dla dowolnej probabilistycznej miary produktowej μ na $\mathbb{X}$,

$$\mu(\{|F - \mathrm{Med}_\mu(F)| \geqslant t\}) \leqslant 4\exp\left(-\frac{t^2}{4\sigma^2}\right) \qquad dla\ t > 0,$$

gdzie

$$\sigma^2 := \sup_{x \in \mathbb{X}} \sum_{i=1}^n a_i(x)^2.$$

Dowód. Dla $m \in \mathbb{R}$ połóżmy $A = \{F \leqslant m\}$, zauważmy, że warunek (20) implikuje, że dla dowolnego $x \in \mathbb{X}$,

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

stąd

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

Zatem dla dowolnego m,

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

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

10 Porównywanie supremów procesów stochastycznych

W kolejnych wykładach zajmiemy się badaniem supremów procesów stochastycznych, czyli zmiennych losowych postaci $\sup_{t\in T} X_t$. Zbiór T nie musi być podzbiorem prostej rzeczywistej, by uniknąć problemów z mierzalnością będziemy zakładać, że zbiór jest przeliczalny, alternatywnie można zakładać ośrodkowość procesu $(X_t)_{t\in T}$.

Przykłady.
i) Norma wektora losowego w ośrodkowej przestrzeni Banacha $\|X\| = \sup_{\varphi} \varphi(X)$, gdzie supremum jest brane po przeliczalnym podzbiorze kuli jednostkowej wybijającym normę wektora.
ii) Norma operatorowa macierzy losowej $\|X\| = \sup_{t,s} \sum_{ij} X_{ij} t_i s_j$, gdzie supremum bierzemy po przeliczalnym gęstym podzbiorze B_2^n.
iii) Supremum procesu empirycznego $\sup_{f\in\mathcal{F}} \sum_{i=1}^n f(X_i)$ - tutaj $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o wartościach w pewnej przestrzeni $\mathbb{X}$, a $\mathcal{F}$ przeliczalną klasą funkcji mierzalnych na $\mathbb{X}$.

10.1 Nierówności symetryzacyjne

Od tej pory $\varepsilon, \varepsilon_1, \varepsilon_2, \ldots$ oznaczają niezależne zmienne losowe takie, że $\mathbf{P}(\varepsilon_i = \pm 1) = 1/2$ (ciąg Bernoulliego), a $g, g_1, g_2, \ldots$ ciąg niezależnych zmiennych losowych o rozkładzie $\mathcal{N}(0,1)$. Będziemy też zakładać, że ciągi (ε_k) i (g_k) są od siebie niezależne i niezależne od pozostałych zmiennych losowych.

Fakt 10.1. *Załóżmy, że $X_1, X_2, \ldots, X_n$ są niezależnymi zmiennymi losowymi o wartościach w przestrzeni $\mathbb{X}$, $\mathcal{F}$ jest przeliczalną klasą funkcji mierzalnych na $\mathbb{X}$ oraz $\mathbf{E} f(X_k) = 0$ dla wszystkich k i $f \in \mathcal{F}$. Wówczas dla dowolnej niemalejącej funkcji wypukłej G na $\mathbb{R}$,*

$$\mathbf{E}G\left(\sup_{f\in\mathcal{F}} \sum_{k=1}^n f(X_k)\right) \leqslant \mathbf{E}G\left(2\sup_{f\in\mathcal{F}} \sum_{k=1}^n \varepsilon_k f(X_k)\right) \leqslant \mathbf{E}G\left(\sqrt{2\pi}\sup_{f\in\mathcal{F}} \sum_{k=1}^n g_k f(X_k)\right). \quad (21)$$

oraz

$$\mathbf{E}G\left(\frac{1}{2}\sup_{f\in\mathcal{F}} \left|\sum_{k=1}^n \varepsilon_k f(X_k)\right|\right) \leqslant \mathbf{E}G\left(\sup_{f\in\mathcal{F}} \left|\sum_{k=1}^n f(X_k)\right|\right) \leqslant \mathbf{E}G\left(2\sup_{f\in\mathcal{F}} \left|\sum_{k=1}^n \varepsilon_k f(X_k)\right|\right)$$
$$\leqslant \mathbf{E}G\left(\sqrt{2\pi}\sup_{f\in\mathcal{F}} \left|\sum_{k=1}^n g_k f(X_k)\right|\right). \quad (22)$$

Dowód. Niech $(Y_1, \ldots, Y_n)$ będzie niezależną kopią ciągu $(X_1, \ldots, X_n)$, niezależną od zmiennych ε_k. Wówczas na mocy nierówności Jensena,

$$\begin{aligned}
\mathbf{E}G\left(\sup_{f\in\mathcal{F}}\sum_{k=1}^n f(X_k)\right) &= \mathbf{E}_X G\left(\sup_{f\in\mathcal{F}}\sum_{k=1}^n (f(X_k)-\mathbf{E}_Y f(Y_k))\right)\\
&\leqslant \mathbf{E}_X G\left(\mathbf{E}_Y \sup_{f\in\mathcal{F}}\sum_{k=1}^n (f(X_k)-f(Y_k))\right)\\
&\leqslant \mathbf{E}G\left(\sup_{f\in\mathcal{F}}\sum_{k=1}^n (f(X_k)-f(Y_k))\right)\\
&= \mathbf{E}G\left(\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k(f(X_k)-f(Y_k))\right)\\
&\leqslant \frac{1}{2}\mathbf{E}G\left(2\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k f(X_k)\right)+\frac{1}{2}\mathbf{E}G\left(2\sup_{f\in\mathcal{F}}\sum_{k=1}^n (-\varepsilon_k) f(Y_k)\right)\\
&= \mathbf{E}G\left(2\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k f(X_k)\right).
\end{aligned}$$

Wykorzystaliśmy powyżej też fakt, że zmienne (X_k, Y_k) są niezależne i mają ten sam rozkład co (Y_k, X_k), zatem dla dowolnego ciągu znaków $\eta_k = \pm 1$, proces $(\sum_{k=1}^n \eta_k(f(X_k) - f(Y_k)))_{f\in\mathcal{F}}$ ma ten sam rozkład co proces $(\sum_{k=1}^n (f(X_k) - f(Y_k)))_{f\in\mathcal{F}}$.

By udowodnić drugą nierówność w (21) zauważamy, że (g_k) ma ten sam rozkład co $(\varepsilon_k|g_k|)$ i $\sqrt{2\pi}\mathbf{E}|g_k| = 2$. Zatem z nierówności Jensena

$$\begin{aligned}
\mathbf{E}G\left(2\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k f(X_k)\right) &= \mathbf{E}G\left(\sqrt{2\pi}\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k \mathbf{E}_g|g_k| f(X_k)\right)\\
&\leqslant \mathbf{E}G\left(\sqrt{2\pi}\sup_{f\in\mathcal{F}}\sum_{k=1}^n \varepsilon_k|g_k| f(X_k)\right)\\
&= \mathbf{E}G\left(\sqrt{2\pi}\sup_{f\in\mathcal{F}}\sum_{k=1}^n g_k f(X_k)\right)
\end{aligned}$$

Druga i trzecia nierówność w (21) wynika z (22) zastosowanego do $\mathcal{F}\cup -\mathcal{F}$. W dowodzie pierwszej nierówności ponownie wykorzystujemy nierówność Jensena:

$$
\begin{aligned}
\mathbf{E}G\left(\frac{1}{2}\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}\varepsilon_k f(X_k)\right|\right) &= \mathbf{E}_{X,\varepsilon}G\left(\frac{1}{2}\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}\varepsilon_k(f(X_k)-\mathbf{E}_Y f(Y_k))\right|\right)\\
&\leqslant \mathbf{E}_{X,\varepsilon}G\left(\frac{1}{2}\mathbf{E}_Y\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}\varepsilon_k(f(X_k)-f(Y_k))\right|\right)\\
&\leqslant \mathbf{E}G\left(\frac{1}{2}\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}\varepsilon_k(f(X_k)-f(Y_k))\right|\right)\\
&= \mathbf{E}G\left(\frac{1}{2}\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}(f(X_k)-f(Y_k))\right|\right)\\
&\leqslant \frac{1}{2}\mathbf{E}G\left(\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}f(X_k)\right|\right)+\frac{1}{2}\mathbf{E}G\left(\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}-f(X_k)\right|\right)\\
&= \mathbf{E}G\left(\sup_{f\in\mathcal{F}}\left|\sum_{k=1}^{n}f(X_k)\right|\right).
\end{aligned}
$$

□

W szczególnym przypadku, gdy $\mathbb{X}=\mathbb{F}$, a $\mathcal{F}$ to klasa funkcjonałów liniowych na $\mathbb{F}$ otrzymujemy.

Wniosek 10.2. *Załóżmy, że $X_1,\ldots,X_n$ są niezależnymi scentrowanymi wektorami losowymi o wartościach w ośrodkowej przestrzeni Banacha $\mathbb{F}$. Wówczas dla dowolnej niemalejącej funkcji wypukłej G na $\mathbb{R}_+$,*

$$
\mathbf{E}G\left(\frac{1}{2}\left\|\sum_{k=1}^{n}\varepsilon_k X_k\right\|\right)\leqslant \mathbf{E}G\left(\left\|\sum_{k=1}^{n}X_k\right\|\right)\leqslant \mathbf{E}G\left(2\left\|\sum_{k=1}^{n}\varepsilon_k X_k\right\|\right)\leqslant \mathbf{E}G\left(\sqrt{2\pi}\left\|\sum_{k=1}^{n}g_k X_k\right\|\right).
$$

10.2 Zasada kontrakcji dla procesów Bernoulliego

Zacznijmy od łatwego faktu zwanego zasadą kontrakcji.

Fakt 10.3. *Załóżmy, że $|\lambda_k|\leqslant 1$ dla $1\leqslant k\leqslant n$, zaś T jest ograniczonym podzbiorem $\mathbb{R}^n$. Wówczas dowolnej wypukłej niemalejącej funckji G na $\mathbb{R}$,*

$$
\mathbf{E}G\left(\sup_{t\in T}\sum_{k=1}^{n}\lambda_k t_k\varepsilon_k\right)\leqslant \mathbf{E}G\left(\sup_{t\in T}\sum_{k=1}^{n}t_k\varepsilon_k\right).
$$

Dowód. Funkcja $(\lambda_1,\ldots,\lambda_n)\mapsto \mathbf{E}G(\sup_{t\in T}\sum_{k=1}^{n}\lambda_k t_k\varepsilon_k)$ jest wypukła na $[-1,1]^n$, więc przyjmuje swoje maksimum w którymś z wierzchołków. □

Kolejna nierówność między procesami Bernoulliego, uogólniająca znacząco poprzedni fakt, została sformułowana i udowodniona przez Talagranda.

Twierdzenie 10.4. *Załóżmy, że dla* $k = 1, \ldots, n$, $\varphi_k \colon \mathbb{R} \to \mathbb{R}$ *są* 1*-lipschitzowskie oraz* $\varphi_k(0) = 0$, *zaś* $G \colon \mathbb{R} \to \mathbb{R}$ *jest funkcją wypukłą i niemalejącą. Wówczas dla dowolnego zbioru ograniczonego* $T \subset \mathbb{R}^n$,

$$\mathbf{E} G\left(\sup_{t\in T}\sum_{k=1}^n \varphi_k(t_k)\varepsilon_k\right) \leqslant \mathbf{E} G\left(\sup_{t\in T}\sum_{k=1}^n t_k\varepsilon_k\right).$$

Dowód. Łatwy argument indukcyjny pokazuje, że wystarczy wykazać, że dla ograniczonego podzbioru $T \subset \mathbb{R}^2$ i funkcji 1-lipschitzowskiej φ na $\mathbb{R}$ takiej, że $\varphi(0) = 0$ zachodzi

$$\mathbf{E} G\left(\sup_{t\in T}(t_1 + \varphi(t_2)\varepsilon)\right) \leqslant \mathbf{E} G\left(\sup_{t\in T}(t_1 + t_2\varepsilon)\right).$$

Wystarczy zatem pokazać, że dla dowolnego $s, t \in T$ prawa strona powyższej nierówności jest większa równa

$$I := \frac{1}{2}(G(t_1 + \varphi(t_2)) + G(s_1 - \varphi(s_2))).$$

Bez straty ogólności możemy też zakładać, że

$$t_1 + \varphi(t_2) \geqslant s_1 + \varphi(s_2) \quad \text{oraz} \quad s_1 - \varphi(s_2) \geqslant t_1 - \varphi(t_2). \tag{23}$$

Rozpatrzymy 4 przypadki.

Przypadek 1. $t_2 \geqslant 0$ i $s_2 \geqslant 0$. Załóżmy wpierw dodatkowo, że $s_2 \leqslant t_2$. Wykażemy, że

$$2I \leqslant G(t_1 + t_2) + G(s_1 - s_2), \quad \text{czyli } G(a) - G(b) \leqslant G(c) - G(d)$$

dla $a := s_1 - \varphi(s_2)$, $b := s_1 - s_2$, $c := t_1 + t_2$, $d := t_1 + \varphi(t_2)$. Z 1-lipschitzowskości φ mamy $|\varphi(s_2)| \leqslant s_2$, skąd wynika, że $a \geqslant b$ oraz, biorąc pod uwagę pierwszą nierówność w (23), $d \geqslant b$. Mamy też (wobec tego, że φ jest 1-lipschitzowska oraz $s_2 \leqslant t_2$)

$$a - b = s_2 - \varphi(s_2) \leqslant t_2 - \varphi(t_2) = c - d.$$

Funkcja $x \mapsto G(x + y) - G(y)$ jest rosnąca dla $y \geqslant 0$, zatem

$$G(a) - G(b) \leqslant G(d + (a - b)) - G(d) \leqslant G(c) - G(d).$$

Jeśli $s_2 \geqslant t_2$ to pokażemy, że

$$2I \leqslant G(t_1 - t_2) + G(s_1 + s_2), \quad \text{czyli } G(a) - G(b) \leqslant G(c) - G(d)$$

dla $a := t_1 + \varphi(t_2)$, $b := t_1 - t_2$, $c := s_1 + s_2$, $d := s_1 - \varphi(s_2)$. Mamy $|\varphi(t_2)| \leqslant t_2$, skąd wynika, że $a \geqslant b$ oraz, biorąc pod uwagę drugą nierówność w (23), $d \geqslant b$. Mamy też (wobec tego, że φ jest 1-lipschitzowska oraz $s_2 \geqslant t_2$)

$$a - b = t_2 + \varphi(t_2) \leqslant s_2 + \varphi(s_2) = c - d.$$

i dalej argumentujemy jak poprzednio.

Przypadek 2. $t_2 \leqslant 0$ i $s_2 \leqslant 0$. Rozumujemy analogicznie jak w przypadku 1.

Przypadek 3. $t_2 \geqslant 0$ i $s_2 \leqslant 0$. Wówczas $\varphi(t_2) \leqslant t_2$ i $-\varphi(s_2) \leqslant -s_2$, stąd

$$2I \leqslant G(t_1 + t_2) + G(s_1 - s_2).$$

Przypadek 4. $t_2 \leqslant 0$ i $s_2 \geqslant 0$. Wówczas $\varphi(t_2) \leqslant -t_2$ i $-\varphi(s_2) \leqslant s_2$, stąd

$$2I \leqslant G(t_1 - t_2) + G(s_1 + s_2).$$

□

10.3 Lemat Slepiana

Celem tej części jest udowodnienie następującego twierdzenia, będącego jednym z wariantów tzw. lematu Slepiana.

Twierdzenie 10.5 (Slepian-Fernique)**.** *Załóżmy, że X i Y są n-wymiarowymi wektorami gaussowskimi o średniej zero oraz*

$$\mathbf{E}|X_i - X_j|^2 \geqslant \mathbf{E}|Y_i - Y_j|^2 \quad \textit{dla } 1 \leqslant i, j \leqslant n.$$

Wówczas

$$\mathbf{E} \max_{i \leqslant n} X_i \geqslant \mathbf{E} \max_{i \leqslant n} Y_i.$$

Idea dowodu polega na rozważeniu procesu

$$Z(t) = \sqrt{t} X + \sqrt{1-t} Y, \quad t \in [0,1]. \tag{24}$$

interpolującego między X i Y. By obliczyć $\frac{d}{dt}\mathbf{E}f(Z(t))$ dla gładkich funkcji f będziemy potrzebować dwóch lematów dotyczących gaussowskiego całkowania przez części.

Lemat 10.6 (Jednowymiarowe gaussowskie całkowanie przez części)**.** *Załóżmy, że $f \in C^1(\mathbb{R}^n)$ oraz $|f(x)| + |f'(x)| \leqslant Ce^{t|x|^2}$ dla pewnego $t < 1/2$. Wówczas*

$$\mathbf{E} g f(g) = \mathbf{E} f'(g) \quad \textit{dla } g \sim \mathcal{N}(0,1).$$

Dowód. Całkując przez części dostajemy:

$$\mathbf{E}gf(g) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -f(x)\frac{d}{dx}e^{-x^2/2}dx = -f(x)e^{-x^2/2}|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} f'(x)e^{-x^2/2}dx = \mathbf{E}f'(g).$$

□

Lemat 10.7 (Wielowymiarowe gaussowskie całkowanie przez części)**.** *Załóżmy, że $f \in C^1(\mathbb{R}^n)$ oraz dla $\varepsilon > 0$ istnieje $C_\varepsilon < \infty$ taki, że $|f(x)| + |\nabla f(x)| \leqslant C_\varepsilon e^{\varepsilon|x|^2}$. Wówczas dla dowolnego n-wymiarowego wektora gaussowskiego X o średniej 0,*

$$\mathbf{E}(X_i f(X)) = \sum_{j=1}^{n} \mathrm{Cov}(X_i, X_j)\mathbf{E}\frac{\partial f}{\partial x_j}(X) \qquad \textit{dla } 1 \leqslant i \leqslant n.$$

Dowód. Wiemy, że X ma ten sam rozkład co AY dla $Y \sim N(0, I_n)$ i pewnego $A \in M_{n,n}$. Stąd

$$\mathbf{E}(X_i f(X)) = \sum_{k=1}^{n} a_{ik}\mathbf{E}(Y_k f(AY)) = \sum_{k=1}^{n} a_{ik}\mathbf{E}(Y_k g(Y)),$$

gdzie $g(x) = f(Ax)$. Stosując warunkowo Lemat 10.7 dostajemy

$$\mathbf{E}(Y_k g(Y)) = \mathbf{E}\frac{\partial g}{\partial x_k}(Y) = \sum_{j=1}^{n} a_{jk}\mathbf{E}\frac{\partial f}{\partial x_j}(AY).$$

By dokończyć dowód wystarczy zauważyć, że

$$\mathbf{E}(X_i f(X)) = \sum_{j,k=1}^{n} a_{ik}a_{jk}\mathbf{E}\frac{\partial f}{\partial x_j}(X) = \sum_{j}(AA^T)_{ij}\mathbf{E}\frac{\partial f}{\partial x_j}(X).$$

□

Wniosek 10.8. *Załóżmy, że X i Y są niezależnymi n-wymiarowymi wektorami gaussowskimi o średniej zero oraz proces $Z(t)$ jest zadany przez (24). Wówczas dla $f \in C^2(\mathbb{R}^n)$ takiej, że dla $\varepsilon > 0$ istnieje $C_\varepsilon < \infty$, $f(x) + |\nabla f(x)| + |\mathrm{Hess} f(x)| \leqslant C_\varepsilon e^{\varepsilon|x|^2}$ zachodzi*

$$\frac{d}{dt}\mathbf{E}f(Z(t)) = \frac{1}{2}\sum_{i,j=1}^{n}(\mathrm{Cov}(X_i, X_j) - \mathrm{Cov}(Y_i, Y_j))\mathbf{E}\frac{\partial^2 f}{\partial x_i \partial x_j}(Z(t)) \qquad \textit{dla } t \in (0, 1).$$

Dowód. Mamy

$$\frac{d}{dt}\mathbf{E}f(Z(t)) = \sum_{i=1}^{n}\mathbf{E}\left(\frac{\partial f}{\partial x_i}(Z(t))\frac{dZ_i(t)}{dt}\right) = \frac{1}{2}\sum_{i=1}^{n}\mathbf{E}\left(\frac{\partial f}{\partial x_i}(Z(t))\left(\frac{X_i}{\sqrt{t}} - \frac{Y_i}{\sqrt{1-t}}\right)\right).$$

Stosując Lemat 10.7 do $2n$-wymiarowego wektora (X,Y) i funkcji $\frac{\partial f}{\partial x_i}(\sqrt{t}x+\sqrt{1-t}y)$ dostajemy

$$\mathbf{E}\left(\frac{\partial f}{\partial x_i}(Z(t))\frac{X_i}{\sqrt{t}}\right)=\sum_{j=1}^{n}\mathrm{Cov}(X_i,X_j)\mathbf{E}\frac{\partial^2 f}{\partial x_i\partial x_j}(Z(t))$$

oraz

$$\mathbf{E}\left(\frac{\partial f}{\partial x_i}(Z(t))\frac{Y_i}{\sqrt{1-t}}\right)=\sum_{j=1}^{n}\mathrm{Cov}(Y_i,Y_j)\mathbf{E}\frac{\partial^2 f}{\partial x_i\partial x_j}(Z(t)).$$

□

Dowód Twierdzenia 10.5. Funkcja $f(x)=\max_i x_i$ nie jest gładka, będziemy zatem ją odpowiednio aproksymować, by móc stosować wyprowadzone powyżej wzory. Określmy dla $\beta>0$

$$f_\beta(x):=\frac{1}{\beta}\log\sum_{i=1}e^{\beta x_i}.$$

Wówczas

$$\max_i x_i\leqslant f_\beta(x)\leqslant\max_i x_i+\frac{\log n}{\beta},$$

zatem wystarczy wykazać, że $\mathbf{E}f_\beta(X)\geqslant\mathbf{E}f_\beta(Y)$ dla dowolnego $\beta>0$.

