Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki

Maciej Bartczak
Nr albumu: 384493

Majoryzacja Schura dla entropii sum niezależnych zmiennych wykładniczych

Praca magisterska
na kierunku MATEMATYKA

Praca wykonana pod kierunkiem
prof. dr. hab. Rafała Latały
Instytut Matematyki

Warszawa, Lipiec 2021

Streszczenie

Praca podejmuje problematykę porównywania entropii kombinacji liniowych niezależnych zmiennych losowych o identycznym rozkładzie. Głównym wynikiem pracy jest nowe twierdzenie mówiące, że w przypadku kombinacji zmiennych wykładniczych z nieujemnymi współczynnikami entropia jest monotoniczna względem porządku Schura dla kwadratów współczynników. Dowód tego faktu przeprowadzony jest z wykorzystaniem tzw. *metody przecinających się gęstości*. W pracy zawarte są również definicje i klasyczne wyniki związane z nierównościami entropijnymi, jak i sugestie odnośnie dalszych możliwych badań w tej dziedzinie.

Słowa kluczowe

entropia, wykładnicze zmienne losowe, majoryzacja Schura

Dziedzina pracy (kody wg programu Socrates-Erasmus)

11.1 Matematyka

Klasyfikacja tematyczna

Primary:

60	Probability theory and stochastic processes
60E	Distribution theory
60E15	Inequalities; stochastic orderings

Secondary:

94	Information and communication theory, circuits
94A	Communication, information
94A17	Measures of information, entropy

Tytuł pracy w języku angielskim

Schur majorization for entropy of sums of exponential random variables

Spis treści

Wprowadzenie

Gaussowskie zmienne losowe często pojawiają się w analizie zagadnień z dziedziny rachunku prawdopodobieństwa jako obiekt graniczny czy też maksymalny. Dla przykładu wystarczy przywołać Centralne Twierdzenie Graniczne oraz fakt że maksymalizują one entropię, tj. wartość $h(X) = -\mathbb{E} \ln f_X(X) = -\int_{\mathbb{R}} f_X(x) \ln f_X(x) dx$ (gdzie f_X to gęstość zmiennej X) w klasie rzeczywistych zmiennych losowych o ustalonej wariancji.

Wobec tych dwóch faktów nasuwa się następujące zagadnienie: co można powiedzieć o entropii kombinacji liniowych niezależnych zmiennych losowych o ustalonym rozkładzie, które zachowują wariancję? Inaczej mówiąc jakie są własności odwzorowania

$$(a_1, ..., a_n) \mapsto h\left(\sum_{j=1}^{n} a_j X_j\right), \tag{1}$$

gdzie X_j to niezależne zmienne losowe o identycznym rozkładzie, a współczynniki $a_j \in \mathbb{R}$ (lub $a_j \geq 0$) są takie, że $\sum_{j=1}^{n} a_j^2 = 1$? Pytaniu temu, sięgającemu czasów przełomowej pracy Claude'a Shannona *A mathematical theory of communication* [1], towarzyszy szereg zagadnień.

Podstawowym problemem jest analiza zachowania ciągu $h\big((X_1 + ... + X_n)/\sqrt{n}\big)$. Jak można by się spodziewać ciąg ten zbiega do wartości $h\big(\mathcal{N}(0, \operatorname{Var} X_1)\big)$, przy minimalnym założeniu istnienia $n_0 \in \mathbb{N}$ takiego, że wartość $h\big((X_1 + ... + X_{n_0})/\sqrt{n_0}\big)$ jest skończona [5]. Dopiero relatywnie niedawno udało się udowodnić że ciąg ten jest rosnący, przy założeniu o skończonej wariancji zmiennych X_j [8].

Mogłoby się wydawać, że skoro zmienna gaussowska maksymalizuje entropię, a kombinacje liniowe o równych współczynnikach $a_j = 1/\sqrt{n}$ zbiegają do zmiennej gaussowskiej z $n \to \infty$ to i wartość odwzorowania (1) dla ustalonego $n \in \mathbb{N}$ będzie maksymalizowana przez równe współczynniki $a_j = 1/\sqrt{n}$. Rozstrzygnięcie tego zagadnienia nie jest jednoznaczne. Nie będzie tak na przykład dla zmiennej losowej o rozkładzie jednostajnym na sumie dwóch oddzielonych od siebie przedziałów, co z dowiedli autorzy pracy *A reverse entropy power inequality for log-concave random vectors* [11]. Mimo tego że wspomniany kontrprzykład wykorzystuje w konstrukcji zmienne losowe o rozkładzie jednostajnym, zagadnienie pozostaje nierozstrzygnięte dla najprostszych jednostajnych zmiennych losowych, tj. skupionych na przedziale $[-a, a]$. Daje się natomiast udzielić odpowiedzi pozytywnej dla zmiennych losowych będących mieszankami gaussowskimi, tj. zmiennymi postaci $R \cdot G$, gdzie R to dowolna zmienna losowa, a G to niezależna standardowa zmienna gaussowska, co jest jednym z wyników pracy *Gaussian mixtures: entropy and geometric inequalities* [12].

Ogólniejszym zagadnieniem związanym z badaniem własności odwzorowania (1) jest identyfikacja porządku $\preceq$ w zbiorze $\{(a_1, ..., a_n) \in \mathbb{R}^n : a_1^2 + ... + a_n^2 = 1\}$ takiego, że

$$(a_1^2, ..., a_n^2) \preceq (b_1^2, ..., b_n^2) \implies h\left(\sum_{j=1}^{n} a_j X_j\right) \geq h\left(\sum_{j=1}^{n} b_j X_j\right). \tag{2}$$

Okazuje się, że naturalny kandydat (co zostanie wyjaśnione w dalszej części pracy), tj. porządek Schura, spełnia implikację (2) dla zmiennych X_j będących mieszankami gaussowskimi [12]. Dla wektorów $a, b \in \mathbb{R}^n$ i ich malejących posortowań $a_1^* \geq ... \geq a_n^*$ oraz $b_1^* \geq ... \geq b_n^*$ porządek Schura zdefiniowany jest jako [4, Definition A.1]

$$a \preceq b \iff \sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j \text{ oraz } \forall_{k=1,...,n} \quad \sum_{j=1}^{k} a_j^* \leq \sum_{j=1}^{k} b_j^*. \tag{3}$$

Równoważna definicja odzwierciedlająca geometryczną naturę porządku Schura zadana jest przez

$$a \preceq b \iff a \in \operatorname{conv}\left\{(b_{\sigma(1)}, ..., b_{\sigma(n)}) : \sigma \text{ jest permutacją zbioru } \{1, ..., n\}\right\}. \tag{4}$$

Rezultatem niniejszej pracy magisterskiej jest uogólnienie tego wyniku na przypadek zmiennych wykładniczych poprzez wykorzystanie *metody przecinających się gęstości* [13]. W związku z tym, że w odróżnieniu od mieszanek gaussowskich które są zmiennymi symetrycznymi, będziemy pracować z zmiennymi wykładniczymi, które to skupione są na dodatniej półprostej, ograniczymy się do dodatnich współczynników, tj. należących do zbioru

$$S_+^{n-1} = \left\{(a_1, \ldots, a_n) : \sum_{j=1}^{n} a_j^2 = 1, \forall_{j=1,\ldots,n}\ a_j \geq 0\right\}.$$

Techniki dowodowe wykorzystane w pracy silnie bazują na tych zaprezentowanych w [14], gdzie posłużyły do wykazania pokrewnego rezultatu, przywołanego jako twierdzenie 17 w niniejszej pracy.

Praca jest ustrukturyzowana w następujący sposób: w rozdziale 1. przywołamy niezbędne pojęcia i klasyczne wyniki związane z analizą form postaci (1), w rozdziale 2. sformułujemy główne twierdzenie oraz udowodnimy szereg lematów potrzebnych do wykazania go, natomiast w rozdziale 3. przedstawimy dowód głównego twierdzenia, wnioski z niego płynące oraz sugestie i obserwacje odnośnie dalszych możliwych badań.

