Wysokowymiarowy Rachunek Prawdopodobieństwa

Rafał Latała

16 października 2018

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

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

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 [4] i monografii Romana Vershynina [5], 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 (X,d) będzie przestrzenią metryczną, zaś A dowolnym podzbiorem X. Dla $t>0$ określamy *t-otoczenie* zbioru A wzorem

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

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

Definicja 2.2. Niech μ będzie miarą probabilistyczną na (X,d). *Funkcją koncentracji* miary μ definiujemy jako

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

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

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

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

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

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

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

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

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

Zauważmy też, że

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

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

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

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

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

to

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

W szczególności

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

Zauważmy, że powyższe oszacowania nie zależą od wymiaru przestrzeni.

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

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

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

W szczególności

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

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

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

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

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

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 (X,d) oraz $f\colon X \to \mathbb{R}$. *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 $F\colon X \to \mathbb{R}$,*

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

oraz

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

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

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

Dodając powyższą nierówność do poprzedniej otrzymamy ostatnią część faktu. □

Przypomnijmy definicję funkcji lipschitzowskiej

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

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

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

Analogicznie można zdefiniować funkcje lipschitzowskie między przestrzeniami metrycznymi.

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

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

oraz

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

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

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

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

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

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

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

□

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

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

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

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

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

W szczególności

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

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

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

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

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

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

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

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

3 Nierówności izoperymetryczne

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

3.1 Klasyczna izoperymetria

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

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

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

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

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

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 przyjąc, 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 $x \in \mathbb{R}$

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

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

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

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 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*

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

gdzie

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

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

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

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

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

zatem

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

□

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 $g_{k,n}(x) = c_{k,n}^{-1}\tilde{g}_{n,k}(x)$, gdzie $\tilde{g}_{n,k} = (\frac{n-|x|^2}{n})^{(n-k-2)/2}\mathbb{1}_{\{|x|\leqslant\sqrt{n}\}}$ 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 [1]). □

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

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 metrycznych X i Y. Powiemy, że funkcja borelowska $T\colon X \to Y$ *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 $A \in \mathcal{B}(Y)$.

Szczególnie wygodny jest transport lipschitzowski.

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

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

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 na ćwiczeniach.

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,*

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

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

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

to dla $t \geqslant 0$

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

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

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

4.2 Nierówność Azumy

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

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

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

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

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

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

Liczymy

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

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

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

Teza twierdzenia wynika z Faktu 4.2. $\square$

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

4.3 Zastosowania nierówności Azumy

Wniosek 4.5. *Niech (X_i, d_i) będą przetrzeniami metrycznymi, $X = X_1 \times \cdots \times X_n$ z odległością l_1, to znaczy $d(x,y) = \sum_{i=1}^n d_i(x_i, y_i)$ dla $x, y \in X$ oraz niech $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ będzie produktem miar probabilistycznych μ_i na X_i. Wówczas dla dowolnej funkcji 1-lipschitzowskiej F na X*

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

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

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

Dowód. Na mocy Faktu 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

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

stąd

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

i teza wynika z Twierdzenia 4.3. □

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

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

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

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

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

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

w szczególności

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

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

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

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

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

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

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

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

5 Nierówność Poincaré

5.1 Definicja i podstawowe własności

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

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

gdzie

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

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

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

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

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

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

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

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

Fakt 5.5. *Nierówność Poincaré jest równoważna nierówności*

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

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

Dowód. Ponieważ

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

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

By udowodnić przeciwne oszacowanie zauważmy, że

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

Stąd

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

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

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

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

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

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

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

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

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

Iterując tę nierówność n razy dostajemy

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

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

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

Zauważmy, że

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

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

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

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

5.3 Tensoryzacja

Fakt 5.9. *Załóżmy, że μ_i są miarami probabilistycznymi na X_i, $X=X_1\times\ldots\times X_n$ oraz $\mu=\mu_1\otimes\mu_2\otimes\cdots\otimes\mu_n$. Wówczas dla dowolnej funkcji $f\in L^2(X,\mu)$*

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

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

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

gdzie ostatnia nierówność wynika np. z nierówności Jensena. □

Literatura

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

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

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

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

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