Bez straty ogólności możemy zakładać, że wektory X i Y są niezależne. Zdefiniujmy $Z(t)$ wzorem (24), zauważmy, że $Z(1)=X$, $Z(0)=Y$, wystarczy zatem iż pokażemy $\frac{d}{dt}\mathbf{E}f_\beta(Z(t))\geqslant 0$ dla $t\in(0,1)$.

Prosty rachunek pokazuje, że dla $1\leqslant i,j\leqslant n$,

$$\frac{\partial f_\beta(x)}{\partial x_i}=\frac{e^{\beta x_i}}{\sum_k e^{\beta x_k}}=:p_i(x),\qquad\frac{\partial^2 f_\beta(x)}{\partial x_i\partial x_j}=\beta(\delta_{ij}p_i(x)-p_i(x)p_j(x)).$$

Stosując Wniosek 10.8 dostajemy

$$\begin{aligned}\frac{d}{dt}\mathbf{E}f_\beta(Z(t))=&\frac{\beta}{2}\sum_{i=1}^{n}(\mathrm{Var}(X_i)-\mathrm{Var}(Y_i))\mathbf{E}(p_i(Z(t))(1-p_i(Z(t)))\\&-\frac{\beta}{2}\sum_{i\neq j}(\mathrm{Cov}(X_i,X_j)-\mathrm{Cov}(Y_i,Y_j))\mathbf{E}(p_i(Z(t))p_j(Z(t))).\end{aligned}$$

Zauważmy jednak, że $1-p_i(x)=\sum_{j\neq i}p_j(x)$, stąd dla dowolnych liczb a_i

$$\sum_{i=1}^{n}a_ip_i(x)(1-p_i(x))=\sum_{i\neq j}a_ip_i(x)p_j(x)=\sum_{i\neq j}a_jp_i(x)p_j(x).$$

Wykorzystując powyższą tożsamość otrzymujemy

$$\frac{d}{dt}\mathbf{E}f_\beta(Z(t))=\frac{\beta}{4}\sum_{i\neq j}(\mathbf{E}|X_i-X_j|^2-\mathbf{E}|Y_i-Y_j|^2)\mathbf{E}(p_i(Z(t))p_j(Z(t)))\geqslant 0.$$

□

11 Metoda łańcuchowa I - szacowania supremów procesów przy pomocy entropii metrycznej

11.1 Entropia metryczna

Zacznijmy od ważnej definicji liczb pokryciowych.

Definicja 11.1. Niech (T,d) będzie przestrzenią metryczną. Dla $\varepsilon > 0$ przez $N(T,d,\varepsilon)$ oznaczamy najmniejszą liczbę kul otwartych o promieniu ε, które pokrywają T, tzn.

$$N(T,d,\varepsilon) := \inf\Big\{N\colon\ T \subset \bigcup_{i=1}^{N} B(x_i,\varepsilon) \text{ dla pewnych } x_1,\ldots,x_N \in T\Big\}.$$

Uwaga 11.2. Możemy zdefiniować

$$S(T,d,\varepsilon) := \sup\Big\{N\colon\ \ \text{istnieją } x_1,\ldots,x_N \in T,\ d(x_i,x_j) \geqslant \varepsilon \text{ dla } i \neq j\Big\},$$

wtedy $N(T,d,\varepsilon) \leqslant S(T,d,\varepsilon) \leqslant N(T,d,\varepsilon/2)$.

Uwaga 11.3. Często rozważa się *liczby entropijne* zdefiniowane jako

$$e_n(T,d) := \inf\{\varepsilon > 0\colon\ N(T,d,\varepsilon) \leqslant 2^n\}.$$

11.2 Górne oszacowania entropijne

Załóżmy, że φ jest funkcją Younga na $[0,\infty)$, tzn. φ jest wypukłe, ściśle rosnące oraz $\varphi(0) = 0$. Przyjmijmy też, że na T jest określona metryka d taka, że

$$\mathbf{E}\varphi\left(\frac{|X_t - X_s|}{d(t,s)}\right) \leqslant 1 \quad \text{dla } t,s \in T,\ t \neq s. \tag{25}$$

Przez $\Delta(T) = \Delta(T,d)$ będziemy oznaczali średnicę przestrzeni metrycznej (T,d).

Kolejne twierdzenie pokazuje jak szacować suprema procesów przy pomocy entropii metrycznej. Udowodnili je niezależnie, uogólniając wcześniejszy wynik Dudleya z 1967 roku (Wniosek 11.7) Kôno i Pisier w 1980 roku.

Twierdzenie 11.4 (Kôno-Pisier)**.** *Jeśli proces $(X_t)_{t\in T}$ spełnia warunek* (25)*, to dla dowolnego $t_0 \in T$,*

$$\mathbf{E}\sup_{s,t\in T}(X_s - X_t) \leqslant 2\mathbf{E}\sup_{t\in T}|X_t - X_{t_0}| \leqslant 8\int_0^{\Delta(T)} \varphi^{-1}(N(t,d,\varepsilon))d\varepsilon.$$

Dowód. Oczywiście możemy założyć, że prawa strona postulowanego oszacowania jest skończona. Niech $\varepsilon_k = 2^{-k}\Delta(T)$ dla $k = 0, 1, \dots$ i niech $T_0 = \{t_0\}$ a dla $k \geqslant 1$, T_k będzie podzbiorem T mocy $N(t, d, \varepsilon_k)$ takim, że $T \subset \bigcup_{t\in T_k} B(t, \varepsilon_k)$. Możemy zatem znaleźć funkcję $u_k\colon T \to T_k$ taką, że $d(t, u_k(t)) \leqslant \varepsilon_k$.

Ustalmy zbiór skończony $S \subset T$, niech $\delta := \inf_{t,s\in S, t\neq s} d(t, s)$ i wybierzmy k_0 takie, że $\varepsilon_{k_0} \leqslant \delta/2$. Wtedy każda kulka $B(t, \varepsilon_{k_0})$ zawiera conajwyżej jeden punkt z S więc $|S| \leqslant N(T, d, \varepsilon_{k_0})$. Zdefinujmy przekształcenia π_k na S wzorami

$$\pi_{k_0}(t) = t \quad \text{oraz} \quad \pi_k(t) := u_k \circ u_{k+1} \circ \ldots \circ u_{k_0-1} \ \text{ dla } 0 \leqslant k \leqslant k_0 - 1.$$

Zauważmy, że $\pi_{k_0}(S) = S$ i $\pi_k(S) \subset T_k$, zatem $|\pi_k(S)| \leqslant N(T, d, \varepsilon_k)$ dla $0 \leqslant k \leqslant k_0$.

Mamy $\pi_0(t) = t_0$, zatem

$$\begin{aligned}\sup_{t\in S}|X_t - X_{t_0}| &= \sup_{t\in S}\left|\sum_{k=1}^{k_0}(X_{\pi_k(t)} - X_{\pi_{k-1}(t)})\right| \leqslant \sum_{k=1}^{k_0}\sup_{t\in S}|X_{\pi_k(t)} - X_{\pi_{k-1}(t)}| \\ &= \sum_{k=1}^{k_0}\sup_{t\in\pi_k(S)}|X_t - X_{u_{k-1}(t)}| \leqslant \sum_{k=1}^{k_0}\varepsilon_{k-1}\sup_{t\in\pi_k(S)}\frac{|X_t - X_{u_{k-1}(t)}|}{d(t, u_{k-1}(t))}.\end{aligned}$$

Mamy

$$\begin{aligned}\mathbf{E}\sup_{t\in\pi_k(S)}\frac{|X_t - X_{u_{k-1}(t)}|}{d(t, u_{k-1}(t))} &\leqslant \mathbf{E}\varphi^{-1}\left(\sum_{k\in\pi_k(S)}\varphi\left(\frac{|X_t - X_{u_{k-1}(t)}|}{d(t, u_{k-1}(t))}\right)\right) \\ &\leqslant \varphi^{-1}\left(\mathbf{E}\sum_{k\in\pi_k(S)}\varphi\left(\frac{|X_t - X_{u_{k-1}(t)}|}{d(t, u_{k-1}(t))}\right)\right) \leqslant \varphi^{-1}(|\pi_k(S)|) \\ &\leqslant \varphi^{-1}(N(T, d, \varepsilon_k)).\end{aligned}$$

gdzie druga nierówność wynika z wklęsłości funkcji φ^{-1}, a trzecia z założenia (25). Otrzymujemy zatem

$$\mathbf{E}\sup_{t\in S}|X_t - X_{t_0}| \leqslant \sum_{k=1}^{k_0}\varepsilon_{k-1}\varphi^{-1}(N(T, d, \varepsilon_k)) = 2\sum_{k=1}^{k_0}\varepsilon_k\varphi^{-1}(N(T, d, \varepsilon_k)).$$

By zakończyć dowód wystarczy zauważyć, że

$$\int_0^{\Delta(T)}\varphi^{-1}(N(t, d, \varepsilon))d\varepsilon \geqslant \sum_{k=1}^{k_0+1}\int_{\varepsilon_k}^{\varepsilon_{k-1}}\varphi^{-1}(N(T, d, \varepsilon_{k-1})) = \frac{1}{2}\sum_{k=1}^{k_0+1}\varepsilon_{k-1}\varphi^{-1}(N(T, d, \varepsilon_{k-1}))$$

oraz

$$\mathbf{E}\sup_{t\in T}|X_t - X_{t_0}| = \sup\left\{\mathbf{E}\sup_{t\in T}|X_t - X_{t_0}|\colon\ S \subset T,\ |S| < \infty\right\}.$$

□

Uwaga 11.5. Oszacowanie z Twierdzenia 11.3 można rozszerzyć na procesy nieośrodkowe pod warunkiem, że zdefiniujemy

$$\mathbf{E}\sup_{s,t\in T}(X_s - X_t) := \sup\left\{\mathbf{E}\sup_{s,t\in S}(X_t - X_s)\colon\ S\subset T,\ |S|<\infty\right\}.$$

Uwaga 11.6. Jeśli proces X_t jest symetryczny (tzn. ma ten sam rozkład co $(-X_t)_{t\in T}$), to

$$\mathbf{E}\sup_{t,s\in T}(X_t - X_s) = \mathbf{E}\sup_{t\in T} X_t + \mathbf{E}\sup_{s\in T}(-X_s) = 2\mathbf{E}\sup_{t\in T} X_t.$$

Kolejny wynik to wniosek z Twierdzenia 11.4, który był udowodniony pierwotnie (w nieco innym sformułowaniu) przez Dudleya.

Wniosek 11.7 (Dudley)**.** *Załóżmy, że $(X_t)_{t\in T}$ jest scentrowanym procesem gaussowskim oraz $d(t,s) = (\mathbf{E}|X_t - X_s|^2)^{1/2}$. Wówczas*

$$\mathbf{E}\sup_{t\in T} X_t = \frac{1}{2}\mathbf{E}\sup_{s,t\in T}(X_s - X_t) \leqslant C\int_0^{\Delta(T)}\sqrt{\ln N(T,d,\varepsilon)}d\varepsilon = C\int_0^{\infty}\sqrt{\ln N(T,d,\varepsilon)}d\varepsilon,$$

gdzie C jest stałą numeryczną (można przyjąć $C = 4\sqrt{8/3}(1+\sqrt{\ln 3/\ln 2}) \leqslant 15$).

Dowód. Ostatnia równość wynika stąd, że $N(T,d,\varepsilon) = 1$ dla $\varepsilon > \Delta(T)$. Niech $\tilde{d}(s,t) = \sqrt{8/3}d(s,t)$ oraz $\psi_2(x) = \exp(x^2) - 1$. Wykorzystując fakt, że $\mathbf{E}\exp(\lambda g^2) = (1-2\lambda)^{-1/2}$ dla $g\sim N(0,1)$ i $\lambda < 1/2$ nietrudno udowodnić, że dla $s\neq t$,

$$\mathbf{E}\psi_2\left(\frac{|X_t - X_s|}{\tilde{d}(t,s)}\right) = 1.$$

Stąd z Twierdzenia 11.4 otrzymujemy

$$\begin{aligned}\mathbf{E}\sup_{s,t\in T}(X_s - X_t) &\leqslant 8\int_0^{\sqrt{8/3}\Delta(T)}\psi_2^{-1}\left(N\left(T,\tilde{d},\varepsilon\right)\right)d\varepsilon\\ &= 8\sqrt{\frac{8}{3}}\int_0^{\Delta(T)}\sqrt{\log(N(T,d,\varepsilon)+1)}.\end{aligned}$$

Zauważmy, że $\log(n+1)/\log n \leqslant \log 3/\log 2$ dla $n\geqslant 2$, ponadto $N(T,d,\varepsilon)\geqslant 2$ dla $d < \Delta(T)/2$, zatem

$$\int_0^{\Delta(T)}\sqrt{\log(N(T,d,\varepsilon)+1)}d\varepsilon \leqslant \left(\sqrt{\frac{\log 3}{\log 2}}+1\right)\int_0^{\infty}\sqrt{\log N(T,d,\varepsilon)}d\varepsilon.$$

□

11.3 Minoryzacja Sudakowa dla procesów gaussowskich

W tej części będziemy zakładać, że X_t jest scentrowanym procesem gaussowskim oraz

$$d(t,s) = (\mathbf{E}|X_t - X_s|^2)^{1/2} \text{ dla } t,s \in T.$$

Oszacowania Dudleya nie można w ogólnej sytuacji odwrócić. Prawdziwe jest słabsze oszacowanie udowodnione przez Sudakowa.

Twierdzenie 11.8 (Minoryzacja Sudakowa)**.** *Załóżmy, że $(X_t)_{t\in T}$ jest scentrowanym procesem gaussowskim. Wówczas*

$$\frac{1}{4}\sup_{\varepsilon>0} \varepsilon\sqrt{\log N(T,d,\varepsilon)} \leqslant \mathbf{E}\sup_{t\in T} X_t.$$

Dowód. Załóżmy, że $N(T,d,\varepsilon) \geqslant N$, wówczas istnieje $S \subset T$ takie, że $|S| = N$ oraz $\|X_t - X_s\|_2 \geqslant \varepsilon$ dla $t,s \in S$, $t \neq s$. Położmy $Y_t = \varepsilon g_t/\sqrt{2}$, gdzie $(g_t)_{t\in S}$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$. Twierdzenie Slepiana-Fernique'a implikuje, że

$$\mathbf{E}\sup_{t\in T} X_t \geqslant \mathbf{E}\max_{t\in S} X_t \geqslant \mathbf{E}\max_{t\in S} Y_t = \frac{\varepsilon}{\sqrt{2}}\mathbf{E}\max_{t\in S} g_t \geqslant \frac{1}{4}\varepsilon\sqrt{\log n},$$

gdzie ostatnia nierówność wynika z poniższego Lematu 11.10. □

Lemat 11.9. *Jeśli $g \sim \mathcal{N}(0,1)$, to*

$$\frac{t}{\sqrt{2\pi}(t^2+1)}e^{-t^2/2} \leqslant \mathbf{P}(g \geqslant t) \leqslant \frac{1}{\sqrt{2\pi}t}e^{-t^2/2}.$$

Dowód. Górna nierówność wynika z szacowania

$$\int_t^\infty e^{-x^2/2}dx \leqslant \int_t^\infty \frac{x}{t}e^{-x^2/2}dx = \frac{1}{t}e^{-t^2/2}.$$

By udowodnić dolną definiujemy funkcję

$$f(t) := (t^2+1)\int_t^\infty e^{-x^2/2}dx - te^{-t^2/2}$$

i pokazujemy, że $\lim_{t\to\infty} f(t) = 0$ oraz na podstawie udowodnionego już górnego oszacowania

$$f'(t) = 2t\int_t^\infty e^{-x^2/2} - 2e^{-t^2/2} \leqslant 0.$$

□

Lemat 11.10. *Załóżmy, że $g_1, g_2, \ldots, g_n$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$. Wówczas*

$$\mathbf{E}\max_{k\leqslant n} g_k \geqslant \frac{1}{2\sqrt{2}}\sqrt{\log n}.$$

Dowód. Dla $n = 1$ nierówność jest oczywista. Dla $2 \leqslant n \leqslant 12$ mamy $\log n \leqslant 8/\pi$ i

$$\mathbf{E} \max_{k \leqslant n} g_k \geqslant \mathbf{E} \max\{g_1, g_2\} = \mathbf{E} g_1 + \mathbf{E}(g_2 - g_1)_+ = \frac{1}{2} \mathbf{E}|g_2 - g_1| = \frac{1}{\sqrt{\pi}} \geqslant \frac{1}{2\sqrt{2}} \sqrt{\log n}.$$

Dla $n \geqslant 13$ pokazujemy, że

$$\begin{aligned} \mathbf{P}\left(\max_{i \leqslant n} g_i \leqslant \sqrt{\log n}\right) \leqslant \mathbf{P}(g \leqslant \sqrt{\log n})^n &\leqslant \left(1 - \frac{\sqrt{\log n}}{\sqrt{2\pi}(1 + \log n)} e^{-(\log n)/2}\right)^n \\ &\leqslant \exp\left(-\frac{\sqrt{n \log n}}{\sqrt{2\pi}(1 + \log n)}\right) \leqslant e^{-1}, \end{aligned}$$

przy czym ostatnie szacowanie wynika stąd, że wobec monotoniczności $x/\log x$ na $[e, \infty)$,

$$\sqrt{n/\log n} \geqslant \sqrt{13/\log 13} \geqslant 5 \geqslant \sqrt{2\pi}(1 + \pi/8) \geqslant \sqrt{2\pi}(1 + 1/\log n).$$

Zatem

$$\begin{aligned} \mathbf{E} \max_{i \leqslant n} g_i &\geqslant \mathbf{E}(\sqrt{\log n} \mathbb{1}_{\{\max_i g_i \geqslant \sqrt{\log n}\}} + g_1 \mathbb{1}_{\{\max_i g_i \leqslant 0\}}) \\ &\geqslant \sqrt{\log n} \mathbf{P}(\max_i g_i \geqslant \sqrt{\log n}) - (\mathbf{E} g_1^2)^{1/2} \mathbf{P}(\max_i g_i \leqslant 0)^{1/2} \\ &\geqslant (1 - e^{-1}) \sqrt{\log n} - 2^{-n/2} \geqslant \frac{1}{2\sqrt{2}} \sqrt{\log n}. \end{aligned}$$

□

Poniższy lemat pokazuje, że oszacowanie z Lematu 11.10 jest optymalne z dokładnością do stałej.

Lemat 11.11. *Załóżmy, że zmienne $X_1, \ldots, X_n$ spełniają warunek subgaussowskości*

$$\mathbf{E} \exp(\lambda X_i) \leqslant \exp(a\lambda^2) \quad \text{dla } 1 \leqslant i \leqslant n,\ \lambda > 0.$$

Wówczas

$$\mathbf{E} \max_{1 \leqslant i \leqslant n} X_i \leqslant 2\sqrt{a \log n}.$$

Zauważmy, że w lemacie nie ma założenia niezależności oraz, że zmienne $\mathcal{N}(0, 1)$ spełniają założenia z $a = 1/2$.

