Notatki do wykładu
Elementy teorii macierzy losowych.

Radosław Adamczak

16 lipca 2010

Streszczenie

Niniejsze notatki zawierają skrótowe przedstawienie zagadnień prezentowanych w ramach wykładu prowadzonego na Wydziale MIM UW w semestrze zimowym 2009/2010. Ponieważ pisane były na bieżąco oraz nigdy nie zostały poddane dokładnej korekcie, z prawdopodobieństwem bliskim 1 występują w nich literówki oraz (miejmy nadzieję, że drobne) nieścisłości. Czytelników, którzy zauważą usterki lub mają jakieś inne uwagi, uprzejmie proszę o kontakt mailowy na adres `radamcz@mimuw.edu.pl`.

Na ostatnich stronach notatek można znaleźć listę zadań rozwiązywanych na ćwiczeniach wraz z zadaniami egzaminacyjnymi.

1 Czym są macierze losowe? Wybrane przykłady zastosowań.

Intuicyjnie macierz losowa to po prostu macierz, której współczynniki zostały wygenerowane w pewien losowy sposób, np. poprzez rzuty kostką.

Bardziej formalnie, przez macierz losową rozumiemy zmienną losową o wartościach w przestrzeni $\mathcal{M}_K(m,n)$ macierzy m na n nad ciałem K (w dalszym ciągu ograniczymy się do $K = \mathbb{R}$ oraz $K = \mathbb{C}$, niemniej rozważa się również macierze losowe o współczynnikach z ciał dyskretnych lub kwaternionowych), a więc funkcję mierzalną

$$A\colon (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathcal{M}_K(m,n), \mathcal{B}),$$

gdzie $(\Omega, \mathcal{F}, \mathbb{P})$ jest pewną przestrzenią probabilistyczną, zaś $\mathcal{B}$ jest σ-ciałem zbiorów borelowskich (traktujemy przestrzeń $\mathcal{M}_K(m,n)$ jako podprzestrzeń K^{nm}). Teoria macierzy losowych zajmuje się badaniem własności tego typu obiektów, gdy $m, n \to \infty$. Oczywiście powyższa definicja jest zbyt ogólna, by mogła stanowić podstawę interesującej teorii. W praktyce rozważa się konkretne typy macierzy losowych, pojawiające się w zastosowaniach (np. w fizyce, statystyce czy informatyce). Poniżej przedstawionych zostanie pokrótce kilka przykładowych typów macierzy losowych i ich zastosowań. Lista ta nie jest oczywiście kompletna, a przykłady mają jedynie charakter wprowadzający.

1.1 Analiza składowych głównych

Niech $X, X_1, X_2, \ldots, X_n$ będzie ciągiem niezależnych wektorów losowych o tym samym rozkładzie w $\mathbb{R}^d$. O pojedynczym wektorze możemy myśleć jako o zbiorze danych charakteryzujących losowy element pewnej populacji (np. w badaniach medycznych poszczególne współrzędne wektora X_i mogą odpowiadać wzrostowi, wadze, ciśnieniu, etc. i-tego pacjenta czy też rocznemu przychodowi, wydatkom, liczbie osób, etc. w gospodarstwie domowym wylosowanym do badania statystycznego).

Jednym z podstawowych problemów analizy statystycznej jest opisanie wzajemnych zależności między różnymi charakterystykami elementu populacji najlepiej w jak najprostszy sposób, ale przy niezbyt dużej utracie informacji. Jedną z pierwszych metod użytych w tym celu jest tzw. *analiza składowych głównych*. Jako miarę zależności między zmiennymi losowymi przyjmuje się w tej metodzie kowariancję, a sama metoda sprowadza się do znalezienia nowej bazy w $\mathbb{R}^d$, w której poszczególne współrzędne wektora X byłyby nieskorelowane.

Niech zatem C będzie macierzą kowariancji wektora X, tzn. $C := \mathbb{E}(X - \mathbb{E}X)(X - \mathbb{E}X)^T$ (wektory dla wygody będziemy zapisywać w formie kolumnowej). Jest to macierz symetryczna i nieujemnie określona, zatem posiada d (licząc z krotnościami) rzeczywistych wartości własnych $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_d$ oraz w pewnej bazie ortogonalnej $e_1, \ldots, e_d$ ma postać

$$\begin{bmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & \ldots & 0 & \ddots & 0 \\ 0 & \ldots & 0 & 0 & \lambda_d \end{bmatrix}.$$

Jeżeli od pewnego i, kolejne wartości własne zaczynają szybko maleć, przyjmuje się, że odpowiadające im współrzędne (w nowej bazie) reprezentują szum, a istotna informacja zawarta jest w początkowych współrzędnych. Nową reprezentacją wektora X jest rzut ortogonalny PX na $\mathrm{span}(e_1, \ldots, e_i)$. Uzyskuje się więc w ten sposób pewną kompresję danych (można wykazać, że jest to najlepsza kompresja liniowa, w sensie średniokwadratowym, tzn. że minimalizuje wyrażenie $\mathbb{E}|X - QX|$ po wszystkich przekształceniach liniowych Q rzędu i). Następnie, nowe zmienne poddaje się dalszej analizie, aby wyciągnąć z nich wnioski na temat populacji.

Nie będziemy wnikać w szczegóły metodologiczne oraz w problem interpretacji nowych zmiennych, przekonamy się natomiast, że w praktyce analiza składowych głównych prowadzi do zagadnień matematycznych związanych z macierzami losowymi.

Zauważmy, że aby znaleźć wartości i wektory własne macierzy C, musimy znać rozkład wektora X, który z reguły jest dla nas niedostępny, analiza statystyczna ma na celu właśnie zbadanie tego rozkładu na podstawie danych empirycznych. Możemy jednak przybliżyć macierz C na podstawie $X_1, \ldots, X_n$, korzystając z prawa wielkich liczb. Mamy

$$\mathbb{E}X \simeq \frac{1}{n}\sum_{i=1}^{n} X_i$$

oraz

$$C \simeq \tilde{C} = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^T, \quad \text{gdzie} \quad Y_i = X_i - \frac{1}{n}\sum_{i=1}^{n} X_i$$

(powyższy estymator jest obciążony, ale nie będziemy tu wnikać w subtelności natury statystycznej).

Zauważmy, że $\tilde{C}$ jest **macierzą losową**.

Jeżeli wymiar d naszych danych jest stosunkowo niski i dysponujemy dużą próbką, powyższe przybliżenia są wystarczająco dobre. Jednak współcześnie dość często mamy do czynienia z sytuacjami, gdy wymiar d jest duży, porównywalny z n (np. jeśli X_i reprezentują dane genetyczne). W naturalny sposób pojawia się więc następujące pytanie:

- jakie warunki muszą być spełnione, żeby powyższa metoda dostarczała nam informacji na temat wektorów i wartości własnych macierzy C, czyli kiedy macierz $\tilde{C}$ jest wystarczająco dobrym przybliżeniem C?

1.2 Poziomy energetyczne w fizyce jądrowej

W fizyce kwantowej układy są opisane przez równanie Schrödingera, postaci

$$H\psi = \lambda\psi,$$

gdzie ψ jest elementem ośrodkowej przestrzeni Hilberta, opisującym stan układu, zaś H jest operatorem samosprzężonym na tej przestrzeni.

Liczby λ dla których powyższe równanie ma nietrywialne rozwiązanie to oczywiście wartości własne H. Fizycznie odpowiadają one poziomom energetycznym, które mogą być przyjmowane przez układ.

Jednym z podstawowych problemów przy opisie układów kwantowych jest scharakteryzowanie możliwych poziomów energetycznych. Niestety, w wielu sytuacjach równanie Schrödingera jest zbyt skomplikowane, by można je było rozwiązać, co sugeruje, że podobnie jak w przypadku termodynamiki, być może trzeba uciec się do opisu statystycznego. Sytuacja jest tu zresztą bardziej skomplikowana niż w fizyce klasycznej, gdyż często nie jest nawet znana dokładna postać operatora H.

Pomysłem Wignera, który pod koniec lat pięćdziesiątych dwudziestego wieku dał początek rozwojowi teorii macierzy losowych w fizyce, było, aby spojrzeć na „typowy" operator samosprzężony H, poprzez ograniczenie się do skończonego (ale dużego) wymiaru oraz wprowadzenie miary na wszystkich macierzach symetrycznych w tym wymiarze. Z rozważań fizycznych wynika, że miara ta powinna być niezmiennicza ze względu na ortogonalną zmianę bazy, dodatkowo wprowadzono założenie, że współczynniki macierzy są niezależne (mod. symetria macierzy).

Okazuje się, że te założenia istotnie ograniczają klasę rozpatrywanych macierzy losowych, w szczególności ich współczynniki muszą być zmiennymi gaussowskimi. Choć początkowe wyniki teoretyczne dot. zachowania wartości własnych tych macierzy losowych w dużej skali nie pokrywały się z danymi eksperymentalnymi, bardziej szczegółowe rezultaty dotyczące lokalnego zachowania wartości własnych okazały się być zaskakująco zgodne z doświadczeniem.

Spowodowało to duże zainteresowanie modelami opartymi na macierzach losowych wśród fizyków. Często są one świadomie używane jako pewnego rodzaju „czarne skrzynki", które mają opisywać złożone układy, dla których reguły ewolucji nie są znane. W innych przypadkach zgodność wyników eksperymentalnych z istniejącym modelem macierzy losowej lub też jej brak są wskazówkami do zaproponowania modelu fizycznego, opisującego dane zjawisko.

1.3 Compressed sensing

Kolejny przykład zastosowania macierzy losowych dotyczy pomiarów dyskretnych sygnałów wysokowymiarowych. Wyobraźmy sobie, że chcemy odtworzyć wektor $x \in \mathbb{R}^N$, a potrafimy ustalić (zmierzyć) wartość dowolnego funkcjonału liniowego $\langle y, x\rangle$ ($y \in \mathbb{R}^N$). Sytuacja taka ma często miejsce w praktyce, co więcej wymiar sygnału (N) może być bardzo duży (np. gdy wektor x koduje obraz, tego typu zastosowania pojawiają się np. w medycynie). Naturalnie, każdy kolejny pomiar wiąże się z pewnym kosztem, należy więc w miarę możliwości wykorzystać dodatkową wiedzę o strukturze sygnału, aby zmniejszyć liczbę pomiarów. W wielu zagadnieniach (np. we wspomnianym wyżej problemie obrazowania medycznego czy przy przetwarzaniu dźwięku) można znaleźć bazę w $\mathbb{R}^N$ o tej własności, że wszystkie lub istotna większość analizowanych wektorów ma w niej bardzo małą liczbę (w stosunku do N) niezerowych współrzędnych (na tym spostrzeżeniu opiera się wiele metod kompresji obrazów, np. jpeg, bazą o pożądanych własnościach jest w tym przypadku odpowiednia baza falkowa). Przypuśćmy dla uproszczenia, że wektor x jest już wyrażony w takiej bazie i oznaczmy liczbę jego niezerowych współrzędnych przez $m << N$. Gdybyśmy znali nośnik wektora x, do zidentyfikowania go wystarczyłoby nam oczywiście m pomiarów (wzdłuż kierunków odpowiadających odpowiednim współrzędnym). Nośnik ten jest jednak z reguły nieznany, często zresztą może nim być dowolny zbiór współrzędnych o mocy co najwyżej m. Jak w takim razie wyznaczyć x przy pomocy niewielkiej liczby pomiarów?

Pytanie to dało początek burzliwie rozwijającej się w ostatnich latach dziedzinie „compressed sensing". Jedną z popularnych metod wyznaczania x jest zmierzenie składowych tego wektora w n ($n \simeq m$) niezależnych kierunkach (co matematycznie odpowiada przemnożeniu x przez pewną macierz A o wymiarach n na N), a następnie wybranie spośród wszystkich wektorów zgodnych z wynikiem pomiarów, wektora $\hat{x}$, minimalizującego normę ℓ_1. Formalnie,

$$\hat{x} = \operatorname{argmin}\{\|y\|_1 \colon y \text{ t.że } Ax = Ay\},$$

gdzie $\|y\|_1 = |y_1| + \ldots + |y_N|$.

Bardziej intuicyjnym podejściem może się wydawać znalezienie wśród wektorów zgodnych z pomiarami wektora o najmniejszym nośniku, czyli minimalizacja $\|y\|_0 := |\{i\colon y_i \neq 0\}|$, problem ten jest jednak trudny algorytmicznie, podczas gdy powyższa metoda (nazywana czasami metodą *basis pursuit*) może być efektywnie zaimplementowana (ponieważ $\|y\|_1$ jako norma jest funkcją wypukłą). Co więcej, przy odpowiednim doborze macierzy A daje ten sam wynik dla wektorów o co najwyżej m niezerowych współrzędnych. Dokładniej, istnieje stała uniwersalna $c > 0$ taka że dla dowolnych $n \leqslant N$ można znaleźć macierz A o tej własności, że ilekroć x ma co najwyżej $m = cn/\log(2N/n)$ współrzędnych niezerowych, zachodzi równość $\hat{x} = x$. Innymi słowy, liczba pomiarów niezbędnych aby wyznaczyć sygnał o m niezerowych współrzędnych wzrasta co najwyżej o czynnik logarytmiczny (zauważmy, że gdy m jest proporcjonalne do N, tracimy jedynie czynnik stały).

Jaki jest związek powyższego zagadnienia z macierzami losowymi? Jak dotąd jedyne znane konstrukcje macierzy A o dobrych własnościach rekonstrukcji sygnałów są zrandomizowane. Macierzami o najlepszych własnościach są macierze o niezależnych współczynnikach $\mathcal{N}(0,1)$ lub macierze znaków losowych (współczynniki są niezależne o rozkładzie symetrycznym na $\{+1,-1\}$). Nieco gorszą rekonstrukcję można uzyskać przy pomocy cząstkowej macierzy Fouriera, czyli macierzy o n wierszach wylosowanych z pełnej macierzy Fouriera. Tego typu macierze mają z kolei dobre własności numeryczne.

Podobnie jak w poprzednich przykładach, dowód powyższych faktów wiąże się z analizą wartości własnych odpowiednich macierzy losowych związanych z A.

2 Macierze Wignera. Zbieżność miary spektralnej

Materiał przedstawiony w tym rozdziale został zaczerpnięty w dużym stopniu z pozycji [G] oraz [AGZ]. Idea kombinatorycznego dowodu głównego twierdzenia pochodzi od Wignera, może być ona jednak sformalizowana na wiele sposobów. Poniższe przedstawienie jest modyfikacją i jak mi się wydaje pewnym uproszczeniem formalizmów ze wspomnianych książek.

Macierzami Wignera nazywamy symetryczne (hermitowskie) macierze losowe, których współczynniki na i powyżej przekątnej są niezależnymi zmiennymi losowymi. Macierze te zostały wprowadzone w pierwszych pracach Wignera, zaś przedstawione poniżej twierdzenie dotyczące granicznego rozkładu ich miary spektralnej można uważać za początek asymptotycznej teorii macierzy losowych. W rozważaniach fizycznych często dodatkowo zakłada się, że współczynniki mają rozkład normalny (rzeczywisty lub zespolony), co daje tym macierzom dodatkową symetrię (niezmienniczość rozkładu ze względu na sprzężenia macierzami ortogonalnymi lub unitarnymi).

2.1 Wstępne definicje i założenia

Rozważmy nieskończoną tablice trójkątną $(X_{ij})_{1\leqslant i\leqslant j<\infty}$ niezależnych zmiennych losowych. Dla dowolnej liczby całkowitej $N \geqslant 1$ zdefiniujmy macierz

$$A_N = [X_{ij}]_{i,j\leqslant N},$$

gdzie dla $i > j$ przyjmujemy $X_{ij} := X_{ji}$.

Macierz A_N jest symetryczna, zatem posiada N (licząc z krotnościami) rzeczywistych wartości własnych $\lambda_1(A_N) \leqslant \lambda_2(A_N) \leqslant \ldots \leqslant \lambda_N(A_N)$. Podstawowym pytaniem teorii macierzy losowych jest zachowanie asymptotyczne wartości własnych dla $N \to \infty$. Pojawia się tu pewien problem formalny, mianowicie wymiar wektora złożonego z wartości własnych zmienia się wraz z N, więc *a priori* nie jest jasne w jaki sposób powinniśmy tego typu wektory porównywać (co to znaczy, że wektory są bliskie, lub że ciąg wartości własnych zbiega). Problem ten rozwiązuje się kodując informację o ciągu $(\lambda_i(A))_{i\leqslant N}$ przy pomocy dyskretnej miary probabilistycznej.

Definicja 1. *Miarą spektralną macierzy symetrycznej (hermitowskiej) $A \in \mathcal{M}(N,N)$ nazwiemy* *miarę probabilistyczną L_N na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, daną wzorem,*

$$L_A = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(A)},$$

gdzie δ_x jest miarą Diraca skupioną w x.

Uwaga Dla każdego zbioru borelowskiego $I \subseteq \mathbb{R}$,

$$L_A(E) = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}_{\{\lambda_i(A)\in I\}} = \frac{|\{i \leqslant N \colon \lambda_i(A) \in I\}|}{N}.$$

Miara spektralna zbioru I informuje nas więc ile (jaka część) wartości własnych macierzy A należy do zbioru I. Oczywiście na podstawie miary spektralnej możemy odtworzyć ciąg $\lambda_i(A)$ (dystrybuanta L_A ma skoki dokładnie w punktach $\lambda_i(A)$, zaś wielkość skoku wyznacza krotność wartości własnej). Co więcej, L_A jest elementem przestrzeni borelowskich miar probabilistycznych na $\mathbb{R}$, na której możemy rozpatrywać znaną z rachunku prawdopodobieństwa topologię słabej zbieżności. Rozwiązuje to problem porównywania ciągów wartości własnych macierzy o różnych wymiarach, po utożsamieniu ich z miarami probabilistycznymi wszystkie takie ciągi są tego samego „typu".

Aby móc sformułować interesujące twierdzenie dotyczące asymptotycznego zachowania wartości własnych macierzy A_N, musimy narzucić dodatkowe ograniczenia na zmienne losowe X_{ij}. Załóżmy zatem, że

(A1) dla dowolnych $i \leqslant j$, $\mathbb{E}X_{ij} = 0$,

(A2) dla dowolny $i < j$, $\mathbb{E}X_{ij}^2 = 1$,

(A3) dla dowolnego $k \in \mathbb{N}$, $m_k := \sup_{i,j} \mathbb{E}|X_{ij}|^k < \infty$.

Okazuje się, że aby opisać asymptotyczne zachowanie wartości własnych macierzy A_N wygodnie jest je przeskalować i rozpatrywać macierze $N^{-1/2}A_N$. Intuicyjnie, skalowanie $N^{-1/2}$ można wyjaśnić faktem, że jeżeli rozpatrzymy macierz bez diagonali, to z założenia (A2), drugi moment normy euklidesowej dowolnego wiersza/kolumny wynosi $\sqrt{N-1} \simeq \sqrt{N}$.

Oznaczmy zatem przez $\lambda_1^N \leqslant \lambda_2^N \leqslant \ldots \leqslant \lambda_N^N$ oraz L_N wartości własne oraz miarę spektralną macierzy $N^{-1/2}A_N$.

Zauważmy, że λ_i^N są zmiennymi losowymi (formalnie wynika to z ciągłości wartości własnych macierzy symetrycznej jako funkcji macierzy), zaś L_N miarą losową na prostej.

Okazuje się, że przy powyższych założeniach, miary L_N są słabo zbieżne p.n. Zanim sformułujemy odpowiednie twierdzenie, zdefiniujmy miarę graniczną.

Definicja 2. *Rozkładem Wignera nazwiemy miarę probabilistyczną σ na $\mathbb{R}$, o gęstości danej wzorem*

$$g(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x).$$

Uwaga Ze względu na postać gęstości, miara Wignera jest często nazywana *semicircle law.*

Twierdzenie 1. *Jeśli $(X_{ij})_{i\leqslant j<\infty}$ są niezależnymi zmiennymi losowymi, spełniającymi założenia (A1)-(A3), to z prawdopodobieństwem 1, miara L_n zbiega słabo do miary σ, czyli*

$$\mathbb{P}(L_n \xrightarrow{D} \sigma) = 1.$$

W szczególności, dla dowolnego przedziału I,

$$\frac{|i \leqslant N \colon \lambda_i^N \in I|}{N} \to \frac{1}{2\pi}\int_I \sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x)dx \text{ p.n.}$$

Uwaga Druga część twierdzenia wynika z pierwszej i absolutnej ciągłości σ względem miary Lebesgue'a (jest to standardowy fakt dot. słabej zbieżności).

Zanim udowodnimy powyższe twierdzenie, musimy przypomnieć kilka dość standardowych faktów kombinatorycznych oraz dotyczących słabej zbieżności miar. Zostaną one zebrane w następnych dwóch rozdziałach wraz z informacją, gdzie można znaleźć ich dowody.

2.2 Liczby Catalana

Definicja 3. *Dla $k \in \mathbb{N}$, k-tą liczbę Catalana C_k definiujemy wzorem*

$$C_k = \frac{\binom{2k}{k}}{k+1} = \frac{(2k)!}{(k+1)!k!}$$

Liczby Catalana mają duże znaczenie w kombinatoryce. Poniżej przedstawiamy kilka ich wybranych interpretacji kombinatorycznych oraz podstawowych własności. Dowody pozostawiamy jako ćwiczenie.

Fakt 1. *1. C_k jest*

a) liczbą poprawnych nawiasowań złożonych z k par nawiasów $(,)$.

b) liczbą ścieżek Dycka długości $2k$, tzn. ciągów $x = (x_0, x_1, \ldots, x_{2k})$, takich że
- $x_0 = x_{2k} = 0$,
- $x_i \in \mathbb{N}$, $i = 0, 1, \ldots, 2k$,
- $|x_{i+1} - x_i| = 1$, $i = 0, 1, \ldots, 2k-1$.

c) liczbą nieprzecinających się *podziałów zbioru $\{1, \ldots, 2k\}$ na zbiory dwuelementowe, tzn. takich podziałów $\mathcal{P}$, dla których* **nie** *istnieją liczby $1 \leqslant a < b < c < d \leqslant k$, takie że $\{a, c\}, \{b, d\} \in \mathcal{P}$.*

d) liczbą triangulacji $(n+2)$-kąta wypukłego przy użyciu nieprzecinających się przekątnych.

2. Dla $k \geqslant 1$, $C_k = C_0 C_{k-1} + C_1 C_{K-2} + \ldots + C_{k-1} C_0$.

Aby przedstawić kolejną interpretację kombinatoryczną liczb Catalana, musimy wprowadzić dodatkową definicję.

Definicja 4. *Drzewem zorientowanym nazwiemy drzewo (czyli graf spójny, acykliczny), w którym wyróżniony został korzeń (a zatem także struktura „genealogiczna") oraz liniowy porządek na zbiorze bezpośrednich potomków dowolnego wierzchołka v.*

Przez izomorfizm dwóch drzew zorientowanych rozumiemy bijekcję między ich zbiorami wierzchołków, przeprowadzającą korzeń, na korzeń oraz zachowującą relacje sąsiedztwa i porządku.

Fakt 2. *Z dokładnością do izomorfizmu, liczba drzew zorientowanych o $k+1$ wierzchołkach (k krawędziach) wynosi C_k.*

Do dowodu Twierdzenia 1 będziemy potrzebowali związku liczb Catalana z miarą Wignera. Opisuje je

Fakt 3. *Dla $k \in \mathbb{N}$ zachodzą następujące równości:*

$$\int_{\mathbb{R}} x^{2k} d\sigma(x) = C_k,$$

$$\int_{\mathbb{R}} x^{2k+1} d\sigma(x) = 0.$$

2.3 Zbieżność momentów, a słaba zbieżność

Dowód Twierdzenia 1, który przedstawimy, oparty jest na związku między słabą zbieżnością miar probabilistycznych, a zbieżnością ich momentów. Poniższe dwa fakty są dość standardowe, ich dowody można znaleźć w większości podręczników rachunku prawdopodobieństwa.

Twierdzenie 2. *Dowolna miara probabilistyczna μ o zwartym nośniku na $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ jest wyznaczona jednoznacznie przez ciąg swoich momentów, $\int_{\mathbb{R}} x^k d\mu(x)$, $k = 0, 1, 2, \ldots$*

Twierdzenie 3. *Niech μ, μ_n, $n \in \mathbb{N}$ będą borelowskimi miarami probabilistycznymi na $\mathbb{R}$, o wszystkich momentach skończonych. Jeśli dla dowolnego $k \in \mathbb{N}$ zachodzi*

$$\lim_{n\to\infty} \int_{\mathbb{R}} x^k d\mu_n(x) = \int_{\mathbb{R}} x^k d\mu(x)$$

oraz miara μ jest wyznaczona jednoznacznie przez swoje momenty, to $\mu_n \xrightarrow{D} \mu$.

2.4 Zbieżność średniej miary spektralnej – twierdzenie Wignera

Aby zrozumieć w jaki sposób twierdzenia opisane w poprzednim paragrafie mogą zostać użyte do dowodu Twierdzenia 1, spróbujmy ustalić związek pomiędzy momentami miary L_A, a współczynnikami macierzy A.

Z definicji miary L_A mamy

$$\int_R x^k dL_A(x) = \frac{1}{N} \sum_{i=1}^{N} \lambda_i(A)^k.$$

Macierz A, jako macierz symetryczna, diagonalizuje się w pewnej bazie ortogonalnej, tzn. istnieje przekształcenie $U \in O_N$, takie że $A = UDU^{-1}$, gdzie $D = \mathrm{Diag}(\lambda_1(A), \ldots, \lambda_N(A))$. Zatem $A^k = UD^kU^{-1}$, skąd wynika, że $\mathrm{tr}\, A^k = \mathrm{tr}\, D^k = \lambda_1(A)^k + \ldots + \lambda_N(A)^k$. Widzimy więc, że dla dowolnego $k \in \mathbb{N}$ zachodzi równość

$$\int_{\mathbb{R}} x^k dL_A(x) = \frac{1}{N} \mathrm{tr}\, A^k. \tag{1}$$

Wyrażając ślad A^k przy pomocy współczynników macierzy A (przyjmijmy, że $A = [a_{ij}]_{i,j=1}^N$), dostajemy

$$\int_{\mathbb{R}} x^k dL_A(x) = \frac{1}{N} \sum_{i_1,\ldots,i_k=1}^{N} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} a_{i_k i_1}. \tag{2}$$

W badanym przez nas modelu losowym współczynniki są niezależnymi zmiennymi losowymi (z dokładnością do symetrii macierzy), co znacznie ułatwia analizę skomplikowanej sumy z ostatniej równości i pozwala zredukować dowód Twierdzenia 1 do rozważań kombinatorycznych. Poniższe twierdzenie, udowodnione przez Wignera, jest pierwszym krokiem w stronę dowodu Twierdzenia 1.

Twierdzenie 4 (Wigner)**.** *Przy założeniach Twierdzenia 1, dla dowolnego $k \in \mathbb{N}$ zachodzi*

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E}\mathrm{tr}\, (\frac{1}{\sqrt{N}} A_N)^k = \int_{\mathbb{R}} x^k d\sigma(x).$$

Uwaga Powyższe twierdzenie może być zinterpretowane jako zbieżność *średniej miary spektralnej* macierzy $N^{-1/2} A_N$ do rozkładu Wignera. Rzeczywiście, jeśli zdefiniujemy miarę $\bar{L}_N = \mathbb{E} L_N$ (tzn. $\bar{L}_N(I) = \mathbb{E} L_N(I)$ dla dowolnego zbioru borelowskiego $I \subseteq \mathbb{R}$), to jak łatwo sprawdzić,

$$\int_{\mathbb{R}} x^k d\bar{L}_N(x) = \frac{1}{N} \mathbb{E}\mathrm{tr}\, (\frac{1}{\sqrt{N}} A_N)^k.$$

Zatem twierdzenie Wignera, w połączeniu z Twierdzeniami 2 i 3, implikują, że $\bar{L}_n \xrightarrow{D} \sigma$.

Dowód Twierdzenia 4. Oznaczmy

$$B_{k,N} = \frac{1}{N}\mathbb{E}\mathrm{tr}\,(\frac{1}{\sqrt{N}}A_N)^k = \frac{1}{N^{k/2+1}} \sum_{i_1,\ldots,i_k=1}^{N} \mathbb{E}(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}),$$

gdzie w drugiej równości użyliśmy (2) oraz liniowości wartości oczekiwanej.

Idea dowodu polega na pogrupowaniu indeksów $\mathbf{i} = (i_1, \ldots, i_k)$ w rozłączne zbiory i pokazaniu, że wkład części z tych zbiorów do powyższej sumy jest zaniedbywalny, podczas gdy każdy z pozostałych zbiorów daje asymptotycznie wkład równy $N^{k/2+1}$. W tym celu wygodnie jest wprowadzić dodatkowe oznaczenia.

- Indeks $\mathbf{i} = (i_1, \ldots, i_k)$ będziemy utożsamiać z zamkniętą ścieżką $i_1 \to i_2 \to \ldots \to i_k \to i_1$ w pełnym grafie o wierzchołkach $1, 2, \ldots, N$.
- Na indeksach wprowadzamy relację równoważności

  $$\mathbf{i} = (i_1, \ldots, i_k) \sim_N \mathrm{j} = (j_1, \ldots, j_k)$$

  wtedy i tylko wtedy, gdy istnieje bijekcja $f \colon \{1, \ldots, N\} \to \{1, \ldots, N\}$, taka że $f(i_l) = j_l$ dla wszystkich $l \leqslant k$.

  Łatwo zauważyć, że dla każdego indeksu $\mathbf{i} \in \{1, \ldots, N\}^k$ istnieje indeks $\mathrm{j} \in \{1, \ldots, k\}^k$, taki że $\mathbf{i} \sim_N \mathrm{j}$. Innymi słowy, klasy abstrakcji relacji $\sim_N$ zależą tak naprawdę tylko od k. W szczególności istnieje podzbiór $\mathcal{K} \subseteq \{1, \ldots, k\}$ o tej własności, że dla dowolnego N, dowolna klasa abstrakcji relacji $\sim_N$ ma dokładnie jeden element wspólny z $\mathcal{K}$. Na potrzeby dalszej części dowodu ustalmy jeden taki zbiór (nie będzie on istotny w dowodzie, ale pozwoli nieco uprościć notację).

  Od tej pory będziemy pisali $\mathbf{i} \sim \mathrm{j}$ zamiast $\mathbf{i} \sim_N \mathrm{j}$ (w świetle powyższych uwag nie prowadzi to do niejednoznaczności).
- Z dowolną ścieżką $\mathbf{i}$ możemy związać graf $G(\mathbf{i})$ o **zbiorze** wierzchołków $V(\mathbf{i}) = \{i_1, \ldots, i_k\}$ oraz **multizbiorze** krawędzi $E(\mathbf{i}) = \{(i_j, i_{j+1})\}_{j=1,\ldots,k}$, gdzie przyjmujemy $i_{k+1} = i_1$.