Rozdział 1

Niezbędne pojęcia i klasyczne wyniki

1.1. Entropia i jej własności

Definicja 1. *Niech X będzie ciągłą rzeczywistą zmienną losową o gęstości $f : \mathbb{R} \to \mathbb{R}$. Entropią zmiennej X nazywamy wartość*

$$h(X) = -\mathbb{E} \ln f(X) = -\int_{\mathbb{R}} f(x) \ln (f(x))\, dx.$$

Na potrzeby wyznaczania wartości całek tego typu przyjmujemy konwencję $0 \cdot \ln(0) = 0$ oraz $\lambda \cdot \ln(0) = -\infty$ dla $\lambda > 0$. Warto dodatkowo zauważyć, że oznaczenie $h(X)$ jest pewnym nadużyciem notacyjnym, gdyż entropia jest funkcją rozkładu zmiennej X, a nie jej realizacji. W związku z tym, że jest to szeroko przyjęte oznaczenie i my będziemy je stosować. Entropia spełnia następujące własności:

$$h(aX + b) = h(X) + \ln |a|, \quad \text{dla } a, b \in \mathbb{R} \text{ i } a \neq 0, \tag{5}$$

$$h(X) = \inf \left\{ -\int_{\mathbb{R}} f(x) \ln (g(x))\, dx : g \text{ - gęstość prawdopodobieństwa} \right\}. \tag{6}$$

Dowód. Niech $a, b \in \mathbb{R}$, $a \neq 0$ i g będzie gęstością zmiennej $aX + b$, wówczas $g(x) = f(\frac{x-b}{a})\frac{1}{|a|}$. Stosując podstawienie $y = (x - b)/a$ otrzymujemy

$$\begin{aligned} h(aX + b) &= -\int_{\mathbb{R}} f\left(\frac{x-b}{a}\right) \frac{1}{|a|} \ln \left(f\left(\frac{x-b}{a}\right) \frac{1}{|a|} \right) dx \\ &= -\int_{\mathbb{R}} f(y) \frac{1}{|a|} \ln \left(f(y) \frac{1}{|a|} \right) |a| dy = -\int_{\mathbb{R}} f(y) \ln \left(f(y) \frac{1}{|a|} \right) dy \\ &= h(X) - \int_{\mathbb{R}} f(y) \ln \left(\frac{1}{|a|} \right) dy = h(X) + \ln |a|, \end{aligned}$$

co dowodzi własności (5). Niech g będzie gęstością prawdopodobieństwa pewnej ciągłej zmiennej losowej. W celu udowodnienia własności (6) musimy pokazać, że zachodzi

$$-\int_{\mathbb{R}} f(x) \ln f(x) dx \leq -\int_{\mathbb{R}} f(x) \ln g(x) dx.$$

Bez straty ogólności możemy założyć, że supp $g \subseteq$ supp f, gdyż w przeciwnym przypadku wartość wyrażenia po prawej stronie wynosi $+\infty$. Korzystając z nierówności $\ln x \leq x - 1$ uzyskujemy, że

$$\int_{\mathbb{R}} f \ln g - \int_{\mathbb{R}} f \ln f = \int_{\operatorname{supp} f} f \ln \frac{g}{f} \leq \int_{\operatorname{supp} f} f \left(\frac{g}{f} - 1 \right) = 0.$$

Co więcej, możemy zauważyć że w przypadku gdy $h(X)$ przyjmuje skończoną wartość, nierówność jest ostra dla gęstości g takiej że $g \neq f$ na pewnym zbiorze dodatniej miary. $\square$

1.2. Przypadki graniczne

Następujące twierdzenia obrazują graniczne zachowanie entropii sum $(X_1 + \cdots + X_n)/\sqrt{n}$ niezależnych zmiennych losowych o identycznym rozkładzie oraz ekstremalność rozkładu normalnego.

Twierdzenie 1. *[8] Niech X będzie ciągłą zmienną losową całkowalną z kwadratem, wówczas dla $X_j \overset{iid}{\sim} X$ dla każdego $n \in \mathbb{N}$ zachodzi*

$$h\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) \leq h\left(\frac{X_1 + \cdots + X_{n+1}}{\sqrt{n+1}}\right).$$

Twierdzenie 2. *[5] Niech X będzie ciągłą zmienną losową. Jeśli dla $X_j \overset{iid}{\sim} X$ istnieje $n_0 \in \mathbb{N}$ takie, że $h\big((X_1 + \cdots + X_{n_0})/\sqrt{n_0}\big)$ jest skończona, to*

$$h\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right) \xrightarrow{n\to\infty} h\left(\mathcal{N}(0, \operatorname{Var} X)\right).$$

Twierdzenie 3. *Niech X będzie rzeczywistą ciągłą zmienną losową o gęstości $f : \mathbb{R} \to \mathbb{R}$ i skończonej wariancji σ^2, a G niech będzie jednowymiarową zmienną gaussowską o tej samej wariancji, wówczas*

$$h(X) \leq h(G) = \frac{1}{2}\ln(2\pi e\sigma^2),$$

a równość zachodzi wtedy i tylko wtedy gdy X ma rozkład gaussowski.

Dowód. Bez straty ogólności załóżmy że $\mathbb{E}X = \mathbb{E}G = 0$. Wówczas

$$\begin{aligned} h(X) &\overset{(6)}{\leq} -\int_{\mathbb{R}} f(x)\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{1}{2}x^2/\sigma^2}\right)dx = -\int_{\mathbb{R}} f(x)\left(\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) - \frac{x^2}{2\sigma^2}\right)dx \\ &= \ln\left(\sqrt{2\pi\sigma^2}\right) + \frac{1}{2} = \frac{1}{2}\ln\left(2\pi e\sigma^2\right). \end{aligned}$$

Co więcej, zgodnie z obserwacją z dowodu własności (6) nierówność jest ostra w przypadku gdy X ma inny rozkład niż gaussowski. $\square$

W analogiczny sposób można dowieść, że wśród ciągłych rzeczywistych zmiennych losowych skupionych na dodatniej półprostej i o ustalonym pierwszym momencie rozkład wykładniczy jest maksymalizatorem entropii.

Fakty tego rodzaju są konsekwencją znacznie silniejszego twierdzenia, tj. nierówności Pinskera-Csiszára-Kullbacka, która pozwala szacować pewną metrykę miedzy miarami poprzez tzw. *entropią względną*.

1.3. Klasyczne nierówności

Definicja 2 (norma i metryka całkowitej wariacji). *Niech μ będzie miarą z znakiem, a Σ niech będzie σ-ciałem zbiorów mierzalnych miary μ. Wówczas normą całkowitej wariacji (ang. total variation) nazywamy wartość*

$$\|\mu\|_{TV} = \sup_{A\in\Sigma} |\mu(A)|.$$

Dla zmiennych losowych X, Y o wartościach w przestrzeni mierzalnej $(E, \mathcal{E})$ i rozkładach μ_X, μ_Y metryką całkowitej wariacji nazwiemy wartość

$$d_{TV}(X, Y) = \|\mu_X - \mu_V\|_{TV} = \sup\left\{|\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| : A \in \mathcal{E}\right\}.$$

Definicja 3 (entropia względna, dywergencja Kullbacka-Leiblera)**.** *Dla ciągłych zmiennych losowych X i Y o gęstościach f_X i f_Y ich entropią względną nazwiemy wartość*

$$D_{KL}(X||Y) = \int_{\mathbb{R}} f_X(x) \ln\left(\frac{f_X(x)}{f_Y(x)}\right) dx.$$

Sformułujemy teraz nierówność Pinskera-Csiszára-Kullbacka. Ograniczymy się do ciągłych rozkładów prawdopodobieństwa, gdyż takie są przedmiotem tej pracy, jednak można rozszerzyć definicję entropii względnej i uzyskać analog tej nierówności dla ogólniejszej klasy rozkładów.

Twierdzenie 4 (nierówność Pinskera-Csiszára-Kullbacka [10, Lemma 2.5])**.** *Niech X i Y będą ciągłymi rzeczywistymi zmiennymi losowymi, wówczas zachodzi zachodzi*

$$d_{TV}(X, Y) \leq \sqrt{\frac{1}{2} D_{KL}(X||Y)}.$$

W szczególności, gdy Y ma rozkład gaussowski o pierwszych dwóch momentach równych momentom zmiennej X, zachodzi

$$d_{TV}(X, Y) \leq \sqrt{\frac{1}{2} D_{KL}(X||Y)} = \sqrt{\frac{1}{2}(h(Y) - h(X))}.$$

Kolejne zastosowanie entropii, znajdujące użytek np. w statystyce, obrazuje nierówność dająca dolne ograniczenie na błąd kwadratowy tzw. estymatora badanej zmiennej losowej.

Twierdzenie 5. *[6, Theorem 8.6.6] Niech będzie dana ciągła rzeczywista zmienna losowa X oraz jest estymator $\hat{X}$, tj. zmienna losowa postaci $\hat{X} = f(X_1, \ldots, X_n)$, gdzie $X_j \overset{iid}{\sim} X$ a f jest pewną ustaloną funkcją $f : \mathbb{R}^n \to \mathbb{R}$. Wówczas zachodzi*

$$\mathbb{E}(X - \hat{X})^2 \geq \frac{1}{2\pi e} e^{2h(X)}.$$

Przytoczymy jeszcze jedno twierdzenie, które co prawda nie ukazuje dodatkowych zastosowań entropii, lecz daje pełniejszy obraz badanego zagadnienia.

Twierdzenie 6 (Nierówność Shannona-Stama/Entropy Power Inequality [3])**.** *Niech dane będą X i Y, niezależne rzeczywiste ciągłe zmienne losowe oraz stała $\lambda \in [0, 1]$, wówczas*

$$h\left(\sqrt{\lambda}X + \sqrt{1-\lambda}Y\right) \geq \lambda h(X) + (1 - \lambda)h(Y),$$

pod warunkiem że powyższe entropie istnieją.