Dowód. Dla $\lambda > 0$,

$$\mathbf{E} \max_{i \leqslant n} X_i \leqslant \frac{1}{\lambda} \log \mathbf{E} e^{\lambda \max_{i \leqslant n} X_i} \leqslant \frac{1}{\lambda} \log \sum_{i \leqslant n} \mathbf{E} e^{\lambda X_i} \leqslant \frac{\log n}{\lambda} + a\lambda.$$

Optymalizując powyższą nierówność po $\lambda > 0$ dostajemy tezę. □

Do szacowania supremów procesów gaussowskich z dołu będziemy potrzebowali wzmocnionego oszacowania Sudakowa.

Twierdzenie 11.12. *Załóżmy, że* $(X_t)_{t\in T}$ *jest scentrowanym procesem gaussowskim oraz* $t_1,\ldots,t_N \in T$, $\varepsilon > 0$ *spełniają warunek* $d(t_i,t_j) \geqslant \varepsilon$ *dla* $i \neq j$. *Wówczas*

$$\mathbf{E}\sup_{t\in T} X_t \geqslant \frac{1}{8}\varepsilon\sqrt{\log N} + \min_{i\leqslant N}\mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t,$$

gdzie $\alpha > 0$ *jest pewną stałą uniwersalną (można przyjąć* $\alpha = 1/(8\sqrt{2})$*).*

Dowód. Określmy zmienne losowe

$$Y_i := \sup_{d(t,t_i)\leqslant\alpha\varepsilon}(X_t - X_{t_i}) - \mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon}(X_t - X_{t_i}) = \sup_{d(t,t_i)\leqslant\alpha\varepsilon}(X_t - X_{t_i}) - \mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t.$$

Koncentracja procesów gaussowskich implikuje, że $\mathbf{E}\exp(\lambda Y_i) \leqslant \exp(\lambda^2\alpha^2\varepsilon^2)$ dla $\lambda \in \mathbb{R}$. Stąd Lemat 11.11 (zastosowany do zmiennych $X_i = -Y_i$) implikuje

$$\mathbf{E}\max_{i\leqslant N}(-Y_i) \leqslant \alpha\varepsilon\sqrt{2\log N}.$$

Ponieważ

$$\max_i(a_i + b_i + c_i) \geqslant \max_i a_i - \max_i(-b_i) + \min_i c_i,$$

więc

$$\begin{aligned}
\mathbf{E}\sup_{t\in T} X_t &\geqslant \mathbf{E}\max_{i\leqslant N}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t = \mathbf{E}\max_{i\leqslant N}\left(X_{t_i} + Y_i + \mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t\right) \\
&\geqslant \mathbf{E}\max_{i\leqslant N} X_{t_i} - \mathbf{E}\max_i(-Y_{t_i}) + \min_{i\leqslant N}\mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t \\
&\geqslant \frac{\varepsilon}{4}\sqrt{\log N} - \alpha\varepsilon\sqrt{2\log N} + \min_{i\leqslant N}\mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t \geqslant \frac{\varepsilon}{8}\sqrt{\log N} + \min_{i\leqslant N}\mathbf{E}\sup_{d(t,t_i)\leqslant\alpha\varepsilon} X_t,
\end{aligned}$$

o ile np. $\alpha = 1/(8\sqrt{2})$. □

11.4 Stacjonarne procesy gaussowskie

Definicja 11.13. Proces $(X_t)_{t\in T}$ nazywamy stacjonarnym, jeśli istnieje grupa G działająca na T taka, że
i) działanie G jest tranzytywne, tzn. dla $t,s \in T$ istnieje $g \in G$ takie, że $g(t) = s$
ii) dla dowolnego $g \in G$ proces $(X_t)_{t\in T}$ ma ten sam rozkład co $(X_{g(t)})_{t\in T}$.

Uwaga 11.14. W przypadku, gdy (X_t) jest scentrowanym procesem gaussowskim bby dowieść warunku ii) definicji wystarczy sprawdzić, że $\mathrm{Cov}(X_t,X_s) = \mathrm{Cov}(X_{g(t)},X_{g(s)})$ dla $s,t \in T$, $g \in G$.

Przykład. Niech c_k będzie ciągiem sumowalnym z kwadratem i określmy

$$X_t := \sum_k c_k(g_k \sin(kt) + g_k' \cos(kt))$$

gdzie $g_1, g_1', g_2, \ldots$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$, a $t \in \mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$. Tu $\mathbb{T}$ działa na siebie poprzez dodawanie i wystarczy zauważyć, że

$$\mathrm{Cov}(X_t, X_s) = \sum_k c_k^2 \cos(k(t-s)) = \mathrm{Cov}(X_{t+u}, X_{s+u}).$$

Twierdzenie 11.15 (Fernique)**.** *Załóżmy, że $(X_t)_{t\in T}$ jest stacjonarnym procesem gaussowskim. Wówczas*

$$c\int_0^\infty \sqrt{\log N(T,d,\varepsilon)}d\varepsilon \leqslant \mathbf{E}\sup_{t\in T} X_t \leqslant C\int_0^\infty \sqrt{\log N(T,d,\varepsilon)}d\varepsilon,$$

gdzie $0 < c < C < \infty$ są stałymi uniwersalnymi (można np. przyjąć $c = (192\sqrt{2})^{-1}$, $C = 15$)

Dowód. Twierdzenie da się wywnioskować ze znacznie ogólniejszego twierdzenia Talagranda o mierze majoryzującej, ale pokażemy bardziej bezpośredni dowód.

Dla uproszczenia notacji ustalmy $t_0 \in T$ i określmy $B(\varepsilon) = B(t_0, \varepsilon)$ - kula jednostkowa o środku w t_0 i promieniu ε. Zauważmy, że stacjonarność implikuje w szczególności, że

$$\mathbf{E}\sup_{s\in B(t,\varepsilon)} X_s = \mathbf{E}\sup_{s\in B(\varepsilon)} X_s \quad \text{dla } t \in T.$$

Niech $\alpha \leqslant 1/2$ będzie stałą z Twierdzenia 11.12. Zauważmy, że jeśli $t \in B(\alpha^{n+1})$ to $B(t,\alpha^{n+3}) \subset B(\alpha^n)$ stąd Twierdzenie 11.12 implikuje

$$\mathbf{E}\sup_{t\in B(\alpha^n)} X_t \geqslant \frac{1}{8}\alpha^{n+2}\sqrt{\log N(B(\alpha^{n+1}), d, \alpha^{n+2})} + \mathbf{E}\sup_{t\in B(\alpha^{n+3})} X_t.$$

Iterując poprzednią nierówność dostajemy dla dowolnego n

$$\mathbf{E}\sup_{t\in T} \geqslant \frac{1}{8}\sum_{j\geqslant 0}\alpha^{n+3j+2}\sqrt{\log N(B(\alpha^{n+3j+1}), d, \alpha^{n+3j+2})}$$

stąd stosując tę nierówność dla $n, n+1, n+2$ mamy i dobierając $n \in \mathbb{Z}$ tak by $\alpha^{n+2} \geqslant \Delta(T)$ dostajemy

$$\mathbf{E}\sup_{t\in T} \geqslant \frac{1}{24}\sum_{k\geqslant n+1}\alpha^{k+1}\sqrt{\log N(B(\alpha^k), d, \alpha^{k+1})} = \frac{1}{24}\sum_{k\in\mathbb{Z}}\alpha^{k+1}\sqrt{\log N(B(\alpha^k), d, \alpha^{k+1})}.$$

Zbiór T da się pokryć kulami $B(t_i, \alpha^k)$, $i = 1, \ldots, N(T, d, \alpha^k)$ a każdą z tych kul da się pokryć $N(B(t_i, \alpha^k), d, \alpha^{k+1}) = N(B(\alpha^k), d, \alpha^{k+1})$ kulami o promieniu α^{k+1}, więc

$$N(T, d, \alpha^{k+1}) \leqslant N(T, d, \alpha^k) N(B(\alpha^k), d, \alpha^{k+1}).$$

Stąd

$$\begin{aligned}
\mathbf{E} \sup_{t \in T} X_t &\geqslant \frac{1}{24} \sum_{k \in \mathbb{Z}} \alpha^{k+1} \sqrt{\log N(T, d, \alpha^{k+1})} - \frac{1}{24} \sum_{k \in \mathbb{Z}} \alpha^{k+1} \sqrt{\log N(T, d, \alpha^k)} \\
&= \frac{1}{24} (1 - \alpha) \sum_{k \in \mathbb{Z}} \alpha^{k+1} \sqrt{\log N(T, d, \alpha^{k+1})}.
\end{aligned}$$

Z drugiej strony

$$\begin{aligned}
\int_0^\infty \sqrt{\log N(T, d, \varepsilon)} &\leqslant \sum_{k \in \mathbb{Z}} \int_{\alpha^{k+1}}^{\alpha^k} \sqrt{\log N(T, d, \alpha^{k+1})} d\varepsilon \\
&= \frac{1 - \alpha}{\alpha} \sum_{k \in \mathbb{Z}} \alpha^{k+1} \sqrt{\log N(T, d, \alpha^{k+1})}.
\end{aligned}$$

□

12 Miary majoryzujące

Oszacowania entropijne omawiane w poprzednim rozdziale są bardzo użyteczne w zastosowaniach, ponieważ istnieje szereg narzędzi do szacowania z góry liczb pokryciowych (szczególnie w przypadku, gdy metryka jest euklidesowa). Jednak, chociaż w wielu przykładach, Twierdzenie 11.4 prowadzi do dobrych oszacowań, to w ogólności nie można go odwrócić. Trudność ta jest związana z tym, że entropia metryczna traktuje równomiernie całą przestrzeń metryczną, nie rozróżniając miejsc w których jest ona bardziej lub mniej zagęszczona. Fernique zaproponował nowy sposób szacowania, za pomocą tzw. miar majoryzujących (czyli odpowiednio dobranych miar probabilistycznych na T), a Talagrand wykazał, że w przypadku gaussowskim oszacowanie Fernique'a daje się odwrócić. Obecnie, częściej niż miar majoryzujących, używa się bardziej kombinatorycznego podejścia za pomocą ciągów podziałów przestrzeni, ale zaczniemy od klasycznego podejścia.

12.1 Oszacowania z góry

W przypadku procesów subgaussowskich (tzn. takich, które spełniają warunek (25) z funkcją Younga $\varphi = \psi_2$) oszacowanie górne z użyciem miar majoryzujących udowodnił Fernique. Oszacowanie to było potem uogólniane, między innymi przez Talagranda. Ostateczne sformułowanie, bez dodatkowych warunków wzrostu nakładanych na funkcje Younga, wykazał w 2006 roku Bednorz.

Twierdzenie 12.1 (Bednorz)**.** *Załóżmy, że φ jest funkcją Younga, (T,d) jest przestrzenią metryczną, a proces $(X_t)_{t\in T}$ spełnia warunek (25). Wówczas dla dowolnej miary probabilistycznej μ na (T,d), której nośnik jest gęsty w T oraz przeliczalnego podzbioru $T_0\subset T$,*

$$\mathbf{E}\sup_{s,t\in T_0}(X_s-X_t)\leqslant 32\sup_{t\in T}\int_0^{\Delta(T)}\varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon.$$

Uwaga 12.2. Talagrand wykazał, że dla dowolnej przestrzeni metrycznej (T,d) i dowolnej funkcji Orlicza φ istnieje miara probabilistyczna μ na (T,d) taka, że

$$\sup_{t\in T}\int_0^{\Delta(T)}\varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon\leqslant 4\int_0^{\Delta(T)}\varphi^{-1}(N(t,d,\varepsilon))d\varepsilon,$$

zatem (z dokładnością do stałych) Twierdzenie 12.1 jest silniejsze niż Twierdzenie 11.4.

By skrócić notację zdefiniujmy

$$\sigma_\mu(t):=\int_0^{\Delta(T)}\varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon\quad\text{oraz}\quad S(T,\mu)=S(T,\mu,\varphi):=\sup_{t\in T}\sigma_\mu(t).$$

Kluczem do dowodu Twierdzenia 12.1 jest następujące deterministyczne szacowanie.

Twierdzenie 12.3. *Dla dowolnej funkcji Younga φ i miary probabilistycznej μ na (T,d) istnieje miara probabilistyczna ν na $T\times T$ taka, że*

$$\sup_{t\in T}\left|f(t)-\int fd\mu\right|\leqslant 16S(T,\mu)\left(\frac{2}{3}+\frac{1}{3}\int_{T\times T}\varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right)d\nu(u,w)\right).$$

Dowód twierdzenia 12.1. . Ustalmy $t_0\in T$ i połóżmy $Y_t:=X_t-X_{t_0}$. Wówczas $Y_s-Y_t=X_s-X_t$ oraz $\mathbf{E}|Y_t|\leqslant d(t,t_0)+\mathbf{E}\varphi(|Y_t|/d(t,t_0))/\varphi(1)<\infty$. Możemy zatem zakładać, że proces $(X_t)_{t\in T}$ jest całkowalny.

Załóżmy najpierw, że σ-ciało $\mathcal{F}$ jest skończone. Sklejając wszystkie elementy wchodzące w skład jednego atomu (i wyrzucając atomy o mierze zerowej) możemy zakładać, że przestrzeń zdarzeń elementarnych Ω jest skończona i $\mathbf{P}(\{\omega\})>0$ dla $\omega\in\Omega$. Zauważmy, że dla $s,t\in T$,

$$|X_s(\omega)-X_t(\omega)|\leqslant d(s,t)\varphi^{-1}(1/\mathbf{P}(\{\omega\})),$$

w szczególności X_t ma ciągłe trajektorie. Stąd stosując Twierdzenie 12.3 do każdej trajektorii z osobna dostajemy po odcałkowaniu

$$\begin{aligned}\mathbf{E}\sup_{s,t\in T_0}|X_s-X_t|&\leqslant 2\mathbf{E}\sup_{t\in T}\left|X_t-\int X_ud\mu(u)\right|\\&\leqslant 32S(T,\mu)\left(\frac{2}{3}+\frac{1}{3}\int_{T\times T}\mathbf{E}\varphi\left(\frac{|X_u-X_w|}{d(u,w)}\right)d\nu(u,w)\right)\leqslant 32S(T,\mu).\end{aligned}$$

Jeśli T_0 jest skończonym podzbiorem T, to istnieje rosnący ciąg skończonych σ-ciał $(\mathcal{F}_n)_{n\geqslant 1}$ taki, że $\sigma((X_t)_{t\in T_0}) = \sigma(\bigcup_n \mathcal{F}_n)$. Określmy $X_t^n := \mathbf{E}(X_t|\mathcal{F}_n)$, wówczas X_t^n zbiega do X_t p.n. i w L^1, ponadto na mocy nierówności Jensena

$$\mathbf{E}\varphi\left(\frac{|X_n(t)-X_n(s)|}{d(s,t)}\right) \leqslant \mathbf{E}\varphi\left(\frac{|X(t)-X(s)|}{d(s,t)}\right) \leqslant 1.$$

Stąd

$$\mathbf{E}\sup_{s,t\in T_0}|X_s - X_t| = \lim_n \mathbf{E}\sup_{s,t\in T_0}|X_s^n - X_t^n| \leqslant 32S(T,\mu).$$

Jeśli T_0 jest przeliczalne, to jest wstępującą sumą zbiorów skończonych i proste przejście graniczne pokazuję tezę w tym przypadku. $\square$

By udowodnić Twierdzenie 12.3 możemy zakładać, że $S(T,\mu)<\infty$, co w szczególności implikuje, że $\Delta(T)<\infty$ (bo $S(T,\mu)\geqslant \Delta(T)\varphi(1)$). Niech $k_0\in\mathbb{Z}$ spełnia

$$4^{k_0}\leqslant \varphi^{-1}(1) < 4^{k_0+1}.$$

Określmy dla $t\in T$,

$$r_{k_0}(t) := \Delta(T) \quad \text{oraz} \quad r_k(t) := \min\left\{\varepsilon\geqslant 0\colon\ \varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)\leqslant 4^k\right\} \quad \text{dla } k>k_0.$$

Lemat 12.4. *i) Funkcje r_k są 1-Lipschitzowskie na (T,d).*
ii) Dla $t\in T$ zachodzi

$$\sum_{k\geqslant k_0} r_k(t)4^k \leqslant \frac{4}{3}\sigma_\mu(t).$$

iii) Dla $m\geqslant k_0$ i $t\in T$,

$$\sum_{k=k_0}^{m-1} 4^k \sum_{i=k}^{m} 2^{i-k} r_i(t) \leqslant \frac{8}{3}\sigma_\mu(t).$$

Dowód. i) Funkcja r_{k_0} jest stała, a dla $k>k_0$ mamy $B(s,\varepsilon+d(s,t))\supset B(t,\varepsilon)$, skąd łatwo wynika, że $r_k(s)\leqslant r_k(t)+d(s,t)$.

ii) Mamy

$$\begin{aligned}
\frac{3}{4}\sum_{k\geqslant k_0} r_k(t)4^k &= \sum_{k\geqslant k_0} r_k(t)(4^k-4^{k-1}) \leqslant \sum_{k\geqslant k_0}(r_k(t)-r_{k+1}(t))4^k + \limsup_{k\to\infty} r_k(t)4^k \\
&\leqslant \sum_{k\geqslant k_0}\int_{r_{k+1}(t)}^{r_k(t)} \varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon + \limsup_{k\to\infty}\int_0^{r_k(t)}\varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon \\
&\leqslant \int_0^{\Delta(T)}\varphi^{-1}\left(\frac{1}{\mu(B(t,\varepsilon))}\right)d\varepsilon.
\end{aligned}$$

iii) Liczymy

$$\sum_{k=k_0}^{m-1} 4^k \sum_{i=k}^{m} 2^{i-k} r_i = \sum_{k=k_0}^{m-1} \sum_{i=k}^{m} 2^{k-i} 4^i r_i \leqslant \sum_{j=0}^{\infty} 2^{-j} \sum_{i \geqslant k_0} 4^i r_i = 2 \sum_{i \geqslant k_0} 4^i r_i.$$

i korzystamy z ii). □

Określmy dla $k \geqslant k_0$ operatory liniowe S_k, działające na ograniczonych funkcjach borelowskich na T, wzorem

$$S_k f(t) := \fint_{B_k(t)} f(u) d\mu(u) = \frac{1}{\mu(B_k(t)} \int_{B_k(t)} f(u) d\mu(u), \quad \text{gdzie } B_k(t) := B(t, r_k(t)).$$

Kolejny lemat podsumowuje proste własności S_k, które przydadzą nam się później.

Lemat 12.5. *Dla $k \geqslant k_0$,*
i) $S_k 1 = 1$,
ii) $S_k f \leqslant S_k g$ *dla* $f \leqslant g$, *w szczególności* $|S_k f| \leqslant S_k |f|$,
iii) $S_k S_{k_0} f = S_{k_0} f = \int_T f d\mu$,
iv) $\lim_{k\to\infty} S_k f(t) = f(t)$ *dla funkcji ciągłych* f *i* $t \in T$,

Dowód. Warunki i)-iii) są oczywiste, a iv) wynika stąd, że $\lim_{k\to\infty} r_k(t) = 0$. □

Zanim przejdziemy do dowodu Twierdzenia 12.1 wykażemy jeszcze jeden fakt.

Lemat 12.6. *Zachodzą następujące oszacowania:*
i) $S_i r_j \leqslant r_i + r_j$ *dla* $i, j \geqslant k_0$,
ii) $S_m S_{m-1} \cdots S_{k+1} r_k \leqslant \sum_{i=k}^{m} 2^{i-k} r_i$ *dla* $m > k \geqslant k_0$.