  Zauważmy, że graf $G(\mathbf{i})$ może mieć krawędzie wielokrotne oraz krawędzie postaci (i, i).
- Szkieletem grafu G nazwiemy graf $\tilde{G} = (\tilde{V}, \tilde{E})$, otrzymany z G przez zignorowanie krotności i orientacji krawędzi. Zauważmy, że dla każdego $\mathbf{i}$ graf $\tilde{G}(\mathbf{i})$ jest spójny. co więcej, jeżeli $\mathbf{i} \sim \mathrm{j}$, to grafy $G(\mathbf{i})$, $G(\mathrm{j})$ (a więc także $\tilde{G}(\mathbf{i})$ i $\tilde{G}(\mathrm{j})$) są izomorficzne.
- Oznaczmy dodatkowo $P(\mathbf{i}) = \mathbb{E}(X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1})$.
- Dla niezorientowanej krawędzi e grafu $\tilde{G}(\mathbf{i})$, zdefiniujmy krotność e względem $\mathbf{i}$ (ozn. $d_{\mathbf{i}}(e)$) jako liczbę krawędzi grafu $G(\mathbf{i})$ odpowiadających e (liczba $d_{\mathbf{i}}(e)$ mówi nam więc ile razy niezorientowana krawędź e została odwiedzona w ścieżce $i_1 \to i_2 \to \ldots \to i_k \to i_1$.)

Możemy teraz zapisać

$$B_{k,N} = \frac{1}{N^{k/2+1}} \sum_{a \in \mathcal{K}} \sum_{\substack{\mathbf{i} \in \{1,\ldots,N\}^k \\ \mathbf{i} \sim a}} P(\mathbf{i}) \tag{3}$$

Zauważmy, że dzięki równości $X_{ij} = X_{ji}$, dla dowolnej niezorientowanej krawędzi $e = \{i, j\}$ grafu pełnego o N wierzchołkach możemy zdefiniować $X_e = X_{ij}$ (zwróćmy uwagę, że dopuszczamy tu przypadek $i = j$).

Ponadto, z niezależności zmiennych $(X_{ij})_{i \leqslant j}$ oraz symetrii macierzy A_N, $P(\mathbf{i}) = \prod_{e \in \tilde{G}(\mathbf{i})} \mathbb{E} X_e^{d_{\mathbf{i}}(e)}$.

W szczególności, z założenia $\mathbb{E}X_{ij} = 0$, jeżeli ścieżka $\mathbf{i}$ przechodzi przez pewną krawędź tylko raz, to $P(\mathbf{i}) = 0$. W równaniu (3) możemy się więc ograniczyć do sumowania po ścieżkach a, w których każda krawędź ma krotność przynajmniej 2. Wynika stąd w szczególności, że graf $\tilde{G}(a) = (\tilde{V}(a), \tilde{E}(a))$ ma co najwyżej $\lfloor k/2 \rfloor$ krawędzi.

Przypomnijmy teraz klasyczny lemat z teorii grafów.

Lemat 1. *Dowolny graf spójny G o zbiorze wierzchołków V i zbiorze (niezorientowanych) krawędzi E spełnia nierówność*

$$|V| \leqslant |E| + 1.$$

Co więcej, równość zachodzi wtedy i tylko wtedy, gdy G jest drzewem.

Z powyższego lematu wynika, że

$$|\tilde{V}(a)| \leqslant |\tilde{E}(a)| + 1 \leqslant \lfloor k/2 \rfloor + 1. \tag{4}$$

Zauważmy, że $l = |\tilde{V}(a)|$ jest liczbą różnych wartości przyjmowanych przez ciąg a. Zatem dla dużych N istnieje

$$N(N-1)\dots(N-l+1) \leqslant N^l$$

indeksów $\mathbf{i} \in \{1, \dots, N\}^k$, takich że $\mathbf{i} \sim a$. Ponadto, z ograniczoności momentów zmiennych X_{ij} (założenie (A3)) wynika, że $P(\mathbf{i})$ jest ograniczone przez stałą zależną tylko od k. Wynika stąd, że jeżeli $l < k/2 + 1$, to

$$\frac{1}{N^{k/2+1}} \sum_{\substack{\mathbf{i} \in \{1,\dots,N\}^k \\ \mathbf{i} \sim a}} P(\mathbf{i}) \leqslant M(k) N^{l-k/2-1} \to 0 \text{ dla } N \to \infty.$$

Z powyższej obserwacji i nierówności (4) wynika w szczególności, że dla k nieparzystego

$$\lim_{N\to\infty} B_{k,N} = 0 = \int_{\mathbb{R}} x^k d\sigma(x).$$

Ponadto, jeżeli k jest parzyste ($k = 2m$), to jedyne ścieżki a, które mogą mieć asymptotyczny wpływ na $B_{k,N}$ to takie, że

$$|\tilde{V}(a)| = m + 1 = |\tilde{E}(a)| + 1.$$

Zgodnie z drugą częścią lematu, powyższa równość implikuje, że $\tilde{G}(a)$ jest drzewem. Ponadto, kolejność w jakiej odwiedzane są wierzchołki $\tilde{G}(a)$ w ścieżce $a = (a_1, \dots, a_k)$ zadaje na tym drzewie orientację (przyjmujemy, że korzeniem jest a_1, zaś wśród bezpośrednich potomków dowolnego wierzchołka większe są te, które odwiedzane są w pierwszej kolejności). Ponieważ każde drzewo zorientowane można przejść na dokładnie jeden sposób, tak aby

- wystartować z korzenia,
- odwiedzić wszystkie wierzchołki
- dla każdego wierzchołka kolejność w jakiej odwiedzane są jego dzieci wyznaczała ciąg malejący,
- na koniec powrócić do korzenia,

wynika stąd, że liczba istotnych ścieżek $a \in \mathcal{K}$ jest równa liczbie drzew zorientowanych o $m+1$ wierzchołkach, czyli C_m. Oznaczmy zbiór tych ścieżek przez $\mathcal{A}$. Zauważmy teraz, że z założenia $\mathbb{E}X_{ij}^2 = 1$ dla $i \neq j$ (założenie (A2)), dla dowolnego $a \in \mathcal{A}$ oraz $\mathbf{i} \sim a$, $P(i) = 1$.

To kończy dowód Twierdzenia Wignera, gdyż

$$\begin{aligned} \frac{1}{N^{k/2+1}} \sum_{a \in \mathcal{K}} \sum_{\substack{\mathbf{i} \in \{1,\dots,N\}^k \\ \mathbf{i} \sim a}} P(\mathbf{i}) &= \frac{1}{N^{m+1}} \sum_{a \in \mathcal{A}} \sum_{\substack{\mathbf{i} \in \{1,\dots,N\}^k \\ \mathbf{i} \sim a}} P(\mathbf{i}) + \mathcal{O}(N^{-1}) \\ &= \frac{1}{N^{m+1}} C_m N(N-1)\cdots(N-m) + \mathcal{O}(N^{-1}), \end{aligned}$$

czyli

$$\lim_{N\to\infty} B_{k,N} = C_m = \int_{\mathbb{R}} x^k d\sigma(x).$$

$\square$

2.5 Dowód Twierdzenia 1

Z Twierdzenia 3 oraz równości (1), aby udowodnić Twierdzenie 1, wystarczy wykazać, że

$$\mathbb{P}\Big(\forall_{k\in\mathbb{N}} \frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k \overset{N\to\infty}{\to} \int_{\mathbb{R}} x^k d\sigma(x)\Big) = 1.$$

Korzystając z twierdzenia Wignera oraz przeliczalnej addytywności miary, powyższy fakt możemy zredukować do

$$\mathbb{P}\Big(\Big|\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k - \mathbb{E}\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big| \overset{N\to\infty}{\to} 0\Big) = 1$$

dla $k \in \mathbb{N}$.

Z lematu Borela-Cantellego wynika, że aby wykazać powyższą równość, wystarczy udowodnić, że dla dowolnego $\varepsilon > 0$,

$$\sum_{N=1}^{\infty} \mathbb{P}\Big(\Big|\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k - \mathbb{E}\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big| > \varepsilon\Big) < \infty.$$

Z kolei z nierówności Czebyszewa wynika, że zbieżność powyższego szeregu jest implikowana przez

$$\sum_{N=1}^{\infty} \operatorname{Var}\Big(\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big) < \infty. \tag{5}$$

Zbieżność szeregu (5) można uzyskać metodą kombinatoryczną, podobną do użytej w dowodzie twierdzenia Wignera.

Zauważmy najpierw, że z (2) oraz wzoru na wariancję sumy zmiennych losowych ($\operatorname{Var}(\sum_i Y_i) = \sum_{ij} \operatorname{Cov}(X_i, X_j)$), zachodzi równość

$$\operatorname{Var}\Big(\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big) = \frac{1}{N^{k+2}} \sum_{\mathbf{i},\mathrm{j}\in\{1,\ldots,N\}^k} \Big(P(\mathbf{i},\mathrm{j}) - P(\mathbf{i})P(\mathrm{j})\Big),$$

gdzie dla $\mathbf{i} = (i_1, \ldots, i_k)$, $\mathrm{j} = (j_1, \ldots, j_k)$,

$$P(\mathbf{i},\mathrm{j}) = \mathbb{E}(X_{i_1 i_2} X_{i_2,i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1} X_{j_1 j_2} X_{j_2,j_3} \cdots X_{j_{k-1} j_k} X_{j_k j_1})$$

oraz

$$P(\mathbf{i}) = \mathbb{E}(X_{i_1 i_2} X_{i_2,i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}).$$

Podobnie jak w dowodzie twierdzenia Wignera, $P(\mathbf{i},\mathrm{j}) - P(\mathbf{i})P(\mathrm{j})$ jest ograniczone przez stałą zależną tylko od k. Możemy więc spróbować ponownie pogrupować indeksy $\mathbf{i},\mathrm{j}$ w rozłączne zbiory (których liczba zależy tylko od k) i wykazać, że wkład każdego ze zbiorów do sumy jest niewielki w porównaniu z N^{k+2}. Konstrukcja odpowiedniego podziału jest zbliżona do konstrukcji z dowodu twierdzenia Wignera, zaprezentujemy ją więc nieco mniej szczegółowo niż w tamtym dowodzie.

Tym razem definiujemy relację równoważności na parach indeksów:

$$(\mathbf{i},\mathrm{j}) \sim (\mathbf{i}',\mathrm{j}')$$

wtedy i tylko wtedy gdy istnieje różnowartościowe odwzorowanie f zbioru $\{1,\ldots,N\}$ w siebie, takie że dla $l \leqslant k$, $f(i_l) = i'_l$ oraz $f(j_l) = j'_l$. Podobnie jak poprzednio, liczba klas abstrakcji tej relacji zależy tylko od k i możemy wybrać system reprezentantów klas abstrakcji $\mathcal{L} \subseteq \{1,\ldots,2k\}^2$.

Z każdą parą indeksów $(\mathbf{i},\mathrm{j})$ możemy związać graf $H(\mathbf{i},\mathrm{j})$ będący sumą grafów $G(\mathbf{i}), G(\mathrm{j})$ zdefiniowanych w dowodzie twierdzenia Wignera (dodajemy zbiory wierzchołków oraz multizbiory krawędzi). Zauważmy, że jeżeli grafy $\tilde{G}(\mathbf{i}), \tilde{G}(\mathrm{j})$ nie dzielą krawędzi, to z niezależności, $P(\mathbf{i},\mathrm{j}) - P(\mathbf{i})P(\mathrm{j}) = 0$. Wyrażenie to jest oczywiście równe 0 także wtedy, gdy graf $\tilde{H}(\mathbf{i},\mathrm{j})$ (szkielet grafu H) posiada krawędź krotności 1 (czyli taką, której w grafie $H(\mathbf{i},\mathrm{j})$ odpowiada tylko jedna krawędź skierowana).

Wszystkie powyższe własności pary indeksów $(\mathbf{i}, \mathrm{j})$ zależą tylko od jej klasy abstrakcji w relacji $\sim$. Widzimy zatem, że jeśli zapiszemy

$$\mathrm{Var}\Big(\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big) = \frac{1}{N^{k+2}} \sum_{(a,b)\in\mathcal{L}} \sum_{\substack{(\mathbf{i},\mathrm{j}):\\ (\mathbf{i},\mathrm{j})\sim(a,b)}} \Big(P(\mathbf{i},\mathrm{j}) - P(\mathbf{i})P(\mathrm{j})\Big),$$

niezerowy wkład do sumy będą miały tylko te pary (a, b), w których grafy $G(a), G(b)$ dzielą krawędź (a zatem $\tilde{H}(a, b)$ jest spójny) oraz każda krawędź $\tilde{H}(a, b)$ ma co najmniej dwa odpowiedniki w grafie $H(a, b)$ (skąd $\tilde{H}(a, b)$ ma co najwyżej k krawędzi). Z lematu 1 wynika zatem, że graf $\tilde{H}(a, b)$ ma co najwyżej $k + 1$ wierzchołków.

To już implikuje, że $|\{(\mathbf{i}, \mathrm{j}) \in \{1, \ldots, N\}^2 : (\mathbf{i}, \mathrm{j}) \sim (a, b)\}| \leqslant N^{k+1}$, czyli że

$$\mathrm{Var}\Big(\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big) = \mathcal{O}(N^{-1}).$$

Oszacowanie to jest jednak dla nas nieco za słabe, gdyż szereg harmoniczny jest rozbieżny. Aby je poprawić, zauważmy, że sytuacja w której wkład pary $(a, b) \in \mathcal{L}$ do wariancji jest niezerowy, a graf $\mathcal{H}(a, b)$ ma $k+1$ wierzchołków jest niemożliwa. Rzeczywiście, implikowałoby to, że graf ten jest drzewem, w którym każda krawędź ma dokładnie 2 odpowiedniki w grafie $H(a, b)$. Jednak grafy $\tilde{G}(a), \tilde{G}(b)$ jako spójne podgrafy $\tilde{H}(a, b)$ również musiałyby być wówczas drzewami. Ponieważ ścieżki $a_1 \to a_2 \to \ldots \to a_k \to a_1$, $b_1 \to b_2 \to \ldots \to b_k \to b_1$ odwiedzają wszystkie wierzchołki grafów $\tilde{G}(a)$, $\tilde{G}(b)$, każda z krawędzi tych grafów występuje w odpowiadającej mu ścieżce przynajmniej dwa razy. Stąd jednak wynika, że wspólna krawędź tych grafów występuje w tych ścieżkach łącznie przynajmniej 4 razy, co daje sprzeczność z faktem, że ma dokładnie dwa odpowiedniki w $H(a, b)$.

Zatem $|\{(\mathbf{i}, \mathrm{j}) \in \{1, \ldots, N\}^2 : (\mathbf{i}, \mathrm{j}) \sim (a, b)\}| \leqslant N^k$, co implikuje, że

$$\mathrm{Var}\Big(\frac{1}{N^{k/2+1}} \operatorname{tr} A_N^k\Big) = \mathcal{O}(N^{-2})$$

i pociąga (5), kończąc dowód Twierdzenia 1.

2.6 Osłabienie założeń. Uogólnienia

2.7 Założenia dotyczące całkowalności

Założenia Twierdzenia 1 można osłabiać na wiele sposobów, np. poprzez rezygnację z założeń (A3) lub (A1). To na ile możemy ograniczyć założenia dot. całkowalności zmiennych X_{ij} zależy od klasy ciągów zmiennych losowych, które rozważamy. My ograniczymy się do ciągów i.i.d. i wykażemy, że w tym wypadku do zbieżności miary spektralnej wystarczy całkowalność z kwadratem.

Twierdzenie 5. *Załóżmy, że zmienne $(X_{ij})_{1\leqslant i\leqslant j<\infty}$ są niezależne, o tym samym rozkładzie, takim że*

$$\mathbb{E}X_{ij} = 0,\ \mathbb{E}X_{ij}^2 = 1.$$

Wówczas

$$L_n \xrightarrow{D} \sigma \text{ p.n.}$$

Dowód powyższego twierdzenia będzie oparty na następującym lemacie.

Lemat 2 (Nierówność Hoffmana-Wielandta)**.** *Niech A, B będą symetrycznymi macierzami $N \times N$ o wartościach własnych odp. $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$ oraz $\gamma_1 \leqslant \gamma_2 \leqslant \ldots \leqslant \gamma_N$. Wówczas*

$$\sum_{i=1}^{N} |\lambda_i - \gamma_i|^2 \leqslant \operatorname{tr}(A - B)^2. \tag{6}$$

Dowód. Ponieważ $\operatorname{tr} A^2 = \sum_{i=1}^N \lambda_i^2$, $\operatorname{tr} B^2 = \sum_{i=1}^N \gamma_i^2$ oraz $\operatorname{tr} AB = \operatorname{tr} BA$, łatwo pokazać, że (6) jest równoważne nierówności

$$\operatorname{tr} AB \leqslant \sum_{i=1}^N \lambda_i \gamma_i.$$

Możemy bez straty ogólności założyć, że A jest macierzą diagonalną, zaś $B = UDU^T$, gdzie $D = Diag(\gamma_1, \ldots, \gamma_N)$, a $U = [u_{ij}]_{i,j\leqslant N} \in O_N$.

Zatem

$$\operatorname{tr} AB = \sum_{i,j\leqslant N} \lambda_i \gamma_j u_{ij}^2,$$

skąd wynika, że aby udowodnić lemat, wystarczy wykazać, że

$$\sum_{i=1}^N \lambda_i \gamma_i = \sup_{[v_{ij}]-\text{ podw.stochastyczna}} \sum_{i,j\leqslant N} \lambda_i \gamma_j v_{ij},$$

gdzie supremum jest po wszystkich macierzach podwójnie stochastycznych $N \times N$ (czyli macierzach o nieujemnych współczynnikach, których suma w każdym wierszu i każdej kolumnie wynosi 1).

To można wykazać na kilka sposobów. Pierwszy z nich, który pozostawiamy jako ćwiczenie sprowadza się do pokazania, że punktami ekstremalnymi zbioru macierzy symetrycznych są dokładnie macierze permutacji, dla których nierówność jest łatwa.

Drugi sposób sprowadza się do indukcji po rozmiarze macierzy. Dla $N = 1$, teza jest trywialna. Załóżmy więc, że zachodzi dla $N - 1$. Jeżeli $v_{11} = 1$, to $v_{1j} = v_{j1} = 0$ dla dowolnego j i możemy zastosować założenie indukcyjne. W przeciwnym wypadku istnieją $k, l > 0$, takie że $v_{k1}, v_{1l} > 0$. Niech $v = \min(v_{k1}, v_{1l})$. Zdefiniujmy macierz $[\tilde{v}_{ij}]$ wzorem

$$\tilde{v}_{11} = v_{11} + v,\ \tilde{v}_{k1} = v_{k1} - v,\ \tilde{v}_{1l} = v_{1l} - v,\ \tilde{v}_{kl} = v_{kl} + v,$$

oraz $\tilde{v}_{ij} = v_{ij}$ dla pozostałych par i, j.

Nowa macierz jest nadal macierzą bistochastyczną. Ponadto,

$$\sum_{ij} \lambda_i \gamma_j \tilde{v}_{ij} - \sum_{ij} \lambda_i \gamma_j v_{ij} = v(\lambda_1 - \lambda_k)(\gamma_1 - \gamma_l) \geqslant 0.$$

Zauważmy, że łączna liczba zer w pierwszym wierszu i pierwszej kolumnie macierzy $[\tilde{v}_{ij}]$ zmniejszyła się w porównaniu do macierzy $[v_{ij}]$. Kontynuując opisaną powyżej procedurę uzyskamy więc w skończonej liczbie kroków macierz w której na miejscu $(1,1)$ stoi jedynka, co jak zauważyliśmy powyżej pozwala na zastosowanie założenia indukcyjnego. $\square$

Dowód Twierdzenia 5. Niech $C > 0$ będzie dowolną liczbą rzeczywistą, taką by rozkład $X_{ij}\mathbf{1}_{\{|X_{ij}|\leqslant C\}}$ był niezdegenerowany.

Zdefiniujmy

$$X_{ij}^{(C)} = \frac{X_{ij}\mathbf{1}_{\{|X_{ij}|\leqslant C\}} - \mathbb{E}X_{ij}\mathbf{1}_{\{|X_{ij}|\leqslant C\}}}{\sqrt{\operatorname{Var}(X_{ij}\mathbf{1}_{\{|X_{ij}|\leqslant C\}})}}.$$

Zmienne $X_{ij}^{(C)}$ spełniają założenia Twierdzenia 1, zatem miara spektralna $L_N^{(C)}$ macierzy $A_N^{(C)} = \frac{1}{\sqrt{N}}[X_{ij}^{(C)}]_{i,j\leqslant N}$ prawie na pewno zbiega słabo do miary Wignera. Oznaczmy wartości własne tej macierzy przez $\lambda_i^{(C),N}$.

Rozważmy teraz dowolną funkcję Lipschitzowską $f\colon \mathbb{R} \to \mathbb{R}$ i oznaczmy przez L jej stałą Lipschitza. Mamy

$$
\begin{aligned}
\Big| \int_{\mathbb{R}} f(x) dL_N^{(C)}(x) - \int_{\mathbb{R}} f(x) dL_n(x) \Big| &= \frac{1}{N} \Big| \sum_{i=1}^{N} f(\lambda_i^N) - f(\lambda_i^{(C),N}) \Big| \leqslant \frac{L}{N} \sum_{i=1}^{N} \Big| \lambda_i^N - \lambda_i^{(C),N} \Big| \\
&\leqslant \frac{L}{\sqrt{N}} \Big(\sum_{i=1}^{N} (\lambda_i^N - \lambda_i^{(C),N})^2 \Big)^{1/2} \\
&\leqslant L \sqrt{\frac{1}{N} \mathrm{tr} \Big(\frac{1}{\sqrt{N}} A_N^{(C)} - \frac{1}{\sqrt{N}} A_N \Big)^2} \\
&= L \sqrt{\frac{1}{N^2} \sum_{i,j \leqslant N} (X_{ij} - X_{ij}^{(C)})^2},
\end{aligned}
$$

przy czym druga nierówność wynika z nierówności Cauchy'ego-Schwarza, zaś trzecia z Lematu 2. Zauważmy, że z mocnego prawa wielkich liczb, przy $N \to \infty$, prawa strona powyższej nierówności dąży p.n. to $L\sqrt{\mathbb{E}(X_{12} - X_{12}^{(C)})^2}$, co jak łatwo sprawdzić zbiega do 0, dla $C \to \infty$. W połączeniem ze zbieżnością miary spektralnej $L_N^{(C)}$ i faktem, że słabą zbieżność wystarczy sprawdzać na funkcjach Lipschitzowskich, pozwala to w prosty sposób zakończyć dowód twierdzenia. □

Uwaga o współczynnikach zespolonych Do tej pory ograniczaliśmy się do badania miary spektralnej macierzy symetrycznych o rzeczywistych współczynnikach. W rozważaniach fizycznych, często istotną rolę odgrywają macierze hermitowskie o współczynnikach zespolonych. W zasadzie wszystkie twierdzenia wykazane do tej pory można udowodnić także dla tego typu macierzy, czasami może się to wiązać z niewielkimi komplikacjami technicznymi (właśnie dla ich uniknięcia skoncentrowaliśmy się na przypadku rzeczywistym). Aby zilustrować jakie założenia o zmiennych losowych czynione są w przypadku zespolonym, poniżej podajemy przykładowo odpowiednik Twierdzenia 1.

Twierdzenie 6. *Załóżmy, że $(Z_{ij})_{1 \leqslant i \leqslant j < \infty}$ są niezależnymi zespolonymi zmiennymi losowymi, takimi że*

- *dla dowolnych $i < j$, $\mathbb{E} Z_{ij} = 0$, $\mathbb{E}|Z_{ij}|^2 = 1$,*
- *dla dowolnego i, $Z_{ii} \in \mathbb{R}$ p.n., $\mathbb{E} Z_{ii} = 0$,*
- *dla dowolnego $k \in \mathbb{N}$, $m_k := \sup_{ij} \mathbb{E}|Z_{ij}|^k < \infty$.*

Zdefiniujmy dodatkowo $Z_{ij} = \bar{Z}_{ji}$ dla $i > j$ i oznaczmy przez L_N miarę macierzy $N^{-1/2}[Z_{ij}]_{i,j \leqslant N}$. Wówczas, z prawdopodobieństwem 1 zachodzi zbieżność

$$L_N \xrightarrow{D} \sigma \text{ dla } N \to \infty.$$

2.8 Założenia dotyczące scentrowania zmiennych

Spróbujemy teraz osłabić założenie $\mathbb{E} X_{ij} = 0$. Z punktu widzenia kombinatorycznego dowodu Twierdzenia 1 było ono niezwykle ważne, pozwalało bowiem wyłączyć z rozważań wszystkie ścieżki, w których pewna krawędź była odwiedzona jedynie raz. Okazuje się jednak, że nie odgrywa ono istotnej roli, twierdzenie pozostaje prawdziwe, jeżeli wszystkie zmienne mają wspólną średnią m i wariancję 1, a nawet gdy średnie poszczególnych zmiennych są różne, pod warunkiem, że macierz średnich ma niewielki rząd. Dokładniej, zachodzi następujące

Twierdzenie 7. *Niech $(X_{ij})_{1 \leqslant i \leqslant j < \infty}$ będą zespolonymi zmiennymi losowymi, takimi że $X_{ii} \in \mathbb{R}$ p.n. Niech*

$$B_N = \mathbb{E} A_N = [\mathbb{E} X_{ij}]_{i,j \leqslant N},$$

gdzie przyjmujemy $X_{ij} = \bar{X}_{ji}$ dla $i > j$. Zdefiniujmy ponadto macierze $A_N = [X_{ij}]_{i,j \leqslant N}$ oraz $\tilde{A}_N = A_N - B_N$.

Załóżmy, że
$$L_{\frac{1}{\sqrt{N}}\tilde{A}_N} \xrightarrow{D} \sigma \text{ p.n.}$$
(np., że $X_{ij} - \mathbb{E}X_{ij}$ są niezależne mają jednostajnie ograniczone momenty dowolnego rzędu oraz oraz $\mathbb{E}|X_{ij} - \mathbb{E}X_{ij}|^2 = 1$). Jeśli
$$\lim_{N\to\infty} \frac{\text{rank } B_N}{N} = 0,$$
to $L_{\frac{1}{\sqrt{N}}A_N} \xrightarrow{D} \sigma$ p.n.

Wniosek Jeżeli zmienne X_{ij} są rzeczywiste, o tej samej wartości oczekiwanej, to $L_{N^{-1/2}A_N} \to \sigma$ p.n.

Dowód wniosku. Wystarczy zauważyć, że w tym wypadku wszystkie współczynniki macierzy B_N są równe, a więc ma ona rząd co najwyżej 1. □

Okazuje się, że za Twierdzeniem 7 nie kryją się żadne fakty probabilistyczne dotyczące zachowania zmiennych X_{ij}. Wynika ono z następującego twierdzenia z algebry liniowej.

Twierdzenie 8. *Jeśli A, B są macierzami hermitowskimi $N \times N$, zaś F_A, F_B dystrybuantami ich miar spektralnych, to*
$$\sup_{t\in\mathbb{R}} |F_A(t) - F_B(t)| \leqslant \frac{1}{N}\text{rank } (A - B).$$

Twierdzenie 7 wynika z powyższego, gdyż słaba zbieżność miar probabilistycznych jest równoważna zbieżności dystrybuant w punktach ciągłości dystrybuanty miary granicznej.

Dalszy ciąg niniejszego paragrafu poświęcimy na dowód Twierdzenia 8, po drodze formułując kilka faktów z algebry liniowej przydatnych w analizie wartości własnych macierzy losowych. Poniżej przez B^* będziemy oznaczali sprzężenie hermitowskie macierzy lub wektora o współczynnikach zespolonych.

Lemat 3 (Courant-Fischer)**.** *Niech A będzie macierzą hermitowską $N\times N$, o wartościach własnych $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$. Wówczas*
$$\lambda_k = \min_{x_1,\ldots,x_{N-k}\in\mathbb{C}^N} \max_{\substack{x\in\mathbb{C}^N\setminus\{0\} \\ x\perp x_1,\ldots,x_{N-k}}} \frac{x^*Ax}{x^*x}.$$

Uwaga Jak łatwo zobaczyć z dowodu powyższego twierdzenia, w przypadku macierzy o współczynnikach rzeczywistych, $\mathbb{C}^N$ może zostać zastąpione przez $\mathbb{R}^N$.

Dowód. Bez straty ogólności możemy założyć, że $A = Diag(\lambda_1, \ldots, \lambda_N)$ (gdyż macierze zmiany bazy do bazy wektorów własnych są unitarne, a więc zachowują iloczyn skalarny).

Ustalmy najpierw dowolne $x_1, \ldots, x_{N-k} \in \mathbb{C}^N$, Chcemy znaleźć wektor jednostkowy $x \perp x_1, \ldots, x_{N-k}$, taki że $x^*Ax \geqslant \lambda_k$.

Zauważmy, że istnieje wektor $x \perp x_1, \ldots, x_{N-k}$, którego współrzędne od pierwszej do $(k-1)$-ej są równe 0. Rzeczywiście, niech $P\colon \mathbb{C}^N \to \mathbb{C}^N$ oznacza rzut ortogonalny na podprzestrzeń H rozpiętą przez ostatnie $N-k+1$ wektorów bazowych. Ponieważ span$(Px_1, \ldots, Px_{N-k})$ ma wymiar co najwyżej $N-k < N-k+1 = \dim(H)$, istnieje wektor jednostkowy $x \in H$, ortogonalny do $Px_1, \ldots, Px_{N-k}$. Mamy
$$\langle x, x_i\rangle = \langle x, Px_i\rangle = 0.$$

Oznaczmy i-tą współrzędną x przez $x(i)$. Mamy
$$x^*Ax = \sum_{i=1}^N x(i)^2\lambda_i = \sum_{i=k}^N x(i)^2\lambda_i \geqslant \lambda_k \sum_{i=k}^N x(i)^2 = \lambda_k.$$

Aby zakończyć dowód lematu wystarczy zauważyć, że jeżeli $x_1, \ldots, x_{N-k}$ są wektorami bazowymi odpowiadającymi $\lambda_{k+1}, \ldots, \lambda_N$, to

$$\max_{\substack{x \in \mathbb{C}^N \setminus \{0\} \\ x \perp x_1, \ldots, x_{N-k}}} \frac{x^* A x}{x^* x} = \lambda_k.$$

□

Wnioskiem z powyższego lematu jest tzw. twierdzenie Lidskiiego.

Lemat 4 (Twierdzenie Lidskiiego)**.** *Niech A będzie macierzą hermitowską $N \times N$, $z \in \mathbb{C}^N, \eta \in \{-1, +1\}$. Zdefiniujmy macierz*

$$B = A + \eta z z^*$$

i oznaczmy przez $\lambda_i(A)$ (odp. $\lambda_i(B)$) wartości własne macierzy A (odp. B) ustawione w porządku rosnącym.

Zachodzą następujące nierówności

a) $\lambda_1(A) \leqslant \lambda_2(B) \leqslant \lambda_3(A) \leqslant \lambda_4(B) \leqslant \ldots$,

b) $\lambda_1(B) \leqslant \lambda_2(A) \leqslant \lambda_3(B) \leqslant \lambda_4(A) \leqslant \ldots$

Dowód. Ustalmy k. Z Lematu 3 wynika, że istnieją $x_1, \ldots, x_{N-k} \in \mathbb{C}^N$, takie że dla każdego niezerowego $x \perp x_1, \ldots, x_{N-k}$ zachodzi nierowność

$$\lambda_k(B) \geqslant \frac{x^*(A + \eta z z^*)x}{x^* x}.$$

W szczególności, dla dowolnego niezerowego $x \perp x_1, \ldots, x_{N-k}, z$, mamy

$$\lambda_k(B) \geqslant \frac{x^* A x}{x^* x}.$$

Zatem

$$\lambda_k(B) \geqslant \max_{\substack{x \in \mathbb{C}^N \setminus \{0\} \\ x \perp x_1, \ldots, x_{N-k}, z}} \frac{x^* A x}{x^* x} \geqslant \min_{x_1, \ldots, x_{N-k+1} \in \mathbb{C}^N} \max_{\substack{x \in \mathbb{C}^N \setminus \{0\} \\ x \perp x_1, \ldots, x_{N-k+1}}} \frac{x^* A x}{x^* x} = \lambda_{k-1}(A).$$

Z symetrii $\lambda_k(A) \geqslant \lambda_{k-1}(B)$, co kończy dowód lematu. □

Do dowodu Twierdzenia 8 potrzebny nam jeszcze jeden elementarny fakt z algebry liniowej.