Dzięki któremu to uzyskujemy dolne oszacowanie wartości odwzorowania (1), mianowicie

$$h\left(\sum_{j=1}^{n} a_j X_j\right) \geq h(X_1). \tag{7}$$

1.4. Porządek Schura

Definicja 4. *[4, Definition A.1] Dla wektorów $a, b \in \mathbb{R}^n$ oraz ich malejących posortowań $a_1^* \geq ... \geq a_n^*$ i $b_1^* \geq ... \geq b_n^*$ powiemy, że wektor b majoryzuje wektor a w porządku Schura, co będziemy oznaczać jako $a \preceq b$, jeśli*

$$\sum_{j=1}^{n} a_j = \sum_{j=1}^{n} b_j \text{ oraz } \forall_{k=1,\dots,n} \quad \sum_{j=1}^{k} a_j^* \leq \sum_{j=1}^{k} b_j^*.$$

Jak wcześniej wspomnieliśmy, porządek Schura jest naturalnym kandydatem na porządek zachowujący monotoniczność entropii kombinacji liniowych niezależnych zmiennych losowych o identycznym rozkładzie (porównującym kwadraty współczynników). Jest tak ze względu na szereg własności porządku Schura:

- Elementami maksymalnymi w zbiorze $\{(a_1, \dots, a_n) : \sum_{j=1}^n a_j^2 = 1, \forall_{j=1,\dots,n}\ a_j \geq 0\}$ są wektory bazy standardowej w $\mathbb{R}^n$, czyli np.

  $$(1, 0, \dots, 0),$$

  co jest w zgodzie z oszacowaniem (7) wynikającym z nierówności Shannona-Stama.

- Elementem minimalnym w tym samym zbiorze jest wektor

  $$\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right),$$

  co jest w zgodzie z wynikiem pracy [12] oraz postulowanym rezultatem.

- Porządek Schura jest niezmienniczy na permutacje kolejności współczynników, tak samo jak entropia kombinacji liniowych niezależnych zmiennych losowych o identycznym rozkładzie.

1.5. Pozostałe definicje

W pracy przyjmujemy następujące konwencje i oznaczenia

1. Standardowy iloczyn skalarny w $\mathbb{R}^n$ będziemy oznaczać poprzez $\langle \cdot, \cdot \rangle$.
2. Mówiąc o mierzalności funkcji $f : \mathbb{R}^n \to \mathbb{R}$ będziemy mieć na myśli mierzalność względem standardowych σ-ciał borelowskich.

Definicja 5. *Niech będzie dana funkcja mierzalna $f : \mathbb{R} \to \mathbb{R}$. Punkt $x \in \mathbb{R}$ nazwiemy punktem zmiany znaku funkcji f na dodatni jeśli istnieją $u, v \in \mathbb{R}$ spełniające $u < x < v$ i takie, że funkcja f jest dodatnia prawie wszędzie na przedziale (x, v), niedodatnia prawie wszędzie na przedziale (u, x) oraz ujemna na pewnym podzbiorze przedziału (u, x) dodatniej miary. Punkt $x \in \mathbb{R}$ nazwiemy punktem zmiany znaku funkcji f jeśli jest jest punktem zmiany znaku na dodatni dla funkcji f lub $-f$.*

Warto skomentować, że definicja punktu zmiany znaku ma na celu przede wszystkim ujednoznacznienie tego pojęcia w przypadku funkcji, które np. zerują się na przedziałach niezerowej długości.

Definicja 6. *Dodatnią częścią sfery jednostkowej $S^{n-1} \subset \mathbb{R}^n$ nazwiemy zbiór*

$$S_+^{n-1} = \left\{(a_1, \dots, a_n) : \sum_{j=1}^{n} a_j^2 = 1, \forall_{j=1,\dots,n}\ a_j \geq 0\right\}.$$

Rozdział 2

Lematy niezbędne do wykazania głównego twierdzenia

2.1. Lematy techniczne

Lemat 7. *[4, Lemma B.1] Niech dane będą wektory $x, y \in \mathbb{R}^n$ takie, że $x \preceq y$. Wówczas wektor x można uzyskać z wektora y za pomocą skończonej ilości operacji postaci*

$$T^{\lambda}_{j,k}(z_1, \ldots, z_n) = (z_1, \ldots, z_{j-1}, \lambda z_j + (1-\lambda)z_k, z_{j+1}, \ldots, z_{k-1}, \lambda z_k + (1-\lambda)z_j, z_{k+1}, \ldots, z_n),$$

dla odpowiednich $\lambda \in [0,1]$ oraz $1 \leq j < k \leq n$.

Dowód. Operacja $T^0_{j,k}$ odpowiada permutacji współrzędnych j i k, zatem bez straty ogólności możemy założyć, że ciągi x_j i y_j nie są swoimi permutacjami oraz że zachodzi $x_1 \geq \cdots \geq x_n$ oraz $y_1 \geq \cdots \geq y_n$.

Niech $j = \max\{i : x_i < y_i\}$ oraz $k = \min\{i > j : x_i > y_i\}$. Taki indeks j istnieje, ponieważ $x \neq y$. Indeks k również musi istnieć, gdyż w przeciwnym przypadku mielibyśmy $x_i = y_i$ dla $i > j$. W połączeniu z nierównością $\sum_{i=1}^{j-1} x_i \leq \sum_{i=1}^{j-1} y_i$ wynikającą z założenia $x \preceq y$ oraz nierównością $x_j < y_j$ wynikającą z definicji j prowadziłoby to do ostrej nierówności $\sum_{i=1}^{n} x_i < \sum_{i=1}^{n} y_i$. Brak równości przeczyłby założeniu $x \preceq y$. Zauważmy, że wybór indeksów gwarantuje nierówności $y_j > x_j \geq x_k > y_k$.

Pokażemy, że da się wybrać taką operację $T^{\lambda}_{j,k}$, że wektory $y' = T^{\lambda}_{j,k}y$ oraz x pokrywają się na o jednej więcej współrzędnej niż wektory y oraz x. W tym celu wprowadźmy $\delta = \min(y_j - x_j, x_k - y_k)$ oraz $1 - \lambda = \delta/(y_j - y_k) \in [0, 1/2]$. Niech $y' = T^{\lambda}_{j,k}y = \lambda y + (1-\lambda)T^0_{j,k}y$, wówczas

$$y'_i = \begin{cases} y_i & \text{dla } i \notin \{j,k\}, \\ \lambda y_k + (1-\lambda)y_j & \text{dla } i = k, \\ \lambda y_j + (1-\lambda)y_k & \text{dla } i = j. \end{cases}$$

W przypadku gdy $\delta = y_j - x_j$ zachodzi

$$y'_j = \frac{x_j - y_k}{y_j - y_k} y_j + \frac{y_j - x_j}{y_j - y_k} y_k = x_j,$$

natomiast w przypadku gdy $\delta = x_k - y_k$ mamy

$$y'_k = \frac{y_j - x_k}{y_j - y_k} y_k + \frac{x_k - y_k}{y_j - y_k} y_j = x_k.$$

Czyli dla każdych z dwóch możliwych wartości δ wektor $y' = T^{\lambda}_{j,k}y$ posiada co najmniej o jedną więcej współrzędną niż wektor y, która pokrywa się wartością z współrzędną wektora x. Pozostaje wykazać, że $x \preceq y'$, by za pomocą kolejnych operacji $T^{\lambda}_{j,k}$ zrównać wektor y' z wektorem x. Istotnie

- dla $1 \leq \nu < j$ mamy $\sum_{i=1}^{\nu} y'_i = \sum_{i=1}^{\nu} y_i \geq \sum_{i=1}^{\nu} x_i$, ponieważ $x \preceq y$,
- zachodzi $y'_j = x_j$ albo $x_j = y'_k \leq y'_j$. Nierówność jest prawdziwa gdyż y'_j jest średnią ważoną tych samych wartości co y'_k lecz z większą wagą przy większej wartości. Zatem i $\sum_{i=1}^{j} y'_i \geq \sum_{i=1}^{j} x_i$,
- dla $j < \nu < k$ mamy $\sum_{i=1}^{\nu} y'_i = \sum_{i=1}^{j} y'_i + \sum_{i=j+1}^{\nu} x_i \geq \sum_{i=1}^{\nu} x_i$,
- ostatecznie dla $\nu \geq k$, ponieważ $y'_j + y'_k = y_j + y_k$, mamy $\sum_{i=1}^{\nu} y'_i = \sum_{i=1}^{\nu} y_i \geq \sum_{i=1}^{\nu} x_i$,
- naturalnie, z tego samego powodu, zachowana jest równość $\sum_{i=1}^{\nu} y'_i = \sum_{i=1}^{\nu} x_i$.