Dowód. i) Lipschitzowskości r_j implikuje $r_j(u) \leqslant r_i(t) + r_j(t)$ dla $u \in B_j(t)$.

ii) Udowodnimy oszacowanie przez indukcję po m. Dla $m = k+1$ z i) mamy $S_{k+1} r_k \leqslant r_{k+1} + r_k \leqslant r_k + 2r_{k+1}$. By wykazać krok indukcyjny załóżmy, że ii) zachodzi dla $m > k \geqslant k_0$. Założenie indukcyjne, liniowość S_{m+1} i część i) implikują

$$S_{m+1} S_m \cdots S_{k+1} r_k \leqslant S_{m+1} \sum_{i=k}^{m} 2^{i-k} r_i = \sum_{i=k}^{m} 2^{i-k} S_{m+1} r_i \leqslant \sum_{i=k}^{m} 2^{i-k} (r_i + r_{m+1}) \leqslant \sum_{i=k}^{m+1} 2^{i-k} r_i.$$

□

Dowód Twierdzenia 12.3. Mamy

$$\begin{aligned}\left|f(t) - \int_T f d\mu\right| &= \lim_{m\to\infty} \left|S_m f - S_m S_{m-1}\cdots S_{k_0} f\right|(t)\\ &= \lim_{m\to\infty}\left|\sum_{k=k_0}^{m-1} S_m S_{m-1}\cdots S_{k+1}(I-S_k)f\right|(t)\\ &\leqslant \lim_{m\to\infty}\sum_{k=k_0}^{m-1} S_m S_{m-1}\cdots S_{k+2}\left|S_{k+1}(I-S_k)f\right|(t). \end{aligned}\tag{26}$$

Dla $k \geqslant k_0$ zachodzi

$$\begin{aligned}\left|S_{k+1}(I-S_k)f\right|(t) &= \left|⨍_{B_{k+1}(t)} ⨍_{B_k(u)} (f(u)-f(w))d\mu(u)d\mu(w)\right|\\ &\leqslant ⨍_{B_{k+1}(t)} ⨍_{B_k(u)} |f(u)-f(w)| d\mu(u)d\mu(w).\end{aligned}$$

Mamy $\varphi(xy) \geqslant x\varphi(y)$ dla $x \geqslant 1$, $y \geqslant 0$, zatem

$$x \leqslant 1 + \frac{\varphi(xy)}{\varphi(y)} \quad \text{dla } x \geqslant 0, y > 0. \tag{27}$$

Stąd dla $u \neq w$

$$\frac{|f(u)-f(w)|}{4^{k+1}d(u,w)} \leqslant 1 + \frac{1}{\varphi(4^{k+1})}\varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right).$$

Zauważmy, że dla $w \in B_k(u)$ zachodzi $d(u,w) \leqslant r_k(u)$, ponadto, z definicji r_{k+1} wynika, że $\mu(B_{k+1}(t)) \geqslant 1/\varphi(4^{k+1})$. Zatem

$$|f(u)-f(w)| \leqslant r_k(u)4^{k+1} + \mu(B_{k+1}(t))r_k(u)4^{k+1}\varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right) \quad \text{dla } w \in B_k(u)$$

i

$$\left|S_{k+1}(I-S_k)f\right|(t) \leqslant 4^{k+1}S_{k+1}r_k(t) + \int_T r_k(u)4^{k+1} ⨍_{B_k(u)} \varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right)d\mu(w)d\mu(u).$$

Stąd, wobec Lematu 12.6,

$$\begin{aligned}&S_m S_{m-1}\cdots S_{k+2}\left|S_{k+1}(I-S_k)f\right|(t)\\ &\qquad\leqslant 4^{k+1}\sum_{i=k}^{m} 2^{i-k} r_i(t) + 4\int_T r_k(u)4^k ⨍_{B_k(u)} \varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right)d\mu(w)d\mu(u).\end{aligned}$$

Powyższe oszacowanie, (26) oraz Lemat 12.4 implikują

$$\left|f(t) - \int_T f d\mu\right| \leqslant \frac{32}{3}\sigma_\mu(t) + 4\sum_{k=k_0}^{\infty}\int_T r_k(u)4^k \fint_{B_k(u)} \varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right) d\mu(w)d\mu(u).$$

Niech ν będzie miarą probabilistyczną na $T \times T$ daną wzorem

$$\nu(A) := \frac{1}{M}\sum_{k=k_0}^{\infty}\int_T r_k(u)4^k \fint_{B_k(u)} \mathbb{1}_A(u,w)d\mu(w)d\mu(u),$$

gdzie

$$M := \sum_{k=k_0}^{\infty}\int_T r_k(u)4^k d\mu(u) \leqslant \frac{4}{3}\int_T \sigma_\mu(u)d\mu(u) \leqslant \frac{4}{3}S(\mu,T).$$

Wówczas

$$\begin{aligned}\left|f(t) - \int_T f d\mu\right| &\leqslant \frac{32}{3}\sigma_\mu(t) + 4M\int_{T\times T}\varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right) d\nu(u,w)\\ &\leqslant 16S(T,\mu)\left(\frac{2}{3} + \frac{1}{3}\int_{T\times T}\varphi\left(\frac{|f(u)-f(w)|}{d(u,w)}\right) d\nu(u,w)\right).\end{aligned}$$

□

Twierdzenie 12.1 implikuje w szczególności oszacowanie supremów procesów gaussowskich, udowodnione pierwotnie przez Fernique'a.

Wniosek 12.7 (Fernique)**.** *Załóżmy, że $(X_t)_{t\in T}$ jest scentrowanym procesem gaussowskim oraz $d(t,s) = (\mathbf{E}|X_t - X_s|^2)^{1/2}$. Wówczas*

$$\mathbf{E}\sup_{t\in T} X_t \leqslant C\sup_{t\in T}\int_0^{\Delta(T)}\sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon = C\sup_{t\in T}\int_0^{\infty}\sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon,$$

gdzie C jest stałą numeryczną (można przyjąć $C = 48\sqrt{8/3}\sqrt{\ln 3/\ln 2} \leqslant 100$).

Dowód. Postępujemy jak w dowodzie Wniosku 11.7. Stosując Twierdzenie 12.1 do $\varphi = \psi_2$ i metryki $\sqrt{8/3}d$ dostajemy

$$\mathbf{E}\sup_{t\in T} X_t = \frac{1}{2}\mathbf{E}\sup_{s,t\in T}(X_s - X_t) \leqslant 16\sqrt{\frac{8}{3}}\sup_{t\in T}\int_0^{\Delta(T)}\sqrt{\ln\left(1 + \frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon.$$

Zauważmy, że

$$\int_0^{\Delta(T)}\sqrt{\ln\left(1 + \frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon \leqslant \sqrt{\ln 3}\Delta(T) + \sqrt{\frac{\ln 3}{\ln 2}}\int_0^{\Delta(T)}\sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon.$$

Ponadto dla $\varepsilon < \Delta(T)/2$ istnieje $t \in T$ taki, że $\mu(B(t,\varepsilon)) \leqslant 1/2$, więc

$$\sup_{t\in T}\int_0^{\Delta(T)}\sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon \geqslant \frac{1}{2}\sqrt{\ln 2}\Delta(T).$$

□

12.2 Dwustronne szacowania supremów procesów gaussowskich

W tej sekcji będziemy rozważać scentrowane procesy gaussowskie $(X_t)_{t\in T}$. Twierdzenie Fernique'a (Wniosek 12.7) mówi, że dla takich procesów i metryki $d(t,s) = \|X_t - X_s\|_2$ zachodzi

$$\mathbf{E}\sup_{t\in T} X_t \leqslant C\tilde{\gamma}_2(T,d),$$

gdzie

$$\tilde{\gamma}_2(T,d) := \inf\left\{\sup_{t\in T}\int_0^\infty \sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon\colon\ \mu \text{ miara probabilistyczna na } T\right\}.$$

Często, wygodniej niż z miarę probabilistyczną na T konstruować ciąg rozbić T.

Definicja 12.8. Mówimy, że $(\mathcal{A}_n)_{n\geqslant 0}$ jest *dopuszczalnym ciągiem rozbić* T, jeśli jest to ciąg rosnący (tzn. $\mathcal{A}_{n+1}$ jest podrozbiciem $\mathcal{A}_n$), $\mathcal{A}_0 = \{T\}$ oraz $|\mathcal{A}_n| \leqslant N_n := 2^{2^n}$ dla $n \geqslant 1$.

Dla przestrzeni metrycznej (T,d) określamy $\gamma_2(T,d)$ wzorem

$$\gamma_2(T,d) := \inf\left\{\sup_{t\in T}\sum_{n=0}^\infty 2^{n/2}\Delta(A_n(t))\colon\ (\mathcal{A}_n)_{n\geqslant 0} \text{ dopuszczalny ciąg rozbić } T\right\},$$

gdzie $A_n(t)$ oznacza taki zbiór z rozbicia $\mathcal{A}_n$ dla którego $t \in A_n(t)$.

Fakt 12.9. *Dla dowolnej przestrzeni metrycznej (T,d), $\tilde{\gamma}_2(T,d) \leqslant \sqrt{2}\gamma_2(T,d)$.*

Dowód. Ustalmy dopuszczalny ciąg podziałów $(\mathcal{A}_n)_{n\geqslant 0}$ przestrzeni T. By wykazać tezę wystarczy skonstruować miarę probabilistyczną μ na T taką, że

$$\sup_{t\in T}\int_0^{\Delta(T)} \sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon \leqslant \sqrt{2}\sup_{t\in T}\sum_{n=0}^\infty 2^{n/2}\Delta(A_n(t)).$$

Wybierzmy $T_n \subset T$ dla $n = 1,2\ldots$ takie, że $|T_n| \leqslant N_n$ i T_n zawiera po jednym punkcie z każdego ze zbiorów należących do rozbicia $\mathcal{A}_n$. Wówczas

$$\sum_{n\geqslant 1}|T_n|e^{-2^n} \leqslant 4e^{-2} + 16e^{-4} + \sum_{k\geqslant 8} 2^k e^{-k} \leqslant 1,$$

więc istnieje miara probabilistyczna taka, że $\mu(\{t\}) \geqslant \exp(-2^n)$ dla $t \in T_n$. Stąd dla $n \geqslant 1$, $\mu(B(t,\Delta(A_n(t)))) \geqslant \exp(-2^n)$ zatem

$$\begin{aligned}\int_0^{\Delta(T)} \sqrt{\ln\left(\frac{1}{\mu(B(t,\varepsilon))}\right)}d\varepsilon &= \sum_{n=0}^\infty \int_{\Delta(A_{n+1}(T))}^{\Delta(A_n(T))} \sqrt{\ln\left(\frac{1}{\mu(B(t,\Delta(A_{n+1}(t))))}\right)}d\varepsilon \\ &\leqslant \sum_{n=1}^\infty 2^{(n+1)/2}\Delta(A_n(t)) \leqslant \sqrt{2}\sum_{n=0}^\infty 2^{n/2}\Delta(A_n(t)).\end{aligned}$$

$\square$

Twierdzenie 12.10 (Talagrand)**.** *Niech $(X_t)_{t\in T}$ będzie scentrowanym, ośrodkowym procesem gaussowskim oraz $d(t,s) := \|X_t - X_s\|_2$. Wówczas $\gamma_2(T,d) \leqslant C\mathbf{E}\sup_{t\in T} X_t$.*

Dowód, który pokażemy poniżej pochodzi od van Handela. Zanim przedstawimy kluczowy fakt na którym on się opiera, będziemy potrzebowali kilku definicji.

Definicja 12.11. Dla $n \geqslant 0$ i przestrzeni metrycznej (T,d) definiujemy

$$e_n(T) = e_n(T,d) := \inf\{r > 0\colon\ N(T,d,r) < N_n = 2^{2^n}\}.$$

Nietrudno zauważyć, że $\frac{1}{2}\Delta(T) \leqslant e_0(T) \leqslant \Delta(T)$, $N(T,d,r) < N_n$ dla $r > e_n(T)$, zaś dla $r < e_n(T)$, $N(T,d,r) \geqslant N_n$. W szczególności istnieją punkty $x_1, \ldots, x_{N_n} \in T$ takie, że $d(x_i,x_j) \geqslant \frac{1}{2}e_n(T)$ dla $1 \leqslant i < j \leqslant N_n$.

Fakt 12.12 (van Handel)**.** *Załóżmy, że funkcje $r_n\colon\ T \to [0,\infty)$, $n = 0,1,\ldots$ spełniają warunek*

$$\forall_{n\geqslant 0}\forall_{A\subset T}\ e_n(A) \leqslant \frac{1}{6}\Delta(A) + \sup_{t\in A} r_n(t). \tag{28}$$

Wówczas

$$\gamma_2(T) \leqslant 70\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2} r_n(t).$$

Dowód. Dla dopuszczalnego ciągu podziałów $\mathcal{A} = (\mathcal{A}_n)_{n\geqslant 0}$ przestrzeni T określmy

$$\gamma_2(\mathcal{A}) := \sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}\Delta(A_n(t)).$$

Zauważmy, że funkcje $\tilde{r}_n := \min\{r_n, \Delta(T)\}$ spełniają (28), więc możemy bez straty ogólności zakładać, że $r_n \leqslant \Delta(T)$.

Określimy w sposób indukcyjny dopuszczalny ciąg podziałów $(\mathcal{A}_n)_{n\geqslant 0}$. Kładziemy $\mathcal{A}_0 = \mathcal{A}_1 := \{T\}$. Załóżmy, że mamy określony podział $\mathcal{A}_n$ dla $n \geqslant 1$ taki, że $|\mathcal{A}_n| \leqslant N_n$, skonstruujemy jego podpodział $\mathcal{A}_{n+1}$. Ustalmy w tym celu zbiór $A \in \mathcal{A}_n$ i podzielmy go najpierw na zbiory $A^1, \ldots, A^{n+1}$ dane wzorami:

$$\begin{aligned} A^i &:= \{t \in A\colon\ 2^{-i}\Delta(T) < r_{n-1}(t) \leqslant 2^{1-i}\Delta(T)\} \quad \text{dla } 1 \leqslant i \leqslant n,\\ A^{n+1} &:= \{t \in A\colon r_{n-1}(t) \leqslant 2^{-n}\Delta(T)\}. \end{aligned}$$

Wówczas jak łatwo zauważyć

$$r_{n-1}(s) \leqslant 2r_{n-1}(t) + 2^{-n}\Delta(T) \quad \text{dla } s,t \in A^i,\ 1 \leqslant i \leqslant n+1,$$

zatem z założenia (28)

$$e_{n-1}(A^i) \leqslant \frac{1}{6}\Delta(A) + 2r_{n-1}(t) + 2^{-n}\Delta(T) \quad \text{dla } t \in A^i,\ 1 \leqslant i \leqslant n+1,$$

w szczególności każdy A^i da się podzielić na mniej niż N_{n-1} zbiorów A^{ij} takich, że

$$\Delta(A^{ij}) \leqslant \frac{1}{3}\Delta(A) + 4r_{n-1}(t) + 2^{1-n}\Delta(T).$$

Określmy

$$\mathcal{A}_{n+1} := \left\{A^{ij}\colon\ A \in \mathcal{A}_n, i \leqslant n+1, j < N_n\right\}.$$

Wówczas

$$|\mathcal{A}_{n+1}| \leqslant N_n(n+1)N_{n-1} \leqslant N_{n+1}.$$

Zauważmy, że na mocy indukcyjnej konstrukcji

$$\Delta(A_{n+1}(t)) \leqslant \frac{1}{3}\Delta(A_n(t)) + 4r_{n-1}(t) + 2^{1-n}\Delta(T) \quad \text{dla } t \in T,\ n \geqslant 1.$$

Stąd

$$\begin{aligned}\sum_{n=0}^{\infty} 2^{n/2}\Delta(A_n(t)) &\leqslant (1+\sqrt{2})\Delta(T) + \sum_{n=2}^{\infty} 2^{n/2}\left(\frac{1}{3}\Delta(A_{n-1}(t)) + 4r_{n-2}(t) + 2^{2-n}\Delta(T)\right)\\ &\leqslant \frac{\sqrt{2}}{3}\sum_{n=1}^{\infty} 2^{n/2}\Delta(A_n(t)) + 8\sum_{n=0}^{\infty} 2^{n/2}r_n(t) + (5+3\sqrt{2})\Delta(T).\end{aligned}$$

Zauważmy, że (28) implikuje

$$\Delta(T) \leqslant 2e_0(T) \leqslant \frac{1}{3}\Delta(T) + 2\sup_{t\in T} r_0(t),$$

stąd

$$\Delta(T) \leqslant 3\sup_{t\in T} r_0(t)$$

Zatem

$$\sum_{n=0}^{\infty} 2^{n/2}\Delta(A_n(t)) \leqslant \frac{\sqrt{2}}{3}\sum_{n=0}^{\infty} 2^{n/2}\Delta(A_n(t)) + (23+9\sqrt{2})\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}r_n(t).$$

Zatem biorąc supremum obu stron po t otrzymujemy

$$\gamma_2(\mathcal{A}) \leqslant \frac{\sqrt{2}}{3}\gamma_2(\mathcal{A}) + (23+9\sqrt{2})\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}r_n(t).$$

Stąd

$$\gamma_2(T) \leqslant \gamma_2(\mathcal{A}) \leqslant \frac{3}{3-\sqrt{2}}(23+9\sqrt{2})\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}r_n(t) \leqslant 70\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2}r_n(t).$$

$\square$

Załóżmy, że $(X_t)_{t\in T}$ jest scentrowanym procesem gaussowskim takim, że $\mathbf{E}\sup_{t\in T} X_t < \infty$. Określmy

$$g(A) := \mathbf{E}\sup_{t\in A} X_t, \quad A\subset T$$

oraz

$$K(a,t) := \inf_{r\geqslant 0}\{ar + g(T) - g(B(t,r))\} \quad a>0, t\in T.$$

Ponadto dla $\delta>0$ niech $r_\delta(a,t)\in[0,,\infty)$ spełnia nierówność

$$ar_\delta(a,t) + g(T) - g(B(t,r_\delta(a,t))) \leqslant K(a,t).$$

Lemat 12.13. *Dla $\varepsilon>0$ i ciągu δ_n zachodzi*

$$\sup_{t\in T}\sum_{n=0}^{\infty} 2^{n/2} r_{\delta_n}(\varepsilon 2^{n/2},t) \leqslant \frac{2+\sqrt{2}}{\varepsilon}\left(g(T) + \sum_{n=0}^{\infty}\delta_n\right).$$

Dowód. Mamy $K(b,t)\leqslant br_\delta(a,t) + g(T) - g(B(t,r_\delta(a,t)))$, zatem

$$\delta + K(a,t) - K(b,t) \geqslant (a-b)r(a,t) + \delta \quad a,b,\delta>0, t\in T,$$

stąd

$$\begin{aligned}\sum_{n=0}^{m}\varepsilon 2^{n/2}(1-2^{-1/2})r_{\delta_n}(\varepsilon 2^{n/2},t) &\leqslant \sum_{n=0}^{m}\Big(K(\varepsilon 2^{n/2},t) - K(\varepsilon 2^{(n-1)/2},t) + \delta_n\Big)\\ &= K(\varepsilon 2^{m/2},t) - K(\varepsilon 2^{-1/2},t) + \sum_{n\geqslant 0}\delta_n \leqslant g(T) + \sum_{n\geqslant 0}\delta_n,\end{aligned}$$

gdzie ostatnia nierówność wynika z szacowania $0\leqslant K(a,t)\leqslant g(T)$. □

Dowód Twierdzenia 12.10. Wykażemy wpierw, że dla $\varepsilon,\delta>0$ i $n\geqslant 0$,

$$e_n(A) \leqslant C\varepsilon\Delta(A) + \left(\frac{2}{\alpha} + C\varepsilon\right)\sup_{t\in A} r_\delta(\varepsilon 2^{n/2},t) + C2^{-n/2}\delta, \tag{29}$$

gdzie C jest stałą uniwersalną, zaś α stałą z Twierdzenia 11.12.

Określmy

$$\sigma := \sup_{t\in A} r_\delta(\varepsilon 2^{n/2},t), \quad r := \sigma + \Delta(A).$$

By udowodnić (29) wystarczy rozpatrzeć przypadek, gdy $\sigma < \frac{\alpha}{2}e_n(T)$. Na mocy definicji $e_n(T)$, istnieją $t_i\in T$, $1\leqslant i\leqslant N_n$ takie, że $d(t_i,t_j) > e_n(T)/2$ dla $i\neq j$. Twierdzenie 11.12 implikuje, że

$$g\left(\bigcup_{i=1}^{N_n} B(t_i,\sigma)\right) \geqslant \frac{1}{16}e_n(T)\sqrt{\log N_n} + \min_{i\leqslant N_n} g(B(t_i,\sigma)).$$

Stąd istnieje $k \leqslant N_n$ takie, że

$$g(T) - g\left(\bigcup_{i=1}^{N_n} B(t_i, \sigma)\right) \leqslant g(T) - g(B(t_k, \sigma)) - \frac{\sqrt{\log 2}}{16} 2^{n/2} e_n(T).$$

Mamy

$$\begin{aligned} g(T) - g(B(t_k, \sigma)) &\leqslant g(T) - g(B(t_k, r_\delta(\varepsilon 2^{n/2}, t_k))) \leqslant K(\varepsilon 2^{n/2}, t_k) + \delta \\ &\leqslant \varepsilon 2^{n/2} r + g(T) - g(B(t_k, r)) + \delta \leqslant \varepsilon 2^{n/2} r + g(T) - g\left(\bigcup_{i=1}^{N_n} B(t_i, \sigma)\right) + \delta, \end{aligned}$$

gdzie ostatnia nierówność wynika stąd, że $\bigcup_{i \leqslant N_n} B(t_i, \sigma) \subset B(t_k, r)$. Porównując ostatnie dwa oszacowania dostajemy

$$e_n(A) \leqslant \frac{16}{\sqrt{\log 2}} \left(\varepsilon r + 2^{-n/2} \delta\right)$$

i nierówność (29) zachodzi z $C = 16/\sqrt{\log 2}$.