Lemat 5. *Dowolna macierz hermitowska A rzędu k ma przedstawienie postaci*

$$A = \sum_{i=1}^{k} \eta_i z_i z_i^*,$$

gdzie $z_i \in \mathbb{C}^N$, $\eta_i \in \{-1, 1\}$.

Dowód. Istnieje macierz unitarna U oraz macierz diagonalna $D = Diag(\lambda_1, \ldots, \lambda_k, \underbrace{0, \ldots, 0}_{N-k})$, gdzie $\lambda_i \in \mathbb{R} \setminus \{0\}$, takie że $A = UDU^*$. Oczywiście

$$D = \sum_{i=1}^{k} \eta_i (\sqrt{|\lambda_i|} e_i)(\sqrt{|\lambda_i|} e_i)^*,$$

gdzie $\eta_i = \operatorname{sgn} \lambda_i$, zaś e_i jest i-tym wektorem bazy. Zatem $A = \sum_{i=1}^{k} \eta_i z_i z_i^*$, gdzie $z_i = \sqrt{|\lambda_i|} U e_i$. □

Dowód Twierdzenia 8. Niech $k = \text{rank}\,(A - B)$. Z lematu,

$$A - B = \sum_{i=1}^{k} \eta_i z_i z_i^*,$$

zatem iterując Twierdzenie Lidskiiego otrzymujemy, że $\lambda_i(A) \leqslant \lambda_{i+k}(B)$ oraz $\lambda_i(B) \leqslant \lambda_{i+k}(A)$ dla $i \leqslant N - k$.

Rozważmy dowolne i oraz $t < \lambda_i$. Mamy

$$F_B(t) \leqslant \frac{i + k - 1}{N},$$

gdyż jeśli $i+k \leqslant N$, to $\lambda_{i+k}(B) \geqslant \lambda_i(A) > t$, zaś w przeciwnym wypadku nierówność jest spełniona trywialnie.

Podobnie, jeśli $t \geqslant \lambda_i(A)$, to

$$F_B(t) \geqslant \frac{i - k}{N},$$

gdyż jeśli $i - k > 0$, to $\lambda_{i-k}(B) \leqslant \lambda_i(A) \leqslant t$.

Zatem dla $t < \lambda_1$ mamy $|F_A(t) - F_B(t)| = F_B(t) \leqslant kN^{-1}$. Dla $t \in [\lambda_i, \lambda_{i+1})$, mamy

$$F_A(t) - F_B(t) = \frac{i}{N} - F_B(t) \in \Big[-\frac{k}{N}, \frac{k}{N} \Big].$$

Wreszcie dla $t \geqslant \lambda_N$ mamy $|F_A(t) - F_B(t)| = 1 - F_B(t) \leqslant 1 - (N - k)N^{-1} = kN^{-1}$. □

3 Macierze gaussowskie – podstawowe własności

Niniejszy rozdział bazuje na pozycjach [M] oraz [AGZ]. Wyprowadzenie gęstości macierzy niezmienniczych o niezależnych współczynnikach jest uproszczeniem dowodu przedstawionego w [M]. Dowód wzoru na gęstość wartości własnych macierzy GOE można znaleźć w obu wspomnianych źródłach, aczkolwiek [M] omija większość istotnych zagadnień analitycznych, których przedstawienie poniżej zostało oparte na pozycji [AGZ]

Jednym z podstawowych postulowanych przez fizyków założeń dot. macierzy losowych jest, aby ich rozkład był niezmienniczy ze względu na sprzężenia macierzami ortogonalnymi lub unitarnymi (czyli nie zależał od wyboru bazy ortonormalnej, w której są reprezentowane). Macierze o tych własnościach stanowią bardzo szeroką klasę modeli, badaną w ogólności przy użyciu metod analitycznych. Okazuje się jednak, że jeżeli do założenia niezmienniczości dodamy założenie, że zmienne na i powyżej przekątnej są niezależne (ze względu na symetrię macierzy jest to najsilniejsze założenie o niezależności na jakie możemy sobie pozwolić), klasa dopuszczalnych rozkładów istotnie się zawęża. W szczególności współczynniki macierzy muszą mieć rozkład gaussowski. W następnych rozdziałach udowodnimy ten fakt oraz pokażemy, że tego typu specjalne macierze mają dużo dobrych własności analitycznych i probabilistycznych. Jedną z nich jest możliwość znalezienia jawnego wzoru na gęstość wartości własnych.

Podobnie jak w poprzednich paragrafach, dowody ograniczymy do macierzy o współczynnikach rzeczywistych, natomiast w przypadku zespolonym zasygnalizujemy tylko jak zmieniają się odpowiednie twierdzenia.

3.1 Gaussian Orthogonal Ensemble (GOE)

Definicja 5. *Powiemy, że macierz losowa*

$$H = \begin{bmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & \ddots & \dots & H_{2N} \\ \vdots & \dots & \ddots & \vdots \\ H_{N1} & \dots & \dots & H_{NN} \end{bmatrix}$$

jest macierzą GOE jeśli

- *dla* $i, j \leqslant N$, $H_{ij} = H_{ji}$ *p.n.*,
- $(H_{ij})_{1 \leqslant i \leqslant j \leqslant N}$ *są niezależne*,
- *dla każdego* i, $H_{ii} \sim \mathcal{N}(0, 2)$,
- *dla dowolnych* $i < j$, $H_{ij} \sim \mathcal{N}(0, 1)$.

Uwagi:

1. Powyżej przyjmujemy konwencję, że drugim parametrem rozkładu gaussowskiego jest wariancja, w szczególności jeśli $g \sim \mathcal{N}(0, 1)$, to $\sqrt{2}g \sim \mathcal{N}(0, 2)$.
2. Macierz symetryczną $N \times N$ możemy utożsamić z wektorem w $\mathbb{R}^{N(N+1)/2}$. Przy takim utożsamieniu, rozkład P_N macierzy H ma gęstość względem miary Lebesgue'a, daną wzorem

$$\begin{aligned} g(H) &= \frac{1}{2^{N/2}} \frac{1}{(2\pi)^{N(N+1)/4}} \exp\Big(- \frac{1}{4} \sum_{i \leqslant N} H_{ii^2} - \frac{1}{2} \sum_{1 \leqslant i \leqslant j \leqslant N} H_{ij}^2 \Big) \\ &= \frac{1}{2^{N/2}} \frac{1}{(2\pi)^{N(N+1)/4}} \exp\Big(- \frac{1}{4} \operatorname{tr} H^2 \Big). \end{aligned}$$

3. Jak łatwo sprawdzić, rozkład macierzy GOE nie zależy od wyboru bazy ortonormalnej, tzn. dla dowolnej macierzy $U \in O_N$,
$$H \sim U^T H U,$$

3.1.1 Charakteryzacja macierzy GOE

Poniższe twierdzenie implikuje, że symetryczna macierz losowa o niezależnych współczynnikh (mod. symetria) i z gęstością niezmienniczą na sprzężenia macierzami ortogonalnymi musi być z dokładnościa do skalowania i przesunięcia macierzą GOE.

Twierdzenie 9. *Załóżmy, że symetryczna macierz losowa* $H = [H_{ij}]_{i,j \leqslant N}$ *spełnia warunki*

(i) $(H_{ij})_{i,j \leqslant N}$ *są niezależnymi zmiennymi losowymi*,

(ii) dla każdej macierzy $U \in O_n$, $H \sim U^T H U$.

Wówczas macierz H *jest albo stała p.n. i równa* $\lambda \mathrm{Id}$ *dla pewnego* $\lambda \in \mathbb{R}$, *albo ma gęstość postaci*

$$g(H) = \exp(-a \operatorname{tr} H^2 - b \operatorname{tr} H - c),$$

gdzie $a > 0, b, c \in \mathbb{R}$.

Dowód. Wykażemy najpierw, że jeżeli macierz H spełnia założenia twierdzenia, to zmienne H_{ij} dla $i \neq j$ mają wszystkie ten sam rozkład i podobnie, wszystkie zmienne H_{ii} mają ten sam rozkład.

W tym celu wystarczy ograniczyć się do sprzężeń macierzami permutacji. Niech dla $k < l$, $U(k, l)$ oznacza macierz transpozycji $k \leftrightarrow l$ (czyli w terminach bazy $U(k, l)e_k = e_l, U(k, l)e_l = e_k$ oraz $U(kl)e_i = e_i$ dla $i \neq k, l$).

Oznaczmy współczynniki macierzy $U(k, l)^T H U(k, l)$ przez a_{ij}. Mamy

a) $a_{kk} = H_{ll}$,

b) $a_{kj} = H_{lj}$ dla $j \neq k, l$,

c) $a_{ik} = H_{il}$ dla $i \neq k, l$.

Z punktu a) wynika, że wszystkie elementy na przekątnej mają ten sam rozkład, z b) równość rozkładów elementów leżących w tej samej kolumnie, poza elementem na przekątnej, zaś z c) równość rozkładów elementów z tego samego wiersza, leżących poza przekątną. Stąd już wynika równość rozkładów odpowiednich współczynników.

Załóżmy teraz dodatkowo, że rozkład macierzy H ma gładką i ściśle dodatnia gęstość g względem miary Lebesgue'a na $\mathbb{R}^{N(N+1)/2}$. Z niezależności i udowodnionej równości odpowiednich rozkładów gęstość ta jest postaci

$$g(H) = \prod_{1 \leqslant i \leqslant j \leqslant N} f_{ij}(H_{ij})$$

gdzie f_{ij} są gęstościami prawdopodobieństwa na prostej (wiemy już, że $f_{ii} = f_{jj}$ oraz $f_{ij} = f_{kl}$ dla dowolnych $i < j$, $k < l$, ale dla skrócenia zapisu nie będziemy uwzględniać tego w notacji).

Aby znaleźć funkcje f_{ij}, wystarczy ograniczyć się do podmacierzy $[H_{ij}]_{i,j \leqslant 2}$. Łatwo wykazać, że również jej rozkład jest niezmienniczy ze względu na sprzężenia macierzami ortogonalnymi (wynika to z faktu, że dowolną izometrię $\mathbb{R}^2$ można rozszerzyć do izometrii $\mathbb{R}^N$). Załóżmy zatem, że $N = 2$. Rozważmy szczególną postać izometrii

$$U_\theta = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix},$$

czyli obrót o kąt $\theta \in \mathbb{R}$.

Zdefiniujmy $F_\theta \colon \mathbb{R}^3 \to \mathbb{R}^3$ wzorem

$$F_\theta(H) = U_\theta^T H U_\theta.$$

Korzystając z powyższej definicji, łatwo sprawdzić, że

$$\begin{aligned} \frac{d}{d\theta} F_\theta(H) &= \frac{dU_\theta^T}{d\theta} H U_\theta + U_\theta^T H \frac{dU_\theta}{d\theta} \\ &= \frac{dU_\theta^T}{d\theta} U_\theta F_\theta(H) + F_\theta(H) U_\theta^T \frac{dU_\theta}{d\theta} \\ &= A f_\theta(H) + F_\theta(H) A^T, \end{aligned} \tag{7}$$

gdzie

$$A = \frac{dU_\theta^T}{d\theta} U_\theta = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$

Z twierdzenia o całkowaniu przez podstawienie, wynika że dla każdego θ gęstość macierzy $F_\theta^{-1}(H) = U_\theta H U_\theta^T$ jest równa

$$g(F_\theta(H)) |\det DF_\theta(H)| = g(F_\theta(H))$$

(jakobian jest równy 1, gdyż przekształcenie F_θ^{-1} zachowuje normę Hilberta-Schmidta macierzy).

Z drugiej strony, z niezmienniczości rozkładu H względem sprzężeń macierzami ortogonalnymi, rozkład $F_\theta^{-1}(H)$ jest równy rozkładowi H. Stąd $g(F_\theta(H)) = g(H)$ p.n. względem miary Lebesgue'a. Z gładkości g wynika, że równość ta musi zachodzić wszędzie (dopełnienie zbioru miary zero jest gęste). Zatem

$$\frac{dg(F_\theta(H))}{d\theta} = 0.$$

Lewą stronę powyższej równości możemy policzyć korzystając z (7). Otrzymujemy

$$\sum_{1 \leqslant i \leqslant j \leqslant 2} \frac{f'_{ij}(F_\theta(H)_{ij})}{f_{ij}(F_\theta(H)_{ij})} g(F_\theta(H)) (AF_\theta(H) + F_\theta(H)A^T)_{ij} = 0,$$

gdzie przyjmujemy konwencję, że dla macierzy B, przez B_{ij} oznaczamy jej współczynniki. Korzystając z założenia, że g jest dodatnia oraz faktu, że F_θ jest na $\mathbb{R}^3$, otrzymujemy stąd

$$\sum_{1 \leqslant i \leqslant j \leqslant 2} \frac{f'_{ij}(H_{ij})}{f_{ij}(H_{ij})} (AH + HA^T)_{ij} = 0$$

dla dowolnej symetrycznej macierzy H. Mamy

$$AH + HA^T = \begin{bmatrix} -2H_{12} & -H_{22} + H_{11} \\ -H_{22} + H_{11} & 2H_{12} \end{bmatrix},$$

zatem z powyższej równości otrzymujemy

$$-2\frac{f_{11}'(H_{11})}{f_{11}(H_{11})}H_{12} + \frac{f_{12}'(H_{12})}{f_{12}(H_{12})}(H_{11} - H_{22}) + 2\frac{f_{22}'(H_{22})}{f_{22}(H_{22})}H_{12} = 0,$$

co dla $H_{11} \neq H_{22}$ możemy zapisać jako

$$\Big(\frac{f_{11}'(H_{11})}{f_{11}(H_{11})} - \frac{f_{22}'(H_{22})}{f_{22}(H_{22})}\Big)\frac{1}{H_{11} - H_{22}} = \frac{f_{12}'(H_{12})}{f_{12}(H_{12})}\frac{1}{2H_{12}}. \tag{8}$$

Zauważmy, że lewa strona powyższej równości zależy tylko od H_{11}, H_{22}, zaś prawa tylko od H_{12}. Wynika stąd, że obie strony muszą być stałe. Oznaczając

$$-2a = \frac{f_{12}'(H_{12})}{f_{12}(H_{12})}\frac{1}{2H_{12}},$$

otrzymujemy

$$\frac{d}{dH_{12}}\log f_{12}(H_{12}) = -4aH_{12},$$

skąd

$$f_{12}(H_{12}) = C\exp(-2aH_{12}^2).$$

Przekształcając lewą stronę (8), otrzymujemy

$$\frac{f_{11}'(H_{11})}{f_{11}(H_{11})} + 2aH_{11} = \frac{f_{22}'(H_{22})}{f_{22}(H_{22})} + 2aH_{22}$$

skąd ponownie otrzymujemy, że obie strony muszą być stałe. Zatem dla pewnej stałej b,

$$\frac{d}{dH_{11}}\log f_{11}(H_{11}) = -2aH_{11} + b,$$

czyli

$$f_{11}(H_{11}) = D\exp(-aH_{11}^2 - bH_{11}).$$

Możemy teraz wrócić do ogólnego przypadku macierzy dowolnych rozmiarów i napisać

$$g(H) = \exp\Big(-a\Big(\sum_i H_{ii}^2 + 2\sum_{i<j} H_{ij}^2\Big) - b\sum_i H_{ii} - c\Big) = \exp(-a\mathrm{tr}\, H^2 - b\mathrm{tr}\, H - c),$$

dla pewnej stałej normalizującej c. Zauważmy, że musimy mieć $a > 0$, gdyż w przeciwnym przypadku otrzymana funkcja nie byłaby całkowalna.

Wzór na rozkład macierzy H udało nam się znaleźć przy założeniu, że posiada on gładką i dodatnią gęstość. Ogólny przypadek można łatwo sprowadzić do powyższego, poprzez dodanie do macierzy H niezależnej macierzy gaussowskiej. Niech zatem G będzie macierzą GOE, niezależną od H. Zdefiniujmy

$$H_\varepsilon = H + \varepsilon G.$$

Można sprawdzić, że dla $\varepsilon > 0$ macierz H_ε ma gładką i ściśle dodatnią gęstość. Jest także niezmiennicza na sprzężenia macierzami ortogonalnymi i ma niezależne współczynniki (mod. symetria). Zatem, jak wykazaliśmy, ma gęstość postaci (9), czyli jej współczynniki są zmiennymi gaussowskimi. Zauważmy, że a we wzorze (9) odpowiada $1/(2\sigma^2)$ gdzie σ^2 to wspólna wariancja zmiennych na przekątnej. Ponadto $b\sigma^2$ jest wspólną średnią tych zmiennych. Oznaczmy przez m_ε i σ_ε^2 średnią i wariancję zmiennej $(H_\varepsilon)_{11}$. Z oczywistej zbieżności $H_\varepsilon \to H$ dla $\varepsilon \to 0$ otrzymujemy w szczególności $(H_\varepsilon)_{11} \xrightarrow{D} H_{11}$. Ale słaba zbieżność zmiennych gaussowskich jest równoważna zbieżności ich parametrów, co więcej granica jest zawsze gaussowska. Zatem, jeżeli $\sigma_\varepsilon^2 \to 0$, macierz H jest deterministyczną macierzą diagonalną, w przeciwnym wypadku posiada gęstość postaci (9). □

3.1.2 Rozkład łączny wartości własnych macierzy GOE

Twierdzenie 10. *Niech H będzie macierzą GOE ($N \times N$), zaś $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$ jej wartościami własnymi. Wówczas gęstość wektora losowego $(\lambda_1, \ldots, \lambda_N)$ wyraża się wzorem*

$$h(\lambda_1, \ldots, \lambda_N) = D_N \prod_{i<j} |\lambda_i - \lambda_j| \exp\Big(-\frac{1}{4} \sum_{i=1}^{N} \lambda_i^2 \Big) \mathbf{1}_{\{\lambda_1 < \ldots < \lambda_N\}},$$

gdzie D_N jest pewną stałą normującą.

Wnioski

1. Wartości własne macierzy GOE z prawdopodobieństwem 1 są jednokrotne.
2. Niech e_i oznacza jednostkowy wektor własny macierzy H, odpowiadający λ_i, wybrany tak, aby jego pierwszy niezerowy wyraz był dodatni (zauważmy, że ponieważ wartości własne mają krotność 1, definicja ta jest jednoznaczna). Wówczas $(e_1, \ldots, e_N)$ jest niezależny od $(\lambda_1, \ldots, \lambda_N)$.
3. Macierz $[e_1|e_2|\ldots|e_N]$ (czyli macierz o kolumnach e_i) ma taki sam rozkład jak macierz $[u_{ij}\mathrm{sgn}(u_{1j})]_{i,j\leqslant N}$, gdzie $U = [u_{ij}]$ jest macierzą rozłożoną względem miary Haara na grupie ortogonalnej O_N.

Uwaga W wyrażeniu $\Delta(\lambda_1, \ldots, \lambda_N) = \prod_{i<j} |\lambda_i - \lambda_j|$ możemy rozpoznać wyznacznik Vandermonde'a

$$\det \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \ldots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \lambda_2^2 & \ldots & \lambda_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_N & \lambda_N^2 & \ldots & \lambda_N^{N-1} \end{bmatrix}.$$

Dowód Twierdzenia 10. Najprościej rzecz ujmując, dowód sprowadza się do wzoru na całkowanie przez podstawienie i bazuje na właściwej parametryzacji zbioru macierzy symetrycznych przy użyciu ich wartości własnych i dodatkowych parametrów. Parametryzacja taka nie jest możliwa dla całego zbioru macierzy symetrycznych, niemniej można ją uzyskać po odrzuceniu domkniętego zbioru miary zero. Poniżej przedstawimy te własności parametryzacji, które są kluczowe dla dowodu. Natomiast samą, dość techniczną, konstrukcję przykładowej parametryzacji podamy później (wybór konkretnej parametryzacji nie ma znaczenia dla dowodu twierdzenia).

Macierz symetryczną H o wartościach własnych $\lambda_1 \leqslant \ldots \leqslant \lambda_N$ możemy zapisać w postaci $H = UDU^T$, gdzie $U \in O_N$, zaś $D = Diag(\lambda_1, \ldots, \lambda_N)$. Jeżeli $\lambda_1 < \lambda_2 < \ldots < \lambda_N$, to zapis ten jest prawie jednoznaczny (dla każdej wartości własnej odpowiadający jej jednostkowy wektor własny możemy wybrać na dwa sposoby).

Okazuje się, że można znaleźć podzbiór otwarty $\tilde{O}_N \subset O_N$, który daje się gładko sparametryzować pewnym podzbiorem otwartym $A \subseteq \mathbb{R}^{n(n-1)/2}$, taki że zbiór macierzy symetrycznych, które nie mają przedstawienia postaci $UDiag(\lambda_1, \ldots, \lambda_N)U^T$, gdzie $\lambda_1 < \ldots < \lambda_N$ oraz $U \in \tilde{O}_n$ jest domkniętym zbiorem miary zero. Istnieje zatem gładkie przekształcenie $T\colon A \times \Delta \overset{1-1}{\to} \mathcal{M}^{sym}(N, N)$, (gdzie $\Delta = \{(\lambda_1, \ldots, \lambda_N) \in \mathbb{R}^N \colon \lambda_1 < \ldots < \lambda_N\}$), takie że

$$T(U, \lambda) = T_1(U)Diag(\lambda_1, \ldots, \lambda_N)T_1(U)^T$$

oraz $\mathcal{M}^{sym}(N, N)\backslash T(U \times \Delta)$ jest domkniętym zbiorem miary 0.

Rozważmy teraz dowolny zbiór otwarty $B \subseteq A \times \Delta$. Mamy

$$\mathbb{P}(T^{-1}(H) \in B) = \mathbb{P}(H \in T(B)) = \int_{T(B)} g(H)dH = \int_B g(T(U, \lambda))| \det DT| d\lambda dU, \tag{9}$$

gdzie g jest gęstością macierzy GOE. Ponieważ $g(H) = c_N \exp(-4^{-1}\mathrm{tr}\, H^2)$, zachodzi równość

$$g(T(U,\lambda)) = c_N \exp\Big(-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2\Big).$$

Przyjrzyjmy się dokładnie macierzy DT. Oznaczmy $[h_{ij}] = T(U,\lambda)$. Przy ustalonym U, odwzorowanie T jest liniową funkcją λ, więc $\frac{\partial T}{\partial \lambda}$ nie zależy od λ, zaś $\frac{\partial h_{ij}}{\partial u_r}$ dla $r = 1, 2, \ldots, N(N-1)/2$ również jest liniową funkcją λ. Wynika stąd, że $\det DT$ jest przy ustalonym U wielomianem zmiennej λ stopnia co najwyżej $N(N-1)/2$.

Zauważmy ponadto, że wzór $T(U,\lambda) = T_1(U)Diag(\lambda_1,\ldots,\lambda_N)T_1(U)^T$ pozwala rozszerzyć T na $A \times \mathbb{R}^N$. Wówczas, dla λ takich, że $\lambda_i = \lambda_j$ przy pewnych $i \neq j$, przekształcenie T nie jest $1-1$ w otoczeniu punktu (U,λ). Wynika to stąd, że przestrzeń wektorów własnych dla pewnej wartości własnej jest przynajmniej dwuwymiarowa i możemy w niej lekko obrócić bazę wektorów własnych. Ponieważ $\tilde{O}_n$ jest zbiorem otwartym, parametryzowanym przez A, istnieją punkty (U',λ) dowolnie bliskie (U,λ), takie że $T(U',\lambda) = T(U,\lambda)$. Zatem $DT(U,\lambda)$ nie może być odwracalne, czyli $\det DT = 0$, skąd wynika, że $\det DT$, jako wielomian zmiennej λ jest podzielny przez $\prod_{i<j}(\lambda_i - \lambda_j)$. Ale, jak wiemy, $\det DT$ jest wielomianem stopnia $N(N-1)/2$, skąd wynika, że

$$\det DT(U,\lambda) = f(U)\prod_{i<j}(\lambda_i - \lambda_j)$$

dla pewnej funkcji f. Podstawiając to wyrażenie do wzoru (9), otrzymujemy

$$\mathbb{P}(T^{-1}(H) \in B) = \int_B c_N \prod_{i<j}|\lambda_i - \lambda_j| \exp\Big(-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2\Big) f(U) d\lambda dU$$

dla dowolnego zbioru otwartego w $A \times \Delta$. Z lematu o $\pi - \lambda$ układach wynika, że równość ta zachodzi dla dowolnego zbioru borelowskiego, zatem gęstość $T^{-1}(H)$ jest postaci

$$c_N \prod_{i<j}|\lambda_i - \lambda_j| \exp\Big(-\frac{1}{4}\sum_{i=1}^{N}\lambda_i^2\Big) f(U),$$

skąd po odcałkowaniu względem U dostajemy gęstość wektora wartości własnych. □

Konstrukcja parametryzacji użytej w dowodzie Twierdzenia 10

Zbiór $\tilde{O}_N$, wykorzystany w dowodzie twierdzenia można zdefiniować np. jako zbiór wszystkich macierzy ortogonalnych, których elementy diagonalne są ściśle dodatnie, a wszystkie minory niezerowe. Oczywiście jest to zbiór otwarty.

Wykażemy najpierw następujący

Lemat 6. *Zbiór $\tilde{O}_N$ posiada gładką parametryzację pewnym otwartym podzbiorem $\mathbb{R}^{N(N-1)/2}$.*

Dowód. Dla $U \in \tilde{O}_N$, zdefiniujmy

$$F(U) = \Big(\frac{U_{12}}{U_{11}}, \ldots, \frac{U_{1N}}{U_{11}}, \frac{U_{23}}{U_{22}} \ldots, \frac{U_{2N}}{U_{22}}, \ldots, \frac{U_{N-2,N-1}}{U_{N-2,N-2}}, \frac{U_{N-2,N}}{U_{N-2,N-2}}, \frac{U_{N-1,N}}{U_{N-1,N-1}}\Big) = \Big(\frac{U_{ij}}{U_{ii}}\Big)_{1\leqslant i<j\leqslant N}.$$

Niech $A = F(\tilde{O}_N)$. Wykażemy, że A jest otwarty, oraz, że F ma gładką odwrotność $S = F^{-1}\colon A \to \tilde{O}_N$.

Postać przekształcenia S możemy otrzymać poprzez indukcję. Niech $(v_{ij})_{i<j} = F(U)$, gdzie $U = [U_{ij}]_{i,j\leqslant N} \in O_N$. Zauważmy, że warunki $\sum_{j=1}^{N} U_{ij}^2 = 1$ oraz $U_{ii} > 0$ (wynikające z definicji zbioru $\tilde{O}_N$), implikują, że

$$U_{ii} = \frac{1}{\sqrt{1 + \sum_{j\neq i}^{N} \frac{U_{ij}^2}{U_{ii}^2}}}. \tag{10}$$

Pozwala to wyznaczyć U_{11} na podstawie $(v_{1j})_{1<j\leqslant N}$. Dalsze wyrazy macierzy U możemy wyznaczyć wiersz po wierszu, indukcyjnie. Załóżmy, że znamy już U_{ij} dla $i \leqslant k$, $j \leqslant N$. Macierz U jest ortogonalna, więc jej $(k+1)$–wszy wiersz jest prostopadły do poprzednich, czyli dla $i \leqslant k$,

$$\sum_{j=1}^{N} U_{ij} U_{k+1,j} = 0.$$

Dzieląc powyższą równość przez $U_{k+1.k+1}$, oraz korzystając z faktu, że $U_{k+1,j}/U_{k+1,k+1} = v_{k+1,j}$ dla $j > k+1$, otrzymujemy układ równań na $x_j = U_{k+1,j}/U_{k+1,k+1}$, $j \leqslant k$:

$$\sum_{j=1}^{k} U_{ij} x_j = -1 - \sum_{j=k+2}^{N} U_{ij} v_{k+1,j}$$

dla $i = 1, \ldots, k$. Ponieważ macierz U ma z założenia niezerowe wiodące minory główne, układ ten ma dokładnie jedno rozwiązanie. Znając x_j oraz $v_{k+1,j}$, możemy wyznaczyć $U_{k+1,k+1}$ z równości (10), co następnie pozwala na wyznaczenie współczynników $U_{k+1,j}$, $j \neq k+1$.

Aby udowodnić, że A jest otwarty, a S gładkie, zauważmy, że powyższa konstrukcja może być przeprowadzona dla $v = (v_{ij})_{i<j} \in \mathbb{R}^{N(N-1)/2}$, jeśli wyznaczniki macierzy zadających układy równań są w każdym kroku odwracalne. Zbiór macierzy odwracalnych jest otwarty, a rozwiązanie układu równań, którego współczynniki tworzą macierz odwracalną, lokalnie jest gładką funkcją tych współczynników, możemy więc przez indukcję wykazać, że zbiór tych wektorów v, dla których możemy przeprowadzić powyższą konstrukcję jest otwarty, a macierz U, przez nią zdefiniowana jest gładką funkcją v. Aby wektor v należał do A, dodatkowo wszystkie minory otrzymanej macierzy muszą być niezerowe, warunek ten oczywiście zadaje zbiór otwarty. □

Twierdzenie 11. *Zbiór macierzy symetrycznych H, nie posiadających reprezentacji postaci*

$$H = UDiag(\lambda_1, \ldots, \lambda_N)U^T,$$

gdzie $\lambda_1 < \lambda_2 < \ldots < \lambda_N$, $U \in \tilde{O}_N$, jest miary Lebesgue'a zero (jak zwykle utożsamiamy macierze symetryczne z wektorami z $\mathbb{R}^{N(N+1)/2}$).

Powyższe twierdzenie udowodnimy, znajdując wielomiany, które znikają na macierzach nie posiadających odpowiedniej reprezentacji i korzystając z następującego lematu, którego dowód zostawiamy jako ćwiczenie.

Lemat 7. *Niech P będzie dowolnym niezerowym wielomianem n zmiennych. Wówczas zbiór $\{x \in \mathbb{R}^n : P(x) = 0\}$ ma miarę Lebesgue'a zero.*

Przypomnijmy teraz dwa klasyczne pojęcia algebraiczne dotyczące wielomianów jednej zmiennej.

Definicja 6. *Wyróżnikiem wielomianu rzeczywistego $P(x) = a_n x^n + \ldots + a_1 x + a_0$ nazwiemy liczbę*

$$\Delta(P) = a_n^{2n-1} \prod_{i<j} (\lambda_i - \lambda_j)^2,$$

gdzie $\lambda_1, \ldots, \lambda_n$ są pierwiastkami P.

Uwaga Z reguły wyróżnik definiuje się jako $a_n^{2n-2} \prod_{i<j} (\lambda_i - \lambda_j)^2$ (w szczególnym przypadku wielomianów stopnia 2 daje to znany ze szkoły wzór $\Delta = b^2 - 4ac$).