Czyli faktycznie zachodzi $x \preceq y'$, a zatem po co najwyżej n krokach wektor y zastanie przetransformowany na wektor x. □

Lemat 8. *Niech funkcja $f : \mathbb{R} \to \mathbb{R}$ będzie dana wzorem $f(x) = \sum_{j=1}^{n} a_j e^{b_j x}$ dla pewnych stałych $a_j, b_j \in \mathbb{R}$ oraz $n \in \mathbb{N}$. Wówczas funkcja f jest stale równa zero albo ma co najwyżej $n-1$ pierwiastków.*

Dowód. Lemat jest oczywisty dla $n = 1$. W celu wykazania tezy z wykorzystaniem indukcji załóżmy nie wprost, że dla pewnego $n \in \mathbb{N}$ ważone sumy $n-1$ funkcji wykładniczych mogą mieć co najwyżej $n-2$ pierwiastków lub są stale równe 0, ale pewna funkcja f postaci $f(x) = a_1 e^{b_1 x} + \cdots + a_n e^{b_n x}$ ma już n pierwiastków. Bez straty ogólności załóżmy, że $a_1 \neq 0$. Wówczas

$$f(x) = 0 \iff \tilde{f}(x) := a_1 + a_2 e^{(b_2-b_1)x} + \cdots + a_n e^{(b_n-b_1)x} = 0.$$

Na mocy twierdzenia Rolle'a, $\tilde{f}'$ pochodna funkcji $\tilde{f}$ posiada co najmniej $n-1$ pierwiastków, jednak zgodnie z założeniem jako ważona suma $n-1$ funkcji wykładniczych może ich mieć co najwyżej $n-2$. Nie może być ona również stale równa zero, gdyż założyliśmy, że $a_1 \neq 0$. Wobec tego hipoteza indukcyjna musi być prawdziwa. □

Lemat 9. *Niech będą dane $n \in \mathbb{N}$ i funkcja mierzalna $f : \mathbb{R} \to \mathbb{R}$ takie, że zachodzi $\forall_{k=0,\ldots,n-1} \int_{\mathbb{R}} x^k f(x)dx = 0$. Wówczas funkcja f zmienia znak co najmniej n razy bądź jest stale równa 0.*

Dowód. Załóżmy nie wprost, że $x_1, \ldots, x_k$ są jedynymi punktami zmiany znaku funkcji f dla pewnego $k < n$. Niech wielomian P będzie zadany wzorem $P(x) = \prod_{j=1}^{k}(x - x_j)$, wówczas na mocy założenia $\int_{\mathbb{R}} P(x)f(x)dx = 0$. Z drugiej strony, ponieważ funkcje P i f współdzielą punkty zmiany znaku, ich iloczyn Pf, który nie może być stale równy zero, nie zmienia znaku a zatem całka $\int_{\mathbb{R}} P(x)f(x)dx$ nie może wynosić zero. Wobec uzyskanej sprzeczności wnioskujemy że faktycznie musi zachodzić $k \geq n$. □

2.2. Własności rozkładu wykładniczego

W kolejnych lematach będziemy wykorzystywać pojęcie splotu funkcji, a dokładniej splotu funkcji rzeczywistych określonych na dodatniej półprostej. Dla uniknięcia niejasności przypominamy definicję oraz przyjmujemy oznaczenia

$$(f * g)(x) = f(x) * g(x) = \int_0^x f(y)g(x-y)dy = \int_0^x f(y-x)g(x)dy.$$

Lemat 10. *Niech będzie dana trzykrotnie różniczkowalna funkcja $g : [0, \infty) \mapsto \mathbb{R}$ i liczba rzeczywista $\lambda \geq 0$ takie, że $g' + \lambda \geq 0$, $g'' \leq 0$, $g''' \geq 0$. Dodatkowo niech funkcja $h : [0, \infty) \mapsto \mathbb{R}$ będzie zdefiniowana poprzez równanie $e^{h(x)} = e^{g(x)} * (e^{-\lambda x} 1_{[0,\infty)}(x)) = \int_0^x e^{g(y)} e^{-\lambda(x-y)} dy$. Wówczas również funkcja h spełnia warunki $h' + \lambda \geq 0$, $h'' \leq 0$, $h''' \geq 0$.*

Dowód. Najpierw sprowadzimy dowód do przypadku $\lambda = 0$. Załóżmy, że lemat jest w istocie prawdziwy dla $\lambda = 0$ i rozpiszmy

$$\begin{aligned} e^{h(x)} &= e^{g(x)} * \left(e^{-\lambda x} 1_{[0,\infty]}(x)\right) = \int_0^x e^{g(y)} e^{-\lambda(x-y)} dy = e^{-\lambda x} \int_0^x e^{g(y)+\lambda y} dy \\ &= e^{-\lambda x} \cdot \left(e^{g(x)+\lambda x} * e^{-0 \cdot x} 1_{[0,\infty)}(x)\right) =: e^{-\lambda x} e^{h_0(x)}. \end{aligned}$$

Funkcja $\tilde{g}(x) = g(x) + \lambda x$ spełnia założenia lematu z $\lambda = 0$, a zatem funkcja h_0 spełnia warunki $h_0' \geq 0$, $h_0'' \leq 0$, $h_0''' \geq 0$. Wobec tego $h' + \lambda = h_0' \geq 0$, $h'' = h_0'' \leq 0$, $h''' = h_0''' \geq 0$. W analogiczny sposób można zaobserwować, że dodanie stałej do funkcji g zachowuje prawdziwość tezy. Obserwacja ta przyda się w dalszej części dowodu.

Teraz wykażemy lemat w przypadku $\lambda = 0$. Niech $I(x) = \int_0^x e^{g(y)} dy$, wówczas

$$I'(x) = e^{g(x)}, \quad I''(x) = e^{g(x)} g'(x), \quad I'''(x) = e^{g(x)} \left(g''(x) + g'(x)^2\right)$$

oraz

$$h'(x) = (\ln I(x))' = \frac{e^{g(x)}}{I(x)} \geq 0$$

ze względu na dodatniość licznika i mianownika. Ponadto

$$h''(x) = (\ln I(x))'' = \frac{1}{I(x)^2} \left(I(x) I''(x) - I'(x)^2\right) = \frac{1}{I(x)^2} \left(I(x) e^{g(x)} g'(x) - e^{2g(x)}\right) \leq 0,$$

gdyż

$$I(x) e^{g(x)} g'(x) - e^{2g(x)} \leq 0 \iff I(x) g'(x) \leq e^{g(x)}$$

oraz dzięki $g'' \leq 0$ zachodzi

$$I(x) g'(x) = \int_0^x e^{g(y)} g'(x) dy \leq \int_0^x e^{g(y)} g'(y) dy = e^{g(x)} - e^{g(0)} \leq e^{g(x)}.$$

Co więcej

$$\begin{aligned} h'''(x) = (\ln I(x))''' &= \frac{1}{I(x)^3} \left(2I'(x)^3 + I(x)^2 I(x)''' - 3I(x) I(x)' I(x)''\right) \\ &= \frac{1}{I(x)^3} \left(2e^{3g(x)} + I(x)^2 e^{g(x)} (g''(x) + g'(x)^2) - 3I(x) e^{2g(x)} g'(x)\right) \\ &= \frac{e^{3g(x)}}{I(x)^3} \left(2 + (I(x) e^{-g(x)})^2 (g''(x) + g'(x)^2) - 3(I(x) e^{-g(x)}) g'(x)\right). \end{aligned}$$

Ustalmy $x \geq 0$. Zgodnie z wcześniejszą obserwacją bez straty ogólności możemy odjąć od funkcji g wartość $g(x)$. Dodatkowo ustalmy stałe $a = g'(x) \geq 0$ oraz $b = -g''(x) \geq 0$. Wówczas dowód nierówności $h''' \geq 0$ sprowadza się do wykazania, że

$$f(I(x)) \geq 0 \quad \text{dla } f(t) = t^2(a^2 - b) - 3ta + 2.$$

Załóżmy chwilowo, że $b > 0$ i wyprowadźmy oszacowania na wartość $I(x)$. Mamy

$$\begin{aligned}
0 \leq I(x) &= \int_0^x e^{g(y)} dy \leq \int_0^x e^{g(x)+g'(x)(y-x)+g''(x)\frac{(y-x)^2}{2}} dy = \int_0^x e^{-ay-by^2/2} dy \\
&= e^{a^2/(2b)} \int_0^x e^{-(y\sqrt{b}+a/\sqrt{b})^2/2} dy \leq e^{a^2/(2b)} \int_0^\infty e^{-(y\sqrt{b}+a/\sqrt{b})^2/2} dy \\
&= e^{a^2/(2b)} \frac{1}{\sqrt{b}} \int_{a/\sqrt{b}}^\infty e^{-u^2/2} du.
\end{aligned}$$

W pierwszej nierówności skorzystaliśmy z założenia $g''' \geq 0$, natomiast w drugiej nierówności zwyczajnie rozszerzyliśmy granice całkowania dla funkcji dodatniej. W ostatnim przejściu dokonaliśmy zamiany zmiennych $u = y\sqrt{b} + a/\sqrt{b}$.

Możemy teraz skorzystać z oszacowania na ogon gaussowski pochodzącego od Sampforda [2], a później ponownie odkrytego przez Szarka i Werner [7], tj.

$$\int_t^\infty e^{-y^2/2} dy \leq \frac{4}{\sqrt{8+t^2}+3t} e^{-t^2/2} \quad \text{dla } t \geq 0,$$

by otrzymać

$$I(x) \leq e^{a^2/(2b)} \frac{1}{\sqrt{b}} e^{-a^2/(2b)} \frac{4}{\sqrt{8+a^2/b}+3a/\sqrt{b}} = \frac{4}{3a+\sqrt{8b+a^2}} =: I_0.$$

Również w przypadku $b = 0$ możemy oszacować $I(x)$ z góry przez I_0, gdyż

$$I(x) \leq \int_0^x e^{-ay} dy = \frac{1}{a}(1-e^{-ax}) \leq \frac{1}{a} = \frac{4}{3a+\sqrt{8\cdot 0+a^2}} = I_0.$$

By dokończyć dowód pozostaje jedynie wykazać, że t_0 – pierwszy pierwiastek $f(t)$ na prawo od 0, spełnia nierówność $t_0 \geq \frac{4}{3a+\sqrt{8b+a^2}} = I_0$. W tym celu należy rozważyć trzy przypadki.