By zakończyć dowód wystarczy wybrać

$$\varepsilon := \frac{1}{6C}, \quad \delta_n := \frac{1}{C} 2^{-n} g(T), \quad r_n := \left(\frac{2}{\alpha} + \frac{1}{6}\right) r_{\delta_n}(\varepsilon 2^{n/2}, t) + C 2^{-n/2} \delta_n$$

i zastosować Fakt 12.12 i Lemat 12.13. □

Wniosek 12.14 (Fernique-Talagrand). *Niech $(X_t)_{t \in T}$ będzie scentrowanym, ośrodkowym procesem gaussowskim oraz $d(t,s) := \|X_t - X_s\|_2$. Wówczas*

$$\mathbf{E} \sup_{t \in T} X_t \sim \gamma_2(T, d) \sim \tilde{\gamma}_2(T, d).$$

12.3 Zmienne i procesy subgaussowskie

Fakt 12.15. *Niech Z będzie zmienną losową. Następujace własności zmiennej Z są równoważne:*
i) Ogony zmiennej Z spełniają

$$\mathbf{P}(|Z| \geqslant t) \leqslant 2 \exp(-t^2/K_1^2) \quad \text{dla } t \geqslant 0.$$

ii) Momenty zmiennej Z spełniają

$$\|Z\|_p = (\mathbf{E}|Z|^p)^{1/p} \leqslant K_2 \sqrt{p} \quad \text{dla } p \geqslant 1.$$

iii) Transformata Laplace'a Z^2 spełnia

$$\mathbf{E} \exp(\lambda^2 Z^2) \leqslant \exp(K_3^2 \lambda^2) \quad \text{dla } |\lambda| \leqslant \frac{1}{K_3}.$$

iv) Zmienna Z ma skończoną normę ψ_2-Orlicza , tzn.

$$\mathbf{E}\exp(Z^2/K_4^2) \leqslant 2.$$

Jeśli dodatkowo $\mathbf{E}Z = 0$ to warunki i)-iv) są równoważne
v) Transformata Laplace'a X^2 spełnia

$$\mathbf{E}\exp(\lambda Z) \leqslant \exp(K_5^2\lambda^2) \quad \textit{dla } \lambda \in \mathbb{R}.$$

Ponadto optymalne stałe dla których zachodzą powyższe nierówności są porównywalne ze sobą z dokładnością do stałej uniwersalnej, tzn. $K_i \leqslant CK_j$ dla $i, j = 1, \ldots, 5$.

Dowód. i)$\Rightarrow$ii) Stosujemy całkowanie przez części:

$$\begin{aligned}\mathbf{E}|Z|^p &= p\int_0^\infty t^{p-1}\mathbf{P}(|Z| \geqslant t)dt \leqslant 2p\int_0^\infty t^{p-1}\exp(-t^2/K_1^2)dt \\ &= pK_1^p\int_0^\infty s^{(p-2)/2}\exp(-s)ds = 2K_1^p\Gamma(p/2+1).\end{aligned}$$

Funkcja Γ jest logarytmicznie wypukła, stąd dla $x \in [1,2]$, $\Gamma(x) \leqslant \max\{\Gamma(1), \Gamma(2)\} = 1$, oraz dla $x \in [k, k+1]$, $k = 1, 2, \ldots$ mamy

$$\Gamma(x+1) = x(x-1)\cdots(x-k+1)\Gamma(x-k+1) \leqslant x^k \leqslant x^x.$$

Zatem $\|Z\|_p \leqslant 2K_1\sqrt{p/2}$ i ii) zachodzi z $K_2 = \sqrt{2}K_1$.

ii)$\Rightarrow$iii) Mamy dla $2e\lambda^2K_2^2 \leqslant 1$,

$$\begin{aligned}\mathbf{E}\exp(\lambda^2Z^2) &= 1 + \sum_{k=1}^\infty \frac{\lambda^{2k}}{k!}\mathbf{E}|Z|^{2k} \leqslant 1 + \sum_{k=1}^\infty \frac{\lambda^{2k}}{k!}K_2^{2k}k^k \leqslant 1 + \sum_{k=1}^\infty (e\lambda^2K_2^2)^k \\ &= 1 + \frac{e\lambda^2K_2^2}{1-e\lambda^2K_2^2} \leqslant 1 + 2e\lambda^2K_2^2 \leqslant \exp(2eK_2^2\lambda^2).\end{aligned}$$

Stąd dostajemy iii) z $K_3 = \sqrt{2e}K_2$.

iii)$\Rightarrow$iv) Oczywiste z $K_4 = K_3/\sqrt{\ln 2}$.

iv)$\Rightarrow$i) Natychmiastowy wniosek z nierówności Czebyszewa z $K_4 = K_1$.

iii)$\Rightarrow$v) Mamy $e^x \leqslant x + e^{x^2}$ dla $x \in \mathbb{R}$ (funkcja $f(x) = x + e^{x^2} - e^x$ spełnia $f(0) = f'(0) = 0$ i $f'' \geqslant 0$), zatem dla $|\lambda| \leqslant K_3^{-1}$ mamy

$$\mathbf{E}\exp(\lambda Z) \leqslant \mathbf{E}(\lambda Z + \exp(\lambda^2Z^2)) \leqslant \exp(K_3^2\lambda^2).$$

Dla $\lambda > K_3^{-1}$ korzystamy z nierówności $2\lambda x \leqslant \lambda^2K_3^2 + x^2K_3^{-2}$ i dostajemy

$$\mathbf{E}\exp(\lambda Z) \leqslant \exp\left(\frac{1}{2}K_3^2\lambda^2\right)\mathbf{E}\exp\left(\left(\frac{Z}{\sqrt{2}K_3}\right)^2\right) \leqslant \exp\left(\frac{1}{2}(K_3^2\lambda^2+1)\right) \leqslant \exp(K_3^2\lambda^2),$$

czyli v) zachodzi z $K_5 = K_3$.

v)$\Rightarrow$i) Wynika z Faktu 4.2 z $K_1 = 2K_5$. $\square$

Definicja 12.16. Jeśli Z spełnia równoważne własności wymienione w Fakcie 12.15, to mówimy, że Z jest subgaussowka.

Uwaga 12.17. Używa się różnych definicji stałej subgaussowskości zmiennej Z. Najczęściej się definiuje tę stałą albo poprzez transformatę Laplace'a jako:

$$\inf\Big\{\sigma > 0\colon\ \mathbf{E}\exp(\lambda(Z - \mathbf{E}Z)) \leqslant \exp(\lambda^2\sigma^2/2) \text{ dla } \lambda \in \mathbb{R}\Big\},$$

albo poprzez normę Orlicza ψ_2:

$$\|Z\|_{\psi_2} := \inf\{t > 0\colon\ \mathbf{E}\exp((Z/t)^2) \leqslant 2\}.$$

Dla procesu subgaussowskiego najwygodniej jest przyjąć definicję opartą o normę Orlicza.

Definicja 12.18. Mówimy, że proces $(X_t)_{t\in T}$ jest subgaussowski względem metryki d, jeśli

$$\mathbf{E}\psi_2\left(\frac{X_t - X_s}{d(t,s)}\right) \leqslant 1 \quad \text{dla } t, s \in T.$$

Tweirdzenia 12.1 (dla $\varphi = \psi_2$, zob. też Fakt 12.9) i 12.10 implikują następujące ważne twierdzenie dotyczące porównywania procesów subgaussowskich i gaussowskich.

Twierdzenie 12.19. *Załóżmy, że X_t jest ośrodkowym procesem gaussowskim o średniej zero. Wówczas dla dowolnego osrodkowego procesu $(Y_t)_{t\in T}$, który jest subgaussowski względem metryki $d(t,s) = \|X_t - X_s\|_2$ zachodzi*

$$\mathbf{E}\sup_{t,s\in T}(Y_t - Y_s) \leqslant C\mathbf{E}\sup_{t,s\in T}(X_t - X_s).$$

Jednym z możliwych przykładów procesów subgaussowskich są procesy kanoniczne postaci $X_t = \langle X, t\rangle = \sum_{i=1}^n t_i X_i$ dla $t \in \mathbb{R}^n$. Proces taki jest subgaussowski względem (wielokrotności) metryki euklidesowej, jeśli $\|\langle X, t\rangle\|_{\psi_2} \leqslant K|t|$. Mówimy wtedy o wektorach subgaussowskich.

Definicja 12.20. Mówimy, że n-wymiarowy wektor losowy $X = (X_1, \ldots, X_n)$ jest *subgaussowski*, jeśli

$$\|X\|_{\psi_2} := \sup_{|t|=1}\|\langle t, X\rangle\|_{\psi_2} < \infty.$$

Normę $\|X\|_{\psi_2}$ nazywamy *stałą subgaussowskości* wektora X.

Przykład 1. Jeśli $X_1, X_2, \ldots, X_n$ są niezależnymi subgaussowskimi zmiennymi losowymi o sredniej zero to $X = (X_1, \ldots, X_n)$ jest subgaussowski oraz

$$\|X\|_{\Psi_2} \sim \max_i \|X\|_{\psi_2}.$$

Przykład 2. Jeśli wektor X jest jednostajnie rozłożony na S^{n-1}, to X jest $Cn^{-1/2}$ subgaussowski.

12.4 Oszacowania przez ciąg metryk L_p

Nie wszystkie procesy stochastyczne mają przyrosty kontrolowane tylko przez jedną metrykę. Jednym ze sposobów radzenia sobie z tym problemem jest uogólnienie kombinatorycznej definicji funkcjonału γ_2 na przypadek, gdy na n-tej partycji rozpatrujemy odległość L_p z $p = 2^n$. Przypomijmy, że dla zmiennej losowej Z i $p \geqslant 1$ kładziemy $\|Z\|_p := (\mathbf{E}|Z|^p)^{1/p}$.

Definicja 12.21. Niech $(X_t)_{t\in T}$ będzie ośrodkowym procesem stochastycznym, którego przyrosty mają wszystkie momenty skończone. Określmy

$$\Delta_n(A) := \sup\{\|X_t - X_s\|_{2^n} \colon\ t, s \in A\}, \quad A \subset T, n = 0, 1, \ldots .$$

Definiujemy

$$\gamma_X(T) := \inf_{\mathcal{A}} \sup_{t\in T} \sum_{n=0}^{\infty} \Delta_n(A_n(t)),$$

gdzie infimum jest brane po wszystkich dopuszczalnych ciągach podziałów $\mathcal{A} = (\mathcal{A}_n)_{n\geqslant 0}$ zbioru T.

Fakt 12.22. *i) Jeśli (X_t) jest scentrowanym procesem gaussowskim, to $c\gamma_2(T) \leqslant \gamma_X(T) \leqslant \gamma_2(T)$.*
ii) Jeśli (X_t) jest procesem subgaussowskim względem metryki d, to $\gamma_X(T) \leqslant \sqrt{2}\gamma_2(T, d)$.

Dowód. Dla procesu gaussowkiego o średniej zero mamy $c\sqrt{p}\|X_t - X_s\| \leqslant \|X_t - X_s\|_p \leqslant \sqrt{p}\|X_t - X_s\|_2$ dla $p \geqslant 1$, a dla procesu subgaussowskiego $\|X_t - X_s\|_p \leqslant \sqrt{2p}d(t, s)$ dla $p \geqslant 1$ (zobacz Fakt 12.15 i jego dowód by dostać stałą $\sqrt{2}$). $\square$

Twierdzenie 12.23. *Niech $(X_t)_{t\in T}$ będzie ośrodkowym procesem stochastycznym, którego przyrosty mają wszystkie momenty skończone. Wówczas dla dowolnego $p \geqslant 1$,*

$$\left\| \sup_{t,s\in T} |X_s - X_t| \right\|_p \leqslant 48\gamma_X(T) + 256 \sup_{t,s\in T} \|X_s - X_t\|_p.$$

Dowód. Wystarczy udowodnić oszacowanie dla zbiorów skończonych. Niech $2^{k_0-1} \leqslant 2p \leqslant 2^{k_0}$ dla pewnego $k_0 = 2, 3, \ldots$ oraz wybierzmy $k_1 \geqslant k_0$ takie, że $N_{k_1} = 2^{2^{k_1}} \geqslant |T|$.

Ustalmy $\varepsilon > 0$ i wybierzmy dopuszczalny ciąg podziałów $\mathcal{A}_n$ zbioru T taki, że

$$\sup_{t\in T} \sum_{n=0}^{\infty} \Delta_n(A_n(t)) \leqslant (1 + \varepsilon)\gamma_X(T).$$

Możemy przyjąć, że $A_{k_1}(t) = \{t\}$. Wybierzmy zbiory $T_k \subset T$ takie, że $T_0 = \{t_0\}$ oraz $|T_k| \leqslant N_k$ i T_k zawiera po jednym punkcie z każdego ze zbiorów k-tej partycji $\mathcal{A}_k$ dla $k = 0, \ldots, k_1$. Niech dla $0 \leqslant k \leqslant k_1$, $\pi_k(t) \in T_k \cap A_k(t)$, w szczególności $\pi_0(t) = t_0$ oraz $\pi_{k_1}(t) = t$.

Z nierówności trójkąta w L_p dostajemy

$$\Big\|\sup_{t,s\in T}|X_s - X_t|\Big\|_p \leqslant 2\Big\|\sup_{t\in T}|X_t - X_{\pi_{k_0}(t)}|\Big\|_p + \Big\|\sup_{s,t\in T}|X_{\pi_{k_0}(s)} - X_{\pi_{k_0}(t)}|\Big\|_p.$$

Zauważmy, że $\pi_{k+1}(t) \in A_{k+1}(t) \subset A_k(t)$, więc

$$M := \sup_{t\in T}\sum_{k=k_0}^{k_1-1}\|X_{\pi_{k+1}(t)} - X_{\pi_k(t)}\|_{2^k} \leqslant \sup_{t\in T}\sum_{k=k_0}^{k_1-1}\Delta_k(A_k(t)) \leqslant (1+\varepsilon)\gamma_X(T).$$

Dla $u \geqslant 16$ szacujemy

$$\begin{aligned}
&\mathbf{P}\Big(\sup_{t\in T}|X_t - X_{\pi_{k_0}(t)}| \geqslant uM\Big)\\
&\quad\leqslant \mathbf{P}\Big(\sup_{t\in T}\sum_{k=k_0}^{k_1-1}|X_{\pi_{k+1}(t)} - X_{\pi_k(t)}| \geqslant uM\Big)\\
&\quad\leqslant \mathbf{P}\Big(\exists_{k_0\leqslant k\leqslant k_1-1}\exists_{t\in T}\ |X_{\pi_{k+1}(t)} - X_{\pi_k(t)}| \geqslant u\|X_{\pi_{k+1}(t)} - X_{\pi_k(t)}\|_{2^k}\Big)\\
&\quad\leqslant \sum_{k=k_0}^{k_1-1}\sum_{s\in T_{k+1}}\sum_{s'\in T_k}\mathbf{P}(|X_s - X_{s'}| \geqslant u\|X_s - X_{s'}\|_{2^k})\\
&\quad\leqslant \sum_{k=k_0}^{k_1-1}|T_{k+1}||T_k|u^{-2^k} \leqslant \sum_{k=k_0}^{k_1-1}\Big(\frac{8}{u}\Big)^{2^k} \leqslant 2\Big(\frac{8}{u}\Big)^{2^{k_0}} \leqslant 2\Big(\frac{8}{u}\Big)^{2p}.
\end{aligned}$$

Całkując przez części dostajemy

$$\begin{aligned}
\mathbf{E}\sup_{t\in T}|X_t - X_{\pi_{k_0}(t)}|^p &\leqslant (8M)^p\Big(2^p + p\int_2^\infty u^{p-1}2u^{-2p}du\Big)\\
&= (8M)^p(2^p + 2^{1-2p}) \leqslant (24M)^p \leqslant (24(1+\varepsilon)\gamma_X(T))^p.
\end{aligned}$$

Ponadto

$$\begin{aligned}
\mathbf{E}\sup_{t,s\in T}|X_{\pi_{k_0}(t)} - X_{\pi_{k_0}(s)}|^p &\leqslant \sum_{t',s'\in T_{k_0}}\mathbf{E}|X_{t'} - X_{s'}|^p \leqslant |T_{k_0}|^2\sup_{t,s\in T}\mathbf{E}|X_t - X_s|^p\\
&\leqslant 2^{8p}\sup_{s,t\in T}\|X_t - X_s\|_p^p.
\end{aligned}$$

Powyższe szacowania implikują

$$\Big\|\sup_{t,s\in T}|X_s - X_t|\Big\|_p \leqslant 48(1+\varepsilon)\gamma_X(T) + 2^8\sup_{s,t\in T}\|X_t - X_s\|_p$$

i teza wynika z dowolności $\varepsilon > 0$. $\square$

Uwaga 12.24. Ponieważ $\sup_{s,t\in T}\|X_t - X_s\|_1 \leqslant \gamma_X(T)$, więc Twierdzenie 12.23 implikuje, że

$$\mathbf{E}\sup_{t,s\in T}|X_s - X_t| \leqslant 304\gamma_X(T).$$

Wniosek 12.25. *Załóżmy, że $(X_t)_{t\in T}$ jest ośrodkowym procesem subgaussowskim względem metryki d. Wówczas*

$$\mathbf{P}\left(\sup_{s,t\in T}|X_t - X_s| \geqslant C(\gamma_2(T,d) + u\mathrm{diam}(T,d))\right) \leqslant \exp(-u^2) \quad \textit{dla } u > 0.$$

Dowód. Dal $u < 1$ teza łatwo wynika z uwagi i nierówności Czebyszewa (bierzemy $C = 304e$). Dla $u \geqslant 1$ teza wynika z nierówności Czebyszewa i Twierdzenia 12.23 dla $u = \sqrt{p}$ (na mocy Faktu 12.15 $\|X_t - X_s\|_p \leqslant C\sqrt{p}d(t,s)$). □

13 Macierze losowe o subgaussowskich rzędach

W tym rozdziale poznamy kilka silnych szacowań dla pewnej klasy macierzy losowych. Będziemy potrzebowali jej pewnego unormowania. Typowym normowaniem wektora losowego jest założenie jego izotropowości.

Definicja 13.1. Mówimy, że n-wymiarowy wektor losowy X jest *izotropowy*, jeśli ma średnią zero i identycznościową macierz kowariancji, tzn. $\mathbf{E}X_i = 0$ oraz $\mathbf{E}X_iX_j = \delta_{ij}$ dla $i, j \leqslant n$.

Fakt 13.2. *Jeśli wektor X jest izotropowy, to*

$$\mathbf{E}\langle X,t\rangle\langle X,s\rangle = \langle t,s\rangle \quad \textit{dla } t,s\in\mathbb{R}^n.$$

Dowód. Liczymy

$$\mathbf{E}\langle X,t\rangle\langle X,s\rangle = \sum_{ij} t_i s_j \mathbf{E}X_iX_j = \sum_i t_i s_i = \langle t,s\rangle.$$

□

Sprecyzujemy teraz klasę macierzy, które będziemy badać podczas dalszych rozważań.

Definicja 13.3. Powiemy, że macierz losowa $m \times n$ spełnia *założenie o subgaussowskości z parametrem K*, jeśli jej wiersze $A_1, \ldots, A_m$ są niezależnymi, izotropowymi, subgaussowskimi n-wymiarowymi wektorami losowymi oraz

$$\max_i \|A_i\|_{\psi_2} = \max_i \max_{|t|=1} \|\langle A_i, t\rangle\|_{\psi_2} \leqslant K.$$

Uwaga 13.4. Zauważmy, że $e^{x^2}-1 \geqslant x^2$, zatem $\|X\|_{\Psi_2} \geqslant \|X\|_2$ i izotropowość A_i implikuje, że $K \geqslant 1$.