Definicja 7. *Rugownikiem wielomianów $P = a_n x^n + \ldots + a_0$ i $Q = b_m x^m + \ldots + b_0$ nazwiemy liczbę*

$$r(P, Q) = a_n^m b_m^n \prod_{i,j} (\lambda_i - \gamma_j),$$

gdzie λ_i, γ_j to pierwiastki odpowiednio wielomianu P i Q.

Bezpośrednio z definicji wynika, że wyróżnik wielomianu jest równy zero, wtedy i tylko wtedy, gdy ma on pierwiastki wielokrotne. Podobnie, rugownik dwóch wielomianów wynosi zero, wtedy i tylko wtedy, gdy mają one wspólny pierwiastek. Ponadto, jak łatwo wykazać

$$\Delta(P) = (-1)^{n(n-1)/2} r(P, P'),$$

gdzie n to stopień P.

Rugownik wyraża się wielomianowo poprzez współczynniki a_k, b_k. Powyższa równość implikuje, że również wyróżnik jest wielomianem od współczynników a_k (nietrudno wykazać, że jest tak również przy „standardowej" definicji wyróżnika, z potęgą a_n^{2n-2}, nie jest to jednak istotne dla naszych zastosowań).

Fakt 4. *Niech $P = a_n x^n + \ldots + a_0$, $Q = b_m x^m + \ldots + b_0$. Wówczas*

$$r(P,Q) = \det \begin{bmatrix} a_n & a_{n-1} & \ldots & a_1 & a_0 & 0 & \ldots & \ldots & 0 \\ 0 & a_n & a_{n-1} & \ldots & a_1 & a_0 & 0 & \ldots & 0 \\ 0 & 0 & a_n & a_{n-1} & \ldots & a_1 & a_0 & \ldots & 0 \\ \ldots & \ldots & \ldots & & & & & & \\ 0 & \ldots & \ldots & 0 & a_n & a_{n-1} & \ldots & a_1 & a_0 \\ b_m & b_{m-1} & \ldots & \ldots & b_1 & b_0 & 0 & \ldots & 0 \\ 0 & b_m & b_{m-1} & \ldots & \ldots & b_1 & b_0 & \ldots & 0 \\ \ldots & \ldots & \ldots & & & & & & \\ \ldots & \ldots & 0 & b_m & b_{m-1} & \ldots & \ldots & b_1 & b_0 \end{bmatrix}$$

(powyższa macierz ma $n+m$ wierszy, m razy powtarza się w niej odpowiednio przesunięty ciąg a_k, następnie n razy ciąg b_k).

Lemat 8. *Niech H będzie macierzą symetryczną, której wszystkie wartości własne są jednokrotne i która posiada przedstawienie $H = UDU^T$, gdzie D jest macierzą diagonalną, zaś U macierzą ortogonalną, której przynajmniej jeden współczynnik wynosi zero. Wówczas, dla pewnego k, macierz $H^{(k)}$ otrzymana z H przez usunięcie k-tej kolumny i k-tego wiersza ma wspólną wartość własną z H.*

Dowód. Dla dowolnej macierzy A, oznaczmy przez $A^{(k,l)}$ macierz powstałą z A przez usunięcie k-tej kolumny i l-tego wiersza. Załóżmy nie wprost, że H oraz $H^{(k)}$ nie posiadają wspólnej wartości własnej dla żadnego k i rozpatrzmy dowolną wartość własną H (oznaczmy ją przez λ). Niech $A = H - \lambda I$. Z rozwinięcia Laplace'a, $AA^{adj} = \det(A)\mathrm{Id} = 0$, gdzie $A^{adj} = [(-1)^{i+j}\det(A^{(i,j)})]_{i,j\leqslant N}$. Zatem kolumny macierzy A^{adj} są wektorami własnymi H, odpowiadającymi wartości własnej λ. Z założenia, λ ma krotność 1, zatem wszystkie kolumny A są wielokrotnościami pewnego jednostkowego wektora własnego v macierzy A (wyznaczonego z dokładnością do zwrotu). Zauważmy, że $A_{ii}^{adj} = \det(H^{(k)} - \lambda\mathrm{Id}) \neq 0$ (z założenia, że H i $H^{(k)}$ nie dzielą wartości własnej). Wynika stąd, że $v_i \neq 0$. Ponieważ każda kolumna macierzy U jest wektorem własnym H (wyznaczonym z dokładnością do zwrotu), otrzymujemy stąd, że wszystkie współczynniki U są niezerowe, co przeczy założeniom twierdzenia. □

Wniosek Niech $\mathcal{H} \subseteq \mathcal{M}(N,N)$ będzie zbiorem macierzy symetrycznych H, posiadających wielokrotne wartości własne lub reprezentację postaci $H = UDU^T$, gdzie D jest macierzą diagonalną, zaś $U \in O_N$ oraz pewien współczynnik U wynosi zero. Wówczas istnieje wielomian $N(N+1)/2$ zmiennych P, taki że $P(\mathcal{H}) = 0$ (jak zwykle utożsamiamy macierz symetryczną z elementem $\mathbb{R}^{N(N+1)/2}$).

Dowód. Niech $Q(H)$ oznacza wielomian charakterystyczny macierzy H. Jego współczynniki są oczywiście wielomianami od elementów macierzy. Macierz H ma wielokrotne wartości własne wtedy i tylko wtedy, gdy $\Delta(Q(H)) = 0$. Z kolei z lematu, macierz H o jednokrotnych wartościach własnych, posiadająca reprezentację jak w sformułowaniu twierdzenia, spełnia

$$\prod_{k=1}^{N} r(Q(H), Q(H^{(k)})) = 0.$$

Wielomianem spełniającym żądane warunki jest zatem np. $P(H) = \Delta(Q(H)) \prod_{k=1}^{N} r(Q(H), Q(H^{(k)}))$.
$\square$

Aby wyprowadzić z powyższego lematu Twierdzenie 11 potrzebujemy jeszcze znanego wzoru Cauchy'go-Bineta.

Fakt 5. *Niech $A = [a_{ij}] \in \mathcal{M}(m, n)$, $B = [b_{ij}] \in \mathcal{M}(n, m)$. Dla $I \subseteq \{1, \ldots, n\}$, $|I| = m$ oznaczmy przez $A_I = [a_{ij}]_{i \leqslant m, j \in I}$, $B_I = [b_{ij}]_{i \in I, j \leqslant m}$. Wówczas zachodzi równość*

$$\det(AB) = \sum_{\substack{I \subseteq \{1,\ldots,n\} \\ |I|=m}} \det(A_I) \det(B_I).$$

Dowód Twierdzenia 11. Wykażemy fakt nieco mocniejszy niż teza twierdzenia, mianowicie, że miary zero jest zbiór macierzy H, nie posiadających reprezentacji postaci $H = UDiag(\lambda_1, \ldots, \lambda_N)U^T$, gdzie $U \in \tilde{O}_N$ oraz dla dowolnych $I, J \subseteq \{1, \ldots, N\}$,

$$|I| = |J|, I \neq J \implies \prod_{i \in I} \lambda_i \neq \prod_{i \in J} \lambda_i. \tag{11}$$

Dla dowolnego $k = 1, \ldots, N$ oraz zbiorów $I, J \subseteq \{1, \ldots, N\}$, mocy k, oznaczmy $(\bigwedge^k H)_{I,J} = \det H_{I,J}$, gdzie $H_{I,J}$ jest minorem macierzy H wyznaczonym przez zbiór I wierszy i zbiór J kolumn. $\bigwedge^k H$ jest zatem macierzą rozmiarów $\binom{N}{k} \times \binom{N}{k}$. Łatwo sprawdzić, że $(\bigwedge^k H)^T = \bigwedge^k H^T$, zatem jeżeli H jest symetryczna, to również $\bigwedge^k H$ jest symetryczna. Ponadto wzór Cauchy'ego-Bineta implikuje, że $\bigwedge^k (H_1 H_2) = (\bigwedge^k H_1)(\bigwedge^k H_2)$, skąd wynika, że $\bigwedge^k U$ jest macierzą ortogonalną, co więcej jeżeli $H = U^T DU$, to $\bigwedge^k H = (\bigwedge^k U)^T (\bigwedge^k D) \bigwedge^k U)$, oraz $\bigwedge^k D$ jest macierzą diagonalną, której elementy diagonalne, to wszystkie iloczyny k spośród wartości własnych macierzy H.

Jeżeli zatem wartości własne macierzy H nie spełniają warunku (11), to dla pewnego k, macierz $\bigwedge^k D$ ma wielokrotne wartości własne. Z kolei, jeśli warunek ten jest spełniony, ale pewien minor macierzy U jest zerowy, to dla pewnego k, macierz $\bigwedge^k U$ ma zerowy współczynnik. Z lematu 8 wynika zatem, że zbiór macierzy nie posiadających żądanej reprezentacji zawiera się w zbiorze miejsc zerowych pewnego wielomianu, a zatem - z Lematu 7 - ma miarę zero. $\square$

3.2 Gaussian Unitary Ensemble

Definicja 8. *Macierzą GUE nazywamy macierz losową postaci*

$$H = \begin{bmatrix} X_{11} & \frac{X_{12}+iY_{12}}{\sqrt{2}} & \ldots & \frac{X_{1N}+iY_{1N}}{\sqrt{2}} \\ \frac{X_{12}-iY_{12}}{\sqrt{2}} & \ddots & \ldots & \frac{X_{2N}+iY_{2N}}{\sqrt{2}} \\ \vdots & \ldots & \ddots & \vdots \\ \frac{X_{1N}-iY_{1N}}{\sqrt{2}} & \frac{X_{2N}-iY_{2N}}{\sqrt{2}} & \ldots & X_{NN} \end{bmatrix}, \tag{12}$$

gdzie $X_{ij}, Y_{i,j}$ są niezależnymi zmiennymi $\mathcal{N}(0, 1)$.

Uwagi

1. Macierz H możemy utożsamić z losowym elementem $\mathbb{R}^{N^2}$ (rozpatrując jej współczynniki na przekątnej oraz części rzeczywiste i urojone współczynników nad przekątną). Wówczas jej rozkład ma gęstość
$$\frac{1}{2^{N/2}\pi^{N^2/2}} \exp\Big(-\frac{1}{2}\mathrm{tr}\, H^2\Big).$$
2. Podobnie jak w przypadku macierzy GOE, rozkład macierzy GUE jest niezmienniczy ze względu na wybór bazy ortogonalnej. W tym wypadku oznacza to, że
$$H \sim U^* HU,$$
dla dowolnej macierzy unitarnej.

3. Można wykazać, że jeżeli macierz losowa H jest postaci (12), gdzie X_{ij}, Y_{ij} $i \leqslant j$ są rzeczywistymi niezależnymi zmiennymi losowymi, ma rozkład absolutnie ciągły, niezmienniczy ze względu na sprzężenia macierzami unitarnymi, to jej gęstość jest postaci $g(H) = \exp(-a\mathrm{tr}\, H^2 - b\mathrm{tr}\, H - c)$ dla pewnych $a, b, c \in \mathbb{R}$, $a > 0$.

Również da macierzy GUE można znaleźć gęstość wektora wartości własnych. Dowód poniższego twierdzenia jest zbliżony do dowodu Twierdzenia 10, ale zawiera dodatkowe szczegóły techniczne (ze względu na konieczność sparametryzowania macierzy unitarnych oraz na fakt, że wyznacznik Vandermonde'a pojawia się w gęstości w potędze 2).

Twierdzenie 12. *Gęstość wektora wartości własnych macierzy GUE rozmiarów $N \times N$ wyraża się wzorem*

$$D_N \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} \exp(-\sum_i \lambda_i^2/2)\mathbf{1}_{\{\lambda_1<\lambda_2<\ldots<\lambda_N\}},$$

gdzie D_N jest pewną stałą.

4 Twierdzenie Marczenki-Pastura

Dowód twierdzenia Marczenki-Pastura przedstawiony poniżej jest pewnym przeformułowaniem dowodu z [BS]. W książce tej można znaleźć również dowód kombinatoryczny oraz pewne uogólnienia. Informacje o słabej zbieżności miar niekoniecznie probabilistycznych można znaleźć w książce [F] lub (w ujęciu badziej abstrakcyjnym) w większości podręczników do analizy funkcjonalnej. Podstawowe fakty dotyczące transformaty Stieltjesa są podane w [AGZ] oraz [BS]

Omówimy teraz drugi, po twierdzeniu Wignera, podstawowy fakt dotyczący słabej zbieżności miar spektralnych macierzy losowych – twierdzenie Marczenki Pastura.

Rozważmy tablicę $(X_{ij})_{1\leqslant i,j<\infty}$ niezależnych rzeczywistych zmiennych losowych o tym samym rozkładzie i ustalmy liczbę $y \in (0, \infty)$ oraz ciąg liczb naturalnych N_n, taki że $\lim_{n\to\infty} N_n/n = y$.

Zdefiniujmy macierz losową wymiarów N_n na n,

$$M_n = [X_{ij}]_{i\leqslant N_n, j\leqslant n}.$$

Będziemy zainteresowani zachowaniem miary spektralnej macierzy

$$S = S_n = \frac{1}{n} M_n M_n^T.$$

Uwaga Jak łatwo sprawdzić, macierz S_n jest nieujemnie określona, skąd wynika, że jej miara spektralna skupiona jest na półprostej $[0, \infty)$.

Macierz S ma wymiary $N_n \times N_n$ oraz jest rzędu co najwyżej $\min(N_n, n)$. Wynika stąd, że jeżeli $N_n > n$, 0 jest wartością własną S, krotności przynajmniej $N_n - n$, co dla $y > 1$ jest rzędu $(y-1)n$. Zatem dla $y > 1$ miara spektralna S_n ma dla dużych n atom w zerze o wadze przynajmniej $(y-1)n/N_n \to 1-y^{-1}$. Możemy więc podejrzewać, że asymptotyczne zachowanie miary spektralnej macierzy S_n jest różne dla $y < 1$ oraz $y > 1$. Jest tak w istocie, niemniej przypadki te są ze sobą ściśle powiązane. Zachodzi bowiem następujący fakt, którego dowód pozostawiamy jako ćwiczenie.

Fakt 6. *Dla dowolnej macierzy A, niezerowe wartości własne macierzy AA^T oraz A^TA są równe i mają te same krotności.*

Graniczne zachowanie miary spektralnej macierzy S_n jest opisane przez tzw. rozkład Marczenki-Pastura.

Definicja 9. *Dla $y > 0$ zdefiniujmy $a = a_y = (1 - \sqrt{y})^2$, $b = b_y = (1 + \sqrt{y})^2$. Rozkładem Marczenki-Pastura nazwiemy miarę ν_y o gęstości*

$$g(x) = \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}\mathbf{1}_{[a,b]}(x)$$

na $(0,\infty)$ *oraz atomie w zerze o wadze* $1-1/y$ *dla* $y>1$.

Zachodzi następujące

Twierdzenie 13 (Marczenko-Pastur). *Dla dowolnego* $y>0$ *oraz ciągu liczb naturalnych* N_n, *takiego, że* $N_n/n \to y$ *oraz tablicy* $(X_{ij})_{1\leqslant i,j<\infty}$ *zmiennych losowych i.i.d. o wariancji 1, miara spektralna macierzy* $n^{-1}S_n$ *zbiega prawie na pewno do rozkładu* ν_y.

Podobnie jak twierdzenie Wignera, powyższe twierdzenie można udowodnić metodą momentów, poprzez analizę śladów potęg macierzy S_n. My jednak podamy inny, bardziej analityczny dowód, oparty o transformatę Stieltjesa.

Zanim przejdziemy do omówienia metody dowodu, zasygnalizujmy związek pomiędzy powyższym twierdzeniem Marczenki-Pastura, a problemem przybliżania macierzy kowariancji wektorów losowych. Rozważmy wektor losowy $X=(X_1,\ldots,X_p)^T$. Przypuśćmy, że chcemy przybliżać macierz kowariancji wektora X na podstawie n niezależnych kopii $X^{(1)},\ldots,X^{(n)}$ tego wektora. Załóżmy dla uproszczenia, że współrzędne wektora X są scentrowane. Naturalny estymator macierzy kowariancji ma postać

$$\hat{S}=\frac{1}{n}\sum_{j=1}^{n}X^{(j)}(X^{(j)})^T.$$

Okazuje się, że jeżeli n jest proporcjonalne do p, powyższy estymator nie jest dobrym przybliżeniem macierzy kowariancji. Uzasadnienie, które przedstawiamy poniżej jest raczej nieformalne, jego celem jest jednak raczej przedstawienie związków z twierdzeniem Marczenki-Pastura niż precyzowanie modelu statystycznego lub wyjasnianie, co rozumiemy przez „dobre przybliżenie”. Wyjaśnijmy więc tylko, że odnosi się ono do częstego w statystyce modelu, w którym wymiar danych jest duży (co modelujemy biorąc $p\to\infty$), zaś liczność próbki nie przekracza jego ustalonej wielokrotności.

Załóżmy zatem dodatkowo, że współrzędne X_i są zmiennymi i.i.d., o wariancji 1. Oczywiście wówczas macierz kowariancji wektora X jest równa Id. Zauważmy, że oznaczając $X^{(j)}=(X_{ij})_{i\leqslant p}$, dostajemy

$$(\hat{S})_{kl}=\frac{1}{n}\sum_{j=1}^{n}X_{kj}X_{jl}=\frac{1}{n}(M_nM_n^T)_{kl},$$

gdzie $M_n=[X_{ij}]_{i\leqslant p, j\leqslant n}$. Jeżeli $p,n\to\infty$ w taki sposób, że $p/n\to y\in(0,\infty)$, z Twierdzenia Marchenki-Pastura wynika, że macierz $\hat{S}$ ma z prawdopodobieństwem oddzielonym od zera normę operatorową około $(1+\sqrt{y})^2$, a więc nie może być bliska Id, czyli prawdziwej macierzy kowariancji.

4.1 Podstawowe informacje o transformacie Stieltjesa

Definicja 10. *Transformatą Stieltjesa nieujemnej, skończonej miary borelowskiej* μ *na* $\mathbb{R}$ *nazywamy funkcję* $s_\mu\colon \mathbb{C}^+=\{z\in\mathbb{C}\colon \operatorname{Im} z>0\}\to\mathbb{C}$, *daną wzorem*

$$s_\mu(z)=\int_{\mathbb{R}}\frac{1}{x-z}d\mu(x).$$

Uwaga Całka w powyższej definicji jest dobrze określona, gdyż dla każdego z o dodatniej części urojonej, funkcja podcałkowa jest ograniczona. Ponadto, jak nietrudno wykazać, funkcja s_μ jest analityczna w $\mathbb{C}^+$.

Transformata Stieltjesa jest dość użytecznym narzędziem w badaniu słabej zbieżności miar probabilistycznych, pełniąc rolę podobną do funkcji charakterystycznej (transformaty Fouriera). Szczególne znaczenie transformaty Stieltjesa w badaniu miar spektralnych macierzy losowych wynika z jej dobrych własności algebraicznych, dających wzory rekurencyjne, pozwalające na obniżenie wymiaru rozpatrywanych macierzy (podobnie przydatność funkcji charakterystycznych w badaniu sum niezależnych zmiennych losowych wiąże się z faktem, że funkcja charaktersytyczna takich sum to iloczyn funkcji charakterystycznych składników).

Poniżej przedstawiamy podstawowe fakty dotyczące związków transformaty Stieltjesa ze słabą zbieżnością miar.

Twierdzenie 14 (Transformata Stieltesa jest odwracalna)**.** *Dla dowolnego otwartego przedziału* $I = (a, b)$*, którego końce nie sa atomami* μ*, zachodzi wzór*

$$\mu(I) = \lim_{\varepsilon \to 0} \frac{1}{\pi} \int_I \operatorname{Im} s_\mu(x + i\varepsilon) dx.$$

Dowód. Mamy

$$\begin{aligned} \frac{1}{\pi} \int_a^b \operatorname{Im} s_\mu(x + i\varepsilon) dx &= \frac{1}{\pi} \int_a^b \int_{\mathbb{R}} \operatorname{Im} \frac{1}{\lambda - x - i\varepsilon} d\mu(\lambda) dx \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \int_a^b \frac{\varepsilon}{(\lambda - x)^2 + \varepsilon^2} dx d\mu(\lambda) \\ &= \int_{\mathbb{R}} \frac{1}{\pi} \Big(\arctan((b - \lambda)/\varepsilon) - \arctan((a - \lambda)/\varepsilon) \Big) d\mu(\lambda). \end{aligned}$$

Dla $A \in \mathbb{R}$ mamy $\lim_{\varepsilon \to 0} \arctan(A/\varepsilon) = (\operatorname{sgn} A)\pi/2$, zatem dla $\lambda < a$ oraz $\lambda > b$, funkcja podcałkowa zbiega do 0 przy $\varepsilon \to 0$, zaś dla $\lambda \in (a, b)$ funkcja ta zbiega do $\pi^{-1}(\pi/2 - (-\pi/2)) = 1$. Teza wynika zatem z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej. □

Zanim sformułujemy kolejne twierdzenie, przypomnijmy, że słabą zbieżność miar można określić nie tylko na miarach probabilistycznych, ale ogólniej na miarach nieujemnych skończonych. Dokładniej, mówimy, że ciąg miar μ_n zbiega słabo do miary μ jeżeli

$$\lim_{n \to \infty} \int_{\mathbb{R}} f d\mu_n \to \int_{\mathbb{R}} f d\mu, \tag{13}$$

dla dowolnej funkcji ciągłej i ograniczonej, znikającej w nieskończoności. Jeżeli miary μ_n sa probabilistyczne, oraz μ_n zbiega słabo do μ, to $\mu(\mathbb{R}) \leqslant 1$, przy czym może się oczywiście zdarzyć, że $\mu(\mathbb{R}) < 1$ (w takich sytuacjach μ jest czasami nazywana rozkładem ułomnym). Ponadto, jak łatwo wykazać, jeżeli miara graniczna μ też jest miarą probabilistyczną, to (13) zachodzi dla dowolnej funkcji ciągłej i ograniczonej, co usprawiedliwia stosowanie terminologii *słaba zbieżność* w szerszym kontekście miar niekoniecznie probabilistycznych.

Jednym z podstawowych faktów dotyczących słabej zbieżności jest twierdzenie mówiące, że z każdego ciągu miar probabilistycznych można wybrac podciąg zbieżny (być może do rozkładu ułomnego). Fakt ten (sformułowany w języku dystrybuant) jest często stosowany do dowodu twierdzenia Prochorowa. Wynika on również dość łatwo z tego twierdzenia zastosowanego do prostej uzwarconej nieskończonością. Alternatywnie, dzięki twierdzeniu Riesza, słabą zbiezność miar skończonych można traktować jako szczególny przypadek $*$-słabej zbiezności znanej z teorii przestrzeni Banacha. Prezwartość rodziny miar probabilistycznych jest wówczas szczególnym przypadkiem twierdzenia Banacha-Alaoglu.

Twierdzenie 15. *Niech* μ_n *będzie ciągiem miar probabilistycznych na* $\mathbb{R}$*.*

a) Jeżeli μ_n *zbiega słabo do miary* μ *(niekoniecznie probabilistycznej), to* $s_{\mu_n(z)} \to s_\mu(z)$ *dla dowolnego* $z \in \mathbb{C}^+$*.*

b) Jeżeli $s_{\mu_n}(z) \to s(z)$ *dla dowolnego* $z \in \mathbb{C}^+$*, to istnieje dokładnie jedna miara nieujemna skończona, taka że* $s = s_\mu$*. Ponadto* μ_n *zbiega słabo do* μ*.*

Dowód. Część a) wynika z faktu, że dla ustalonego $z \in \mathbb{C}^+$, funkcja $x \mapsto (x - z)^{-1}$ jest funkcją ograniczoną, znikającą dla $|x| \to \infty$.

Aby wykazać punkt b), zauważmy, że punkt a) oraz odwracalność transformaty Stieltjesa implikują, że wszystkie podciągi słabo zbieżne ciągu μ zbiegają do tej samej granicy (ozn. ją przez μ). Wraz z prezwartością μ_n implikuje to już tezę. □

Zanim przejdziemy do dowodu Twierdzenia 13, wprowadzimy jeszcze pewne użyteczne fakty, których dowód zostawiamy jako ćwiczenie.

Przypomnijmy, że dla dowolnej macierzy A, przez $A^{(k)}$ oznaczamy macierz powstałą z A przez usunięcie k-tego wiersza i k-tej kolumny.

Lemat 9. *Niech $A = [a_{ij}]_{i,j\leqslant N}$ będzie macierzą $N \times N$. Załóżmy, że A są oraz $A^{(k)}$ $(k \leqslant N)$ są odwracalne oraz oznaczmy $H = [h_{ij}]_{i,j\leqslant N} = A^{-1}$. Wówczas*

$$h_{kk} = \frac{1}{a_{kk} - \alpha_k^T (A^{(k)})^{-1}\beta_k},$$

gdzie α_k^T oraz β_k są wektorami otrzymanymi odpowiednio z k-tego wiersza i k-tej kolumny A przez usunięcie k-tej współrzędnej.

W szczególności

$$\operatorname{tr} A^{-1} = \sum_{k=1}^{N} \frac{1}{a_{kk} - \alpha_k^T (A^{(k)})^{-1}\beta_k}$$

Lemat 10. *Jeśli macierz A jest symetryczna i odwracalna oraz $A^{(k)}$ jest odwracalna, to*

$$\operatorname{tr} A^{-1} - \operatorname{tr}(A^{(k)})^{-1} = \frac{1 + \alpha_k^T (A^{(k)})^{-2}\alpha_k}{a_{kk} - \alpha_k^T (A^{(k)})^{-1}\alpha_k}$$

Lemat 11. *Dla dowolnej miary μ i dowolnego $z \in \mathbb{C}^+$, $\operatorname{Im} s_\mu(z) > 0$.*

Lemat 12. *Jeśli μ jest miarą probabilistyczną skupioną na $[0,\infty)$, to dla każdego $z \in \mathbb{C}^+$ oraz $y > 0$,*

$$\operatorname{Im}(y + z - 1 + yzs_\mu(z)) > \operatorname{Im} z.$$

4.2 Dowód twierdzenienia Marczenki-Pastura

Podobnie jak w przypadku twierdzenia Wignera, poprzez zastosowanie Twierdzenia 8, dowód można sprowadzić do przypadku zmiennych scentrowanych, zaś nierówność Hoffmana Wielandta pozwala ograniczyć się do zmiennych ograniczonych. Ponieważ szczegóły techniczne są bardzo podobne do dowodów Twierdzeń 5 i 7, pominiemy je i od razu założymy, że zmienne X_{ij} sa ograniczone i mają średnią zero.

Przy tych założeniach wykażemy, że z prawdopodobieństwem 1, transformaty Stieltjesa (losowych) miar spektralnyh macierzy S_n zbiegają punktowo do transformaty Stieltjesa miary ν_y. Poprzez Twierdzenie 15 implikuje to już tezę Twierdzenia 13.

Oznaczmy miarę spektralną S_n przez L_n, zaś jej transformatę Stieltjesa przez s_n.

Dowód podzielony jest na cztery części. Najpierw wyznaczymy Transformatę Stieltjesa miary ν_y. Następnie wykażemy, że dla każdego $z \in \mathbb{C}^+$, z prawdopodobieństwem 1, $s_n(z) - \mathbb{E}s_n(z) \to 0$. Kolejnym krokiem będzie wykazanie, że dla dowolnego $z \in \mathbb{C}^+$, $\mathbb{E}s_n(z) \to s_{\nu_y}(z)$ (czyli wykorzystamy typowy schemat „zbieżność średnich + koncentracja"). Stąd już wynika, że z prawdopodobieństwem 1, $s_n(z) \to s_{\nu_y}(z)$ dla z z pewnego przeliczalnego gęstego podzbioru $\mathbb{C}^+$. Aby rozszerzyć tę zbieżność na cały zbiór $\mathbb{C}^+$ skorzystamy z ogólnych faktów dotyczących zbieżności funkcji holomorficznych (twierdzenie Vitaliego).

4.2.1 Transformata Steltjesa miar ν_y

Dla uproszczenia notacji oznaczmy $s^y(z) = s_{\nu_y}(z)$. Załóżmy najpierw, że $y < 1$. Z definicji

$$s^y(z) = \int_a^b \frac{1}{x - z} \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)}dx,$$

gdzie $a = (1-\sqrt{y})^2$, $b = (1+\sqrt{y})^2$. Przez podstawienie $x = 1 + y + 2\sqrt{y}\cos\theta$, otrzymujemy

$$\begin{aligned}
s^y(z) &= \int_0^\pi \frac{2}{\pi} \frac{\sin^2\theta}{(1+y+2\sqrt{y}\cos\theta)(1+y+2\sqrt{y}\cos\theta - z)} d\theta \\
&= \int_\pi^{2\pi} \frac{2}{\pi} \frac{\sin^2\theta}{(1+y+2\sqrt{y}\cos\theta)(1+y+2\sqrt{y}\cos\theta - z)} d\theta \\
&= \frac{1}{\pi} \int_0^{2\pi} \frac{\sin^2\theta}{(1+y+2\sqrt{y}\cos\theta)(1+y+2\sqrt{y}\cos\theta - z)} d\theta \\
&= -\frac{1}{4\pi i} \int_0^{2\pi} \frac{(e^{i\theta} - e^{-i\theta})^2}{e^{i\theta}(1+y+\sqrt{y}(e^{i\theta}+e^{-i\theta}))(1+y+\sqrt{y}(e^{i\theta}+e^{-i\theta}) - z)} ie^{i\theta} d\theta.
\end{aligned}$$

W ostatnim wyrażeniu możemy rozpoznać całkę krzywoliniową po okręgu (przypomnijmy, że dla krzywej γ i funkcji f $\int_\gamma f$ definiujemy jako $\int f(r(t))r'(t)$ gdzie r jest parametryzacją krzywej γ). Zatem

$$\begin{aligned}
s^y(z) &= -\frac{1}{4\pi i} \int_{S^1} \frac{(\xi - \xi^{-1})^2}{\xi(1+y+\sqrt{y}(\xi+\xi^{-1}))(1+y+\sqrt{y}(\xi+\xi^{-1}) - z)} d\xi \\
&= -\frac{1}{4\pi i} \int_{S^1} \frac{(\xi^2 - 1)^2}{\xi((1+y)\xi+\sqrt{y}(\xi^2+1))((1+y-z)\xi+\sqrt{y}(\xi^2+1))} d\xi.
\end{aligned}$$

Aby obliczyć powyższą całkę, możemy zastosować teorię residuów. Funkcja podcałkowa ma 5 biegunów, wszystkie rzędu 1. Są nimi $\xi_0 = 0$,

$$\begin{aligned}
\xi_1 &= \frac{-(1+y)-(1-y)}{2\sqrt{y}} = \frac{1}{\sqrt{y}}, \quad \xi_2 = \frac{-(1+y)+(1-y)}{2\sqrt{y}} = -\sqrt{y} \\
\xi_3 &= \frac{-(1+y)+z-\sqrt{(1-y)^2 - 2(1+y)z + z^2}}{2\sqrt{y}}, \quad \xi_4 = \frac{-(1+y)+z+\sqrt{(1-y)^2 - 2(1+y)z + z^2}}{2\sqrt{y}},
\end{aligned}$$

przy czym przyjmijmy, że przez $\sqrt{\cdot}$ oznaczamy gałąź główną pierwiastka. Zauważmy, że $|\xi_1| > 1$, $|\xi_2| < 1$. Ponadto $\xi_3\xi_4 = 1$. Oznaczmy $r = [-(1+y)+z]/2\sqrt{y}$ i zauważmy, że $\xi_4 = r + \sqrt{r^2-1}$. Łatwo sprawdzić, że niezależnie od tego czy r należy do pierwszej czy do drugiej ćwiartki płaszczyzny zespolonej ($r \in \mathbb{C}^+$ z założenia o z), części rzeczywiste r oraz $\sqrt{r^2-1}$ są tych samych znaków, podobnie części urojone. Ponieważ $r \neq 0$, wynika stąd, że $|\xi_4| > |\xi_3|$, a zatem $|\xi_3| < 1$, $|\xi_4| > 1$. W kole jednostkowym funkcja podcałkowa ma zatem trzy bieguny ξ_0, ξ_2, ξ_3. Obliczając residua w tych punktach, otrzymujemy

$$\begin{aligned}
s^y(z) &= -\frac{1}{2}\Big(-\frac{1}{y} - \frac{1}{yz}\sqrt{(1-y)^2 - 2(1+y)z + z^2} - \frac{1-y}{yz}\Big) \\
&= \frac{1-y-z+\sqrt{(1-y-z)^2 - 4yz}}{2yz}.
\end{aligned} \tag{14}$$

Powyższy wzór wyprowadziliśmy przy założeniu, że $y < 1$. Z ciągłości względem zmiennej y, jest on prawdziwy także dla $y = 1$.