1. Funkcja f jest niezdegenerowanym wielomianem kwadratowym, tj. $a^2 - b \neq 0$. W obu podprzypadkach $a^2 - b > 0$ i $a^2 - b < 0$ zachodzi

$$t_0 = \frac{3a - \sqrt{9a^2 - 8(a^2-b)}}{2(a^2-b)} = \frac{3a - \sqrt{a^2+8b}}{2(a^2-b)} = \frac{4}{3a+\sqrt{8b+a^2}} \geq I_0.$$

2. Funkcja f jest liniowa, tj. $a^2 = b \neq 0$. Wówczas $t_0 = 2/(3a)$ i faktycznie

$$I_0 = \frac{4}{3a+\sqrt{8b+a^2}} = \frac{4}{3a+\sqrt{9a^2}} = \frac{2}{3a} \leq t_0.$$

3. Funkcja f jest stale równa 2, co zachodzi w przypadku gdy $a^2 = b = 0$. Co prawda, wówczas nie da się określić wartości t_0, ale nierówność $f(I(x)) = 2 \geq 0$ nadal jest prawdziwa.

□

Wniosek 11. *Niech X_j będą niezależnymi zmiennymi losowymi o rozkładach* $\mathrm{Exp}(\lambda_j)$*, wówczas gęstość rozkładu prawdopodobieństwa $f(x)$ sumy $X_1 + \cdots + X_n$ spełnia nierówności*

$$(\ln f(x))' + \max_j \lambda_j \geq 0, \ (\ln f(x))'' \leq 0, \ (\ln f(x))''' \geq 0.$$

Dowód. Bez straty ogólności załóżmy, że $\lambda_1 \leq \cdots \leq \lambda_n$. Niech f_j będzie gęstością rozkładu sumy $X_1 + \cdots + X_j$. Wykażemy wniosek przez indukcję względem n.

W istocie dla $n = 1$, $\ln f(x) = \ln \lambda_1 - x\lambda_1$ i wszystkie trzy postulowane nierówności są spełnione.

W celu wykazania kroku indukcyjnego załóżmy, że zachodzi

$$(\ln f_{n-1}(x))' + \lambda_{n-1} \geq 0,\ (\ln f_{n-1}(x))'' \leq 0 \text{ i } (\ln f_{n-1}(x))''' \geq 0.$$

Na mocy równości $f_n(x) = \int_0^x f_{n-1}(y)\lambda_n e^{-\lambda_n(x-y)}dy$ i lematu 10 zastosowanego dla $\lambda = \lambda_n$ otrzymujemy, że

$$(\ln f_n(x))' + \lambda_n \geq 0,\ (\ln f_n(x))'' \leq 0 \text{ i } (\ln f_n(x))''' \geq 0.$$

□

Lemat 12. *Niech $n \geq 2$ oraz X_j będą niezależnymi zmiennymi losowymi o rozkładach* $\mathrm{Exp}(\lambda_j)$*, wówczas gęstość rozkładu prawdopodobieństwa $f(x)$ sumy $X_1 + \cdots + X_n$ spełnia*

$$\lim_{x \to 0^+} f(x) = 0.$$

Dowód. Wykażemy lemat poprzez indukcję względem n. Dla $n = 2$ oraz $\lambda_1 = \lambda_2 = \lambda$ mamy $f(x) = c \cdot xe^{-\lambda x}$, co faktycznie zbiega do zera. Natomiast dla $\lambda_1 \neq \lambda_2$ mamy

$$f(x) = c \cdot \int_0^x e^{-\lambda_1 y} e^{-\lambda_2(x-y)} dy = c \cdot \frac{e^{-\lambda_2 x}}{\lambda_2 - \lambda_1}\left(e^{(\lambda_2-\lambda_1)x} - 1\right),$$

co też zbiega do zera. Dla $n > 2$ niech g będzie odpowiednią gęstością sumy $n-1$ zmiennych wykładniczych, wówczas

$$f(x) = \int_0^x g(y)\lambda_n e^{-\lambda_n(x-y)} dy \leq \lambda_n \int_0^x g(y)dy \to 0, \quad \text{dla } x \to 0^+.$$

□

2.3. Metoda przecinających się gęstości

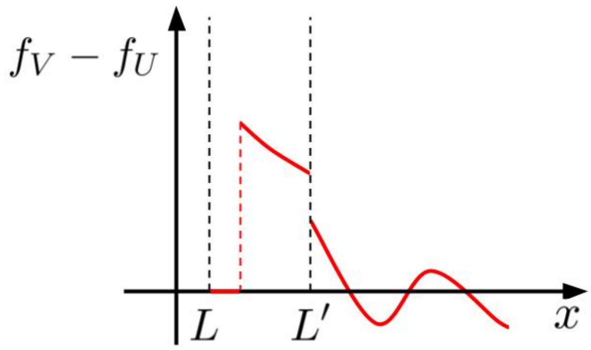
Teraz przedstawimy dowód lematów niezbędnych do wykazania głównego twierdzenia, tzw. *metodę przecinających się gęstości* oraz „zwężającą" własność operacji typu $T_{j,k}^{\lambda}$ z lematu 7.

Lemat 13 (Metoda przecinających się gęstości [13])**.** *Niech U i V będą ciągłymi zmiennymi losowymi o równym pierwszym i drugim momencie oraz takimi, że $U \geq L'$ oraz $V \geq L$ p.n. dla pewnych stałych $L, L' \in \mathbb{R}$ spełniających $L' > L$. Dodatkowo załóżmy że różnica ich gęstości $f_V - f_U$ zmienia znak trzykrotnie na przedziale (L', ∞) i jest dodatnia przed pierwszą zmianą znaku. Wówczas dla dowolnej trzykrotnie różniczkowalnej funkcji $\Phi : (L, \infty) \to \mathbb{R}$ takiej, że $\Phi''' \leq 0$ i $\Phi(L^+) = \infty$ zachodzi*

$$\mathbb{E}\Phi(U) \leq \mathbb{E}\Phi(V),$$

pod warunkiem, że te całki są dobrze określone (dopuszczamy wartości $\pm\infty$).

Dowód. Rysunek 2.1 obrazuje w jaki sposób przykładowe funkcje f_U, f_V oraz stałe L, L' mogą realizować wybrane założenia lematu dotyczące przebiegu funkcji $f_V - f_U$.



Rysunek 2.1: Przykładowy przebieg funkcji $f_V - f_U$.

Zaobserwujmy, że dla dowolnych stałych $a, b, c \in \mathbb{R}$, dzięki założeniu o równości i istnieniu pierwszych dwóch momentów, prawdziwa jest równoważność

$$\begin{aligned}\mathbb{E}\Phi(U) \leq \mathbb{E}\Phi(V) &\iff \int_L^\infty \Phi(x)(f_V(x) - f_U(x))dx \geq 0 \\ &\iff \int_L^\infty (\Phi(x) + ax^2 + bx + c)(f_V(x) - f_U(x))dx \geq 0.\end{aligned}$$

Niech zatem $x_1, x_2, x_3 \in (L', \infty)$, będą punktami zmiany znaku różnicy gęstości $f_V - f_U$ oraz spełniają $x_1 < x_2 < x_3$. Mamy dokładnie tyle swobody ile potrzeba, by dobrać wartości stałych a, b, c tak aby

$$\forall_j \ \Phi(x_j) + ax_j^2 + bx_j + c = 0,$$

co jest równoważne określeniu wielomianu kwadratowego przyjmującego zadane wartości w trzech ustalonych punktach.

Jeśli udowodnimy że x_1, x_2, x_3 są jedynymi trzema punktami zmiany znaku funkcji $\Psi(x) := \Phi(x) + ax^2 + bx + c$ na przedziale (L, ∞) to wykażemy że funkcja podcałkowa $\Phi \cdot (f_V - f_U)$, jako iloczyn funkcji o pokrywających się miejscach zmiany znaku, jest stałego znaku.

Dokładniej mówiąc, założenie $\Phi(L^+) = \infty$, brak zmian znaku funkcji Ψ na przedziale (L, L') oraz zerowanie się f_U na tym samym przedziale implikuje nieujemność funkcji podcałkowej $\Phi \cdot (f_V - f_U)$ na przedziale (L, L').

Natomiast dodatkowo ciągłość funkcji Ψ w punkcie L' oraz dodatniość funkcji $f_V - f_U$ w L^+ gwarantuje, że znaki tych dwóch funkcji o pokrywających się miejscach zmiany znaku również będą się pokrywać, a zatem ich iloczyn, stanowiący funkcję podcałkową $\Phi \cdot (f_V - f_U)$ także będzie nieujemny.