Podamy teraz kilka przykładów klas macierzy spełniających założenia Definicji 13.3. Literami C, c oznaczamy dodatnie stałe skończone, których wartości mogą się zmieniać przy każdym wystąpieniu (jeśli będziemy chcieli ustalić wartość jakiejś stałej będziemy pisać $c_1, C_1, c_2, C_2, \ldots$).

Przykład 1. Macierz $A = (A_{ij})$, której współczynniki są niezależnymi subgaussowskimi zmiennymi losowymi o średniej zero i wariancji 1 spełnia założenia Definicji 13.3 z parametrem $K \leqslant C \max_i \|A_{ij}\|_{\psi_2}$. W szczególności macierz, której wyrazy są niezależnymi zmiennymi $\mathcal{N}(0,1)$ spełnia to założenie z $K = \sqrt{8/3}$.

Przykład 2. Macierz, której wiersze A_i są niezależne i mają rozkład jednostajny na $\sqrt{n}S^{n-1}$ spełnia założenia Definicji 13.3 z parametrem $K \leqslant C$.

Przykład 3. Macierz, której wiersze A_i są niezależne, izotropowe i mają rozkłady spełniające logarytmiczną nierówność Sobolewa z parametrem α spełnia założenia Definicji 13.3 z parametrem $K \leqslant C\alpha$.

Będziemy też używać następujących wielkości dla $T \subset \mathbb{R}^n$,

$$g(T) := \mathbf{E} \sup_{t \in T} \sum_{i=1}^n t_i g_i, \quad \gamma(T) := g(T \cup -T) = \mathbf{E} \sup_{t \in T} \left| \sum_{i=1}^n t_i g_i \right|, \quad R(T) := \sup_{t \in T} |t|,$$

gdzie jak zwykle $g_1, \ldots, g_n$ oznacza niezależne zmienne losowe $\mathcal{N}(0,1)$.

13.1 Odchylenia dla subgaussowskich macierzy losowych

Twierdzenie 13.5. *Załóżmy, że A jest macierzą $m \times n$, która spełnia założenia Definicji 13.3 z parametrem K. Wówczas dla dowonego niepustego ograniczone zbioru $T \subset \mathbb{R}^n$ zachodzi*

$$\mathbf{E} \sup_{t \in T} \big| |At| - \sqrt{m}|t| \big| \leqslant CK^2 \gamma(T).$$

Ponadto

$$\mathbf{P}\left(\sup_{t \in T} \big| |At| - \sqrt{m}|t| \big| \geqslant CK^2(\gamma(T) + uR(T)) \right) \leqslant \exp(-u^2) \quad \text{dla } u > 0.$$

Określmy proces (X_t) wzorem

$$X_t := |At| - \sqrt{m}|t|, \quad t \in \mathbb{R}^n. \tag{30}$$

Kluczowym elementem dowodu Twierdzenia 13.5 jest wykazanie subgaussowskości tego procesu.

Twierdzenie 13.6. *Załóżmy, że macierz losowa A spełnia założenia spełnia założenia Definicji 13.3 z parametrem K oraz proces X_t jest zadany wzorem* (30). *Wówczas*

$$\|X_t - X_s\|_{\psi_2} \leqslant CK^2|t-s| \quad \textit{dla } t,s \in \mathbb{R}^n. \tag{31}$$

Dowód Twierdzenia 13.5. Niech $\tilde{T} = T \cup \{0\}$.

$$\sup_{t\in T} |X_t| = \sup_{t\in T} \big||At| - \sqrt{m}|t|\big| \leqslant \sup_{t,s\in\tilde{T}} (X_t - X_s)$$

oraz

$$g(\tilde{T}) \leqslant 2\gamma(T) \quad \text{i} \quad \sup_{t,s\in\tilde{T}} |t-s| \leqslant 2R(T).$$

Szacowanie wartości oczekiwanej wynika natychmiast z Twierdzenia 13.6 i Twierdzenia 12.19 (zastosowanego do zbioru $\tilde{T}$). Szacowanie prawdopodobieństwa jest konsekwencją Twierdzenia 13.6 oraz Wniosku 12.25 i Twierdzenia 12.10 (również zastosowanych do zbióru $\tilde{T}$). □

By wykazać Twierdzenie 13.6 będziemy potrzebowali kilku lematów.

Lemat 13.7. *Dla dowolnych subgaussowskich zmiennych losowych X, Y zachodzi $\|XY\|_{\psi_1} \leqslant \|X\|_{\psi_2}\|Y\|_{\psi_2}$.*

Dowód. Załóżmy, że $a > \|X\|_{\psi_2}$ i $b > \|Y\|_{\psi_2}$. Wówczas

$$\mathbf{E}\exp\left(\frac{|XY|}{ab}\right) \leqslant \mathbf{E}\exp\left(\frac{X^2}{2a^2} + \frac{Y^2}{2b^2}\right) \leqslant \left(\mathbf{E}\exp\left(\frac{X^2}{a^2}\right)\right)^{1/2}\left(\mathbf{E}\exp\left(\frac{Y^2}{b^2}\right)\right)^{1/2} \leqslant 2,$$

czyli $\|XY\|_{\psi_1} \leqslant ab$. □

Lemat 13.8. *Oszacowanie* (31) *zachodzi dla $s = 0$, tzn.*

$$\big\||At| - \sqrt{m}|t|\big\|_{\psi_2} \leqslant CK^2|t| \quad \textit{dla } t \in \mathbb{R}^n.$$

Dowód. Z uwagi na jednorodność będziemy zakładać, że $|t| = 1$. Na mocy Faktu 12.15 wystarczy udowodnić, że $\mathbf{P}(||At| - \sqrt{m}| \geqslant K^2u) \leqslant 2\exp(-cu^2)$ dla $u > 0$. Rozpatrzymy dwa przypadki.

Przypadek I. $K^2u \leqslant \sqrt{m}$. Mamy

$$\mathbf{P}(||At| - \sqrt{m}| \geqslant K^2u) = \mathbf{P}(||At|^2 - m| \geqslant K^2u||At| + \sqrt{m}|) \leqslant \mathbf{P}(||At|^2 - m| \geqslant K^2u\sqrt{m}).$$

Zauważmy, że

$$|At|^2 - m = \sum_{i=1}^m (|\langle A_i, t\rangle|^2 - 1),$$

zmienne $|\langle A_i, t\rangle|^2 - 1$ są niezależne, mają średnią zero oraz

$$\||\langle A_i, t\rangle|^2 - 1\|_{\psi_1} \leqslant \||\langle A_i, t\rangle|^2\|_{\psi_1} + \|1\|_{\psi_1} \leqslant \|\langle A_i, t\rangle\|_{\psi_2}^2 + \|1\|_{\psi_1} \leqslant 2K^2,$$

gdzie wykorzystaliśmy to, że dla $u > K$,

$$2 \geqslant \mathbf{E}\exp(\langle A_i, t\rangle^2/u^2) \geqslant \exp(\mathbf{E}\langle A_i, t\rangle^2/u^2) = \exp(1/u^2).$$

Stąd nierówność Bernsteina dla zmiennych subwykładniczych (Twierdzenie 4.16) implikuje

$$\begin{aligned}\mathbf{P}(||At| - \sqrt{m}| \geqslant K^2u) \leqslant \mathbf{P}(||At|^2 - m| \geqslant K^2u\sqrt{m}) &\leqslant 2\exp\left(-\frac{K^4u^2m}{16mK^4 + 8K^4u\sqrt{m}}\right) \\ &\leqslant 2\exp\left(-\frac{u^2}{24}\right),\end{aligned}$$

w ostatniej nierówności skorzystaliśmy z tego, że $u \leqslant K^2u \leqslant \sqrt{m}$. □

Przypadek II. $K^2u > \sqrt{m}$. Wówczas

$$\mathbf{P}(||At| - \sqrt{m}| \geqslant K^2u) = \mathbf{P}(|At| \geqslant K^2u + \sqrt{m}) \leqslant \mathbf{P}(|At|^2 - m \geqslant K^4u^2)$$

Ponownie stosując nierówność Bernsteina i to, że $K \geqslant 1$ oraz $m \leqslant K^4u^2$ dostajemy

$$\mathbf{P}(||At| - \sqrt{m}| \geqslant K^2u) \leqslant 2\exp\left(-\frac{K^8u^4}{16mK^4 + 8K^6u^2}\right) \leqslant 2\exp\left(-\frac{u^2}{24}\right).$$

Lemat 13.9. *Oszacowanie* (31) *zachodzi dla* $|t| = |s| = 1$, *tzn.*

$$\||At| - |As|\|_{\psi_2} \leqslant CK^2|t - s| \quad \text{dla } t, s \in S^{n-1}.$$

Dowód. Na mocy Faktu 12.15 i tego, że $\min\{1, 4u\} \leqslant 2\sqrt{u}$ wystarczy udowodnić, że $\mathbf{P}(||At| - |As|| \geqslant uK^2|t - s|) \leqslant 4\exp(-cu^2)$ dla $u > 0$. W tym celu rozpatrzymy dwa przypadki.

Przypadek I. $uK^2 \leqslant 2\sqrt{m}$. Mamy

$$\begin{aligned}\mathbf{P}(||At| - |As|| \geqslant uK^2|t - s|) &= \mathbf{P}(||At|^2 - |As|^2| \geqslant uK^2|t - s|(|At| + |As|)) \\ &= \mathbf{P}(|\langle A(t+s), A(t-s)\rangle| \geqslant uK^2|t - s|(|At| + |As|)) \\ &\leqslant \mathbf{P}\left(|At| \leqslant \frac{1}{2}\sqrt{m}\right) + \mathbf{P}\left(|\langle A(t+s), A(t-s)\rangle| \geqslant \frac{1}{2}\sqrt{m}uK^2|t - s|\right).\end{aligned}$$

Na mocy Lematu 13.8,

$$\mathbf{P}\left(|At| \leqslant \frac{1}{2}\sqrt{m}\right) \leqslant \mathbf{P}\left(||At| - \sqrt{m}| \geqslant \frac{1}{2}\sqrt{m}\right) \leqslant 2\exp(-cm/K^4) \leqslant 2\exp(-cu^2/4).$$

Zauważmy, że

$$\langle A(t+s), A(t-s)\rangle = \sum_{i=1}^{m} \langle A_i, t+s\rangle\langle A_i, t-s\rangle.$$

Zmienne $\langle A_i, t+s\rangle\langle A_i, t-s\rangle$ mają (na mocy Faktu 13.2) średnią $\langle t+s, t-s\rangle = |t|^2 - |s|^2 = 0$ oraz

$$\|\langle A_i, t+s\rangle\langle A_i, t-s\rangle\|_{\psi_1} \leqslant \|\langle A_i, t+s\rangle\|_{\psi_2}\|\langle A_i, t-s\rangle\|_{\psi_2} \leqslant K^2|t+s||t-s| \leqslant 2K^2|t-s|.$$

Stosując nierówność Bernsteina dla zmiennych subwykładniczych (Twierdzenie 4.16) dostajemy

$$\begin{aligned}\mathbf{P}\left(|\langle A(t+s), A(t-s)\rangle| \geqslant \frac{1}{2}\sqrt{m}uK^2|t-s|\right) &\leqslant 2\exp\left(-\frac{mu^2K^4|t-s|^2/4}{16mK^4|t-s|^2 + 4\sqrt{m}uK^4|t-s|^2}\right)\\ &\leqslant 2\exp(-u^2/96),\end{aligned}$$

gdzie w ostatniej nierówności użylismy tego, że $u \leqslant uK^2 \leqslant 2\sqrt{m}$.

Przypadek II. $uK^2 > 2\sqrt{m}$. Wówczas

$$\begin{aligned}\mathbf{P}(||At| - |As|| \geqslant uK^2|t-s|) &\leqslant \mathbf{P}(|A(t-s)| \geqslant uK^2|t-s|)\\ &\leqslant \mathbf{P}\left(||A(t-s)| - \sqrt{m}|t-s|| \geqslant \frac{1}{2}uK^2|t-s|\right)\\ &\leqslant 2\exp(-cu^2)\end{aligned}$$

na mocy Lematu 13.8. □

Dowód Twierdzenia 13.6. Z uwagi na jednorodność, wystarczy wykazać (31) dla $|t| = 1 \leqslant |s|$. Niech $\tilde{s} = s/|s|$, wówczas

$$\|X_t - X_s\|_{\psi_2} \leqslant \|X_t - X_{\tilde{s}}\|_{\psi_2} + \|X_{\tilde{s}} - X_s\|_{\psi_2} = \|X_t - X_{\tilde{s}}\|_{\psi_2} + |\tilde{s} - s|\|X_{\tilde{s}}\|_{\psi_2}.$$

Stosując Lematy 13.8 i 13.9 dostajemy

$$\|X_t - X_s\|_{\psi_2} \leqslant CK^2(|t-\tilde{s}| + |\tilde{s} - s|).$$

By zakończyć dowód wystarczy zauważyć, że trójkąt $\triangle(t, \tilde{s}, s)$ jest rozwartokątny, zatem

$$|t-\tilde{s}| + |\tilde{s}-s| \leqslant \sqrt{2}(|t-\tilde{s}|^2 + |\tilde{s}-s|^2)^{1/2} \leqslant \sqrt{2}|t-s|.$$

□

Wniosek 13.10. *Niech A będzie macierzą losową $m \times n$ spełniającą założenia założenia Definicji 13.3. Wówczas dla $u \geqslant 1$ z prawdopodobieństwem nie mniejszym niż $1 - \exp(-u^2)$ zachodzi*

$$\sqrt{m} - CK^2(\sqrt{n} + u) \leqslant \inf_{|t|=1} |At| \leqslant \sup_{|t|=1} |At| \leqslant \sqrt{m} + CK^2(\sqrt{n} + u).$$

Dowód. Stosujemy Twierdzenie 13.5 do $T = S^{n-1}$ i zauważamy, że $\gamma(T) = \sqrt{n}$ oraz $R(T) = 1$. □

13.2 Lemat Johnsona-Lindenstraussa

Wniosek 13.11. *(Addytywny lemat Johnsona-Lindenstraussa) Niech $T \subset \mathbb{R}^n$, zaś A będzie macierzą losową $m \times n$ spełniającą założenia Definicji 13.3. Wówczas z prawdopodobieństwem 0.99 zachodzi zdarzenie*

$$|t-s| - \delta \leqslant \left| \frac{1}{\sqrt{m}}At - \frac{1}{\sqrt{m}}As \right| \leqslant |t-s| + \delta \quad \textit{dla wszystkich } s,t \in T,$$

gdzie $\delta \leqslant \frac{C}{\sqrt{m}}K^2 g(T)$.

Dowód. Wystarczy zastosować Twierdzenie 13.5 do $T - T$ otrzymując

$$\mathbf{E} \sup_{s,t \in T} \left| \left| \frac{1}{\sqrt{m}}At - \frac{1}{\sqrt{m}}As \right| - |t-s| \right| \leqslant \frac{C}{\sqrt{m}} K^2 \gamma(T-T),$$

zauważyć, że $\gamma(T-T) = 2g(T)$ i skorzystać z nierówności Czebyszewa. □

Wniosek 13.12 (Multyplikatywny Lemat Johnsona-Lindenstraussa)**.** *Załóżmy, że T jest zbiorem skończonym w $\mathbb{R}^n$, a A jest macierzą losową $m \times n$ spełniającą założenia Definicji 13.3. Wówczas z prawdopodobieństwem 0.99 zachodzi*

$$(1-\varepsilon)|t-s| \leqslant \left| \frac{1}{\sqrt{m}}At - \frac{1}{\sqrt{m}}As \right| \leqslant (1+\varepsilon)|t-s|,$$

gdzie $\varepsilon \leqslant CK^2\sqrt{\frac{\log|T|}{m}}$.

Dowód. Niech

$$S := \left\{ \frac{t-s}{|t-s|} : t,s \in T, t \neq s \right\}$$

Zauważmy, że na mocy Lematu 11.11,

$$\gamma(S) = g(S) \leqslant \sqrt{2\log|S|} \leqslant 2\sqrt{\log|T|},$$

zatem Twierdzenie 13.5 zastosowane do zbioru S implikuje

$$\mathbf{E} \sup_{t,s \in T, t \neq s} \left| \frac{|At - As|}{\sqrt{m}|t-s|} - 1 \right| = \frac{1}{\sqrt{m}} \mathbf{E} \sup_{t' \in S} ||At'| - \sqrt{m}|t'|| \leqslant CK^2 \sqrt{\frac{\log|T|}{m}}.$$

□

Uwaga 13.13. Klasyczny Lemat Johnsona Lindenstraussa miał postać jak we Wniosku 13.12, ale dotyczył przypadku, gdy zamiast $\frac{1}{\sqrt{m}}A$ bada się przekształcenia $\sqrt{\frac{n}{m}}P$, gdzie P jest rzutem ortogonalnym na losowo wybraną podprzestrzeń $E \in G_{n,m}$. To jakie się rozpatruje losowe przekształcenie jest mało istotne, kluczowe jest to, że z dużym prawdopodobieństwem jest ono (po obcięciu do zbioru skończonego) bliskie izometrii i ma wartości w przestrzeni nie za wysokiego wymiaru m, który (przy ustalonym ε) jest proporcjonalny do $\log|T|$.

13.3 Losowe przekroje

W tej częsci będziemy badać srednicę losowego przekroju podzbioru $\mathbb{R}^n$. Nasza losowość będzie wyznaczona poprzez branie podprzestrzeni $E := \mathrm{Ker}A$, gdzie A jest macierzą $m \times n$ o niezależnych subgaussowskich rzędach. Zauważmy, że $\dim(E) \geqslant n - m$ oraz jeśli A_i mają ciągłe rozkłady, to $\dim(E) = n - m$ p.n. Co więcej, jeśli współczynniki A są niezależnymi zmiennymi losowymiu $\mathcal{N}(0,1)$, to z niezmienniczości rozkładu gaussowskiego na obroty wynika, że E ma rozkład jednostajny na przestrzeni Grassmana $G_{n,n-m}$.

Twierdzenie 13.14. *Niech $T \subset \mathbb{R}^n$, zaś A będzie macierzą losową $m \times n$ spełniającą założenia Definicji 13.3. Wówczas*

$$\mathbf{E}\mathrm{diam}(T \cap \mathrm{Ker}A) \leqslant \frac{C}{\sqrt{m}} K^2 g(T).$$

Dowód. Wystarczy zauważyć, że na mocy Twierdzenia 13.5,

$$\begin{aligned}\mathbf{E}\sqrt{m}\mathrm{diam}(T \cap \mathrm{Ker}A) &= \mathbf{E} \sup_{t,s \in T \cap \mathrm{Ker}A} \big||At - As| - \sqrt{m}|t-s|\big| \leqslant \mathbf{E} \sup_{t \in T-T} \big||At'| - \sqrt{m}|t'|\big| \\ &\leqslant CK^2\gamma(T-T) = 2CK^2 g(T).\end{aligned}$$

□

Przykład. Zauważmy, że $\mathrm{diam}(B_1^n) = 2$, ale $g(B_1^n) = \mathbf{E}\max_i |g_i| \leqslant \sqrt{2\log(2n)}$, stąd

$$\mathbf{E}\mathrm{diam}(B_1^n \cap E) \leqslant C\sqrt{\frac{\log n}{m}},$$

gdzie średnia jest brana po losowej podprzestrzeni $E \in G_{n,n-m}$. Np dla $m = n/10$ otrzymujemy, że losowy przekrój B_1^n wymiaru $0,9n$ ma średnicę rzędu $\sqrt{\log n/n}$.