Równość (14) zachodzi także dla $y > 1$. Wyprowadzenie jej sprowadza się do uwzględnienia atomu w zerze i obliczenia metodą residuów, tej samej całki, co dla $y < 1$. Szczegóły pominiemy, zaznaczając tylko, że w przypadku $y > 1$, bieguny, które należy uwzględnić we wzorze residuów to ξ_0, ξ_1, ξ_4.

4.2.2 Dowód zbieżności $s_n(z) - \mathbb{E}s_n(z) \to 0$ p.n. (przy ustalonym z).

Podstawowym naszym narzędziem będzie

Twierdzenie 16 (Nierówność Azumy). *Niech $M_0, M_1, \ldots, M_n$ będzie martyngałem względem pewnej filtracji $\mathcal{F}_i$, $M_0 = 0$. Oznaczmy $d_i = M_i - M_{i-1}$ $(i \geqslant 1)$. Wówczas, dla dowolnego $t \geqslant 0$,*

$$\mathbb{P}(|M_n| \geqslant t) \leqslant 2\exp\Big(-\frac{t^2}{2\sum_{i=1}^n \|d_i\|_\infty^2}\Big).$$

Dowód. Bez straty ogólności możemy założyć, że $\|d_i\|_\infty > 0$. Dla $\lambda > 0$ obliczmy

$$\mathbb{E}\exp(\lambda M_n) = \prod_{i=1}^n \|\mathbb{E}(\exp(\lambda d_i)|\mathcal{F}_{i-1})\|_\infty.$$

Mamy $d_i = 2^{-1}(1 - d_i/\|d_i\|_\infty)(-\|d_i\|_\infty) + 2^{-1}(1 + d_i/\|d_i\|_\infty)\|d_i\|_\infty$, zatem z wypukłości funkcji $x \mapsto \exp(\lambda x)$ dostajemy

$$\exp(\lambda d_i) \leqslant 2^{-1}(1 - d_i/\|d\|_\infty)\exp(-\lambda\|d_i\|_\infty) + 2^{-1}(1 + d_i/\|d\|_\infty)\exp(\lambda\|d_i\|_\infty).$$

Z warunku $\mathbb{E}(d_i|\mathcal{F}_{i-1}) = 0$, wynika że

$$\mathbb{E}(\exp(\lambda d_i)|\mathcal{F}_{i-1}) \leqslant \frac{e^{-\lambda\|d_i\|_\infty} + e^{\lambda\|d_i\|_\infty}}{2} \leqslant e^{\lambda^2\|d_i\|_\infty^2/2}.$$

Zatem $\mathbb{E}\exp(\lambda M_n) \leqslant \exp(\lambda^2 \sum_{i=1}^n \|d_i\|_\infty^2/2)$. Z nierówności Czebyszewa,

$$\mathbb{P}(M_n \geqslant t) \leqslant \inf_{\lambda>0} \mathbb{E}\exp(\lambda M_n - \lambda t) \leqslant \inf_{\lambda>0} \exp(\lambda^2 \sum_{i=1}^n \|d_i\|_\infty^2/2 - \lambda t).$$

Aby zakończyć dowód wystarczy teraz zauważyć, że dla $\lambda = t/\sum_i \|d_i\|_\infty^2$ mamy

$$\lambda^2 \sum_{i=1}^n \|d_i\|_\infty^2/2 - \lambda t = -\frac{t^2}{2\sum_{i=1}^n \|d_i\|_\infty^2}$$

oraz skorzystać z faktu, że ciąg $-M_i$ również jest martyngałem. □

Powyższe twierdzenie zastosujemy, aby wykazać, że dla dowolnego $z \in \mathbb{C}^+$ oraz $\varepsilon > 0$,

$$\sum_{n=1}^\infty \mathbb{P}(|s_n(z) - \mathbb{E}s_n(z)| \geqslant \varepsilon) < \infty, \tag{15}$$

skąd poprzez lemat Borela-Cantellego wynika, że $s_n(z) - \mathbb{E}s_n(z) \to 0$ p.n.

Jak łatwo sprawdzić, dla dowolnej odwracalnej macierzy A oraz wektorów α, β takich że $A + \alpha\beta^T$ jest odwracalne, mamy

$$(A + \alpha\beta^T)^{-1} = A^{-1} - \frac{A^{-1}\alpha\beta^T A^{-1}}{1 + \beta^T A^{-1}\alpha}. \tag{16}$$

Zauważmy, że dla $z \in \mathbb{C}^+$, macierze $S_n - z\mathrm{Id}$ są odwracalne (widmo S_n jest rzeczywiste, z tej obserwacji będziemy jeszcze wielokrotnie korzystać w podobnych sytuacjach, nie podając każdorazowo uzasadnienia). Oznaczając wartości własne macierzy S_n przez λ_i możemy napisać

$$s_n(z) = \frac{1}{N_n}\sum_{i=1}^{N_n} \frac{1}{\lambda_i - z} = \frac{1}{N_n}\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1}.$$

Oznaczmy kolumny macierzy M_n przez Y_i $(i = 1, \ldots, n)$ i przypomnijmy, że

$$S_n = \frac{1}{n}\sum_{i=1}^n Y_i Y_i^T.$$

Wektory losowe Y_i są oczywiście stochastycznie niezależne. Możemy zapisać

$$s_n(z) - \mathbb{E}s_n(z) = \frac{1}{N_n}\sum_{k=1}^{n}(\mathbb{E}_k \mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathbb{E}_{k-1}\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1}),$$

gdzie $\mathbb{E}_k = \mathbb{E}(\cdot|Y_1, \ldots, Y_k)$. Zauważmy, że $M_i = \sum_{k=1}^{i}(\mathbb{E}_k \mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathbb{E}_{k-1}\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1})$ jest martyngałem względem naturalnej filtracji generowanej przez Y_i. Oznaczmy $A_k = \frac{1}{n}\sum_{k\neq i} Y_i Y_i^T$ i zauważmy, że $\mathbb{E}_{k-1}\mathrm{tr}\,(A_k - z\mathrm{Id})^{-1} = \mathbb{E}_k \mathrm{tr}\,(A_k - z\mathrm{Id})^{-1}$, a zatem

$$\begin{aligned}
&\mathbb{E}_k \mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathbb{E}_{k-1}\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} \\
&= \mathbb{E}_k(\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathrm{tr}\,(A_k - z\mathrm{Id})^{-1}) - \mathbb{E}_{k-1}(\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathrm{tr}\,(A_k - z\mathrm{Id})^{-1})
\end{aligned}$$

Z równości (16) zastosowanej do $A = A_k - z\mathrm{Id}$, oraz $\alpha = \beta = n^{-1/2}Y_i$, otrzymujemy

$$\begin{aligned}
\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathrm{tr}\,(A_k - z\mathrm{Id})^{-1}) &= \mathrm{tr}\,\Big((A_k - z\mathrm{Id} + \alpha\alpha^T)^{-1} - (A_k - z\mathrm{Id})^{-1}\Big) \\
&= \frac{\mathrm{tr}\,(A_k - z\mathrm{Id})^{-1}\alpha\alpha^T(A_k - z\mathrm{Id})^{-1}}{1 + \alpha^T(A_k - z\mathrm{Id})^{-1}\alpha} \\
&= \frac{\alpha^T(A_k - z\mathrm{Id})^{-2}\alpha}{1 + \alpha^T(A_k - z\mathrm{Id})^{-1}\alpha}
\end{aligned}$$

Oznaczmy $a = \mathrm{Re}\, z$, $b = \mathrm{Im}\, z$. Niech ponadto λ_j będą wartościami własnymi macierzy $A_k - z\mathrm{Id}$ zaś y_j współrzędnymi wektora α w odpowiedniej bazie wektorów własnych. Wówczas,

$$\alpha^T(A_k - z\mathrm{Id})^{-2}\alpha = \Big|\sum_{j=1}^{N}\frac{y_j^2}{(\lambda_j - a - b\mathbf{i})^2}\Big| \leqslant \sum_{j=1}^{N}\frac{y_j^2}{(\lambda_j - a)^2 + b^2}.$$

Z kolei

$$\begin{aligned}
\mathrm{Im}\,\Big(1 + \alpha^T(A_k - z\mathrm{Id})^{-1}\alpha\Big) &= \mathrm{Im}\sum_{j=1}^{N}\frac{y_j^2}{\lambda_j - a - b\mathbf{i}} \\
&= \mathrm{Im}\sum_{j=1}^{N} y_j^2\frac{(\lambda_j - a) + b\mathbf{i}}{(\lambda_j - a)^2 + b^2} \\
&= b\sum_{j=1}^{N}\frac{y_j^2}{(\lambda_j - a)^2 + b^2}.
\end{aligned}$$

Wynika stąd, że $|\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathrm{tr}\,(A_k - z\mathrm{Id})^{-1})| \leqslant b^{-1}$, a zatem $\mathbb{E}_k \mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} - \mathbb{E}_{k-1}\mathrm{tr}\,(S_n - z\mathrm{Id})^{-1} \leqslant 2b^{-1}$. Możemy więc zastosować nierówność Azumy (udowodniliśmy ją dla martyngałów rzeczywistych, ale łatwo sprawdzić, że z dokładnością do stałych przenosi się ona na martyngały o wartościach zespolonych) i otrzymać

$$\mathbb{P}(N_n|s_n(z) - \mathbb{E}s_n(z)| \geqslant t) \leqslant 2\exp(-ct^2b^2/n).$$

Stąd, dla $\varepsilon > 0$,

$$\mathbb{P}(|s_n(z) - \mathbb{E}s_n(z)| \geqslant \varepsilon) \leqslant 2\exp(-c\varepsilon^2 N_n^2 b^2/n),$$

co dla dużych n jest rzędu $2\exp(-c\varepsilon^2 N_n b^2 y)$. Oczywiście wynika stąd zbieżność szeregu (15), co kończy dowód.

4.2.3 Dowód zbieżności $\mathbb{E}s_n(z) \to s^y(z)$.

Krok A. Zauważmy najpierw, że funkcja $s^y(z)$ spełnia następującą tożsamość

$$
\begin{aligned}
\frac{1}{1-y-z-yzs^y(z)} &= \frac{1}{1-y-z-yz\frac{1-y-z+\sqrt{(1-y-z)^2-4yz}}{2yz}} \\
&= \frac{2}{1-y-z-\sqrt{(1-y-z)^2-4yz}} \\
&= \frac{1-y-z+\sqrt{(1-y-z)^2-4yz}}{2yz} = s^y(z). \qquad (17)
\end{aligned}
$$

Dowód będzie opierał się na wykazaniu, że wartość oczekiwana transformaty Stieltjesa empirycznej miary spektralnej spełnia równanie będące dla dużych n nieznacznym zaburzeniem powyższego, co ze względu na stabilność rozwiązań pozwoli wykazać, że jej granicą jest $s^y(z)$. Szczegóły dowodu są niestety dość mało przyjemne rachunkowo.

Chcielibyśmy uzyskać równanie na s_n zbliżone do (17). Zdefiniujmy zatem liczbę $\delta_n = \delta_n(z)$, tak aby spełnione było równanie

$$\mathbb{E}s_n(z) = \frac{1}{1-z-y_n-y_nz\mathbb{E}s_n(z)} + \delta_n(z), \qquad (18)$$

gdzie $Y_n = N_n/n$. Powyższa definicja jest poprawna ze względu na Lemat 12, gwarantujący, że $1-z-y_n-y_nz\mathbb{E}s_n(z) \neq 0$.

Rozwiązując powyższe równanie względem $\mathbb{E}s_n(z)$, otrzymujemy dwa rozwiązania

$$S_1(z) = \frac{1}{2y_nz}\Big(1-z-y_n+y_nz\delta_n(z)+\sqrt{(1-z-y_n-y_nz\delta_n(z))^2-4y_nz}\Big)$$

oraz

$$S_2(z) = \frac{1}{2y_nz}\Big(1-z-y_n+y_nz\delta_n(z)-\sqrt{(1-z-y_n-y_nz\delta_n(z))^2-4y_nz}\Big).$$

Krok B. Wykażemy, że $\mathbb{E}s_n(z) = S_1(z)$. W tym celu wykorzystamy własności transformaty Stieltjesa. Jak łatwo wykazać, $\mathbb{E}s_n(z) \to 0$ dla $\operatorname{Im} z \to \infty$. Zatem z (18) otrzymujemy $\delta_n(z) \to 0$ dla $\operatorname{Im} z \to \infty$. Analizując definicję $S_2(z)$ widzimy, że zbieżność ta implikuje, że $S_2(z) \to -1/y_n$. Zatem dla z o dużej części urojonej mamy $\mathbb{E}s_n(z) = S_1(z)$.

Przypuśćmy, że równość ta nie zachodzi dla wszystkich $z \in \mathbb{C}^+$. Z ciągłości funkcji S_i istnieje $z \in \mathbb{C}^+$, takie że $S_1(z) = S_2(z)$. Wynika stąd, że

$$(1-z-y_n-y_nz\delta_n(z))^2 - 4y_nz = 0,$$

a zatem

$$\mathbb{E}s_n(z) = \frac{1-z-y_n-y_nz\delta_n(z)}{2y_nz}.$$

Wyliczając $\delta_n(z)$ z (18) i podstawiając do powyższego równania, dostajemy

$$\mathbb{E}s_n(z) = \frac{1-z-y_n}{2y_nz} + \frac{1}{2}\mathbb{E}s_n(z) - \frac{1}{2}\frac{1}{1-z-y_n-y_nz\mathbb{E}s_n(z)},$$

co daje

$$\mathbb{E}s_n(z) = \frac{1-z-y_n}{y_nz} + \frac{1}{y_n+z-1+y_nz\mathbb{E}s_n(z)}.$$

Z powyższej równości wynika, że

$$(y_n+z-1+y_nz\mathbb{E}s_n(z))^2 = y_nz.$$

Korzystając z Lematu 12, otrzymujemy stąd równość

$$y_n + z - 1 + y_n z \mathbb{E} s_n(z) = \sqrt{y_n z}. \tag{19}$$

Skorzystamy teraz z następującego lematu (jego dowód, wynikający prosto z Faktu 6 pozostawiamy jako ćwiczenie).

Lemat 13. *Rozważmy dowolną macierz A wymiarów $N \times n$, gdzie $N/n = y$. Niech s oznacza transformatę Stieltjesa miary spektralnej AA^T, zaś $\tilde{s}$ transformatę Stieltjesa miary spektralnej $A^T A$. Wówczas*

$$s(z) = y^{-1}\tilde{s}(z) - \frac{1 - y^{-1}}{z}.$$

Oznaczmy zatem przez $\tilde{s}_n$ transformatę Stieltjesa miary spektralnej macierzy $n^{-1}M_n^T M_n$. Z powyższego lematu dostajemy

$$z\mathbb{E}\tilde{s}_n(z) = y_n - 1 + y_n z \mathbb{E} s_n(z),$$

co w połączeniu z (19) daje

$$z + z\mathbb{E}\tilde{s}_n(z) = \sqrt{y_n z},$$

czyli $1 + \mathbb{E}\tilde{s}_n(z) = \sqrt{y_n}/\sqrt{z}$. To jest jednak niemożliwe, gdyż z Lematu 11, lewa strona ma dodatnią część urojoną, podczas gdy część urojona prawej strony jest ujemna.

Otrzymana sprzeczność dowodzi, że

$$\mathbb{E} s_n(z) = \frac{1}{2y_n z}\Big(1 - z - y_n + y_n z \delta_n(z) + \sqrt{(1 - z - y_n - y_n z \delta_n(z))^2 - 4y_n z}\Big)$$

dla każdego $z \in \mathbb{C}^+$.

Krok C. Aby zakończyć dowód zbieżności $\mathbb{E} s_n(z) \to s^y(z)$, wystarczy wykazać, że $\delta_n(z) \to 0$ dla $n \to \infty$. W tym celu wyrazimy $\delta_n(z)$ w terminach macierzy M_n.

Dla uproszczenia oznaczeń, opuścimy chwilowo część indeksów i będziemy pisać M zamiast M_n oraz N zamiast N_n. W tym kroku dowodu nie będziemy już rozważali zachowania $\delta_n(z)$ jako funkcji zmiennej z, dlatego opuścimy argument i będziemy pisać po prostu δ_n. Wprowadźmy również oznaczenie M^k na macierz powstałą z M przez usunięcie k-tego wiersza (dla odróżnienia od $M^{(k)}$ – macierzy powstałej z M przez usunięcie k-tego wiersza i k-tej kolumny).

Lemat 9 daje

$$s_n(z) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{n^{-1}\alpha_k^T\alpha_k - z - n^{-2}\alpha_k^T (M^k)^T (n^{-1}M^k (M^k)^T - z\mathrm{Id}_{N-1})^{-1} M^k \alpha_k},$$

gdzie α_k^T jest k-tym wierszem M.

Powróćmy do równości (18) i przypomnijmy, że

$$\delta_n = \mathbb{E} s_n(z) - \frac{1}{1 - z - y_n - y_n z \mathbb{E} s_n(z)}.$$

Zatem

$$\begin{aligned}
\delta_n &= \frac{1}{N}\mathbb{E}\Big(\sum_{k=1}^{N} \frac{1}{n^{-1}\alpha_k \alpha^T - z - n^{-2}\alpha_k (M^k)^T (n^{-1}M^k (M^k)^T - z\mathrm{Id}_{N-1})^{-1} M^k \alpha_k^T} - \frac{1}{1 - z - y_n - y_n z \mathbb{E} s_n(z)}\Big) \\
&= -\frac{1}{N}\mathbb{E}\sum_{k=1}^{N} \frac{\varepsilon_k}{(1 - z - y_n - y_n z \mathbb{E} s_n(z))(1 - z - y_n - y_n z \mathbb{E} s_n(z) + \varepsilon_k)},
\end{aligned}$$

gdzie

$$\varepsilon_k = \varepsilon_k(n) = n^{-1}\alpha_k^T \alpha_k - 1 - n^{-2}\alpha_k^T (M^k)^T (n^{-1}M^k (M^k)^T - z\mathrm{Id}_{N-1})^{-1} M^k \alpha_k + y_n + y_n z \mathbb{E} s_n(z).$$

Możemy również napisać

$$
\begin{aligned}
\delta_n =& -\frac{1}{N}\sum_{k=1}^{N}\frac{\mathbb{E}\varepsilon_k}{(1-z-y_n-y_nz\mathbb{E}s_n(z))^2}\\
&+\frac{1}{N}\mathbb{E}\frac{\varepsilon_k^2}{(1-z-y_n-y_nz\mathbb{E}s_n(z))^2(1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_k)}\\
=&\frac{\mathbb{E}\varepsilon_1}{(1-z-y_n-y_nz\mathbb{E}s_n(z))^2}+\mathbb{E}\frac{\varepsilon_1^2}{(1-z-y_n-y_nz\mathbb{E}s_n(z))^2(1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_1)}\\
=:&A_n+B_n,
\end{aligned}
$$

przy czym skorzystalismy z faktu, że zmienne $\varepsilon_k(n)$ mają przy ustalonym n ten sam rozkład (wynika to z niezależności i identyczności rozkładów zmiennych X_{ij}).

$\mathbf{A_n \to 0}$ Z Lematu 12 mianownik jest oddzielony od zera, wystarczy więc wykazać, że $\mathbb{E}\varepsilon_1(n)\to 0$ dla $n\to\infty$.

Obliczmy $\mathbb{E}\varepsilon_1$, korzystając z założeń dotyczących zmiennych losowych X_{ij}. Mamy

$$\frac{1}{n}\mathbb{E}\alpha_1^T\alpha_1 - 1 = \frac{1}{n}\sum_{j=1}^{n}\mathbb{E}X_{1j}^2 = 1-1=0.$$

Oznaczmy $C = (M^1)^T(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}M^1 = [C_{ij}]_{i,j\leqslant n}$. Zauważmy, że M^1 jest niezależne od X_{1i}, a zatem również macierz C jest niezależna od X_{1i}. Zatem, oznaczając przez $\mathbb{E}_1$ warunkową wartość oczekiwaną względem C, z twierdzenia Fubiniego, otrzymujemy

$$\mathbb{E}\alpha_k^T C\alpha_k = \mathbb{E}\sum_{i,j=1}^{n}C_{ij}\mathbb{E}_1X_{1i}X_{1j} = \mathbb{E}\sum_{i=1}^{n}C_{ii} = \mathbb{E}\mathrm{tr}\, C,$$

przy czym w drugiej równości skorzystaliśmy z niezależności zmiennych X_{1i}, X_{1j} dla $i\neq j$ oraz równości $\mathbb{E}X_{1i}=0$, $\mathbb{E}X_{1i}^2=1$.

Zauważmy dalej, że z wzoru $\mathrm{tr}\, EFG = \mathrm{tr}\, FGE$, wynika że

$$
\begin{aligned}
\frac{1}{n}\mathbb{E}\mathrm{tr}\, C &= \mathbb{E}\mathrm{tr}\,(M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}n^{-1}M^1(M^1)^T\\
&= \mathrm{tr}\,\mathrm{Id}_{N-1} + z\mathbb{E}\mathrm{tr}\,(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}\\
&= N-1+z\mathbb{E}\mathrm{tr}\,(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}\\
&= N-1+z\mathbb{E}\mathrm{tr}\,(S_n^{(1)} - z\mathrm{Id}_{N-1})^{-1}.
\end{aligned}
$$

Uwzględniając równość $y_n = N/n$, dostajemy stąd

$$\mathbb{E}\varepsilon_1 = \frac{1}{n} - zn^{-1}\mathbb{E}\Big(\mathrm{tr}\,(S_n^{(1)} - z\mathrm{Id}_{N-1}) - \mathrm{tr}\,(S_n - z\mathrm{Id}_N)\Big).$$

Różnicę śladów w powyższej równości możemy wyliczyć z Lematu 10, skąd dostajemy

$$|\mathbb{E}\varepsilon_1| \leqslant n^{-1} + |z|n^{-1}\mathbb{E}\left|\frac{1+\beta_1^T(S_n^{(1)} - z\mathrm{Id}_{N-1})^{-2}\beta_1}{\beta_1^T\beta_1 - z - \beta_1^T(S_n^{(1)} - z\mathrm{Id}_{N-1})^{-1}\beta_1}\right|,$$

gdzie β_1^T jest wektorem uzyskanym z pierwszego wiersza macierzy S_n przez usunięcie 1-tej współrzędnej.

Oznaczmy $a = \mathrm{Re}\, z$, $b = \mathrm{Im}\, z$. Niech ponadto $\sigma_1,\ldots,\sigma_{N-1}$ oznaczają wartości własne macierzy $S_n^{(1)}$ zaś $(y_1,\ldots,y_{N-1})$ współrzędne wektora β_1 w odpowiedniej bazie wektorów własnych.

Zauważmy, że

$$
\begin{aligned}
|1+\beta_1^T(S_n^{(1)}-z\mathrm{Id}_{N-1})^{-2}\beta_1| &= \Big|1+\sum_{i=1}^{N-1}\frac{y_i^2}{(\sigma_i-z)^2}\Big| \\
&\leqslant 1+\sum_{i=1}^{N-1}\frac{y_i^2}{(\sigma_i-a)^2+b^2},
\end{aligned}
$$

podczas gdy

$$
\begin{aligned}
|\beta_1^T\beta_1 - z - \beta_1^T(S_n^{(1)}-z\mathrm{Id}_{N-1})^{-1}\beta_1| &\geqslant |\mathrm{Im}\,(\beta_k^T\beta_k - z - \beta_k^T(S_n^{(k)}-z\mathrm{Id}_{N-1})^{-1}\beta_k)| \\
&= \Big|-b-\sum_{i=1}^{N-1}\mathrm{Im}\,\frac{y_i^2}{\sigma_i-a-ib}\Big| \\
&= b\Big(1+\sum_{i=1}^{N-1}\frac{y_i^2}{(\sigma_i-a)^2+b^2}\Big).
\end{aligned}
$$

Wynika stąd, że

$$|\mathbb{E}\varepsilon_1| \leqslant n^{-1}+|z|b^{-1}n^{-1},$$

skąd dostajemy, że $\mathbb{E}\varepsilon_1(n)\to 0$ i w konsekwencji $A_n\to 0$ dla $n\to\infty$.

$\mathbf{B_n}\to\mathbf{0}$ Wykażemy najpierw, że $|1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_1|$ jest oddzielone od zera. Pozwoli to na zredukowanie dowodu zbieżności $B_n\to 0$ do $\mathbb{E}|\varepsilon_1(n)|^2\to 0$ dla $n\to\infty$.

Zauważmy, że

$$\mathrm{Im}\,(1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_1) = -\mathrm{Im}\,\Big(-\frac{1}{n}\alpha_1^T\alpha_1+z+\frac{1}{n^2}\alpha_1^T(M^1)^T(n^{-1}M^1(M^1)^T-z\mathrm{Id}_{N-1})M_1\alpha_1\Big)$$

Pierwszy składnik w sumie powyżej jest rzeczywisty, ponadto, jak łatwo wykazać patrząc na rozkład spektralny macierzy $n^{-1}M^k(M^k)^T$ z faktu, że $b=\mathrm{Im}\,z>0$, wynika, że część urojona ostatniego składnika jest nieujemna. Wynika stąd, że $\mathrm{Im}\,(1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_1)<-b$, czyli $|1-z-y_n-y_nz\mathbb{E}s_n(z)+\varepsilon_1|>b$.

Pozostaje zatem wykazać, że $\mathbb{E}|\varepsilon_1|^2\to 0$. Oznaczmy przez $\mathbb{E}_1$ warunkową wartość oczekiwana pod warunkiem $\sigma(\{\alpha_i\}_{i\neq 1})$. Mamy

$$\mathbb{E}|\varepsilon_1|^2 = \mathbb{E}|\varepsilon_1-\mathbb{E}_1\varepsilon_1|^2+\mathbb{E}|\mathbb{E}_1\varepsilon_1-\mathbb{E}\varepsilon_1|^2+|\mathbb{E}\varepsilon_1|^2. \tag{20}$$

Rzeczywiście, korzystając z własności warunkowej wartości oczekiwanej, otrzymujemy, że prawa strona jest równa

$$
\begin{aligned}
&\mathbb{E}|\varepsilon_1|^2-\mathbb{E}(\varepsilon_1\mathbb{E}_1\bar{\varepsilon}_1)-\mathbb{E}(\bar{\varepsilon}_1\mathbb{E}_1\varepsilon_1)+|\mathbb{E}_1\varepsilon_1|^2+\mathbb{E}|\mathbb{E}_1\varepsilon_1|^2-\mathbb{E}(\mathbb{E}_1\varepsilon_1\mathbb{E}\bar{\varepsilon}_1)-\mathbb{E}(\mathbb{E}_1\bar{\varepsilon}_1\mathbb{E}\varepsilon_1)+|\mathbb{E}\varepsilon_1|^2+|\mathbb{E}\varepsilon_1|^2 \\
&=\mathbb{E}|\varepsilon_1|^2,
\end{aligned}
$$

gdyż

$$\mathbb{E}(\varepsilon_1\mathbb{E}_k\bar{\varepsilon}_1)=\mathbb{E}(\bar{\varepsilon}_1\mathbb{E}_k\varepsilon_1)=\mathbb{E}|\mathbb{E}_1\varepsilon_1|^2.$$

Ostatni składnik po prawej stronie (20), jak już wykazaliśmy zbiega do 0 dla $n\to\infty$. Pozostaje oszacować dwa pierwsze składniki.

Oznaczmy przez C macierz $\mathrm{Id}_n-n^{-1}(M^1)^T(n^{-1}M^1(M^1)^T-z\mathrm{Id}_{N-1})^{-1}M^1$. Zauważmy, że macierz ta zależy tylko od $\alpha_i, i\neq 1$. Ponadto $\varepsilon_1=-1-n^{-1}\alpha_1^TC\alpha_1+y_n+y_nz\mathbb{E}s_n(z)$. W połączeniu z niezależnością α_i, implikuje to, że

$$\mathbb{E}_1\varepsilon_k=-1-n^{-1}\sum_{ij}C_{ij}\mathbb{E}X_{1i}X_{1j}+y_n+y_nz\mathbb{E}s_n(z)=-n^{-1}\sum_i C_{ii}+y_n+y_nz\mathbb{E}s_n(z)-1.$$

Zatem

$$\varepsilon_1 - \mathbb{E}_1\varepsilon_1 = \frac{1}{n}\Big(\sum_{i=1}^{n} C_{ii}(|X_{1i}|^2 - 1) + \sum_{i\neq j} C_{ij}X_{1i}X_{1j}\Big).$$

Stąd, używając niezależności X_{1i} oraz równości $\mathbb{E}X_{1i}^2 = 1$, łatwo wykazać, że

$$\begin{aligned}\mathbb{E}_1|\varepsilon_1 - \mathbb{E}_1\varepsilon_1|^2 &= \frac{1}{n^2}\Big(\sum_{i=1}^{n} |C_{ii}|^2(\mathbb{E}X_{1i}^4 - 1) + 2\sum_{i\neq j}|C_{ij}|^2\mathbb{E}X_{1i}^2\mathbb{E}X_{1j}^2\Big)\\ &\leqslant \frac{1}{n^2}\Big(\sum_{i=1}^{n} |C_{ii}|^2 K + 2\sum_{i\neq j}|C_{ij}|^2\mathbb{E}X_{1i}^2\mathbb{E}X_{1j}^2\Big)\\ &\leqslant \frac{\tilde{K}\sum_{i,j}|C_{ii}|^2}{n^2} = \frac{\tilde{K}\operatorname{tr} CC^*}{n^2},\end{aligned}$$

gdzie $K, \tilde{K}$ są pewnymi stałymi (korzystamy z założenia, że zmienne X_{ij} są ograniczone).

Można wykazać (pozostawiamy to jako ćwiczenie), że $\operatorname{tr} CC^*$ jest punktowo ograniczone przez $(n+N)K(z)$, gdzie $K(z)$ jest (nietrudną do wyliczenia) stałą zależną tylko od z. Pokazuje to, że $\mathbb{E}|\varepsilon_1 - \mathbb{E}_1\varepsilon_1|^2 \to \infty$ dla $n \to \infty$.