A zatem pozostaje wykazać, że x_1, x_2, x_3 są jedynymi punktami zmiany znaku funkcji Ψ na przedziale (L, ∞). W istocie tak będzie, gdyż

$$\Psi'(x) = 2ax + b + \Phi'(x) = 2ax + b + \textit{funkcja wklęsła}$$

jako funkcja wklęsła nie może posiadać więcej niż dwóch miejsc zmiany znaku. Wobec czego, jej funkcja pierwotna Ψ może mieć co najwyżej trzy miejsca zmiany znaku, które to już wskazaliśmy. $\square$

Lemat 14. *Załóżmy że $Y = (Y_1, \ldots, Y_n)$ jest wektorem niezależnych zmiennych losowych o identycznym rozkładzie takich, że $Y_1 \geq -L$ p.n. dla pewnego $L \geq 0$. Niech $\Phi : \mathbb{R} \to \mathbb{R}$ będzie funkcją mierzalną ograniczoną z dołu oraz $s_0 \in \mathbb{R}$. Ponadto niech dla każdych $0 \leq d_1 < c_1, c_2 < d_2$ takich, że $d_1^2 + d_2^2 = c_1^2 + c_2^2 \leq 1$ oraz dla każdego $s \geq s_0 + c_1 L + c_2 L$ zachodzi*

$$\mathbb{E}\Phi\left(s + d_1 Y_1 + d_2 Y_2\right) \leq \mathbb{E}\Phi\left(s + c_1 Y_1 + c_2 Y_2\right)$$

Wówczas dla dowolnych $(c_1,\dots,c_n),(d_1,\dots,d_n)\in S_+^{n-1}$ *takich, że* $-L\sum_{j=1}^n c_j\ge s_0$ *zachodzi wynikanie*

$$(c_1^2,\dots,c_n^2)\preceq(d_1^2,\dots,d_n^2)\Rightarrow \mathbb{E}\Phi\left(\sum_{j=1}^n d_jY_j\right)\le \mathbb{E}\Phi\left(\sum_{j=1}^n c_jY_j\right).$$

Dowód. Załóżmy, że $(c_1,\dots,c_n),(d_1,\dots,d_n)\in S_+^{n-1}$ i są takie, że $(c_1^2,\dots,c_n^2)\preceq(d_1^2,\dots,d_n^2)$ oraz $-L\sum_{j=1}^n c_j\ge s_0$. W lemacie 7 dowiedliśmy, że wektor $(c_1^2,\dots,c_n^2)$ można uzyskać z wektora $(d_1^2,\dots,d_n^2)$ poprzez skończoną liczbę operacji postaci

$$T_{j,k}^\lambda(z_1,\dots,z_n)=(z_1,\dots,z_{j-1},\lambda z_j+(1-\lambda)z_k,z_{j+1},\dots,z_{k-1},\lambda z_k+(1-\lambda)z_j,z_{k+1},\dots,z_n),$$

dla odpowiednich $\lambda\in[0,1]$ oraz $1\le j<k\le n$.

Najpierw wykażemy, że przy przyjętych założeniach operacje $T_{j,k}^\lambda$ przeprowadzające wektor $d^2=(d_1^2,\dots,d_n^2)$ na $c^2=(c_1^2,\dots,c_n^2)$ zwiększają średnią funkcji Φ. W tym celu wykażemy, że zachodzi implikacja

$$c^2\preceq(d'')^2,\ (d'')^2=T_{j,k}^\lambda(d')^2\implies \mathbb{E}\Phi\big(\langle d',Y\rangle\big)\le\mathbb{E}\Phi\left(\langle d'',Y\rangle\right) \tag{8}$$

Bez straty ogólności załóżmy zatem, że $\lambda\in(0,1)$ oraz $i=1,j=2$, tj.

$$d_1''=\sqrt{\lambda d_1'^2+(1-\lambda)d_2'^2},\quad d_2''=\sqrt{\lambda d_2'^2+(1-\lambda)d_1'^2}\quad\text{oraz}\quad d_j''=d_j',\ \text{dla } j>2.$$

Wówczas zachodzi $0\le d_1'<d_1'',d_2''<d_2'$ wraz z $d_1'^2+d_2'^2=d_1''^2+d_2''^2\le 1$ oraz $d_1''+d_2''>d_1'+d_2'$, gdyż z nierówności między średnią arytmetyczną i potęgową wynika, że

$$\begin{aligned}d_1''+d_2''&=\sqrt{\lambda d_1'^2+(1-\lambda)d_2'^2}+\sqrt{\lambda d_2'^2+(1-\lambda)d_1'^2}\\&\ge\lambda d_1'+(1-\lambda)d_2'+\lambda d_2'+(1-\lambda)d_1'=d_1'+d_2'.\end{aligned}$$

Wobec tego, widzimy że dla $u,v\in S_+^{n-1}$ zachodzi implikacja,

$$T_{j,k}^\lambda u^2=v^2\implies\sum_{j=1}^n u_j\le\sum_{j=1}^n v_j.$$

tj. zaaplikowanie operacji typu $T_{j,k}^\lambda$ do wektora kwadratów współczynników zwiększa sumę współczynników. Teraz jesteśmy w stanie rozpisać

$$\begin{aligned}\mathbb{E}\Phi(\langle d'',Y\rangle)&=\mathbb{E}\Phi\left(d_1''Y_1+d_2''Y_2+\sum_{j=3}^n d_j'Y_j\right)\\&=\mathbb{E}\left[\mathbb{E}\Phi\left(d_1''Y_1+d_2''Y_2+\sum_{j=3}^n d_j'Y_j\right)\middle|\sum_{j=3}^n d_j'Y_j\right]\\&\ge\mathbb{E}\left[\mathbb{E}\Phi\left(d_1'Y_1+d_2'Y_2+\sum_{j=3}^n d_j'Y_j\right)\middle|\sum_{j=3}^n d_j'Y_j\right]=\mathbb{E}\Phi\left(\langle d',Y\rangle\right).\end{aligned}$$

W nierówności wykorzystaliśmy zakładaną nierówność punktową wewnątrz warunkowej wartości oczekiwanej dla $s=\sum_{j=3}^n d_j'Y_j$, co było możliwe dzięki temu, że

1. Wartość zmiennej warunkującej spełnia

$$\sum_{j=3}^{n} d'_j Y_j \geq s_0 + d''_1 L + d''_2 L \iff -d''_1 L - d''_2 L + \sum_{j=3}^{n} d'_j Y_j \geq s_0,$$

   co jest prawdą, gdyż

$$-d''_1 L - d''_2 L + \sum_{j=3}^{n} d'_j Y_j \geq -L \left(d''_1 + d''_2 + \sum_{j=3}^{n} d'_j \right) \geq -L \sum_{j=1}^{n} c_j \geq s_0,$$

   a przedostatnia nierówność zachodzi ze względu na $c^2 \preceq (d'')^2$ i wcześniejszą obserwację, że zaaplikowanie operacji typu $T_{j,k}^{\lambda}$ do wektora kwadratów współczynników zwiększa sumę współczynników.

2. Wartości oczekiwane zmiennych $\Phi(\langle d', Y \rangle)$ i $\Phi(\langle d'', Y \rangle)$ są dobrze określone, gdyż Φ jako funkcja jest ograniczona dołu, a zatem wartość oczekiwana części ujemnej rozważanych zmiennych jest skończona.

Wybierzmy ciąg operacji postaci $T_{j,k}^{\lambda}$ transformujący d^2 na c^2. Ustalmy jedną z nich i załóżmy że dokonuje transformacji $T_{j,k}^{\lambda}(d')^2 = (d'')^2$. Ponieważ jest to ciąg operacji przeprowadzający d^2 na c^2 i porządek Schura jest przechodni, więc $c^2 \preceq (d'')^2$. Zatem na mocy implikacji (8), $\mathbb{E}\Phi(\langle d', Y \rangle) \leq \mathbb{E}\Phi(\langle d'', Y \rangle)$, co znaczy, że w każdym kroku $T_{j,k}^{\lambda}$ zwiększa średnią funkcji Φ, a zatem ostatecznie otrzymamy

$$\mathbb{E}\Phi(\langle c, Y \rangle) \geq \mathbb{E}\Phi(\langle d, Y \rangle).$$

□

Uwaga 1. *Założenie o ograniczoności funkcji Φ z dołu można zastąpić ogólniejszym, wymagającym jedynie by dla każdego $u \in S_+^{n-1}$*

a) część dodatnia zmiennej $\Phi(\langle u, Y \rangle)$ była całkowalna

lub

b) część ujemna zmiennej $\Phi(\langle u, Y \rangle)$ była całkowalna.

Lemat 14 został sformułowany przy silniejszych założeniach ze względu na to, że tylko w takiej wersji będziemy go wykorzystywać.

Rozdział 3

Główne twierdzenie

3.1. Dowód

Twierdzenie 15. *Niech X_j będą niezależnymi zmienny losowymi o rozkładzie $Exp(1)$. Wówczas dla dowolnych $c, d \in S_+^{n-1}$ zachodzi*

$$(c_1^2, \ldots, c_n^2) \preceq (d_1^2, \ldots, d_n^2) \Rightarrow h\left(\sum_{j=1}^n d_j X_j\right) \leq h\left(\sum_{j=1}^n c_j X_j\right).$$

Dowód. Dla klarowności dowód przedstawimy w podziale na kroki.