Twierdzenie 13.15. *Załóżmy, że $T \subset S^{n-1}$ oraz A jest macierzą losową $m \times n$ spełniającą założenia Definicji 13.3. Wówczas dla $m \geqslant CK^4\gamma(T)^2$*

$$\mathbf{P}(T \cap \ker(A) = \emptyset) \geqslant 1 - \exp(-cm/K^4).$$

Dowód. Oczywiście $R(T) = 1$, więc Twierdzenie 13.5 mówi, że z prawdopodobieństwem przynajmniej $1 - \exp(-u^2)$ mamy

$$\sup_{t \in T} \big||At| - \sqrt{m}|t|\big| \leqslant C_1 K^2(\gamma(T) + u)$$

Załóżmy, że zachodzi powyższe zadarzenie i istnieje $t \in T \cap \ker(A)$. Wówczas

$$\sqrt{m} \leqslant C_1 K^2(\gamma(T) + u).$$

Jeśli $u = \sqrt{m}/(2C_1K^2)$, to dostajemy

$$\sqrt{m} \leqslant C_1K^2(\gamma(T) + \frac{1}{2}\sqrt{m}$$

czyli $m \leqslant 4C_1^2K^4\gamma^2(T)$. □

14 Oszczędne próbkowanie

Ten wykład jest poświęcony zastosowaniom macierzy losowych w teorii oszczędnego próbkowania (ang. compressed sensing). Z konieczności omówimy skrótowo jedynie kilka wybranych zagadnień w możliwie prostych sformułowaniach. Czytelnika szukającego obszerniejszego wprowadzenia w tematykę odsyłamy do notatek z kursu w Marne-la-Vallée [2].

Będziemy zajmować się rozwiązywaniem równania postaci

$$y = Ax \quad x \in T,$$

gdzie A jest macierzą $m \times n$ (z m typowo dużo mniejszym niż n), a T jest podzbiorem $\mathbb{R}^n$. Równanie możemy zapisać jako

$$y_i = \langle A_i, x\rangle, \quad i = 1, \ldots, m,$$

gdzie $A_i \in \mathbb{R}^n$ to wiersze macierzy A.

Wielkość $\langle A_i, x\rangle$ można interpretować jako i-ty pomiar nieznanego wektora x. Przyjmuje się, że znamy zbiór T, dobieramy odpowiednią macierz A tak by odtworzyć na podstawie y wektor x.

Gdy $m < n$ równanie $Ax = y$ ma typowo nieskończenie wiele rozwiązań, będziemy więc zakładać, że zbiór T jest mały. Podobnie jak w poprzednim wykładzie będziemy rozważać macierze losowe A o niezależnych subgaussowskich rzędach.

W dalszej części będziemy używać następującej notacji. Dla $I \subset [n] := \{1, \ldots, n\}$ i $x \in \mathbb{R}^n$ określamy

$$x_I := \left(x_i \mathbb{1}_{\{i\in I\}}\right)_{i=1}^n, \quad I^c := [n] \setminus I.$$

Normę ℓ_q wektora x oznaczamy przez $\|x\|_q$, piszemy $|x|$ zamiast $\|x\|_2$. Dla $q \geqslant 1$, B_q^n oznacza kulę jednostkową w ℓ_q^n, czyli

$$B_q^n = \left\{x \in \mathbb{R}^n : \ \sum_{i=1}^n |x_i|^q \leqslant 1\right\}.$$

Zacznijmy od przypadku, gdy $w = 0$. Najprostszy algorytm prowadzący do rozwiązania naszego równania, to

$$\text{znaleźć } \hat{x} \in T \text{ takie, że } A\hat{x} = y. \tag{32}$$

Twierdzenie 14.1. *Niech A będzie macierzą losową $n \times m$ spełniającą założenia Definicji 13.3. Załóżmy, że x jest ustalonym wektorem z T, $y = Ax$ zaś $\hat{x}$ jest rozwiązaniem (32). Wówczas*

$$\mathbf{E}|\hat{x} - x| \leqslant \frac{C}{\sqrt{m}} K^2 g(T).$$

Dowód. Zauważmy, że $\hat{x} - x \in (T - x) \cap \mathrm{Ker}(A)$, stąd na mocy Twierdzenia 13.14 mamy

$$\mathbf{E}|\hat{x} - x| \leqslant \mathbf{E}\mathrm{diam}((T - x) \cap \mathrm{Ker}(A)) \leqslant \frac{C}{\sqrt{m}} K^2 g(T - x) = \frac{C}{\sqrt{m}} K^2 g(T).$$

□

Uwaga 14.2. i) Inaczej można przeformułować Twierdzenie 14.1 jako

$$\mathbf{E}|\hat{x} - x| \leqslant \varepsilon \mathrm{diam}(T),$$

o ile

$$m \geqslant CK^4 \varepsilon^{-2} d(T), \quad \text{gdzie } d(T) := \frac{g(T)^2}{\mathrm{diam}(T)^2}.$$

ii) Zagadnienie (32) jest złożone obliczeniowo, jeśli T nie jest wypukły. Zauważmy jednak, że $g(\mathrm{conv}(T)) = g(T)$, zatem możemy zawsze zastąpić zbiór T jego uwypukleniem i otrzymać zagadnienie rozsądne obliczeniowo i mające podobne oszacowanie błędu $|\hat{x} - x|$.

14.1 Wektory o małym nośniku

Typowe założenie (bardzo użyteczne w zastosowaniach) mówi, że x ma mały nośnik.

Definicja 14.3. *Nośnikiem* wektora $x \in \mathbb{R}^n$ nazywamy zbiór

$$\mathrm{supp}(x) := \{1 \leqslant i \leqslant n\colon\ x_i \neq 0\}.$$

Normę ℓ_0 wektora x definiujemy jako $\|x\|_0 := |\mathrm{supp}(x)|$. Kładziemy też dla $1 \leqslant p \leqslant n$,

$$\Sigma_p := \{x \in \mathbb{R}^n\colon\ \|x\|_0 \leqslant p\}.$$

Wiedząc, że x ma mały nośnik rozwiązania równania chciałoby się poszukiwać rozwiązując zagadnienie

$$\hat{x} = \mathrm{argmin}\{\|x'\|_0\colon\ Ax' = y\}.$$

Jednak, z uwagi na to, że ℓ_0-norma nie jest normą zagadnienie to jest złożone obliczeniowo. Dlatego wygodniej rozważać normę ℓ_1 i rozwiązywać problem

$$\hat{x} = \mathrm{argmin}\{\|x'\|_1\colon\ Ax' = y\}. \tag{33}$$

Fakt 14.4. *Dla $1 \leqslant p \leqslant n$ zachodzi*

$$\frac{1}{2}(\sqrt{p}B_1^n \cap B_2^n) \subset \mathrm{conv}(\Sigma_p \cap B_2^n) \subset \sqrt{p}B_1^n \cap B_2^n.$$

Dowód. Zauważmy, że

$$\|x\|_1 = \sum_{i \in \mathrm{supp}(x)} |x_i| \leqslant |\mathrm{supp}(x)|^{1/2} \left(\sum_{i=1}^{n} x_i^2 \right)^{1/2} = \|x\|_0^{1/2} |x|,$$

stąd $\Sigma_p \cap B_2^n \subset \sqrt{p} B_1^n$ i łatwo otrzymujemy górne zawieranie.

By udowodnić dolne zawieranie ustalmy $x \in \sqrt{p} B_1^n \cap B_2^n$. Niech I_1 oznacza podzbiór $[n] = \{1, 2, \ldots, n\}$ zawierający indeksy p największych modułów współrzędnych x, I_2 zbiór indeksów kolejnych p co do wielkości modułów współrzędnych itd. Zauważmy, że dla $k \geqslant 2$,

$$\max_{i \in I_k} |x_i| \leqslant \min_{i \in I_{k-1}} |x_i| \leqslant \frac{1}{p} \|x_{I_{k-1}}\|_1,$$

stąd

$$|x_{I_k}| \leqslant \frac{1}{\sqrt{p}} \|x_{I_{k-1}}\|_1.$$

Oczywiście $|x_{I_1}| \leqslant |x| \leqslant 1$, zatem

$$\sum_k |x_{I_k}| \leqslant 1 + \sum_{k \geqslant 2} \frac{1}{\sqrt{p}} \|x_{I_{k-1}}\|_1 \leqslant 1 + \frac{1}{\sqrt{p}} \|x\|_1 \leqslant 2.$$

Ponadto, $x_{I_k} \in \Sigma_p$, zatem

$$x = \sum_{k \geqslant 1} x_{I_k} \in \sum_k |x_{I_k}| \mathrm{conv}(\Sigma_p \cap B_2^n) \subset 2\mathrm{conv}(\Sigma_p \cap B_2^n).$$

□

Fakt 14.5. *Dla* $1 \leqslant p \leqslant n$ *mamy*

$$g(\sqrt{p} B_1^n \cap B_2^n) \leqslant 2 g(\Sigma_p \cap B_2^n) \leqslant 4 \sqrt{p \log \frac{2n}{p}}$$

Dowód. Pierwsza nierówność wynika z dolnego zawierania w Fakcie 14.4. By udowodnić drugą zauważamy wpierw, że

$$g(\Sigma_p \cap B_2^n) = \mathbf{E} \max_{|I|=p} \sqrt{\sum_{i \in I} g_i^2} = \mathbf{E} \sqrt{\sum_{i=1}^{p} (g_i^*)^2},$$

gdzie $g_1^* \geqslant g_2^* \geqslant \ldots \geqslant g_n^*$ oznacza monotoniczne uporządkowanie wektora $(|g_1|, \ldots, |g_n|)$, czyli $g_k^* = k\text{-}\max_i |g_i|$.

Wypukłość funkcji $\exp(\alpha x^2)$ oraz $\exp(x)$ implikuje dla $0 < \alpha < 1/2$,

$$\exp\left(\alpha\left(\mathbf{E}\sqrt{\frac{1}{p}\sum_{i=1}^{p}(g_i^*)^2}\right)^2\right) \leqslant \mathbf{E}\exp\left(\alpha\frac{1}{p}\sum_{i=1}^{p}(g_i^*)^2\right) \leqslant \frac{1}{p}\mathbf{E}\sum_{i=1}^{p}\exp(\alpha(g_i^*)^2)$$
$$\leqslant \frac{1}{p}\mathbf{E}\sum_{i=1}^{n}\exp(\alpha g_i^2) = \frac{n}{p\sqrt{1-2\alpha}}.$$

Biorąc $\alpha = 1/4$ dostajemy $g(\Sigma_p \cap B_2^n) \leqslant 2\sqrt{p\log(2n/p)}$. □

Uwaga 14.6. Można udowodnić, że

$$g(\sqrt{p}B_1^n \cap B_2^n) \geqslant g(\Sigma_p \cap B_2^n) \geqslant c\sqrt{p\log\frac{2n}{p}}.$$

Twierdzenie 14.1 (zastosowane do $x/|x|$ i $T = \Sigma_p \cap B_2^n$ i Fakt 14.5 implikują następujący wniosek.

Wniosek 14.7. *Niech A będzie macierzą losową $m \times n$ spełniającą założenia założenia Definicji 13.3. Niech $x \in \Sigma_p$, $Ax = y$, zaś $\hat{x}$ będzie rozwiązaniem* (33). *Wówczas*

$$\mathbf{E}|\hat{x} - x| \leqslant CK^2\sqrt{\frac{p\log(2n/p)}{m}}|x|.$$

14.2 Dokładna rekonstrukcja wektorów o małym nośniku

Dla macierzy $A \in M_{m\times n}$ i $1 \leqslant p \leqslant n$ określmy

$$\alpha_p(A) := \inf_{x\in\Sigma_p\cap B_2^n}|Ax|, \quad \beta_p(A) := \sup_{x\in\Sigma_p\cap B_2^n}|Ax|, \quad \gamma_p(A) := \frac{\beta_p(A)}{\alpha_p(A)}.$$

Innymi słowy $\alpha_p(A)$ i $\beta_p(A)$ to optymalne stałe takie, że

$$\alpha_p(A)|x| \leqslant |Ax| \leqslant \beta_p(A)|x| \quad \text{dla } x \in \Sigma_p.$$

Definicja 14.8. Powiemy, że macierz A ma *własność dokładnej ℓ_1-rekonstrukcji rzędu p*, jeśli dla dowolnego wektora $x \in \Sigma_p$ rozwiązaniem zagadnienia (33) z $y = Ax$ jest wektor $\hat{x} = x$.

Twierdzenie 14.9. *Załóżmy, że $r > p(1 + \gamma_r^2(A))$. Wówczas A ma własność dokładnej ℓ_1-rekonstrukcji rzędu p.*

Dowód. Niech $s := r - p > \gamma_r^2(A)p$. Dla uproszczenia notacji położymy $\alpha = \alpha_r(A)$, $\beta = \beta_r(A)$. Ustalmy $x \in \Sigma_p$, niech $y = Ax$, $\hat{x}$ będzie rozwiązaniem (33) oraz $h = \hat{x} - x$. Musimy pokazać, że $h = 0$.

Niech π będzie permutacją $[n]$ taką, że

$$|h_{\pi(1)}| \geqslant |h_{\pi(2)}| \geqslant \ldots \geqslant |h_{\pi(n)}|.$$

Innymi słowy $|h_{\pi(l)}| = l\text{-}\max_i |h_i|$. Niech

$$I_0 := \pi(\{1, \ldots, p\}), \quad I_l := \pi(\{p + (l-1)s + 1, p + (l-1)s + 2, \ldots, p + ls\} \cap [n]) \text{ dla } l \geqslant 1,$$

tzn. I_0 zawiera indeksy p największych modułów współrzędnych wektora h, I_1 indeksy kolejnych s co do wielkości modułów współrzędnych itd.

Niech $S := \mathrm{supp}(x)$, wówczas

$$\begin{aligned} \|x\|_1 &\geqslant \|\hat{x}\|_1 = \|x + h\|_1 = \|x_S + h_S\|_1 + \|h_{S^c}\|_1 \geqslant \|x\|_1 - \|h_S\|_1 + \|h_{S^c}\|_1 \\ &\geqslant \|x\|_1 - \|h_{I_0}\|_1 + \|h_{I_0^c}\|_1, \end{aligned}$$

gdzie ostatnia nierówność wynika z tego, że $|S| \leqslant p = |I_0|$ oraz ze sposobu wyboru I_0. Zatem

$$\|h_{I_0^c}\|_1 \leqslant \|h_{I_0}\|_1 \leqslant \sqrt{p}|h_{I_0}|.$$

Następnie zauważmy, że

$$0 = |A\hat{x} - Ax| = |Ah| = \left| A(h_{I_0} + h_{I_1}) + \sum_{l \geqslant 2} A_{h_{I_l}} \right|,$$

zatem wykorzystując definicje $\alpha = \alpha_r(A)$ i $\beta = \beta_r(A)$ i to, że $|I_0 \cup I_1| = r$ oraz $|I_l| \leqslant s \leqslant r$ dostajemy

$$\alpha\,|h_{I_0} + h_{I_1}| \leqslant |A(h_{I_0} + h_{I_1})| \leqslant \sum_{l \geqslant 2} \left| A_{h_{I_l}} \right| \leqslant \sum_{l \geqslant 2} \beta\,|h_{I_l}|\,.$$

Zauważmy, że dla $l \geqslant 2$,

$$|h_{I_l}| \leqslant \sqrt{s} \max_{i \in I_l} |h_i| \leqslant \sqrt{s} \min_{i \in I_{l-1}} |h_i| \leqslant \frac{1}{\sqrt{s}} \|h_{I_{l-1}}\|_1.$$

Stąd

$$\alpha\,|h_{I_0} + h_{I_1}| \leqslant \sum_{l \geqslant 2} \frac{\beta}{\sqrt{s}} \|h_{I_{l-1}}\|_1 = \frac{\beta}{\sqrt{s}} \|h_{I_0^c}\|_1 \leqslant \frac{\beta\sqrt{p}}{\sqrt{s}} |h_{I_0}| \leqslant \frac{\beta\sqrt{p}}{\sqrt{s}}\,|h_{I_0} + h_{I_1}|\,.$$

Ale z naszych założeń wynika, że $\beta\sqrt{p/s} < \alpha$, zatem $h_{I_0} + h_{I_1} = 0$, stąd łatwo otrzymujemy, że $h = 0$. □

Niestety nie są znane deterministyczne macierze losowe takie, że γ_p jest ograniczone oraz m jest porównywalne z p z dokładnością do logarytmów. Okazuje się jednak, że subgaussowska macierz losowa z dużym prawdopodobieństwem spełnia takie warunki.

Twierdzenie 14.10. *Niech A będzie macierzą losową $m \times n$ spełniającą założenia założenia Definicji 13.3 oraz $m \geqslant CK^4 p \log(en/p)$. Wówczas*

$$\mathbf{P}\left(\frac{9}{10}\sqrt{m} \leqslant \alpha_p(A) \leqslant \beta_p(A) \leqslant \frac{11}{10}\sqrt{m}\right) \geqslant 1 - \exp\left(-c\frac{m}{K^4}\right).$$

Dowód. Niech I będzie podzbiorem $[n] = \{1, \ldots, n\}$ mocy p zaś A_I oznacza macierz $m \times p$ powstałą z A przez wybranie kolumn o indeksach z I. Łatwo sprawdzić, że macierz A_I spełnia założenia Definicji 13.3 (z p zamiast n). Wniosek 13.10 implikuje, że z prawdopodobieństwem większym niż $1 - \exp(-u^2)$ zachodzi

$$\sqrt{m} - C_1 K^2(\sqrt{p} + u) \leqslant \inf_{|x|=1} |A_I x| \leqslant \sup_{|x|=1} |A_I x| \leqslant \sqrt{m} + C_1 K^2(\sqrt{p} + u).$$

Zauważmy, że jeśli $\mathrm{supp}(x) \subset I$, to x możemy traktować jako wektor z $\mathbb{R}^I \cong \mathbb{R}^p$ i przy tym utożsamieniu $A_I x = Ax$. Zatem z prawdopodobieństwem większym niż $1 - \binom{n}{p}\exp(-u^2)$ mamy

$$\sqrt{m} - C_1 K^2(\sqrt{p} + u) \leqslant \inf_{x \in \Sigma_p \cap B_2^n} |A_I x| \leqslant \sup_{x \in \Sigma_p \cap B_2^n} |A_I x| \leqslant \sqrt{m} + C_1 K^2(\sqrt{p} + u).$$

Wystarczy teraz zauważyć, że jeśli $m \geqslant 400 C_1^2 K^4 p$ i $u = \sqrt{m}/(20 C_1 K^2)$ to $C_1 K^2(\sqrt{p} + u) \leqslant \sqrt{m}/10$, ponadto wtedy

$$\binom{n}{p}\exp(-u^2) \leqslant \exp\left(p \log(en/p) - \frac{m}{400 C_1^2 K^4}\right) \leqslant \exp\left(-\frac{m}{800 C_1^2 K^4}\right),$$

o ile $m \geqslant 800 C_1^2 K^4 p \log(en/p)$. □

Wniosek 14.11. *Niech A będzie macierzą losową $m \times n$ spełniającą założenia założenia Definicji 13.3 oraz $m \geqslant CK^4 p \log(en/p)$. Wówczas z prawdopodobieństwem przynajmniej $1 - \exp(-cm/K^4)$ macierz A ma własność dokładnej ℓ_1-rekonstrukcji rzędu p.*

Dowód. Stosujemy Twierdzenie 14.10 z $3p$ zamiast p, a następnie Twierdzenie 14.9 z $r = 3p > (1 + (11/9)^2)p$. □

14.3 Algorytm Lasso

W praktyce każdy wykonywany pomiar jest obarczony pewnym błędem. Dlatego naturalnie jest rozpatrywać ogólniejsze równanie postaci

$$y = Ax + w \quad x \in T,$$

bądź równoważnie

$$y_i = \langle A_i, x\rangle + w_i, \quad i = 1, \ldots, m,$$

gdzie $w_i \in \mathbb{R}$ można interpretować jako błąd i-tego pomiaru.

Tutaj musimy zmodyfikować algorytm (33). Jedną z możliwości jest tzw. *algorytm Lasso* postaci

$$\hat{x} = \operatorname{argmin}\{|Ax' - y|\colon\ \|x'\|_1 \leqslant R\}. \tag{34}$$

Twierdzenie 14.12. *Niech A będzie macierzą losową $m\times n$ spełniającą założenia założenia Definicji 13.3, $x \in \Sigma_p$ dla $1 \leqslant p \leqslant n$ oraz $m \geqslant CK^4 p\log(en/p)$. Wówczas rozwiązanie* (34) *z $R = \|x\|_1$ spełnia*

$$\mathbf{P}\left(|\hat{x} - x| \leqslant CK\frac{|w|\sqrt{p\log(en/p)}}{m}\right) \geqslant 1 - 2\exp(-p\log(en/p)).$$

Niech $h = \hat{x} - x$. Dowód twierdzenia rozbijemy na kilka lematów.