Pozostaje nam wykazać, że $\mathbb{E}|\mathbb{E}_1\varepsilon_1 - \mathbb{E}\varepsilon_1|^2 \to 0$ dla $n \to \infty$.

Mamy

$$\begin{aligned}\mathbb{E}_1\varepsilon_1 - \mathbb{E}\varepsilon_1 &= \frac{1}{n}(\operatorname{tr} C - \mathbb{E}\operatorname{tr} C)\\ &= \frac{1}{n}(\operatorname{tr} D - \mathbb{E}\operatorname{tr} D),\end{aligned}$$

gdzie $D = n^{-1}(M^1)^T(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}M^1$. Korzystając ponownie z własności śladu, łatwo wykazać, że

$$\operatorname{tr} D = N - 1 + z\operatorname{tr}(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1},$$

zatem

$$|\mathbb{E}_1\varepsilon_1 - \mathbb{E}\varepsilon_1| \simeq \frac{|z|y}{N-1}|\operatorname{tr}(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1} - \mathbb{E}\operatorname{tr}(n^{-1}M^1(M^1)^T - z\mathrm{Id}_{N-1})^{-1}|.$$

Macierz $n^{-1}M^1(M^1)^T$ jest macierzą otrzymaną z próbki i.i.d. w sposób analogiczny do S_n, posiada jednak $N-1$, a nie N wierszy. Wykazując, że $s_n(z) - \mathbb{E}s_n(z)$ zbiega do zera p.n., pokazaliśmy, że macierz S_n spełnia oszacowanie

$$\mathbb{P}((N-1)^{-1}|\operatorname{tr}(S_n - z\mathrm{Id}_N)^{-1} - \mathbb{E}\operatorname{tr}(S_n - z\mathrm{Id}_N)^{-1}| \geqslant t) \leqslant 2\exp(-c\varepsilon^2N^2b^2/n),$$

gdzie $B = \operatorname{Im} z$. Stosując teraz to oszacowanie do macierzy $n^{-1}M^1(M^1)^T$ (z $N-1$ zamiast N) oraz całkując przez części, z łatwością można wykazać, że

$$\mathbb{E}|\operatorname{tr}(n^{-1}M^1(M^1)^T - z\mathrm{Id}_N)^{-1} - \mathbb{E}\operatorname{tr}(n^{-1}M^1(M^1)^T - z\mathrm{Id}_N)^{-1}|^2 \to 0,$$

co kończy dowód zbieżności $\mathbb{E}s_n(z) \to s^y(z)$.

4.2.4 Dowód zbieżności $\mathbb{P}(\forall_{z\in C^+}, s_n(z) \to s^y(z)) = 1$.

Wiemy już, że dla ustalonego $z \in \mathbb{C}^+$, $\mathbb{P}(s_n(z) \to s^y(z)) = 1$. Zatem z prawdopodobieństwem 1, zbieżność ta zachodzi dla wszystkich $z \in \mathbb{C}^+$ o współrzędnych wymiernych. Oznaczmy zbiór tych ω, dla których jest to prawda przez Ω'. Skorzystamy teraz z klasycznego twierdzenia analizy zespolonej.

Twierdzenie 17 (Twierdzenie Montela)**.** *Niech f_n będzie ciągiem wspólnie ograniczonych funkcji analitycznych w pewnym obszarze U płaszczyzny zespolonej. Wówczas z ciągu f_n można wybrać podciąg niemal jednostajnie zbieżny w U.*

Uwaga Założenia powyższego twierdzenia można bardzo istotnie osłabić, powyższa wersja jest jednak wystarczająca dla naszych zastosowań.

Łatwo wykazać, że transformaty Stieltjesa dowolnej miary probabilistycznej są w obszarze $U_R = \{z \in \mathbb{C}^+ \colon \operatorname{Im} z > R^{-1}\}$ ograniczone przez R. Zatem, przy ustalonym $\omega \in \Omega'$ z dowolnego podciągu ciągu $s_n(z)$ można wybrać podciąg niemal jednostajnie zbieżny w U_R. Funkcja graniczna jest ciągła, więc musi być równa s^y. Wynika stąd, że dla dowolnego $z \in U_R$, $s_n(z) \to s^y(z)$. Ponieważ $\bigcup_{R>0} U_R = \mathbb{C}^+$, otrzymujemy stąd $s_n(z) \to s^y(z)$ dla dowolnego $z \in \mathbb{C}^+$. Kończy to dowód twierdzenia Marczenki-Pastura.

5 Zbieżność największej wartości własnej

Niniejszy rozdział jest kompilacją wielu źródeł. Idea zastosowania twierdzeń porównawczych do analizy zachowania największej wartości własnej w przypadku gaussowskim pochodzi od Gordona. Prosty dowód koncentracji gaussowskiej został zaczerpnięty z [P], zaś dowód lematu Slepiana-Fernique'a jest rozwinięciem szkicu dowodu przedstawionego w [Fer]. Sformułowania ogólnych twierdzeń dot. zbieżności skrajnych wartości własnych zostały zaczerpnięte z [BS]

Miara spektralna macierzy losowych jest użytecznym narzędziem, pozwalającym na badanie tzw. globalnego zachowania ich widma, nie mówi jednak nic o zachowaniu pojedynczych wartości własnych. Zarówno w twierdzeniu Wignera, jak i w twierdzeniu Marczenki-Pastura, graniczna miara spektralna ma zwarty nośnik ($[-2, 2]$ dla twierdzenia Wignera, $[a, b]$ dla twierdzenia Marczenki-Pastura). Ponieważ gęstość miar granicznych jest przy brzegach nośnika niezerowa, wynika stąd, że asymptotycznie największa wartość własna macierzy $N^{-1/2}A_N$ (odp. S_n) wynosi przynajmniej 2 (odp. b). Zauważmy jednak, że zmiana podliniowej (względem wymiaru) liczby wartości własnych ciągu macierzy nie wpływa na słabą zbieżność miar spektralnych (w definicji miary spektralnej normalizujemy przez wymiar, więc zaburzenie podliniowej liczby wartości własnych asymptotycznie nie daje żadnego wkładu do miary spektralnej). Potencjalnie więc największa wartość własna tych macierzy mogłaby rosnąć do nieskończoności.

Okazuje się, że jeżeli zmienne losowe generujące macierz są scentrowane, to przy nieznacznym wzmocnieniu założeń dotyczących całkowalności (z reguły wystarczy zakładać istnienie czwartego momentu), opisana powyżej sytuacja nie zachodzi. Dowody odpowiednich twierdzeń są kombinatoryczne i polegają podobnie jak w przypadku zbieżności miar spektralnych na szacowaniu śladów k-tych potęg rozpatrywanych macierzy losowych. Istotne utrudnienie polega na tym, że aby oszacować normę operatorową macierzy (czyli największą co do modułu wartość własną), trzeba rozpatrywać k rzędu $\log n$, gdzie n jest wymiarem macierzy (przypomnijmy, że w twierdzeniu Wignera rozpatrywaliśmy ustalone k, co pozwalało na łatwe wyeliminowanie pewnych zbiorów ścieżek).

Ponieważ opisane powyżej podejście kombinatoryczne jest dość pracochłonne, a naszym celem jest raczej sygnalizowanie interesujących faktów niż przedstawianie ich w jak największej ogólności, ograniczymy się do macierzy gaussowskich i użyjemy w dowodzie metod bardziej analitycznych. Dodatkową zaletą tego podejścia jestmożliwość uzyskania silniejszych oszacowań na prawdopodobieństwo, co ma znaczenie np. dla zastosowań macierzy losowych w geometrii wypukłej.

Ograniczymy się również do wykazania zbieżności normy operatorowej macierzy w twierdzeniu Marczenki-Pastura, odpowiednie wyniki dla macierzy Wignera oraz wspomniane powyżej ogólniejsze twierdzenia przedstawimy na koniec rozdziału bez dowodów.

Oznaczmy zatem przez $M = [g_{ij}]_{i\leqslant N, j\leqslant n}$ macierz losową o niezależnych współczynnikach o rozkładzie $\mathcal{N}(0, 1)$. Przez $\|M\| = \|M\|_{\ell_2^n \to \ell_2^N}$ będziemy oznaczali normę operatorową macierzy,

$$\|M\|_{\ell_2^n \to \ell_2^N} = \sup\{|Mx| \colon |x| \leqslant 1\},$$

gdzie $|\cdot|$ oznacza standardową normę euklidesową (zarówno w $\mathbb{R}^n$ jak i $\mathbb{R}^N$).

Twierdzenie 18. *Przy powyższych założeniach*

$$\mathbb{E}\|M\| \leqslant \sqrt{N} + \sqrt{n}.$$

Ponadto, dla dowolnego $t > 0$,

$$\mathbb{P}(\|M\| \geqslant \sqrt{N} + \sqrt{n} + t) \leqslant 2\exp(-t^2/2).$$

Jako wniosek otrzymujemy następujące twierdzenie o zachowaniu największej wartości własnej dla macierzy S_n rozpatrywanych w poprzednim rozdziale.

Twierdzenie 19. *Niech M_n, S_n będą określone jak w twierdzeniu Marczenki-Pastura, przy dodatkowym założeniu, że współczynniki macierzy M_n są standardowymi zmiennymi gaussowskimi. Oznaczmy przez λ_n^{max} największą wartość własną macierzy S_n. Wówczas, z prawdopodobieństwem 1,*

$$\lambda_n^{max} \to (1 + \sqrt{y})^2$$

dla $n \to \infty$.

Dowód powyższych twierdzeń odłożymy na później, najpierw wprowadzimy niezbędne narzędzia związane ze zmiennymi gaussowskimi.

5.1 Koncentracja gaussowska i lemat Slepiana-Fernique'a

Oznaczmy przez γ_n standardowy rozkład gaussowski w $\mathbb{R}^n$, zaś przez $G = (g_1, \ldots, g_n)$ wektor losowy o rozkładzie γ_n (czyli g_i są niezależne, o wspólnym rozkładzie $\mathcal{N}(0,1)$, ponadto rozkład G jest rotacyjnie niezmienniczy). Jedną z najbardziej uderzających cech rozkładu γ_n jest tzw. zjawisko koncentracji miary. Może być ono scharakteryzowane na wiele sposobów, w języku zbiorów, nierówności różniczkowych lub nierówności wykładniczych dla funkcji lipschitzowskich. My przedstawimy koncentrację miary właśnie w tym ostatnim sformułowaniu, gdyż to właśnie ono jest najczęściej stosowane w wysokowymiarowym rachunku prawdopodobieństwa.

Twierdzenie 20 (Koncentracja gaussowska)**.** *Niech $f\colon \mathbb{R}^n \to \mathbb{R}$ będzie funkcją 1-lipschitzowską. Wówczas, dla dowolnego $t \geqslant 0$,*

$$\mathbb{P}(|f(G) - \mathbb{E}f(G)| \geqslant t) \leqslant 2\exp(-t^2/2).$$

Uwaga Powyższa nierówność ma bardzo wiele zastosowań, w teorii macierzy losowych, w rachunku prawdopodobieństwa w przestrzeniach Banacha, fizyce statystycznej, a także geometrii wypukłej i kombinatoryce. Jej użyteczność wynika z faktu, że prawa strona jest niezależna od wymiaru n.

Twierdzenie 20 ma bardzo wiele dowodów, zarówno analitycznych, jak i czysto probabilistycznych lub geometrycznych. Jeden z nich polega np. na wykazaniu odpowiedniej nierówności dla miary jednostkowej na sferze S^{N-1} dla dowolnego N a następnie zrzutowaniu tej miary na pierwsze n-współrzędnych i przejściu z N do nieskończoności. Jest to jednak metoda dość złożona, a odpowiednia nierówność na sferze wynika z rozwiązania zagadnienia izoperymetrycznego, o nietrywialnym dowodzie. Dowód, który przedstawimy (pochodzący od G. Pisier) nie daje optymalnej stałej 2 w wykładniku, jest jednak elementarny oraz wykorzystuje podstawowe własności probabilistyczne charakteryzujące miarę gaussowską. Ponieważ optymalna stała nie będzie specjalnie istotna w naszych zastosowaniach ten uproszczony dowód będzie wystarczający dla naszych potrzeb.

Dowód Twierdzenia 20. Łatwo wykazać, że bez straty ogólności możemy założyć, że f jest wszędzie różniczkowalna (ćw.). Dla $x, y \in \mathbb{R}^n$ oraz $\theta \in [0, \pi/2]$, zdefiniujmy

$$x(\theta) = x\sin\theta + y\cos\theta,\ x'(\theta) = x\cos\theta - y\sin\theta.$$

Zauważmy, że $x(\pi/2) = x$, $x(0) = y$. Z założenia o różniczkowalności f, mamy

$$f(x) - f(y) = \int_0^{\pi/2} \langle \nabla f(x(\theta)), x'(\theta)\rangle d\theta.$$

Rozważmy dowolne $\lambda > 0$.

Z nierówności Jensena,

$$\exp(\lambda(f(x)-f(y))) \leqslant \frac{2}{\pi}\int_0^{\pi/2} \exp\Big(\frac{\pi\lambda}{2}\langle \nabla f(x(\theta)), x'(\theta)\rangle\Big)d\theta$$

Zatem, z twierdzenia Fubiniego,

$$\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp(\lambda(f(x)-f(y)))d\gamma_n(x)d\gamma_n(y)$$
$$\leqslant \frac{2}{\pi}\int_0^{\pi/2}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp\Big(\frac{\pi\lambda}{2}\langle Df(x(\theta)), x'(\theta)\rangle\Big)d\gamma_n(x)d\gamma_n(y)d\theta$$

Zauważmy, że dla dowolnego θ, przekształcenie $(x,y)\mapsto(x(\theta),x'(\theta))$ jest liniową izometrią, a więc zachowuje miarę $\gamma_{2n}=\gamma_n\otimes\gamma_n$. Wynika stąd, że powyższa nierówność może być zapisana jako

$$\mathbb{E}\exp(\lambda(f(G_1)-f(G_2))) \leqslant \frac{2}{\pi}\int_0^{\pi/2}\mathbb{E}\exp\Big(\frac{\pi\lambda}{2}\langle\nabla f(G_1), G_2\rangle\Big)d\theta, \tag{21}$$

gdzie G_1, G_2 są niezależnymi standardowymi wektorami gaussowskimi w $\mathbb{R}^n$. Zauważmy, że warunkując względem G_1, możemy wyrazić prawą stronę powyższej nierówności jako

$$\mathbb{E}\exp\Big(\frac{\pi^2\lambda^2}{8}|\nabla f(G_1)|^2\Big),$$

gdyż dla dowolnego $x\in\mathbb{R}^n$ mamy $\mathbb{E}\exp(\langle x, G_2\rangle)=\exp(|x|^2/2)$ (wzór na transformatę Laplace'a zmiennych gaussowskich). To wyrażenie może być z kolei oszacowane przez $\exp(\frac{\pi^2\lambda^2}{8})$, gdyż z warunku Lipschitza, $|\nabla f|\leqslant 1$. Zauważmy teraz, że z nierówności Jensena, zastosowanej warunkowo względem G_1, lewa strona (21) szacuje się z dołu przez $\mathbb{E}\exp(\lambda(f(G_1)-\mathbb{E}f(G_1)))$. Wykazaliśmy zatem, że

$$\mathbb{E}\exp(\lambda(f(G)-\mathbb{E}f(G))) \leqslant \exp(\frac{\pi^2\lambda^2}{8})$$

dla $\lambda>0$, skąd żądana nierówność koncentracyjna (aczkolwiek, jak już wspomnieliśmy, nie z optymalną stałą 2) wynika w standardowy sposób przez zoptymalizowanie wykładniczej nierówności Czebyszewa. □

Twierdzenie 21 (Lemat Slepiana-Fernique'a)**.** *Niech T będzie zbiorem przeliczalnym, zaś $(X_t)_{t\in T}$, $(Y_t)_{t\in T}$ dwoma procesami gaussowskimi na T, takimi że $\mathbb{E}X_t=\mathbb{E}Y_t=0$ dla dowolnego $t\in T$. Załóżmy, że dla dowolnych $s,t\in T$,*

$$\|Y_t-Y_s\|_2 \leqslant \|X_t-X_s\|_2. \tag{22}$$

Wówczas dla dowolnej funkcji wypukłej, rosnącej $f\colon[0,\infty)\to\mathbb{R}$, mamy

$$\mathbb{E}f(\sup_{s,t\in T}(Y_t-Y_s)) \leqslant \mathbb{E}f(\sup_{s,t\in T}(X_t-X_s)). \tag{23}$$

Ponadto

$$\mathbb{E}\sup_{t\in T}Y_t \leqslant \mathbb{E}\sup_{t\in T}X_t. \tag{24}$$

Dowód. Przez przejście graniczne wykorzystujące twierdzenia Lebesgue'a o zbieżności monotonicznej możemy załozyć, że $T=\{1,\ldots,n\}$. Nasze procesy X_t, Y_t będziemy zatem traktować jako scentrowane wektory gaussowskie w $\mathbb{R}^n$: $X=(X_1,\ldots,X_n)$, $Y=(Y_1,\ldots,Y_n)$. Dodatkowo założymy, że są one zdefiniowane na wspólnej przestrzeni probabilistycznej oraz niezależne. Możemy ponadto założyć, że f jest dwukrotnie różniczkowalna i rośnie co najwyżej liniowo (gdyż dowolną funkcję wypukłą rosnącą możemy wyrazić jako supremum funkcji afinicznych rosnących, zaś suprema skończenie wielu funkcji afinicznych rosnących łatwo przybliżyć funkcjami o żądanych własnościach).

Nasze podejście będzie podobne jak w dowodzie nierówności koncentracyjnej. Wprowadźmy parametr $t \in [0,1]$ (nie jest to najlepsze oznaczenie, ze względu na zbiór T w ogólnym sformułowaniu twierdzenia, ponieważ jednak przeszliśmy do przypadku $T = \{1, \ldots, n\}$ nie będzie ono prowadziło do konfliktu oznaczeń) i zdefiniujmy

$$X(t) = (1-t)^{1/2}X + t^{1/2}Y.$$

Zauważmy, że $X(0) = X$, $X(1) = Y$.

Łatwo wykazać, że możemy dodatkowo założyć, że X, Y mają rozkład absolutnie ciągły względem miary Lebesgue'a na $\mathbb{R}^n$ (zostawiamy to jako ćwiczenie). Również $X(t)$ ma więc rozkład absolutnie ciągły.

Zdefiniujmy teraz

$$h(t) = \mathbb{E}f(\max_{i,j\leqslant n}(X_i(t) - X_j(t))) = \int_{\mathbb{R}^n} f(\max_{i,j\leqslant n}(x_i - x_j))g_t(x)dx,$$

gdzie g_t jest gęstością wektora losowego X_t. Aby udowodnić (23) wystarczy wykazać, że h jest funkcją nierosnącą.

Oznaczmy przez $C(t) = [C_{ij}(t)]_{i,j\leqslant n}$ macierz kowariancji zmiennej $X(t)$. Zauważmy, że

$$C_{ij}(t) = \mathbb{E}((1-t)^{1/2}X_i + t^{1/2}Y_i)((1-t)^{1/2}X_j + t^{1/2}Y_j) = (1-t)\mathbb{E}X_iX_j + t\mathbb{E}Y_iY_j,$$

zatem

$$\frac{d}{dt}C_{ij}(t) = \mathbb{E}Y_iY_j - \mathbb{E}X_iX_j.$$

Udowodnimy najpierw, że dla dowolnego $y \in \mathbb{R}^n$ i $t \in [0,1]$,

$$\frac{d}{dt}g_t(x) = \sum_{i,j}(\mathbb{E}Y_iY_j - \mathbb{E}X_iX_j)\frac{\partial^2}{\partial y_i \partial y_j}g_t(x). \tag{25}$$

Skorzystamy w tym celu ze wzoru

$$g_t(x) = \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \exp(i\langle \xi, x\rangle - \frac{1}{2}\langle \xi, C_t\xi\rangle^2)d\xi \tag{26}$$

Jest to szczególny przypadek wzoru na odwrotną transformatę Fouriera. Łatwo go też wykazać bezpośrednio poprzez zamianę zmiennych z faktu, że standardowa gęstość gaussowska jest wektorem własnym transformaty Fouriera). Zatem, różniczkując powyższe równanie na gęstość względem t i przechodząc z różniczkowaniem pod znak całki po prawej stronie (sprawdzenie poprawności tej operacji zostawiamy jako ćwiczenie), dostajemy

$$\frac{d}{dt}g_t(x) = \sum_{i,j}(\mathbb{E}X_iX_j - \mathbb{E}Y_iY_j)\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \xi_i\xi_j \exp(i\langle \xi, x\rangle - \frac{1}{2}\langle \xi, C_t\xi\rangle^2)d\xi.$$

Z drugiej strony, różniczkując obie strony równości (26) względem x_i a następnie względem x_j otrzymujemy

$$\begin{aligned}\frac{\partial^2}{\partial x_i \partial x_j}g_t(x) &= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \frac{\partial^2}{\partial x_i \partial x_j}\exp(i\langle \xi, x\rangle - \frac{1}{2}\langle \xi, C_t\xi\rangle^2)d\xi \\ &= -\frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \xi_i\xi_j \exp(i\langle \xi, x\rangle - \frac{1}{2}\langle \xi, C_t\xi\rangle^2)d\xi,\end{aligned}$$

co razem z poprzednią równością dowodzi tożsamości (25).

Korzystając z założenia, że funkcja f rośnie co najwyżej liniowo, można wykazać, że $h(t)$ jest różniczkowalna oraz

$$\frac{d}{dt}h(t) = \int_{\mathbb{R}^n} f(\max_{i,j\leqslant n}(x_i - x_j))\frac{d}{dt}g_t(x)dx.$$

Stąd

$$\frac{d}{dt}h(t) = \sum_{r,s}(\mathbb{E}Y_rY_s - \mathbb{E}X_rX_s)\int_{\mathbb{R}^n} f(\max_{i,j\leqslant n}(x_i - x_j))\frac{\partial^2}{\partial x_r \partial x_s}g_t(x). \tag{27}$$

Przekształcimy teraz składniki powyższej sumy, całkując przez części. Funkcja podcałkowa jest dwukrotnie różniczkowalna we wszystkich punktach x poza tymi, dla których $\max_i x_i$ lub $\min_i x_i$ jest przyjmowane przez przynajmniej dwie współrzędne x_i, x_j. Ponadto, w obszarze gdzie x_r nie jest ani największą, ani najmniejszą współrzędną x, pochodna cząstkowa $\partial f(\max_{ij}(x_i - x_j))/\partial x_r$ znika. Z kolei np. w obszarze, gdzie x_r jest jedyną największą współrzędną, ta pochodna jest równa $f'(x_r - \min_j x_j)$, podobny wzór zachodzi w obszarze, gdzie x_r jest jedyną najmniejszą współrzędną. Uwzględniając te obserwacje i stosując dwukrotnie wzór na całkowanie przez części, można wykazać następującą równość, której dokładny dowód pozostawiamy jako ćwiczenie (należy uzasadnić, że można całkować przez części na obszarze nieograniczonym wyznaczonym przez przecięcia skończenie wielu półprzestrzeni oraz zauważyć, że przy pierwszym całkowaniu, całki po zbiorach wymiaru $n-1$ występują dwukrotnie z przeciwnymi znakami; dla uzasadnienia przejścia z całkowaniem do zbiorów nieograniczonych należy zauważyć, że funkcja g_t oraz jej pochodne maleją wykładniczo dla $|x| \to \infty$)

$$\begin{aligned}
&\int_{\mathbb{R}^n} f(\max_{i,j\leqslant n}(x_i - x_j))\frac{\partial^2}{\partial x_r^2}g_t(x) \\
&= \int_{\{x_r = \max_i x_i\}} f''(x_r - \min_i x_i)g_t(x)dx + \int_{\{x_r = \min_i x_i\}} f''(\max_i x_i - x_r)g_t(x)dx \\
&\quad + \sum_{s\neq r}\Big(\int_{\{x_r = x_s = \max_i x_i\}}\int f'(x_r - \min_i x_i)g_t(x)d_{n-1}x + \int_{\{x_r = x_s = \min_i x_i\}}\int f'(\max_i x_i - x_r)g_t(x)d_{n-1}x\Big) \\
&= \sum_{s\neq r}\int_{\{x_r = \max_i x_i, x_s = \min_i x_i\}} f''(x_r - x_s)g_t(x)dx + \int_{\{x_r = \min_i x_i, x_s = \max_i x_i,\}} f''(x_s - x_r)g_t(x)dx \\
&\quad + \sum_{s\neq r}\Big(\int_{\{x_r = x_s = \max_i x_i\}} f'(x_r - \min_i x_i)g_t(x)d_{n-1}x + \int_{\{x_r = x_s = \min_i x_i\}} f'(\max_i x_i - x_r)g_t(x)d_{n-1}x\Big)
\end{aligned}$$

przy czym całki $\int d_{n-1}x$ rozumiemy jako całki względem $(n-1)$ wymiarowej miary Haussdorfa (przeskalowanej o $\sqrt{2}$ względem tradycyjnej normalizacji). Podobnie wykazujemy, że dla $r \neq s$,

$$\begin{aligned}
&\int_{\mathbb{R}^n} f(\max_{i,j\leqslant n}(x_i - x_j))\frac{\partial^2}{\partial x_r \partial x_s}g_t(x) \\
&= -\int_{\{x_r = \max_i x_i, x_s = \min_i x_i\}} f''(x_r - x_s)g_t(x)dx - \int_{\{x_r = \min_i x_i, x_s = \max_i x_i\}} f''(x_s - x_r)g_t(x)dx \\
&\quad - \int_{\{x_r = x_s = \max_i x_i\}}\int f'(x_r - \min_i x_i)g_t(x)d_{n-1}x - \int_{\{x_r = x_s = \min_i x_i\}}\int f'(\max_i x_i - x_r)g_t(x)d_{n-1}x.
\end{aligned}$$

Podstawiając powyższe równości do (27) oraz korzystając z faktu, że $\sum_{r\neq s}\gamma_{rs} = \frac{1}{2}\sum_{r\neq s}(\gamma_{rs} + \gamma_{sr})$, otrzymujemy

$$\begin{aligned}
\frac{d}{dt}h(t) &= \frac{1}{2}\sum_{r\neq s}(\mathbb{E}Y_r^2 - 2\mathbb{E}Y_rY_s + \mathbb{E}Y_s^2 - \mathbb{E}X_r^2 + 2\mathbb{E}X_rX_s - \mathbb{E}X_s^2)A_{rs} \\
&= \frac{1}{2}\sum_{r\neq s}(\mathbb{E}(Y_s - Y_r)^2 - \mathbb{E}(X_s - X_r)^2)A_{rs},
\end{aligned}$$

gdzie

$$A_{rs} = \int_{\{x_r = \max_i x_i, x_s = \min_i x_i\}} f''(x_r - x_s) g_t(x) dx + \int_{\{x_r = \min_i x_i, x_s = \max_i x_i,\}} f''(x_s - x_r) g_t(x) dx$$
$$+ \int_{\{x_r = x_s = \max_i x_i\}} f'(x_r - \min_i x_i) g_t(x) d_{n-1}x + \int_{\{x_r = x_s = \min_i x_i\}} f'(\max_i x_i - x_r) g_t(x) d_{n-1}x \geqslant 0,$$

gdyż z założenia wypukłości i monotoniczności f, funkcje podcałkowe sa nieujemne.

Ponieważ z założenia $\mathbb{E}(Y_r - Y_s)^2 \leqslant \mathbb{E}(X_r - X_s)^2$, otrzymujemy stąd $h'(t) \leqslant 0$, co kończy dowód nierówności (23).

Pozostaje wykazać, (24), co jest już nietrudne i sprowadza się do zabiegu formalnego. Zauważmy, że

$$\mathbb{E} \sup_{s,t\in T} (X_t - X_s) = \mathbb{E} \sup_{t\in T} X_t - \mathbb{E} \inf_{s\in T} X_s = \mathbb{E} \sup_{t\in T} X_t + \mathbb{E} \sup_{s\in T} (-X_s) = 2\mathbb{E} \sup_{t\in T} X_t,$$

gdyż rozkłady skończenie-wymiarowe procesu X_t, jako scentrowane rozkłady gaussowskie, są symetryczne, a więc proces $(-X_t)_{t\in T}$ ma taki sam rozkład jak $(X_t)_{t\in T}$. $\square$

Uwaga Lemat Slepiana-Fernique'a jest częścią ogólniejszej teorii, badającej ograniczoność i ciągłość procesów gaussowskich poprzez analizę geometrii zbioru T w (pseudo)metryce L^2 zadanej przez proces

$$d_X(s,t) = \|X_s - X_t\|_2.$$

Teoria ta, mająca początek jeszcze w pracach Kołmogorowa, została daleko rozwinięta przez Dudleya, Fernique'a i Talagranda. Okazuje się, że ograniczoność procesu jest ściśle związana z istnieniem odpowiednich ciągów pokryć zbioru T zbiorami o małych średnicach w metryce d_X.

5.2 Dowód Twierdzeń 18 i 19

Dowód Twierdzenia 18. Niech $\mathcal{M}, \mathcal{N}$ będą przeliczalnymi gęstymi podzbiorami odpowiednio S^{N-1}, S^{n-1}. Zauważmy, że

$$\|M\| = \sup_{x\in\mathcal{M}, y\in\mathcal{N}} \langle My, x\rangle = \sup_{x\in\mathcal{M}, y\in\mathcal{N}} G_{x,y},$$

gdzie

$$G_{x,y} = \sum_{i=1}^{N} \sum_{j=1}^{n} g_{ij} x_i y_j$$

jest procesem gaussowskim na $\mathcal{M} \times \mathcal{N}$.

Główna idea dowodu polega na skonstruowaniu innego procesu gaussowskiego, dla którego wartość oczekiwana supremów jest łatwa do policzenia oraz który dominuje proces G w sensie lematu Slepiana-Fernique'a.

Obliczmy

$$\begin{aligned} d_G((x,y),(x',y'))^2 = \mathbb{E}|G_{x,y} - G_{x',y'}|^2 &= \sum_{i,j} (x_i y_j - x'_i y'_j)^2 \\ &= \sum_{i,j} x_i^2 y_j^2 + \sum_{i,j} (x'_i)^2 (y'_j)^2 - 2\sum_{i,j} x_i x'_i y_j y'_j \\ &= 2 + 2\langle x, x'\rangle\langle y, y'\rangle. \end{aligned}$$

Zdefiniujmy proces gaussowski $\tilde{G} = (\tilde{G}_{x,y})_{(x,y)\in\mathcal{M}\times\mathcal{N}}$ wzorem

$$\tilde{G}_{x,y} = \sum_{i=1}^{N} g_i x_i + \sum_{j=1}^{n} g_{N+j} y_j,$$

gdzie $g_1, \ldots, g_{n+N}$ są niezależnymi standardowymi zmiennymi gaussowskimi. Wówczas

$$\begin{aligned} d_{\tilde{G}}((x,y),(x',y'))^2 = \mathbb{E}|\tilde{G}_{x,y} - \tilde{G}_{x',y'}|^2 &= \sum_{i=1}^{N}(x_i - x_i')^2 + \sum_{j=1}^{n}(y_j - y_j')^2 \\ &= 4 - 2\langle x, x'\rangle - 2\langle y, y'\rangle. \end{aligned}$$

Mamy

$$\begin{aligned} d_{\tilde{G}}((x,y),(x',y'))^2 - d_G((x,y),(x',y'))^2 &= 2 + 2\langle x, x'\rangle\langle y, y'\rangle - 2\langle x, x'\rangle - 2\langle y, y'\rangle \\ &= 2(1 - \langle x, x'\rangle)(1 - \langle y, y'\rangle) \geqslant 0, \end{aligned}$$

gdyż z nierówności Schwarza $\langle x, x'\rangle, \langle y, y'\rangle \leqslant 1$.