1. Na potrzeby skorzystania w dalszej części dowodu z *metody przecinających się gęstości* (lemat 13) wprowadźmy zcentrowane zmienne losowe $Y_j = X_j - \mathbb{E}X_j$. Niech f_c oraz f_d będą gęstościami zmiennych $\sum_{j=1}^n c_j Y_j$ oraz $\sum_{j=1}^n d_j Y_j$. Dzięki niezmienniczości entropii ze względu na przesunięcia (5) (skrótowo *nzm*) oraz własności (6) (skrótowo *max*), w celu udowodnienia twierdzenia wystarczyłoby wykazać, że $-\int f_c \ln f_c \geq -\int f_d \ln f_c$, gdyż wówczas

$$h\left(\sum_{j=1}^n c_j X_j\right) \overset{nzm}{=} -\int f_c \ln f_c \geq -\int f_d \ln f_c \overset{max}{\geq} -\int f_d \ln f_d \overset{nzm}{=} h\left(\sum_{j=1}^n d_j X_j\right).$$

   Innymi słowy, musimy sprawdzić, że dla funkcji $\Phi = -\ln f_c$ zachodzi

$$\mathbb{E}\Phi\left(\sum_{j=1}^n d_j Y_j\right) \leq \mathbb{E}\Phi\left(\sum_{j=1}^n c_j Y_j\right).$$

2. W tym celu wykorzystamy lemat 14. Wprowadźmy stałe $s_0 = -\sum_{j=1}^n c_j$ oraz $\tilde{L} = 1$ w roli L (aby uniknąć przyszłego konfliktu oznaczeń). Przywołajmy założenia lematu:

   - Zachodzi $Y_j \geq -\tilde{L} = -1$ p.n., co jest prawdą.
   - Funkcja f_c, jako splot funkcji ograniczonych będzie ograniczona z góry, wobec czego funkcja $\Phi = -\ln f_c$ będzie ograniczona z dołu. Ponadto funkcja Φ jest mierzalna. Warto zauważyć, że ograniczoność z dołu funkcji Φ można również wywnioskować z tezy wniosku 11.

- Dla każdych $0 \leq d_1 < c_1, c_2 < d_2$ takich, że $d_1^2 + d_2^2 = c_1^2 + c_2^2 \leq 1$ oraz dla $s \geq s_0 + c_1 + c_2 = -\sum_{j=3}^{n} c_j$ ma zachodzić

$$\mathbb{E}\Phi\left(s + d_1 Y_1 + d_2 Y_2\right) \leq \mathbb{E}\Phi\left(s + c_1 Y_1 + c_2 Y_2\right).$$

Prawdziwość ostatniego założenia wykażemy z wykorzystaniem lematu 13.

3. Niech $s \geq s_0 + c_1 + c_2 = -\sum_{j=3}^{n} c_j$. Zdefiniujmy zmienne losowe $U = s + d_1 Y_1 + d_2 Y_2$ i $V = s + c_1 Y_1 + c_2 Y_2$ oraz stałe $L' = \operatorname{essinf} U = s - (d_1 + d_2)$ i $L = s_0 \leq \operatorname{essinf} V = s - (c_1 + c_2)$. Pozostaje nam sprawdzić założenia lematu 13.

   - Dzięki zcentrowaniu zmiennych Y_j i równości $c_1^2 + c_2^2 = d_1^2 + d_2^2$ zmienne U, V mają równe pierwsze i drugie momenty.
   - W istocie, L oraz L' są zdefiniowane tak, by $U \geq L'$ oraz $V \geq L$ p.n.
   - Dla funkcji $\Phi = -\ln f_c$ na mocy wniosku 11 zachodzi $\Phi''' \leq 0$, ponieważ f_c jest przesuniętą gęstością splotu zmiennych wykładniczych.
   - Zachodzi $\Phi(L^+) = \infty$, jest tak ponieważ $f_c(L^+) = 0$, co wynika z lematu 12.
   - $\mathbb{E}\Phi\left(s + d_1 Y_1 + d_2 Y_2\right)$ i $\mathbb{E}\Phi\left(s + c_1 Y_1 + c_2 Y_2\right)$ są dobrze określone. Wynika to z faktu, że funkcja Φ jest ograniczona z dołu oraz, że nośnik zmiennej U, $\operatorname{supp} U = [s - d_1 - d_2, \infty)$, oraz zmiennej V, $\operatorname{supp} V = [s - c_1 - c_2, \infty)$, zawierają się w nośniku funkcji Φ, $\operatorname{supp} \Phi = (-\sum_{j=1}^{n} c_j, \infty)$. Jest tak ponieważ

$$s - d_1 - d_2 \geq s - c_1 - c_2 \geq s_0 + c_1 + c_2 - c_1 - c_2 = -\sum_{j=1}^{n} c_j.$$

   - Ponadto widzimy, że $L' = s - (d_1 + d_2) > -\sum_{j=1}^{n} c_j = L$.
   - Różnica gęstości $f_V - f_U$ zmienia znak dokładnie trzy razy w przedziale (L', ∞) i jest dodatnia w punkcie L', co pokażemy w następnym kroku.

4. Gęstość zmiennej $aX_1 + bX_2$, dla $a \neq b$ zapisuje się jako

$$f_{aX_1+bX_2}(x) = \frac{e^{-x/a} - e^{-x/b}}{a - b} 1_{[0,\infty)}(x),$$

natomiast w przypadku $a = b$ wynosi

$$f_{aX_1+bX_2}(x) = \frac{1}{a^2} x e^{-x/a} 1_{[0,\infty)}(x).$$

Z założenia $L' = \operatorname{essinf} U$ oraz lematu 12 wnioskujemy że $f_U(L') = 0$, a zatem faktycznie różnica gęstości $f_V - f_U$ jest dodatnia w L'

$$f_V(L') - f_U(L') = f_V(L') > 0.$$

Pozostaje wykazać, że różnica gęstości $f_V - f_U$ posiada dokładnie 3 punkty zmiany znaku na przedziale (L', ∞). Dzięki równości dwóch pierwszych momentów zmiennych U, V, na mocy lematu 9 posiada ona co najmniej trzy punkty zmiany znaku.

W przypadku $c_1 \neq c_2$ jest ona postaci

$$(f_V - f_U)(x) = \sum_{j=1}^{4} a_j e^{-\eta_j x},$$

a zatem na mocy lematu 8 nie może mieć więcej niż trzy pierwiastki. Z kolei w przypadku $c_1 = c_2$ jest ona postaci

$$\begin{aligned}(f_V - f_U)(x) &= a_1 e^{-\eta_1 x} + a_2 e^{-\eta_2 x} + a_3 x e^{-\eta_3 x} \\ &= e^{-\eta_3 x}\left(a_1 e^{-(\eta_1-\eta_3)x} + a_2 e^{-(\eta_2-\eta_3)x} + a_3 x\right) =: e^{-\eta_3 x}\tilde{f}(x).\end{aligned}$$

Liczba pierwiastków funkcji $f_V - f_U$ jest równa liczbie pierwiastków funkcji $\tilde{f}$. Rozumując podobnie jak w dowodzie lematu 8, gdyby funkcja $\tilde{f}$ miała więcej niż trzy pierwiastki, to jej pochodna $\tilde{f}'$ na mocy twierdzenia Rolle'a miałaby więcej niż dwa pierwiastki. Przeczyłoby to lematowi 8, gdyż $\tilde{f}'$ jako suma trzech funkcji wykładniczych (jeden z składników to zdegenerowana funkcja wykładnicza równa stałej) nie może mieć więcej niż dwa pierwiastki. Wobec tego różnica gęstości $f_V - f_U$ na przedziale (L', ∞) posiada dokładnie trzy punkty zmiany znaku.

□

3.2. Wnioski

Wniosek 16. *Niech X_j będą niezależnymi zmienny losowymi o rozkładzie* $\mathrm{Gamma}(\alpha, \beta)$ *dla pewnych $\alpha, \beta > 0$, tj. o gęstości $f(x) = c \cdot x^{\alpha-1}e^{-\beta x}$ dla $x > 0$ i $c = \beta^\alpha/\Gamma(\alpha)$. Wówczas dla $\alpha = k \in \mathbb{N}$ i dowolnych $c, d \in S_+^{n-1}$ zachodzi*

$$(c_1^2, \ldots, c_n^2) \preceq (d_1^2, \ldots, d_n^2) \Rightarrow h\left(\sum_{j=1}^n d_j X_j\right) \le h\left(\sum_{j=1}^n c_j X_j\right).$$

Dowód. Bez starty ogólności możemy założyć że $\beta = 1$, gdyż dla $Y \sim \mathrm{Gamma}(\alpha, \beta)$ przeskalowana zmienna $Y' = \beta Y$ ma rozkład $\mathrm{Gamma}(\alpha, 1)$, a przeskalowanie zmiennych w postulowanej nierówności jest przekształceniem równoważnym.