Lemat 14.13. *Zachodzi*

$$\|h\|_1 \leqslant 2\sqrt{p}|h|$$

oraz

$$|Ah|^2 \leqslant 2\langle Ah, w\rangle.$$

Dowód. Niech $S = \operatorname{supp}(x)$. Mamy

$$\|x\|_1 \geqslant \|\hat{x}\|_1 = \|x + h\|_1 = \|x_S + h_S\|_1 + \|h_{S^c}\|_1 \geqslant \|x\|_1 - \|h_S\|_1 + \|h_{S^c}\|_1,$$

zatem

$$\|h\|_1 = \|h_S\|_1 + \|h_{S^c}\|_1 \leqslant 2\|h_S\|_1 \leqslant 2|S|^{1/2}|h_S| \leqslant 2\sqrt{p}|h|.$$

By udowodnić drugie oszacowanie zauważamy, że

$$|w - Ah| = |w + Ax - A\hat{x}| = |y - A\hat{x}| \leqslant |y - Ax| = |w|.$$

Wystarczy teraz podnieść obie strony do kwadratu i zredukować czynnik $|w|^2$. $\square$

Lemat 14.14. *Przy założeniach Twierdzenia 14.12,*

$$\mathbf{P}\left(|Ah|^2 \geqslant \frac{m}{4}|h|^2\right) \geqslant 1 - \exp(-p\log(en/p)).$$

Dowód. Na mocy pierwszego oszacowania z Lematu 14.13,

$$\frac{h}{|h|} \in T_p := 2\sqrt{p}B_1^n \cap S^{n-1}.$$

Fakt 14.5 implikuje

$$R(T_p) = 1, \quad \gamma(T_p) \leqslant 2g(\sqrt{p}B_1^n \cap B_2^n) \leqslant 4\sqrt{p\log(2n/p)}.$$

Stosujemy drugą część Twierdzenia 13.5 z $u = \sqrt{p\log(en/p)}$ i dostajemy, że z prawdopodobieństwem przynajmniej $1 - \exp(-p\log(en/p))$ zachodzi

$$\sup_{t\in T_p} ||At| - \sqrt{m}| \leqslant C_1K^2\sqrt{p\log(en/p)} \leqslant \frac{1}{2}\sqrt{m},$$

gdzie ostatnia nierówność zachodzi o ile $m \geqslant 4C_1^2K^4p\log(en/p)$. Zatem

$$\mathbf{P}\left(|Ah|^2 \geqslant \frac{m}{4}|h|^2\right) \geqslant \mathbf{P}\left(|At| \geqslant \frac{1}{2}\sqrt{m} \text{ dla } t \in T_p\right) \geqslant 1 - \exp(-p\log(en/p)).$$

□

Lemat 14.15. *Przy założeniach Definicji 13.3,*

$$\|\langle At, w\rangle - \langle As, w\rangle\|_{\psi_2} \leqslant CK|w||t-s| \qquad \textit{dla } t, s \in \mathbb{R}^n.$$

Dowód. Korzystając z Faktu 12.15 mamy dla $\lambda \in \mathbb{R}$

$$\begin{aligned}\mathbf{E}\exp(\lambda(\langle At, w\rangle - \langle As, w\rangle)) = \mathbf{E}\exp\left(\sum_{i=1}^m \lambda w_i\langle A_i, t-s\rangle\right) = \prod_{i=1}^m \mathbf{E}\exp(\lambda w_i\langle A_i, t-s\rangle)\\ \leqslant \prod_{i=1} m\exp\left(C\lambda^2 w_i^2\|\langle A_i, t-s\rangle\|_{\psi_2}^2\right) \leqslant \exp\left(C\lambda^2K^2|w|^2|t-s|^2\right)\end{aligned}$$

i teza wynika z ponownego użycia Faktu 12.15. □

Lemat 14.16. *Przy założeniach Twierdzenia 14.12,*

$$\mathbf{P}\left(\langle Ah, w\rangle \leqslant CK|h||w|\sqrt{p\log(en/p)}\right) \geqslant 1 - \exp(-4p\log(en/p)).$$

Dowód. Niech $X_t := \langle At, w\rangle$ oraz T_p będzie jak w dowodzie poprzedniego Lematu. Wówczas

$$\frac{1}{|h|}\langle Ah, w\rangle \leqslant \sup_{t\in T_p} X_t.$$

Na mocy Lematu 14.15 mamy

$$\|X_t - X_s\|_{\psi_2} \leqslant d(t,s) := C_1K|w||t-s| = \|G_t - G_s\|_2,$$

gdzie

$$G_t := C_1K|w|\sum_{i=1}^n t_ig_i$$

Zauważmy, że

$$\operatorname{diam}(T_p, d) \leqslant 2C_1 K|w|$$

oraz na podstawie Twierdzenia 12.10 i Faktu 14.5,

$$\gamma_2(T_p, d) \leqslant C\mathbf{E}\sup_{t\in T} G_t = CK|w|g(T_p) \leqslant CK|w|\sqrt{p\log(en/p)}.$$

Stąd Wniosek 12.25 implikuje dla $u > 0$,

$$\mathbf{P}\left(\langle Ah, w\rangle \geqslant CK|h||w|\left(\sqrt{p\log(en/p)} + u\right)\right) \leqslant \mathbf{P}\left(\sup_{t\in T_p} X_t \geqslant C(\gamma_2(T,d) + u\operatorname{diam}(T,d))\right)$$
$$\leqslant \exp(-u^2)$$

i wystarczy wziąć $u = \sqrt{p\log(en/p)}$. □

Dowód Twierdzenia 14.12. Na podstawie Lematów 14.13, 14.14 i 14.16 z prawdopodobieństwem $1 - 2\exp(-p\log(en/p))$ zachodzi zdarzenie

$$\frac{m}{4}|h|^2 \leqslant |Ah|^2 \leqslant 2\langle Ah, w\rangle \leqslant CK|h||w|\sqrt{p\log(en/p)},$$

które implikuje

$$|h| \leqslant CK\frac{|w|\sqrt{p\log(en/p)}}{m}.$$

□

15 Twierdzenie Dvoretzky'ego

15.1 Prawie izometryczne włożenia przestrzeni

W czasie ostatniego wykładu pokażemy jak można wykorzystać koncentrację dla miar gaussowskich 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 15.1. Powiemy, że przestrzeń ℓ_2^k się wkłada ze stałą A w przestrzeń unormowaną $(E, \|\ \|)$, jeśli istnieją wektory $v_1, \ldots, v_k \in E$ takie, że

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

Uwaga 15.2. Ostatnią definicję można sformułować równoważnie (biorąc $u_i = v_i/\sqrt{A}$) jako istnienie $u_1, \ldots, u_k \in E$ takich, że

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

15.2 Twierdzenie Dvoretzky'ego w wersji Milmana

Zanim sformułujemy twierdzenie będziemy potrzebowali ważnej definicji.

Definicja 15.3. Dla wektora gaussowskiego w przestrzeni unormowanej $(E, \| \ \|)$ postaci $X = \sum_{i=1}^{m} v_i g_i$, gdzie $g_1, \ldots, g_n$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$ określmy

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

Dla przestrzeni unormowanej $(E, \| \ \|)$ określamy

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

Fakt 15.4. *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

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

□

Przykład. Biorąc $X = \sum_{i=1}^{n} e_i g_i$ widzimy, że $\mathbf{E}\|X\|_1 = n\mathbf{E}|g_1| = n(2/\pi)^{1/2}$ oraz $\sigma(X, \| \ \|_1) = \sup_{|t|=1} \|t\|_1 = \sqrt{n}$, zatem $d(\ell_1^n) \geqslant d(X) = \frac{2}{\pi} n$.

Ogólniej, rozważając taki sam X, ale w przestrzeni ℓ_p^n widzimy, że dla $1 \leqslant p \leqslant 2$, $\sigma(X, \| \ \|_p) = n^{1/p-1/2}$ oraz, na mocy Wniosku 6.24, $\mathbf{E}\|X\|_p \geqslant c(\mathbf{E}\|X\|_p^p)^{1/p} \geqslant c' n^{1/p}$. Wykazaliśmy zatem, że

$$d(\ell_p^n) \geqslant d(X) \geqslant cn \quad \text{dla } 1 \leqslant p \leqslant 2.$$

Dla $2 < p < \infty$ mamy $\sigma(X, \| \ \|_p) = 1$. By oszacować $\mathbf{E}\|X\|_p$ zauważmy, ponownie korzystając z Wniosku 6.24, że

$$\mathbf{E}\|X\| + C\sqrt{p} \geqslant (\mathbf{E}\|X\|_p^p)^{1/p} = n^{1/p}(\mathbf{E}|g_1|^p)^{1/p} \geqslant c_1\sqrt{p}n^{1/p},$$

czyli dla $\log n \geqslant C'p$ mamy $d(X) = (\mathbf{E}\|X\|_p)^2 \geqslant \frac{1}{2}c_1^2 p n^{2/p}$. Ponadto dla dowolnego $p \geqslant 2$, $\mathbf{E}\|X\|_p \geqslant \mathbf{E}\|X\|_\infty \geqslant c\log^{1/2}(n+1)$. W ten sposób pokazaliśmy, że

$$d(\ell_p^n) \geqslant d(X) \geqslant \begin{cases} cpn^{2/p} & \text{dla } 2 \leqslant p \leqslant \log(n+1) \\ c\log(n+1) & \text{dla } p \geqslant \log(n+1). \end{cases}$$

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

Twierdzenie 15.5. *Dla dowolnego $\varepsilon > 0$ istnieje stała $c(\varepsilon) > 0$ taka, że dla dowolnej przestrzeni unormowanej E oraz $k \leqslant c(\varepsilon)d(E)$, przestrzeń ℓ_2^k wkłada się ze stałą $1 + \varepsilon$ w E.*

Zanim przejdziemy do dowodu tego twierdzenia przypomnimy definicję ε-sieci i udowodnimy kilka lematów.

Definicja 15.6. Zbiór $A \subset X$ nazywamy ε-siecią przestrzeni metrycznej (X, d), jeśli dla dowolnego $x \in X$ istnieje $y \in A$ takie, że $d(x, y) \leqslant \varepsilon$.

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

Lemat 15.7. *Dla $0 < \varepsilon \leqslant 1$ istnieje ε-sieć N w S^{n-1} taka, że $|N| \leqslant (\frac{3}{\varepsilon})^n$.*

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

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

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

Lemat 15.8. *Niech N będzie ε-siecią w S^{n-1} dla pewnego $\varepsilon \in (0,1)$. Wówczas dla dowolnych wektorów $v_1, \ldots, v_n$ w E,*

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

oraz

$$\inf_{|t|=1}\left\|\sum_{i=1}^n v_it_i\right\| \geqslant \inf_{s\in N}\left\|\sum_{i=1}^n v_is_i\right\| - \frac{\varepsilon}{1-\varepsilon}\sup_{s\in N}\left\|\sum_{i=1}^n v_is_i\right\|.$$

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

$$\left|\left\|\sum_{i=1}^n v_it_i\right\| - \left\|\sum_{i=1}^n v_is_i\right\|\right| \leqslant a\left\|\sum_{i=1}^n v_iu_i\right\| \leqslant \varepsilon \sup_{u\in S^{n-1}}\left\|\sum_{i=1}^n v_iu_i\right\|$$

i łatwo uzyskujemy szacowania z tezy lematu. □

Dowód Twierdzenia 15.5. Bez straty ogólności możemy przyjąć, że $\varepsilon < 1$, dobierzmy $\delta > 0$ takie, że $(1+\delta)/(1-\delta) \leqslant \sqrt{1+\varepsilon}$ oraz $1-\delta-\delta(1+\delta)/(1-\delta) \geqslant 1/\sqrt{1+\varepsilon}$ (np. można przyjąć $\delta = \varepsilon/20$). Niech N będzie δ-siecią w S^{k-1} mocy nie większej niż $(\frac{3}{\delta})^k$. Na podstawie Lematu 15.8 i Uwagi 15.2 wystarczy wykazać, że istnieją wektory $u_1,\ldots,u_k$ w E takie, że

$$1-\delta \leqslant \left\|\sum_{i=1}^k t_iu_i\right\| \leqslant 1+\delta \quad \text{dla } t \in N.$$

Niech $X = \sum_{i=1}^m v_ig_i$ spełnia $d(X) \geqslant d(E)/2$, dla uproszczenia notacji przyjmijmy $d = d(E)$, $\sigma = \sigma(X)$ oraz $M = \mathbf{E}\|X\|$. Niech $F\colon \mathbb{R}^m \to \mathbb{R}$ będzie dane wzorem $F(x) = \|\sum_{i=1}^m v_ix_i\|$. Wówczas na mocy Faktu 15.4 funkcja F jest σ-Lipschitzowska, ponadto $X \sim F(G)$, gdzie $G \sim \gamma_m$. Stąd

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

gdzie ostatnia nierówność jest konsekwencją Twierdzeń 6.6 i 6.8. Niech $X_1,\ldots,X_k$ będą niezależnymi kopiami zmiennej X. Wówczas, dla dowolnego $t \in S^{n-1}$, zmienne $\sum_{i=1}^m t_iX_i$ ma ten sam rozkład co X. Zatem

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

czyli

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

jeśli tylko $1 \leqslant k \leqslant c(\delta)d$, czyli można za u_i przyjąć $X_i(\omega)/M$ dla pewnego ω. □

Uwaga 15.9. Prześledzenie powyższego dowodu pokazuje, że Twierdzenia 15.5 zachodzi z $c(\varepsilon) = c\varepsilon^2\ln^{-1}(1/\varepsilon)$.

15.3 Oszacowania wymiaru euklidesowych przekrojów

By móc stosować Twierdzenie 15.5 potrzebujemy oszacowań z dołu $d(E)$. Zacznijmy od geometrycznego lematu.

Lemat 15.10 (Dvoretzky-Rogers)**.** *Niech E będzie n-wymiarową przestrzenią unormowaną. Wówczas istnieją wektory $v_1, \ldots, v_n \in E$ takie, że $\|\sum_{i=1}^n t_i v_i\| \leqslant |t|$ dla $t \in \mathbb{R}^n$ oraz $\|v_i\| \geqslant \frac{n-i+1}{n}$.*

Dowód. Przypomnijmy, że normę operatora $T\colon \ell_2^n \to E$ określamy wzorem

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

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

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

Wówczas dla dowolnego $\varepsilon > 0$,

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

Mamy jednak przy $\varepsilon \to 0+$,

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

więc

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

Wykazaliśmy w ten sposób, że $\mathrm{tr}(T^{-1}S) \leqslant n\|S\|$ dla dowolnego $S\colon \ell_2^n \to E$. Niech $S = TP$, gdzie P oznacza rzut ortogonalny na podprzestrzeń V. Wówczas

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

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

Lemat 15.11. *Niech ε_i będą niezależnymi symetrycznymi zmiennymi losowymi przyjmującymi wartości ± 1. Wówczas dla dowolnych wektorów $u_1, \ldots, u_n \in E$ zachodzi*

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

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

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

□

Wniosek 15.12. *Dla dowolnej przestrzeni unormowanej E wymiaru n, $d(E) \geqslant \frac{1}{C}\log(n+1)$.*

Dowód. Niech v_i będą takie jak w Lemacie 15.10 oraz $X = \sum_{i=1}^{n} v_i g_i$. Wówczas $\sigma(X) = 1$. Niech (ε_i) będzie ciągiem niezależnych symetrycznymi zmiennych losowych przyjmujących wartości ± 1, niezależnym od (g_i). Wówczas na mocy symetrii g_i,

$$\mathbf{E}\|X\| = \mathbf{E}\left\|\sum_{i=1}^{n} \varepsilon_i v_i g_i\right\| \geqslant \mathbf{E}\max_{i\leqslant n}\|v_i g_i\| \geqslant \frac{1}{2}\mathbf{E}\max_{i\leqslant \lceil n/2\rceil}|g_i| \geqslant \frac{1}{C}\sqrt{\log(n+1)}.$$

□

Definicja 15.13. Dla przestrzeni unormowanej E i $\varepsilon > 0$ określmy

$$d_\varepsilon(E) := \sup\{k\colon\ \ell_2^k \text{ wkłada się ze stałą } 1+\varepsilon \text{ w } E\}.$$

Uwaga 15.14. Twierdzenie 15.5 mówi, że $d_\varepsilon(E) \geqslant c(\varepsilon)d(E)$, nietrudno udowodnić, że $d(E) \geqslant c'(\varepsilon)d_\varepsilon(E)$.

Poniższy fakt pokazuje w szczególności, że logarytmicznej zależności od wymiaru we Wniosku 15.12 nie można poprawić.

Fakt 15.15. *Mamy $d_\varepsilon(\ell_p^n) \leqslant C(1+\varepsilon)^2 p n^{2/p}$ dla $2 \leqslant p \leqslant \log(n+1)$ oraz $d_\varepsilon(\ell_p^n) \leqslant C(1+\varepsilon)^2\log(n+1)$ dla $p \geqslant \log(n+1)$.*

Dowód. Załóżmy napierw, że $p < \infty$ oraz wektory $u_i = (u_i(j))_{j\leqslant n} \in \mathbb{R}^n$, $1 \leqslant i \leqslant k$ spełniają

$$|t| \leqslant \left\|\sum_{i=1}^{k} t_i u_i\right\|_p \leqslant (1+\varepsilon)|t| \quad \text{dla } t \in \mathbb{R}^k.$$

Niech $X = \sum_{i=1}^{k} u_i g_i$, wówczas

$$k = \mathbf{E}|(g_1,\ldots,g_k)|^2 \leqslant \mathbf{E}\|X\|_p^2 \leqslant (\mathbf{E}\|X\|_p^p)^{2/p} = \left(\sum_{j=1}^{n}\mathbf{E}\left|\sum_{i=1}^{k} u_i(j) g_i\right|^p\right)^{2/p}$$
$$= \left(\sum_{j=1}^{n}\mathbf{E}|g_1|^p\left(\sum_{i=1}^{k}|u_i(j)|^2\right)^{p/2}\right)^{2/p} \leqslant Cp\left(\sum_{j=1}^{n}\left(\sum_{i=1}^{k}|u_i(j)|^2\right)^{p/2}\right)^{2/p}.$$

Zauważmy jednak, że dla $1 \leqslant j \leqslant n$,

$$\left(\sum_{i=1}^{k} |u_i(j)|^2\right)^{1/2} = \sup_{|t|=1} \left|\sum_{i=1}^{k} t_i u_i(j)\right| \leqslant \sup_{|t|=1} \left\|\sum_{i=1}^{k} t_i u_i\right\|_p \leqslant (1+\varepsilon).$$

Stąd otrzymujemy $d_\varepsilon(\ell_p^n) \leqslant C(1+\varepsilon)^2 p n^{2/p}$ dla $2 \leqslant p < \infty$. By otrzymać postulowane oszacowanie dla $\log(n+1) \leqslant p \leqslant \infty$ wystarczy zauważyć, że dla $q := \max(2, \log(n+1)) \leqslant p$ zachodzi $\|x\|_p \leqslant \|x\|_q \leqslant e\|x\|_p$. $\square$

Fakt 15.15 oraz Przykład po Fakcie 15.4 pokazują, że

$$d_\varepsilon(\ell_p^n) \sim_\varepsilon d(\ell_p^n) \sim \begin{cases} n & \text{dla } 1 \leqslant p \leqslant 2 \\ p n^{2/p} & \text{dla } 2 \leqslant p \leqslant \log(n+1) \\ \log(n+1) & \text{dla } p \geqslant \log(n+1). \end{cases}$$

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

Twierdzenie 15.16. *Istnieje stała $c(\varepsilon)$ taka, że dla dowolnego n-wymiarowego symetrycznego ciała wypukłego i $k \leqslant c(\varepsilon)\log n$ istnieje k wymiarowa podprzestrzeń V i elipsoida $\mathcal{E}$ w V taka, że*

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

Operacją dualną do włożenia jest rzutowanie. Stąd, standardowy argument pokazuje, że twierdzenie Dvoretzky'ego ma dualną wersję mówiącą, że ciała wypukłe mają rzuty wysokiego wymiaru bliskie elipsoidom.

Wniosek 15.17. *Istnieje stała $c(\varepsilon)$ taka, że dla dowolnego n-wymiarowego symetrycznego ciała wypukłego i $k \leqslant c(\varepsilon)\log n$ istnieje rzut P_V na k-wymiarową podprzestrzeń V i elipsoida $\mathcal{E}$ w V taka, że*

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

Literatura

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

[2] D. Chafaï, O. Guédon, G. Lecué, A. Pajor, *Interactions between compressed sensing random matrices and high dimensional geometry,* Société Mathématique de France, Paris, 2012.

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

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

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

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

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

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