Zatem, z lematu Slepiana-Fernique'a,

$$\mathbb{E} \sup_{x\in\mathcal{M}, y\in\mathcal{N}} G_{x,y} \leqslant \mathbb{E} \sup_{x\in\mathcal{M}, y\in\mathcal{N}} \tilde{G}_{x,y}.$$

Zauważmy jednak, że dla dowolnych $x \in \mathcal{M}, y \in \mathcal{N}$,

$$\tilde{G}_{x,y} \leqslant \sqrt{\sum_{i=1}^{N} g_i^2} + \sqrt{\sum_{j=1}^{n} g_{N+j}^2}.$$

Zatem, z nierówności Jensena,

$$\mathbb{E} \sup_{x\in\mathcal{M}, y\in\mathcal{N}} \tilde{G}_{x,y} \leqslant \sqrt{N} + \sqrt{n},$$

co kończy dowód oszacowania na $\mathbb{E}\|M\|$. Aby wykazać oszacowanie na prawdopodobieństwo, wystarczy zauważyć, że norma operatorowa jest mniejsza niż norma Hilberta-Schmidta macierzy, która po utożsamieniu zbioru macierzy N na n z $\mathbb{R}^{Nn}$ odpowiada standardowej normie euklidesowej. Norma operatorowa, jako funkcja na R^{Nn} jest zatem 1-lipschitzowska, co pozwala zastosować gaussowską nierówność koncentracyjną. □

Dowód Twierdzenia 19. Zauważmy, że

$$\begin{aligned} \lambda_n^{max} &= \sup_{x\in S^{N-1}} \langle S_n x, x\rangle = \frac{1}{n} \sup_{x\in S^{N-1}} \langle M_n M_n^T x, x\rangle \\ &= \frac{1}{n} \sup_{x\in S^{N-1}} |M_n^T x|^2 = \frac{1}{n}\|M_n^T\|^2 = \frac{1}{n}\|M_n\|^2. \end{aligned}$$

Aby wykazać, że λ_n^{max} zbiega p.n. do $(1+\sqrt{y})^2$, wystarczy zatem udowodnić, że $n^{-1/2}\|M_n\|$ zbiega p.n. do $1+\sqrt{y}$. Jak już wspomnieliśmy, oszacowanie z dołu wynika łatwo z Twierdzenia Marczenki-Pastura. Pozostaje zatem wykazać, że

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\|M_n\| \leqslant 1 + \sqrt{y} \text{ p.n.}$$

Z Twierdzenia 18, dla dowolnego $\varepsilon > 0$ mamy

$$\mathbb{P}(n^{-1/2}\|M_n\| \geqslant 1 + \sqrt{N_n/n} + \varepsilon) \leqslant 2\exp(-\varepsilon^2 n/2),$$

skąd poprzez lemat Borela-Cantelli dostajemy, że z prawdopodobieństwem 1,

$$\limsup_{n\to\infty} \frac{1}{\sqrt{n}}\|M_n\| \leqslant \limsup_{n\to\infty}(1 + \sqrt{N_n/n}) = 1 + \sqrt{y},$$

co kończy dowód twierdzenia. □

5.3 Dalsze uwagi

Powyżej udowodniliśmy zbieżność operatorową dla macierzy kowariancji generowanych przez standardowe zmienne gaussowskie. Jak już wspomnieliśmy, fakt ten zachodzi dla dużo szerszej klasy macierzy. Twierdzenia tego typu przy różnych założeniach były udowadniane przez wielu matematyków, np. Gemana, Baia i Yina, Silversteina. Poniżej przedstawiamy optymalne wersje tych twierdzeń, pochodzące z książki [BS].

Twierdzenie 22. *Załóżmy, że $(X_{ij})_{i,j\in\mathbb{N}}$ będą niezależnymi zmiennymi losowymi o tym samym rozkładzie, takimi że $\mathbb{E}X_{ij} = 0$, $\mathbb{E}|X_{ij}|^2 = 1$ oraz $\mathbb{E}|X_{ij}|^4 < \infty$. Niech N_n będzie ciągiem liczb naturalnych, takim że $N_n/n \to y \in (0,\infty)$ dla $n \to \infty$. Zdefiniujmy macierze $M_n = [X_{ij}]_{i\leqslant N_n, j\leqslant n}$ i $S_n = n^{-1}M_nM_n^T$ oraz oznaczmy przez λ_n^{max}, największą wartość własną macierzy S_n. Wówczas, z prawdopodobieństwem 1,*

$$\lambda_n^{max} \to (1+\sqrt{y})^2$$

dla $n \to \infty$. Jeśli ponadto $y \in (0,1)$ oraz λ_n^{min} oznacza najmniejszą wartość własną S_n, to z prawdopodobieństwem 1,

$$\lambda_n^{min} \to (1-\sqrt{y})^2.$$

Jeśli zmienne X_{ij} nie posiadają skończonego czwartego momentu, wówczas z prawdopodobieństwem 1, $\limsup_{n\to\infty} \lambda_1^n = \infty$.

Powyższe twierdzenie mówi zatem również o zbieżności najmniejszej wartości własnej macierzy S_n. Nie wdając się w szczegóły wspomnijmy przy tej okazji, że w przypadku gaussowskim zbieżność najmniejszej wartości własnej może być udowodniona poprzez wzmocnienie lematu Slepiana-Fernique'a, pochodzące od Gordona i dotyczące zmiennych losowych postaci $\inf_s \sup_t G_{st}$ gdzie G_{st} jest procesem gaussowskim indeksowanym dwoma parametrami.

Na zakończenie podajmy jeszcze twierdzenie opisujące zachowanie największej wartości własnych macierzy Wignera.

Twierdzenie 23. *Niech $(X_{ij})_{1\leqslant i\leqslant j<\infty}$ będą niezależnymi zmiennymi losowymi, takimi że $X_{ii} \in \mathbb{R}$ i mają wspólny rozkład oraz X_{ij} dla $i < j$ mają wspólny rozkład. Niech $N^{-1/2}A_N$ będzie ciągiem macierzy Wignera związanych z tablicą (X_{ij}). Wówczas najmniejsza i największa wartość własna macierzy $N^{-1/2}A_N$ zbiegają p.n. odpowiednio do c_1, c_2, wtedy i tylko wtedy, gdy*

(i) $\mathbb{E}X_{11}^2 < \infty$,

(ii) $\mathbb{E}X_{12} = 0$,

(iii) $\mathbb{E}|X_{12}|^2 = \sigma^2$,

(iv) $\mathbb{E}|X_{12}|^4 < \infty$,

(v) $c_1 = -2\sigma$, $c_2 = 2\sigma$.

6 Macierze GUE – lokalne zachowanie wartości własnych

Materiał zawarty w niniejszym rozdziale został w całości zaczerpnięty z [AGZ]. Zmieniona została nieco notacja i układ lematów, zaś większość faktów pomocniczych przedstawiona została w znacznie mniejszej ogólności, jedynie w takim zakresie, w jakim są one niezbędne dla interesujących nas zastosowań, jednak ogólna struktura dowodów dość wiernie odzwierciedla argumentację przedstawioną w książce.

6.1 Wprowadzenie

W niniejszym rozdziale opiszemy kilka podstawowych wyników dotyczących zachowania wartości własnych w małej skali, na przykładzie macierzy GUE. Do tej pory raczej unikaliśmy podawania szczegółów dowodów w przypadku zespolonym, koncentrując się na prostszych przypadkach rzeczywistych i sygnalizując jedynie modyfikacje poszczególnych twierdzeń do przypadku zespolonego. Tym razem okazuje się, że analiza macierzy GUE jest mniej skomplikowana niż odpowiadających im rzeczywistych macierzy GOE, dlatego też skoncentrujemy się na przypadku zespolonym.

Przypomnijmy zatem, że macierz GUE jest postaci

$$H_N = \begin{bmatrix} X_{11} & \frac{X_{12}+iY_{12}}{\sqrt{2}} & \dots & \frac{X_{1N}+iY_{1N}}{\sqrt{2}} \\ \frac{X_{12}-iY_{12}}{\sqrt{2}} & \ddots & \dots & \frac{X_{2N}+iY_{2N}}{\sqrt{2}} \\ \vdots & \dots & \ddots & \vdots \\ \frac{X_{1N}-iY_{1N}}{\sqrt{2}} & \frac{X_{2N}-iY_{2N}}{\sqrt{2}} & \dots & X_{NN} \end{bmatrix}, \tag{28}$$

gdzie $X_{ij}, Y_{i,j}$ są niezależnymi zmiennymi $\mathcal{N}(0,1)$, zaś jej wektor uporządkowanych rosnąco wartości własnych $\lambda = (\lambda_1, \dots, \lambda_N)$ ma gęstość

$$D_N \prod_{i<j} |\lambda_i - \lambda_j|^2 \prod_{i=1}^{n} \exp(-\lambda_i^2/2) \mathbf{1}_{\{\lambda_1 < \lambda_2 < \dots < \lambda_N\}},$$

gdzie D_N jest pewną stałą.

Dla naszych zastosowań wygodniej będzie zapomnieć o uporządkowaniu wartości własnych i rozpatrywać gęstość

$$g(\lambda_1, \dots, \lambda_N) = C_N \prod_{i<j} |\lambda_i - \lambda_j|^2 \exp(-\sum_i \lambda_i^2/2),$$

gdzie $C_N = D_N/N!$ (uwaga: stała C_N nie ma nic wspólnego z tak samo oznaczoną stałą występującą w twierdzeniu o gęstości wartości własnych macierzy GOE). O funkcji g można myśleć jako o gęstości wektora losowego λ otrzymanego poprzez wybór losowej permutacji wartości własnych macierzy H_n.

Uwaga W dalszej części tego rozdziału przez $\lambda_1^N, \dots, \lambda_N^N$ będziemy oznaczać wartości własne macierzy H (a nie $N^{-1/2}H$, jak to miało miejsce w dowodzie twierdzenia Wignera). Czasami dla uproszczenia notacji będziemy opuszczać górny indeks.

Z twierdzenia 23 wynika, że dla dużych N, widmo macierzy H_N skupione jest pomiędzy $-(2+\varepsilon)\sqrt{N}$, a $(2+\varepsilon)\sqrt{N}$. Ponieważ wartości własnych jest łącznie N oraz są one parami różne, możemy się spodziewać, że odległości między sąsiednimi wartościami własnymi są rzędu $N^{-1/2}$. Oczekiwanie to może wzmacniać jeszcze inna heurystyka. Jak wiemy, wartości własne są rozłożone w przybliżeniu względem miary Wignera. Dla ustalonego $x \in (-2,2)$ oraz przedziału I zawierającego x, asymptotycznie mamy $\sigma(I)N$ wartości własnych należących do $\sqrt{N}I$. Abyśmy mogli skorzystać z twierdzenia Wignera, przedział I musi być jednak niezależny od N. Jeśli jednak dla małych t_N zastosujemy formalnie przybliżenie Wignera do przedziału $I = (x - t_N, x + t_N)$, długości zależnej od N, fakt, że $\sigma(I)N \simeq N t_N \pi^{-1}\sqrt{4-x^2}$ sugeruje, że próg $\sqrt{N}t_N$ dla którego przedział $\sqrt{N}I$ nie zawiera już żadnych wartości własnych powinien być rzędu $N^{-1/2}$. Okazuje się, że tak jest w istocie, choć dowód tego faktu, a nawet jego precyzyjne sformułowanie, są nietrywialne. Kolejne rozdziały będą poświęcone twierdzeniom, które formalizują powyższą intuicję. Część z nich przedstawimy z kompletnymi dowodami, w innych przypadkach ograniczymy się do naszkicowania kryjących się za nimi idei.

Twierdzenie 24 (Gaudin-Mehta)**.** *Dla dowolnego zwartego zbioru $A \subseteq \mathbb{R}$,*

$$\lim_{N\to\infty} \mathbb{P}(\sqrt{N}\lambda_1^N \notin A, \dots, \sqrt{N}\lambda_N^N \notin A) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{A^k} \det[K(x_i,x_j)]_{i,j\leqslant k} dx_1 \dots dx_k,$$

gdzie

$$K(x,y) = \frac{\sin(x-y)}{\pi(x-y)},$$

gdzie przyjmujemy, że $K(x,y) = \pi^{-1}$ dla $x = y$.

Zauważmy, że ponieważ K jest funkcją ograniczoną, $\det[K(x_i, x_j)]_{i,j\leqslant k}$ jest co najwyżej rzędu $k^{k/2}$, podczas gdy $k! \simeq \sqrt{2\pi k}(k/e)^k$, co pokazuje, że szereg w powyższym twierdzeniu jest zbieżny. Ponadto jeżeli miara Lebesgue'a zbioru A zbiega do zera, wartość szeregu również zbiega do 0, skąd w szczególności wynika, że dla dowolnego $\varepsilon > 0$ możemy dobrać $\delta \simeq 0$ tak, aby dla dużych N znalezienie wartości własnej w przedziale $(-\delta N^{-1/2}, \delta N^{-1/2})$ było mniejsze od ε.

Powyższe twierdzenie dotyczy zachowania się wartości własnych macierzy H_n w małym otoczeniu zera. Można również sformułować jego odpowiednik dla małych zbiorów w pobliżu innych punktów przedziału $(-2\sqrt{N}, 2\sqrt{N})$.

Twierdzenie 25. *Dla dowolnego $c \in (-2,2)$ i dowolnego zbioru zwartego $A \subseteq \mathbb{R}$,*

$$\lim_{N\to\infty} \mathbb{P}(\sqrt{N}\lambda_1^1, \ldots, \sqrt{N}\lambda_N^N \notin (c\sqrt{N} + A)) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{A^k} \det[K_c(x_i, x_j)]_{i,j\leqslant k} dx_1 \ldots dx_k,$$

gdzie

$$K_c(x,y) = \frac{\sin(2^{-1}\sqrt{4-c^2}(x-y))}{\pi(x-y)}.$$

6.2 Szkic dowodu Twierdzenia Gaudina-Mehty

Dowód Twierdzenia Gaudina-Mehty, który zaprezentujemy, oparty będzie o własności wielomianów Hermita, które sa wielomianami ortogonalnymi dla standardowego rozkładu gaussowskiego i bardzo często pojawiają się w różnych problemach związanych ze zmiennymi gaussowskimi.

Definicja 11. *Dla $n \in \mathbb{N}$ zdefiniujmy $n-$ty wielomian Hermite'a*

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2},$$

zaś n-tą funkcję falową jako

$$\psi_n(x) = \frac{e^{-x^2/4} H_n(x)}{(2\pi)^{1/4}\sqrt{n!}}$$

Wielomiany Hermite'a posiadają wiele dobrych własności, które zbierzemy w poniższym twierdzeniu. Jego dowód zostawiamy jako ćwiczenie.

Twierdzenie 26. *1. H_n jest stopnia n i ma wiodący współczynnik równy 1,*

2. $\{H_k\}_{k\leqslant n}$ jest bazą w przestrzeni wielomianów stopnia co najwyżej n,

3. $\{H_k/\sqrt{k!}\}_{k\geqslant 0}$ jest bazą ortonormalną przestrzeni $L_2(\mathbb{R}, \gamma)$, gdzie γ jest standardową miarą gaussowską, równoważnie $\{\psi_k\}_{k\geqslant 0}$ jest bazą ortonormalną w $L_2(\mathbb{R}, dx)$.

4. $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$.

5. Dla dowolnych $x \neq y$,

$$\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(n-1)!(x-y)}$$

oraz

$$\sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n}\frac{(\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y))}{x-y}.$$

6.

$$\psi_n'(x) = -\frac{x}{2}\psi_n(x) + \sqrt{n}\psi_{n-1}(x)$$

Naszym podstawowym narzędziem będzie wzór wyrażający gęstość g wektora wartości własnych poprzez pewne wyznaczniki związane z wielomianami Hermite'a. Aby go wyprowadzić, będziemy potrzebować następującego ogólnego lematu. Jego dość prosty dowód zostawiamy jako ćwiczenie.

Lemat 14. *Dla dowolnych funkcji $f_1, \ldots, f_N$ oraz $g_1, \ldots, g_N$, całkowalnych z kwadratem, zachodzi wzór*

$$\begin{aligned}
&\frac{1}{N!}\int_{\mathbb{R}^n} \det\Big[\sum_{k=1}^{N} f_k(x_i)g_k(x_j)\Big]_{i,j\leqslant N} dx_1\cdots dx_N\\
&=\frac{1}{N!}\int_{\mathbb{R}^N} \det\Big[\int_{\mathbb{R}} f_i(x_j)\Big]_{i,j\leqslant N} \det\Big[\int_{\mathbb{R}} g_i(x_j)\Big]_{i,j\leqslant N} dx_1\cdots dx_N\\
&=\det\Big[\int_{\mathbb{R}} f_i(x)g_j(x)dx\Big]_{i,j\leqslant N}.
\end{aligned}$$

Twierdzenie 27. *Dla dowolnego $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$,*

$$g(x) = \frac{1}{N!}\det[K_N(x_i, x_j)]_{i,j\leqslant N},$$

gdzie

$$K_N(x, y) = \sum_{k=0}^{N-1} \psi_k(x)\psi_k(y).$$

Dowód. Zauważmy najpierw, że z faktu, że $\{H_k\}_{k\leqslant n}$ stanowią bazę wielomianów stopnia co najwyżej n oraz że współczynnik przy najwyższej potędze w H_k wynosi 1, wynika że

$$\Delta(x_1, \ldots, x_N) = \det[x_i^{j-1}]_{i,j\leqslant N} = \det[H_{j-1}(x_i)]_{i,j\leqslant N}.$$

Rzeczywiście, startując z prawej strony powyższej równości, możemy po kolei dla $j = N, \ldots, 2$ rozbijać j-tą kolumnę macierzy $[H_{j-1}(x_i)]_{i,j\leqslant N}$ na $(x_i^{j-1})_{i=1}^N + (P(x_i))_{i=1}^N$, gdzie P jest wielomianem stopnia mniejszego niż $j-1$. Zatem wektor $(P(x_i))_{i\leqslant N}$ jest kombinacją liniową poprzednich $j-1$ kolumn, co pozwala zastąpić j-tą kolumnę przez $(x_i^{j-1})_{j\leqslant N}$.

Zatem

$$g(x_1, \ldots, x_N) = C_N|\det[H_{j-1}(x_i)]_{i,j\leqslant N}|^2 \exp(-\sum_{i=1}^{N} x_i^2/2) = \tilde{C}_N|\det[\psi_{j-1}(x_i)]_{i,j\leqslant N}|^2$$

dla pewnej stałej $\tilde{C}_N$. Zauważmy teraz, że

$$|\det[\psi_{j-1}(x_i)]_{i,j\leqslant N}|^2 = \det([\psi_{j-1}(x_i)]_{i,j\leqslant N}[\psi_{j-1}(x_i)]_{i,j\leqslant N}^T) = \det[K_N(x_i, x_j)]_{i,j\leqslant N}.$$

Zatem, aby zakończyć dowód, wystarczy wykazać, że $\tilde{C_N} = N!^{-1}$, czyli, że

$$\int_{\mathbb{R}^N} \det[K_N(x_i, x_j)]_{i,j\leqslant N} dx_1\cdots dx_N = N!$$

Z Lematu 14, powyższa całka równa jest

$$N!\det\Big[\int \psi_{i-1}(x)\psi_{j-1}(x)dx\Big]_{i,j\leqslant N} = N!\det \mathrm{Id} = N!,$$

przy czym w pierwszej równości skorzystaliśmy z faktu, że funkcje ψ_i tworzą układ ortonormalny. □

Możemy już wyprowadzić podstawowy wzór, którego analiza pozwoli na udowodnienie twierdzenia Gaudina-Mehty

Twierdzenie 28. *Dla dowolnego zbioru borelowskiego $A \subseteq \mathbb{R}$,*

$$\mathbb{P}(\lambda_1^N, \ldots, \lambda_N^N \in A) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!}\int_{(A^c)^k} \det[K_N(x_i, x_j)]_{i,j\leqslant k} dx_1\cdots dx_k.$$

Uwaga Ponieważ macierz $[K_N(x_i,y_j)]_{i,j\leqslant k}$ jest równa iloczynowi AA^T, gdzie $[\psi_{j-1}(x_i)]_{i\leqslant k, j\leqslant N}$, może być rzędu co najwyżej N, a więc dla $k > N$ jej wyznacznik znika. Wynika stąd, że suma w Twierdzeniu 28 jest tak naprawdę skończona. Twierdzenie zapisujemy jednak w tej formie, gdyż będziemy używać go do przejścia granicznego dla $N \to \infty$, zaś nieskończony szereg postaci jak wyżej jest tak zwanym wyznacznikiem Fredholma, obiektem pojawiającym się w różnych działach matematyki, którego własności są dość dobrze zbadane. Dalsza część dowodu Twierdzenia Gaudina-Mehty będzie się odwoływać właśnie do podstawowych faktów ogólnej teorii wyznaczników Fredholma.

Dowód Twierdzenia 28. Mamy

$$\begin{aligned}\mathbb{P}(\lambda_1^N,\ldots,\lambda_N^N\in A) &= \int_{A^N}\frac{1}{N!}\det[K_N(x_i,x_j)]_{i,j\leqslant N}\\ &= \det\Big[\int_A \psi_{i-1}(x)\psi_{j-1}(x)dx\Big]_{i,j\leqslant N}\\ &= \det\Big(\mathrm{Id}_N - \Big[\int_{A^c}\psi_{i-1}(x)\psi_{j-1}(x)dx\Big]_{i,j\leqslant N}\Big),\end{aligned}$$

przy czym w ostatniej równości użyliśmy faktu, że funkcje ψ_i stanowią układ ortonormalny. Zapisując j-tą kolumnę macierzy $\mathrm{Id} - [\int_{A^c}\psi_{i-1}(x)\psi_{j-1}(x)dx]_{i,j\leqslant N}$ jako $e_i - (\int_{A^c}\psi_{i-1}(x)\psi_{j-1}(x))_{i\leqslant N})$ i korzystając z wieloliniowości wyznacznika, otrzymujemy

$$\mathbb{P}(\lambda_1^N,\ldots,\lambda_N^N\in A) = 1+\sum_{k=1}^N(-1)^k\sum_{1\leqslant r_1<\ldots<r_k\leqslant N}\det\Big[\int_{A^c}\psi_{r_i-1}(x)\psi_{r_j-1}(x)dx\Big]_{i,j\leqslant k},$$

co dalej z Lematu 14 i wzoru Cauchy'ego-Binneta można zapisać jako

$$\begin{aligned}&1+\sum_{k=1}^N\frac{(-1)^k}{k!}\int_{(A^c)^k}\sum_{1\leqslant r_1<\ldots<r_k\leqslant N}(\det[\psi_{r_i-1}(x_j)]_{i,j\leqslant k})^2dx_1\cdots dx_k\\ &=1+\sum_{k=1}^N\frac{(-1)^k}{k!}\int_{(A^c)^k}\det[K_N(x_i,x_j)]_{i,j\leqslant k}dx_1\cdots dx_k\\ &=1+\sum_{k=1}^\infty\frac{(-1)^k}{k!}\int_{(A^c)^k}\det[K_N(x_i,x_j)]_{i,j\leqslant k}dx_1\cdots dx_k.\end{aligned}$$

□

Aby kontynuować, będziemy potrzebować prostego wniosku z nierówności Hadamarda, którego dowód również zostawimy jako ćwiczenie (przypomnijmy, że nierówność Hadamarda mówi, że wyznacznik macierzy jest nie większy niż iloczyn norm wszystkich jej kolumn, czyli w terminach geometrycznych, że spośród wszystkich równoległościanów o zadanych wymiarach, największą objętość ma prostopadłościan).

Lemat 15. *Dla dowolnych ograniczonych funkcji dwóch zmiennych F,G, zachodzi*

$$\Big|\det[F(x_i,x_j)]_{i,j\leqslant n} - \det[G(x_i,x_j)]_{i,j\leqslant n}\Big| \leqslant n^{1+n/2}\|F-G\|_\infty\max(\|F\|_\infty,\|G\|_\infty)^{n-1}.$$

Wprowadźmy teraz następujące oznaczenie. Dla zbioru $A\subseteq\mathbb{R}$ oraz ograniczonej funkcji K na $A\times A$, niech

$$\Delta(K,A) = 1+\sum_{k=1}^\infty\frac{(-1)^k}{k!}\int_{A^k}\det[K(x_i,x_j)]_{i,j\leqslant k}dx_1\cdots dx_k.$$

Liczbę $\Delta(K,A)$ nazwiemy wyznacznikiem Fredholma związanym z jądrem K. Nasza definicja jest jedynie częścią znacznie bogatszej teorii, która pozwala zdefiniować odpowiednik wyznacznika dla pewnej klasy operatorów na dowolnej przestrzeni Hilberta. My nie będziemy tu jednak wprowadzać tej teorii w pełnej ogólności, więc ograniczymy się do tego szczególnego przypadku.

Z poprzedniego lematu łatwo dostajemy

Lemat 16. *Niech K, L będą dwoma funkcjami ograniczonymi na $A \times A$. Wówczas*

$$|\Delta(K,A) - \Delta(L,A)| \leqslant \Big(\sum_{k=1}^{\infty} \frac{k^{1+k/2}|A|^k \max(\|K\|_\infty, \|L\|_\infty)}{k!}\Big)\|K-L\|_\infty.$$

Twierdzenie Gaudina-Mehty można zatem wyrazić jako

$$\lim_{N\to\infty} \mathbb{P}(\sqrt{N}\lambda_1^N, \ldots, \sqrt{N}\lambda_N^N \notin A) = \Delta(K,A),$$

dla dowolnego zbioru zwartego A, gdzie

$$K(x,y) = \frac{\sin(x-y)}{\pi(x-y)}.$$

Ponieważ szereg w Lemacie 16 jest zbieżny, łatwo zauważyć, że dowód Twierdzenia Gaudina-Mehty daje się sprowadzić do wykazania, że

$$\tilde{K}_N(x,y) = \frac{K_N(\frac{x}{\sqrt{N}}, \frac{y}{\sqrt{N}})}{\sqrt{N}} \to K(x,y), \text{ dla } N \to \infty,$$

niemal jednostajnie. Rzeczywiście, z lematu 28 mamy

$$\begin{aligned}
|\mathbb{P}(\sqrt{N}\lambda_1^N, \ldots, \sqrt{N}\lambda_N^N \notin A) - \Delta(K,A)| &= |\Delta(K_N, A/\sqrt{N}) - \Delta(K,A)| \\
&= |\Delta(\tilde{K}_N, A) - \Delta(K,A)| \\
&\leqslant \Big(\sum_{k=1}^{\infty} \frac{k^{1+k/2}|A|^k \max(\|\tilde{K}_N\|_\infty, \|K\|_\infty)^{k-1}}{k!}\Big)\|\tilde{K}_N - K\|_\infty,
\end{aligned}$$

gdzie przez normę $\|\cdot\|_\infty$ rozumiemy normę supremum na $A \times A$.

Jeżeli, $\tilde{K}_N$ zbiega do K niemal jednostajnie, to dla dowolnego zwartego zbioru A oraz dużych N, $\max(\|\tilde{K}_N\|_\infty, \|K\|_\infty) \leqslant 2\|K\|_\infty < \infty$, skąd wynika, że prawa strona powyższej nierówności zbiega do zera.

Pozostaje zatem wykazać zbieżność $\tilde{K}_N$ do K. Dowód będzie się opierał na następującym lemacie.

Lemat 17. *Dla dowolnego $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} \Big| n^{1/4}\psi_{n-k}\Big(\frac{x}{\sqrt{n}}\Big) - \frac{1}{\sqrt{\pi}}\cos\Big(x - \frac{\pi(n-k)}{2}\Big)\Big| = 0, \tag{29}$$

przy czym zbieżność jest niemal jednostajna.

Do dowodu powyższego lematu posłuży nam

Lemat 18.

$$\lim_{n\to\infty} \Big| \frac{\sqrt{n}}{\pi} e^{n/2} \int_{-\infty}^{\infty} x^n \exp(-nx^2/2)\mathbf{1}_{\mathbb{R}_+}(x) \cos\Big(xt - \frac{\pi(n-k)}{2}\Big)x^{-k}dx - \frac{1}{\sqrt{\pi}}\cos\Big(t - \frac{\pi(n-k)}{2}\Big)\Big| = 0, \tag{30}$$

niemal jednostajnie.

Dowód. Zauważmy najpierw, że funkcja $\cos(x - \pi(n-k)/2)$ zależy jedynie od reszty jaką n daje przy dzieleniu przez 4. Dlatego, aby udowodnić tezę lematu, wystarczy wykazać, że

$$\lim_{\substack{n\to\infty \\ n=4l+r}} \frac{\sqrt{n}}{\pi} e^{n/2} \int_{-\infty}^{\infty} x^n \exp(-nx^2/2)\mathbf{1}_{\mathbb{R}_+}(x) \cos\Big(xt - \frac{\pi r}{2}\Big)x^{-k}dx = \frac{1}{\sqrt{\pi}}\cos\Big(t - \frac{\pi r}{2}\Big)$$

dla $r = 0, 1, 2, 3$, przy czym zbieżność jest niemal jednostajna. Metoda, której użyjemy pochodzi od Laplace'a i pozwala udowodnić ogólniejsze fakty postaci

$$\sqrt{n} f(a)^{-n} \int f(x)^n g(x) dx \to \sqrt{-\frac{2\pi f(a)}{f''(a)}} g(a),$$

gdzie a jest maksimum funkcji a. Zamiast formułować ogólne twierdzenie, o wielu założeniach, przedstawimy dowód w tej konkretnej sytuacji, sygnalizując jedynie, że kluczową własnością jest fakt, że istotne w dowodzie jest, że w pewnym otoczeniu punktu a funkcja f jest ściśle wklęsła, zaś poza tym otoczeniem dość szybko maleje.

W pewnym sensie pokazujemy tutaj, że odpowiednio przenormowana funkcja $(f(x)/f(a))^n$, zachowuje się jak jedynka aproksymatywna.

U nas $f(x) = x \exp(-x^2/2) \mathbf{1}_{\mathbb{R}_+}(x)$, $a = 1$, $g(x) = g_t(x) = \cos(xt - \frac{\pi r}{2}) x^{-k}$.