Niech dane będą zatem zmienne losowe $X_{i,j} \overset{iid}{\sim} \mathrm{Exp}(1)$. Prawdą jest że $Y_j \sim X_{j,1} + \cdots + X_{j,k}$. Wprowadźmy stałe $d_{j,i} = d_j/\sqrt{k}$ oraz $c_{j,i} = c_j/\sqrt{k}$ dla $1 \le j \le n$ oraz $1 \le i \le k$. Możemy wówczas zapisać

$$h\left(\sum_{j=1}^n d_j Y_j\right) = h\left(\sqrt{k}\sum_{j=1}^n \sum_{i=1}^k d_{j,i} X_{j,i}\right) = \ln\sqrt{k} + h\left(\sum_{j=1}^n \sum_{i=1}^k d_{j,i} X_{j,i}\right).$$

Wybór współczynników $(c_{j,i})$ oraz $(d_{j,i})$ gwarantuje, że $\sum_{j=1,i=1}^{n,k} d_{j,i}^2 = \sum_{j=1,i=1}^{n,k} c_{j,i}^2 = 1$. Porządek $c^2 \preceq d^2$ jest zachowany gdy potraktujemy $(c_{j,i})$ oraz $(d_{j,i})$ jako wektory z S_+^{nk-1}. Poprzez wykorzystanie głównego twierdzenia otrzymujemy wówczas postulowaną nierówność

$$\ln\sqrt{k} + h\left(\sum_{j=1}^n \sum_{i=1}^k d_{j,i} X_{j,i}\right) \le \ln\sqrt{k} + h\left(\sum_{j=1}^n \sum_{i=1}^k c_{j,i} X_{j,i}\right) = h\left(\sum_{j=1}^n c_j Y_j\right).$$

□

Niestety, z pomocą prezentowanych technik, nie da się obejść założenia $\alpha \in \mathbb{N}$. Istnieje natomiast pokrewny rezultat, silniejszy w tym że dopuszcza niecałkowite wartości α, a słabszy w tym że stwierdza jedynie o maksymalności entropii dla sumy z różnymi wagami.

Twierdzenie 17. *[14] Niech X_j będą niezależnymi zmiennymi losowymi o rozkładzie* Gamma(α, β) *dla pewnych $\alpha, \beta > 0$. Wówczas dla $n \in \mathbb{N}$ i $\alpha \in \mathbb{R}$ takich, że $\alpha \geq 1/n$, oraz dowolnego wektora $d \in S_+^{n-1}$ zachodzi*

$$h\left(\sum_{j=1}^{n} d_j X_j\right) \leq h\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right).$$

3.3. Otwarte pytania

W związku z zaprezentowanymi rezultatami nasuwają się następujące możliwe uogólnienia oraz pokrewne pytania.

1. Czy główne twierdzenie jest prawdziwe dla szerszej klasy zmiennych, tj. innych niż zmienne wykładnicze i mieszanki gaussowskie? W szczególności, czy podobny fakt jest prawdziwy dla zmiennych jednostajnych $U([-a, a])$, bądź też dla zmiennych logwklęsłych?
2. Czy analogicznie jak dla mieszanek gaussowskich, główne twierdzenie jest prawdziwe dla mieszanek rozkładów wykładniczych?
3. Przy założeniach głównego twierdzenia prawdziwa jest nierówność
$$h(X_1 + X_2) \leq h\left(X_1 + \frac{X_2 + \ldots X_{n+1}}{\sqrt{n}}\right),$$
Można zatem zadać sobie zadać pytanie [12] dla jakiej klasy zmiennych losowych prawdziwa będzie nierówność
$$h(X_1 + X_2) \leq h(X_1 + G),$$
gdzie G jest niezależną od X_1 zmienna gaussowską o takiej samej wariancji?

3.4. Alternatywne techniki

Kluczowym komponentem dowodu głównego twierdzenia jest obserwacja, iż gęstość f_c zmiennej „docelowej" (docelowej, bo dochodzimy do niej transformacjami $T_{j,k}^{\lambda}$) spełnia

$$(-\ln f_c)''' \leq 0,$$

co jest niezbędne do skorzystania z *metody przecinających się gęstości.* W lemacie 10 skrupulatnie udowodniliśmy, że dla f będącej gęstością splotu zmiennych wykładniczych zachodzi

$$(\ln f(x))' + \max_j \lambda_j \geq 0, \ \ (\ln f(x))'' \leq 0 \text{ oraz } \ (\ln f(x))''' \geq 0.$$

Być może nie jest przypadkiem, że znak nierówności alternuje dla kolejnych pochodnych. Rozważmy na przykład f będącą splotem n zmiennych Gamma(α, β), wówczas

$$\frac{d^n}{dx^n} \ln f(x) = \frac{d^n}{dx^n} \ln\left(c_{\alpha,\beta} \cdot x^{n\alpha-1} e^{-\beta x}\right) = \begin{cases} (n\alpha - 1)x^{-1} - \beta, & \text{dla } n = 1, \\ (n\alpha - 1)(-1)^{n+1}(n!)x^{-n}, & \text{dla } n > 1. \end{cases}$$

Postać pochodnych daje nadzieję co do tego, że analog wniosku 11 będzie prawdziwy dla ogólniejszej klasy rozkładów. Ponadto, można zaobserwować że dla $n\alpha - 1 \geq 0$, tj. $\alpha \geq 1/n$ również zachodzi $(\ln f)''' \geq 0$, co jest w koincydencji z założeniami twierdzenia 17. Nasuwa się zatem następujące pytanie.

Pytanie 1. *Niech X będzie ciągłą rzeczywistą zmienną losową. Dla jakich rozkładów zmiennej losowej X, dla każdego $u \in \mathbb{R}^n$ lub $\mathbb{R}^n_+$ gęstość f_u sumy $u_1X_1 + \ldots u_nX_n$ spełnia*

a) $(\ln f_u)''' \geq 0$*?*

Można również zastanawiać się kiedy spełniony jest silniejszy warunek

b) $(-1)^{n+1}(\ln f_u)^{(n)} \geq 0$ *dla każdego* $n \geq 2$*?*

Postawione pytanie może budzić skojarzenia z teorią tzw. funkcji *całkowicie monotonicznych* [9, Podrozdział XIII.4], gdzie przedmiotem rozważań są funkcje $\varphi : (0, \infty) \to \mathbb{R}$ spełniające własność

$$\forall_{n\in\mathbb{N}}\ \forall_{x>0}\ (-1)^n\varphi^{(n)}(x) \geq 0.$$

Być może dałoby się za pomocą tej teorii uprościć i wzmocnić lemat 10 udzielając odpowiedzi na następujące ogólne pytanie odnośnie zachowywania własności całkowitej monotoniczności przy splataniu funkcji.

Pytanie 2. *Niech dane będą funkcje $f, g, h : \mathbb{R} \to \mathbb{R}$ i niech f będzie funkcją całkowicie monotoniczną. Dla jakich funkcji g i przy jakich ewentualnych dodatkowych założeniach odnośnie funkcji f, funkcja h zadana równaniem*

$$e^{-h(x)} = e^{-f(x)} * e^{-g(x)} = \int_{\mathbb{R}} e^{-f(x-y)}e^{-g(y)}dy = \int_{\mathbb{R}} e^{-f(y)}e^{-g(x-y)}dy$$

jest całkowicie monotoniczna? Co w przypadku gdy będziemy rozpatrywać jedynie f i g takie, że e^{-f} i e^{-g} są gęstościami prawdopodobieństwa? Co w przypadku gdy ograniczymy dziedzinę do przedziału $(0, \infty)$?

Bibliografia

[1] Shiri Artstein, Keith M. Ball, Franck Barthe, Assaf Naor, *Solution of Shannon's problem on the monotonicity of entropy*, Journal of the American Mathematical Society 17 (2004), s. 975–982.

[2] Keith Ball, Piotr Nayar i Tomasz Tkocz, *A reverse entropy power inequality for log-concave random vectors*, Studia Mathematica 235 (2016), s. 17–30.

[3] Andrew R. Barron, *Entropy and the central limit theorem*, The Annals of Probability 14 (1986), s. 336–342.

[4] Maciej Bartczak, Piotr Nayar i Szymon Zwara, *Sharp variance-entropy comparison for nonnegative gaussian quadratic forms*, preprint, arXiv:2005.11705.

[5] Thomas M. Cover i Joy A. Thomas, *Elements of information theory*, John Wiley & Sons, New York, 1991.

[6] Alexandros Eskenazis, Piotr Nayar i Tomasz Tkocz, *Gaussian mixtures: entropy and geometric inequalities*, Annals of Probability 46 (2018), s. 2908–2945.

[7] Alexandros Eskenazis, Piotr Nayar i Tomasz Tkocz, *Sharp comparison of moments and the log-concave moment problem*, Advances in Mathematics 334 (2018), s. 389–416.

[8] William Feller, *Wstęp do rachunku prawdopodobieństwa, tom II*, Wydawnictwo Naukowe PWN, Warszawa, 2009.

[9] Albert W. Marshall, Ingram Olkin i Barry C. Arnold, *Inequalities: theory of majorization and its applications*, Springer, New York, 2011.

[10] Michael R. Sampford, *Some inequalities on Mill's ratio and related functions*, The Annals of Mathematical Statistics 24 (1953), s. 130–132.

[11] Claude E. Shannon, *A mathematical theory of communication*, The Bell System Technical Journal 27 (1948), s. 379–423.

[12] Aart J. Stam, *Some inequalities satisfied by the quantities of information of Fisher and Shannon*, Information and Control 2 (1959), s. 101–112.

[13] Stanislaw J. Szarek i Elisabeth Werner, *A nonsymmetric correlation inequality for Gaussian measure*, Journal of Multivariate Analysis 68 (1999), s. 193–211.

[14] Alexandre B. Tsybakov, *Introduction to Nonparametric Estimation*, Springer, New York, 2009.