Jak już wspomnieliśmy, $f(x) = xe^{-x^2}$ przyjmuje maksimum w $x = 1$. Mamy $f'(x) = (1 - x^2 \exp(-x^2)) \mathbf{1}_{\mathbb{R}^+}(x)$ oraz $f''(x) = (x^3 - 3x) \exp(-x^2) \mathbf{1}_{\mathbb{R}^+}(x)$. W szczególności $f(1) = e^{-1/2}$, $f'(1) = 0$, $f''(1) = -2e^{-1/2}$, f jest monotoniczna na $(-\infty, 1)$ i $(1, \infty)$ oraz istnieje $c > 0$ (dokładna wartość tej stałej nie ma dla nas znaczenia), takie że dla $|s - 1| \leqslant c$,

$$f(s) \leqslant f(1) - c|1 - s|^2. \tag{31}$$

Dla każdego n zapiszmy całkę w (30) jako $I_1(n)g(a) + I_2(n) + I_3(n)$, gdzie

$$\begin{aligned}
I_1(n) &= \int_{|x-1| \leqslant ck^{1/4}/n^{1/4}} f(x)^n dx \\
I_2(n) &= \int_{|x-1| \leqslant ck^{1/4}/n^{1/4}} f(x)^n (g(x) - g(1)) dx \\
I_3(n) &= \int_{|x-1| > ck^{1/4}/n^{1/4}} f(x)^n g(x) dx.
\end{aligned}$$

Zauważmy, że dla $n \geqslant k$,

$$\begin{aligned}
I_3(n) &\leqslant \sup_{x\colon |x-1| > ck^{1/4}/n^{1/4}} f(x)^{n-k} \int_{|x-1| > ck^{1/4}/n^{1/4}} f(x)^k g(x) dx \\
&\leqslant \Big(f(1) - c^2 \frac{k^{1/2}}{n^{1/2}} \Big)^{n-k} \int_{\mathbb{R}} f(x)^k g(x) dx \\
&\leqslant C f(1)^{n-k} \Big(1 - c^2 \frac{k^{1/2}}{n^{1/2} f(1)} \Big)^{n-k} \\
&\leqslant C f(1)^{n-k} \exp\Big(- c^2 \frac{k^{1/2}(n-k)}{n^{1/2} f(1)} \Big),
\end{aligned}$$

gdzie C jest pewną stałą. Skorzystaliśmy tu z (31) oraz postaci funkcji f, g (aby oszacować całkę, zauważmy, że osobliwość w zerze funkcji g jest kompensowana przez funkcję f^k).

Zatem

$$\sqrt{n} f(1)^{-n} I_3(n) \leqslant C f(1)^{-k} \sqrt{n} \exp\Big(1 - c^2 \frac{k^{1/2}(n-k)}{n^{1/2} f(1)} \Big) \to 0,$$

dla $n \to \infty$. Zbieżność ta jest oczywiście jednostajna względem t, gdyż w naszych oszacowaniach po prostu oszacowaliśmy funkcję cos przez stała (t występuje u nas jedynie jako argument cosinusa).

Przejdźmy do $I_2(n)$ i zauważmy, że funkcja $g(x)$ jest w okolicach punktu 1 funkcją lipschitzowską, przy czym dla dowolnego $T > 0$ stała Lipschitza oraz otoczenie 1 mogą zostać wybrane wspólnie dla wszystkich $|t| < T$. Zatem

$$I_2(n) \leqslant Cc \frac{k^{1/4}}{n^{1/4}} I_1(n),$$

Jeżeli zatem wykażemy, że $\sqrt{n}f(a)^{-n}I_1(n)$ jest zbieżne, z powyższej nierówności otrzymamy niemal jednostajną zbieżność $\sqrt{n}f(a)^{-n}I_2(n)$ do zera.

Pozostaje nam więc wykazać, że

$$\sqrt{n}f(a)^{-n}I_1(n) \to \sqrt{-\frac{2\pi f(a)}{f''(a)}}.$$

Jest to najbardziej subtelny fragment dowodu. Zdefiniujmy funkcję

$$h(r) = \int_0^1 (1-r)(\log f)''(1+rt)dr.$$

Zapiszemy teraz funkcję $\log f(x)$ z wzoru Taylora z resztą w postaci całkowej. Ponieważ $(\log f)''$ znika w punkcie 1, mamy

$$\log f(x) = \log f(1) + \int_1^x (x-u)(\log f)''(u)du,$$

co po zamianie zmiennych $u = 1 + (x-1)r$ daje

$$\log f(x) = \log f(1) + \int_0^1 (1-r)(\log f)''(1+(x-1)r)dr(x-1)^2 = h(x-1)(x-1)^2,$$

czyli

$$f(x) = f(1)\exp(h(x-1)(x-1)^2).$$

Kolejny raz zamieniając zmienne, tym razem w całce definiującej $I_1(n)$, dostajemy

$$\begin{aligned} f(1)^{-n}\sqrt{n}I_1(n) &= \int_{|x|\leqslant ck^{1/4}n^{1/4}} \exp(h(x/\sqrt{n})x^2)dx \\ &= \int_{\mathbb{R}} \exp(h(x/\sqrt{n})x^2)\mathbf{1}_{\{|x|\leqslant ck^{1/4}n^{1/4}\}}dx \end{aligned}$$

Mamy

$$h(0) = \frac{f''(1)}{2f(1)} < 0,$$

ponadto funkcja h jest ciągła, zatem funkcja podcałkowa powyżej jest nie większa niż $\exp(-\varepsilon x^2)$ dla pewnego $\varepsilon > 0$. Z twierdzenia Lebesgue'a o zbieżności zmajoryzowanej, dostajemy więc

$$f(1)^{-n}\sqrt{n}I_1(n) \to \int_{\mathbb{R}} \exp(-h(0)x^2)dx = \sqrt{\frac{-2\pi f(1)}{f''(1)}}.$$

Z ograniczoności cosinusa daje to jednostajną zbieżność $f(1)^{-n}\sqrt{n}g(a)I_1(n)$ do żądanej granicy, kończąc dowód lematu (Uwaga: zbieżność w lemacie jest *niemal* jednostajna, gdyż składnik odpowiadający I_2 jest zbieżny niemal jednostajnie). □

Dowód Lematu 17. Dowód będzie się opierał na triku znanym nam już z dowodu lematu Slepiana-Fernique'a. Funkcję $\exp(-x^2/2)$ wyrazimy jako transformatę Fouriera gęstości gaussowskiej. Daje to

$$H_{n-k}(x)e^{-x^2/2} = (-1)^{n-k}\frac{d^{n-k}}{dx^{n-k}}e^{-x^2/2} = \frac{1}{2\pi}\int_{\mathbb{R}} (i\xi)^{n-k}e^{-i\xi x-\xi^2/2}d\xi,$$

co po przejściu do funkcji ψ_{n-k} daje

$$\psi_{n-k}(x) = \frac{e^{x^2/4}}{(2\pi)^{3/4}\sqrt{(n-k)!}}\int_{\mathbb{R}} (i\xi)^{n-k}e^{-i\xi x-\xi^2/2}d\xi$$

i wreszcie

$$n^{1/4}\psi_{n-k}\Big(\frac{x}{\sqrt{n}}\Big) = \frac{i^{n-k}e^{x^2/4n}n^{1/4}}{(2\pi)^{3/4}\sqrt{(n-k)!}}\int_{\mathbb{R}}(i\xi)^{n-k}e^{-i\xi x/\sqrt{n}-\xi^2/2}d\xi.$$

Zamieniając zmienne w całce i korzystając ze wzoru Stirlinga, otrzymujemy

$$\begin{aligned}
n^{1/4}\psi_{n-k}\Big(\frac{x}{\sqrt{n}}\Big) &= \frac{e^{x^2/4n}n^{3/4+(n-k)/2}}{(2\pi)^{3/4}\sqrt{(n-k)!}}\int_{\mathbb{R}}(i\xi)^{n-k}e^{-i\xi x-\xi^2 n/2}d\xi \\
&= \frac{e^{x^2/4n}n^{3/4+n/2}}{(2\pi)^{3/4}\sqrt{n!}}\int_{\mathbb{R}}(i\xi)^{n-k}e^{-i\xi x-\xi^2 n/2}d\xi(1+o(1)) \\
&= \frac{e^{n/2}\sqrt{n}}{(2\pi)}\int_{\mathbb{R}}(i\xi)^{n-k}e^{-i\xi x-\xi^2 n/2}d\xi(1+o(1)) \\
&= \frac{e^{n/2}\sqrt{n}}{(2\pi)}\int_{\mathbb{R}}e^{-\xi^2 n/2}\xi^{n-k}\mathrm{Re}\, i^{n-k}e^{-i\xi x}d\xi(1+o(1)) \\
&= 2\frac{e^{n/2}\sqrt{n}}{(2\pi)}\int_0^\infty \xi^n e^{-\xi^2 n/2}\cos\Big(\xi x-\frac{\pi(n-k)}{2}\Big)\xi^{-k}d\xi(1+o(1)),
\end{aligned}$$

przy czym ostatnią równość uzyskujemy przez rozpatrzenie wszystkich możliwych reszt jakie $n-k$ może dawać przy dzieleniu przez 4 i skorzystanie z podstawowych własności funkcji trygonometrycznych (parzystość, nieparzystość, przesunięcie fazy między sinusem i cosinusem). Dowód możemy teraz zakończyć korzystając z Lematu 18.

□

Teraz możemy już wykazać zbieżność $\tilde{K}_N$ do K. Zauważmy, że wystarczy wykazać zbieżność jednostajną na $\{(x,y)\in[-M,M]^2, x\neq y\}$ dla dowolnego $M>0$. Zbieżność jednostajna na $[-M,M]^2$ wynika już z ciągłości funkcji $\tilde{K}_N$ oraz K. Z Twierdzenia 26 mamy

$$\begin{aligned}
\tilde{K}_N(x,y) =&\sqrt{N}\frac{\psi_N(x/\sqrt{N})\psi_{N-1}(y/\sqrt{N})-\psi_{N-1}(x/\sqrt{N})\psi_N(y/\sqrt{N})}{x-y} \\
=&\psi_{N-1}(y/\sqrt{N})\int_0^1\psi_N'(tx/\sqrt{N}+(1-t)y/\sqrt{N})dt \\
&-\psi_N(y/\sqrt{N})\int_0^1\psi_{N-1}'(tx/\sqrt{N}+(1-t)y/\sqrt{N})dt \\
=&\psi_{N-1}(y/\sqrt{N})\int_0^1(\sqrt{N}\psi_{N-1}(z_t)-\frac{z_t}{2}\psi_N(z_t))dt \\
&-\psi_N(y/\sqrt{N})\int_0^1\sqrt{N-1}\psi_{N-2}(z_t)-\frac{z_t}{2}\psi_{N-1}(z_t))dt,
\end{aligned}$$

gdzie $z_t=tx/\sqrt{N}+(1-t)y/\sqrt{N}$. Korzystając teraz z Lematu 17 dla $k=0,1,2$, otrzymujemy, że dla dużych N

$$\begin{aligned}
\tilde{K}_N(x,y) =&\frac{1}{\pi}\Big(\cos\Big(y-\frac{\pi(N-1)}{2}\Big)\int_0^1\cos\Big(tx+(1-t)y-\frac{\pi(N-1)}{2}\Big)dt \\
&-\cos\Big(y-\frac{\pi N}{2}\Big)\int_0^1\cos\Big(tx+(1-t)y-\frac{\pi(N-2)}{2}\Big)dt\Big)+o(1) \\
=&\frac{1}{\pi(x-y)}\Big(\cos\Big(y-\frac{\pi(N-1)}{2}\Big)\sin\Big(x-\frac{\pi(N-1)}{2}\Big)-\cos\Big(y-\frac{\pi(N-1)}{2}\Big)\sin\Big(y-\frac{\pi(N-1)}{2}\Big) \\
&-\cos\Big(y-\frac{\pi N}{2}\Big)\sin\Big(x-\frac{\pi(N-2)}{2}\Big)+\cos\Big(y-\frac{\pi N}{2}\Big)\sin\Big(y-\frac{\pi(N-2)}{2}\Big)\Big)+o(1) \\
=&\frac{\sin(x-y)}{\pi(x-y)}+o(1)
\end{aligned}$$

Uwaga Dowód Twierdzenia 25 przebiega w znacznej części tak samo jak Twierdzenia 24, różni się jedynie samą końcówką, gdyż w ogólnym przypadku konieczne jest zbadanie asymptotyki przesuniętego jądra $\tilde{K}_N$. W ogólności jest to nieco bardziej skomplikowane niż w przypadku $c = 0$.

6.3 Luka w spektrum wokół zera

Kolejnym zagadnieniem, którym się zajmiemy, będzie dokładniejsze zbadanie rozmiaru luki w spektrum macierzy GUE wokół punktu 0. Wyznaczymy graniczne zachowanie prawdopodobieństwa, że żadna z wartości własnych nie znajduje się w przedziale $(-t/(2\sqrt{N}), t/(2\sqrt{N}))$. Oczywiście, problem ten można potraktować jako szczególny przypadek twierdzenia Gaudina-Mehty, jednak skomplikowana postać wyznacznika Fredholma występująca w tym twierdzeniu nie pozwala na natychmiastowe odczytanie zależności od t. Naszym celem będzie udowodnienie następującego twierdzenia, które podaje tę zależność bardziej bezpośrednio (aczkolwiek nie w postaci jawnych wzorów, a jedynie poprzez podanie równania różniczkowego spełnionego przez funkcję graniczną).

Twierdzenie 29 (Jimbo-Miwa-Mori-Sato). *Dla dowolnego $t \geqslant 0$,*

$$\lim_{N\to\infty} \mathbb{P}(\sqrt{N}\lambda_1^N, \ldots, \sqrt{N}\lambda_N^N \notin (-t/2, t/2)) = 1 - F(t),$$

gdzie

$$F(t) = 1 - \exp\Big(\int_0^t \frac{\sigma(x)}{x} dx\Big),$$

zaś $\sigma = \sigma(t)$ jest rozwiązaniem tzw równania Painlevé V:

$$(t\sigma'')^2 + 4(t\sigma' - \sigma)(t\sigma' - \sigma + (\sigma')^2) = 0,$$

w szczególności dla $t \simeq 0$,

$$\sigma(t) = -\frac{t}{\pi} - \frac{t^2}{\pi^2} + O(t^3).$$

Jeśli zdefiniujemy zmienne losowe

$$T_N = \min_{i \leqslant n} |\lambda_i^N|,$$

powyższe twierdzenie możemy przeformułowac jako

$$\lim_{N\to\infty} \mathbb{P}(T_N \leqslant t) = F(t).$$

Gdybyśmy wiedzieli, że $F(t)$ jest dystrybuantą miary probabilistycznej, oznaczałoby to, że ciąg T_N jest słabo zbieżny. Asymptotyka funkcji σ podana w twierdzeniu pozwala łatwo wykazać, że $F(0+) = 0$. Można również wykazać, że $F(\infty) = 1$, wykracza to jednak poza ramy naszego wykładu. Oczywiście funkcja F jest monotoniczna i ciągła na $(0, \infty)$, więc wobec powyższych uwag rzeczywiście jest dystrybuantą pewnego rozkładu prawdopodobieństwa.

6.3.1 Dodatkowe własności wyznaczników Fredholma

Dowód Twierdzenia 29 będzie opierał się na dokładnej analizie prawej strony równości w twierdzeniu Gaudina-Mehty. Aby ją przeprowadzić, zdefiniujemy pewne dodatkowe pojęcia związane z ogólną teorią wyznaczników Fredholma. Podobnie jak poprzednio, wszystkie definicje i własności wprowadzimy jedynie w bardzo szczególnym przypadku, potrzebnym do naszych zastosowań.

Ustalmy zatem $t_1 < t_2$ oraz ciągłą, ograniczoną funkcję K na (t_1, t_2). Oznaczmy

$$\Delta_{t_1,t_2}(K) = \Delta(K, (t_1, t_2)).$$

Wprowadźmy jeszcze jedno oznaczenie. Dla ograniczonych funkcji $K, L\colon (t_1, t_2)^2 \to \mathbb{R}$ zdefiniujmy $K \circ L\colon : (t_1, t_2)^2 \to \mathbb{R}$, wzorem

$$(K \circ L)(x, y) = \int_{t_1}^{t_2} K(x, z)L(z, y)dz$$

Uwaga Jeżeli utożsamimy K, L z operatorami jądrowymi na przedziale (t_1, t_2), to $K \circ L$ odpowiadać będzie ich złożeniu. Uzasadnia to użycie symbolu $\circ$ oraz określenie „złożenie K i L”, którego będziemy używać dla funkcji $K \circ L$.

Zanim zdefiniujemy pojęcie rezolwenty, potrzebne nam do dowodu Twierdzenia 29, wprowadźmy jeszcze jedno oznaczenie, upraszczające notację. Dla jądra K oraz dowolnej liczby naturalnej dodatniej, oznaczmy

$$K\begin{pmatrix} x_1 \dots x_n \\ y_1 \dots y_n \end{pmatrix} = \det[K(x_i, y_j)]_{i,j \leqslant n}.$$

Ponadto zdefiniujmy funkcje

$$H_n(x,y) = \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K\begin{pmatrix} x\, x_1 \dots x_n \\ y\, x_1 \dots x_n \end{pmatrix} dx_1 \dots dx_n, n \geqslant 0$$

oraz dodatkowo oznaczmy $H_0(x,y) = K(x,y)$.

Definicja 12. *Sprzężeniem Fredholma ograniczonego, ciągłego jądra $K\colon (t_1,t_2)^2 \to \mathbb{R}$ nazwiemy funkcję*

$$H(x,y) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} H_n(x,y).$$

Ponadto, jeśli $\Delta_{t_1,t_2}(K) \neq 0$, zdefiniujmy rezolwentę K, wzorem

$$R(x,y) = \frac{H(x,y)}{\Delta_{t_1,t_2}(x,y)}.$$

Uwaga Korzystając z Lematu 15 łatwo wykazać, że powyższa definicja jest poprawna. Szczegóły zostawiamy jako ćwiczenie.

Kolejny lemat odegra kluczową rolę w dowodzie Twierdzenia 29.

Lemat 19. *Niech $H(x,y)$ będzie sprzężeniem Fredholma jądra $K\colon (t_1,t_2)^2 \to \mathbb{R}$. Wówczas*

$$K \circ H = H - \Delta_{t_1,t_2}(K) \cdot K = H \circ K.$$

Dowód. Wykażemy jedynie pierwszą równość, dowód drugiej jest analogiczny. Łatwo wykazać (korzystając z ograniczoności K), że lewa strona jest w punkcie (x,y) równa

$$\begin{aligned} &\sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{t_1}^{t_2} K(x,z) H_n(z,y) dz \\ &= \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K(x,z) K\begin{pmatrix} z\, x_1 \dots x_n \\ y\, x_1 \dots x_n \end{pmatrix} dz dx_1 \cdots dx_n, \end{aligned}$$

podczas gdy prawa strona jest równa

$$\sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} \left(K\begin{pmatrix} x\, x_1 \dots x_n \\ y\, x_1 \dots x_n \end{pmatrix} - K\begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} K(x,y) \right) dx_1 \cdots dx_n.$$

Zatem, aby udowodnić lemat, wystarczy wykazać, że

$$\begin{aligned} &\frac{1}{(n-1)!} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K(x,z) K\begin{pmatrix} z\, x_1 \dots x_{n-1} \\ y\, x_1 \dots x_{n-1} \end{pmatrix} dz dx_1 \cdots dx_{n-1} \\ &= -\frac{1}{n!} \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} \left(K\begin{pmatrix} x\, x_1 \dots x_n \\ y\, x_1 \dots x_n \end{pmatrix} - K\begin{pmatrix} x_1 \dots x_n \\ x_1 \dots x_n \end{pmatrix} K(x,y) \right) dx_1 \cdots dx_n \end{aligned}$$

lub równoważnie

$$\int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K\binom{x\, x_1 \dots x_n}{y\, x_1 \dots x_n} dx_1 \cdots dx_n = \tag{32}$$

$$\int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K\binom{x_1 \dots x_n}{x_1 \dots x_n} K(x,y) dx_1 \cdots dx_n - n \int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K(x,z) K\binom{z\, x_1 \dots x_{n-1}}{y\, x_1 \dots x_{n-1}} dz dx_1 \cdots dx_{n-1} \tag{33}$$

Stosując rozwinięcie Laplace'a względem pierwszego wiersza, otrzymujemy

$$\begin{aligned} K\binom{x\, x_1 \dots x_n}{y\, x_1 \dots x_n} &= K(x,y) K\binom{x_1 \dots x_n}{x_1 \dots x_n} + \sum_{i=1}^{n} (-1)^i K(x, x_i) K\binom{x_1\, x_2 \dots\ x_i\ \ x_{i+1} \dots x_n}{y\ \ x_1 \dots x_{i-1}\, x_{i+1} \dots x_n} \\ &= K(x,y) K\binom{x_1 \dots x_n}{x_1 \dots x_n} - \sum_{i=1}^{n} K(x, x_i) K\binom{x_i\, x_1 \dots\ x_i\ \ x_{i+1} \dots x_n}{y\ \ x_1 \dots x_{i-1}\, x_{i+1} \dots x_n}. \end{aligned}$$

przy czym w drugiej równości skorzystaliśmy z antysymetrii wyznacznika. Całkując powyższą równość dostajemy (32), gdyż z symetrii całki odpowiadające wszystkim składnikom sumy po prawej stronie są równe

$$\int_{t_1}^{t_2} \cdots \int_{t_1}^{t_2} K(x,z) K\binom{z\, x_1 \dots x_{n-1}}{y\, x_1 \dots x_{n-1}} dz dx_1 \cdots dx_{n-1}.$$

□

6.3.2 Szkic dowodu Twierdzenia 29

Poniżej przedstawimy zarys dowodu Twierdzenia 29, koncentrując się raczej na formalnej stronie dość skomplikowanych obliczeń niż na zachowaniu pełnej ścisłości. W szczególności, nie będziemy każdorazowo precyzyjnie uzasadniać poprawności przejść granicznych, wynikającej dość łatwo ze standardowych twierdzeń analizy matematycznej (np. o związku zbieżności jednostajnej pochodnych z różniczkowalnością granicy).

Powracając do oznaczeń z Twierdzenia Gaudina-Mehty, oznaczmy

$$K(x,y) = \frac{\sin(x-y)}{\pi(x-y)}.$$

Będziemy rozpatrywać wyznaczniki i rezolwenty Fredholma związane z jądrem K, i przedziałem (t_1, t_2). Zależność od parametrów t_i będziemy zapisywać w dolnym indeksie. Naszym celem jest znalezienie równania różniczkowego na

$$\sigma(t) = t \frac{d}{dt} \Delta_{-t/2, t/2}(K).$$

Niech $R_{t_1,t_2}(x,y)$ oznacza rezolwentę jądra K, zdefiniowaną w poprzednim rozdziale. Pierwszym krokiem dowodu jest obliczenie pochodnych cząstkowych funkcji $R_{t_1,t_2}(x,y)$ oraz pewnych innych funkcji z nią związanych względem parametrów t_i, przy ustalonych wartościach x, y.

Lemat 20. *Dla $i = 1, 2$ zachodzą równości,*

$$\frac{\partial \log \Delta_{t_1,t_2}}{\partial t_i} = (-1)^{i+1} R_{t_1,t_2}(t_i, t_i) \tag{34}$$

oraz

$$\frac{\partial R_{t_1,t_2}(x,y)}{\partial t_i} = (-1)^i R_{t_1,t_2}(x, t_i) R(t_i, y). \tag{35}$$

Dowód. Zauważmy, że

$$\frac{\partial}{\partial t_i}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}\det[K(x_j,x_k)]_{j,k\leqslant n}dx_1\dots dx_n$$
$$= (-1)^i\sum_{k=1}^{n}\int_{t_1}^{t_2}\cdots\int_{t_1}^{t_2}K\Big(\begin{matrix}x_1\dots x_{k-1}\ t_i\ x_{k+1}\dots x_n\\ x_1\dots x_{k-1}\ t_i\ x_{k+1}\dots x_n\end{matrix}\Big)dx_1\cdots dx_{k-1}dx_{k+1}\cdots dx_n$$
$$= (-1)^i nH_{n-1}(t_i,t_i),$$

gdzie H_{n-1} jest funkcją pojawiającą się w definicji H_{t_1,t_2} – sprzężenia Fredholma funkcji K (rozpatrywanego jako jądro całkowe na zbiorze (t_1,t_2)). H_{n-1} zależy więc także od przedziału (t_1,t_2), czego nie będziemy jednak zaznaczać w notacji.

Stąd, korzystając z Lematu 15, ograniczoności funkcji K oraz standardowych twierdzeń analizy, otrzymujemy, że

$$\frac{\partial\Delta_{t_1,t_2}}{\partial t_i} = (-1)^{i+1}H_{t_1,t_2}(t_i,t_i),$$

co łatwo implikuje równość (34).

Aby udowodnić (35) przypomnijmy, że z Lematu 19, dla dowolnych $x,y\in(t_1,t_2)$ zachodzi równość

$$R_{t_1,t_2}(x,y) = K(x,y)+\int_{t_1}^{t_2}K(x,z)R_{t_1,t_2}(z,y)dz = K(x,y)+\int_{t_1}^{t_2}R_{t_1,t_2}(x,z)K(z,y)dz. \tag{36}$$

Z definicji $R_{t_1,t_2}(x,y)$ łatwo wykazać, że funkcja ta jest różniczkowalna względem t_i (ponownie pokazujemy, że szereg pochodnych jest niemal jednostajnie zbieżny). Korzystając z powyższej równości wykażemy, że

$$\begin{aligned}\frac{\partial R_{t_1,t_2}(x,y)}{\partial t_i} &= \frac{\partial}{\partial t_i}\int_{t_1}^{t_2}K(x,z)R_{t_1,t_2}(z,y)dz\\ &= (-1)^iK(x,t_i)R_{t_1,t_2}(t_i,y)+\int_{t_1}^{t_2}K(x,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dz.\end{aligned} \tag{37}$$

Dowód powyższego wzoru przeprowadzimy tylko dla $i=2$, przypadek $i=1$ jest analogiczny. Mamy

$$\frac{1}{h}\Big(\int_{t_1}^{t_2+h}K(x,z)R_{t_1,t_2+h}(z,y)dz-\int_{t_1}^{t_2}K(x,z)R_{t_1,t_2}(z,y)dz\Big)$$
$$= \int_{t_1}^{t_2+h}K(x,z)\frac{R_{t_1,t_2+h}(z,y)-R_{t_1,t_2}(z,y)}{h}dz+\frac{1}{h}\int_{t_2}^{t_2+h}K(x,z)R_{t_1,t_2}(z,y)dz.$$

Z ciągłości $R_{t_1,t_2}(x,z)$ względem z wynika, że drugi składnik zbiega przy $h\to 0$ do $K(x,t_2)R_{t_1,t_2}(t_2,y)$. Z kolei pierwszy składnik jest dla małych h równy

$$\int_{t_1}^{t_2+1}K(x,z)\frac{R_{t_1,t_2+h}(z,y)-R_{t_1,t_2}(z,y)}{h}\mathbf{1}_{(t_1,t_2+h)}(z)dz.$$

Funkcja podcałkowa zbiega p.n. do $K(x,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_2}\mathbf{1}_{(t_1,t_2)}(z)$. Aby zakończyć dowód (37) wystarczy więc wykazać, że ilorazy różnicowe $(R_{t_1,t_2+h}(z,y)-R_{t_1,t_2}(z,y))h^{-1}$ są wspólnie ograniczone. Wykazanie tego nie jest trudne, gdyż ponownie z definicji $H_{t_1,t_2}(x,y)$ wynika, że pochodna cząstkowa $H_{t_1,t_2}(x,y)$ względem t_2 jest ograniczona dla x,y,t_1,t_2 ograniczonych.

Z równości (37) dostajemy

$$\begin{aligned}\int_{t_1}^{t_2}R_{t_1,t_2}(x,s)\frac{\partial R_{t_1,t_2}(s,y)}{\partial t_i}ds =&(-1)^iR_{t_1,t_2}(t_i,y)\int_{t_1}^{t_2}R_{t_1,t_2}(x,s)K(s,t_i)ds\\ &+\int_{t_1}^{t_2}\int_{t_1}^{t_2}R_{t_1,t_2}(x,s)K(s,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dzds\end{aligned}$$

Dodając powyższą równość do (37) otrzymujemy (35), gdyż, ponownie z równości (36), mamy

$$(-1)^i R_{t_1,t_2}(t_i,y)\int_{t_1}^{t_2} R_{t_1,t_2}(x,s)K(s,t_i)ds + (-1)^i K(x,t_i)R_{t_1,t_2}(t_i,y) = (-1)^i R_{t_1,t_2}(x,t_i)R_{t_1,t_2}(t_i,y)$$

oraz

$$\begin{aligned}
&\int_{t_1}^{t_2}\int_{t_1}^{t_2} R_{t_1,t_2}(x,s)K(s,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dzds + \int_{t_1}^{t_2} K(x,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dz \\
&= \int_{t_1}^{t_2}\Big(\int_{t_1}^{t_2} R_{t_1,t_2}(x,s)K(s,z)ds + K(x,z)\Big)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dz = \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)\frac{\partial R_{t_1,t_2}(z,y)}{\partial t_i}dz.
\end{aligned}$$

□

Oznaczmy teraz $f(x) = \pi^{-1/2}\sin x$ oraz zdefiniujmy funkcje

$$Q_{t_1,t_2}(x) = f(x) + \int_{t_1}^{t_2} R_{t_1,t_2}(x,y)f(y)dy$$

$$P_{t_1,t_2}(x) = f'(x) + \int_{t_1}^{t_2} R_{t_1,t_2}(x,y)f'(y)dy$$

Lemat 21. *Zachodzą równości*

$$R_{t_1,t_2}(x,y) = R_{t_1,t_2}(y,x) = \frac{Q_{t_1,t_2}(x)P_{t_1,t_2}(y) - Q_{t_1,t_2}(y)P_{t_1,t_2}(x)}{x-y}, \tag{38}$$

$$R_{t_1,t_2}(x,x) = \frac{\partial Q_{t_1,t_2}(x)}{\partial x}P_{t_1,t_2}(x) - \frac{\partial P_{t_1,t_2}(x)}{\partial x}Q_{t_1,t_2}(x), \tag{39}$$

$$\frac{\partial Q_{t_1,t_2}(x)}{\partial t_i} = (-1)^i R_{t_1,t_2}(x,t_i)Q_{t_1,t_2}(t_i),\ i=1,2, \tag{40}$$

$$\frac{\partial P_{t_1,t_2}(x)}{\partial t_i} = (-1)^i R_{t_1,t_2}(x,t_i)P_{t_1,t_2}(t_i),\ i=1,2. \tag{41}$$

Dowód. Równość (39) wynika z (38) poprzez twierdzenie de L'Hopitala, zaś równości (40) i (41) z Lematu 20, przykładowo

$$\begin{aligned}
\frac{\partial Q_{t_1,t_2}(x)}{\partial t_2} &= \frac{\partial}{\partial t_2}\int_{t_1}^{t_2} R_{t_1,t_2}(x,y)f(y)dy \\
&= R_{t_1,t_2}(x,t_2)f(t_2) + \int_{t_1}^{t_2}\frac{\partial R_{t_1,t_2}(x,y)}{\partial t_2}f(y)dy \\
&= R_{t_1,t_2}(x,t_2)f(t_2) + \int_{t_1}^{t_2} R_{t_1,t_2}(x,t_2)R(t_2,y)f(y)dy \\
&= R_{t_1,t_2}(x,t_2)Q_{t_1,t_2}(t_2),
\end{aligned}$$

przy czym drugą równość dowodzi się podobnie jak (35) w Lemacie 20.

Pozostaje zatem wykazać (38). Ponownie użyjemy Lematu 19. Oznaczmy $\tilde{R}_{t_1,t_2}(x,y) = (x-y)R_{t_1,t_2}(x,y)$ oraz $\tilde{K}_{t_1,t_2}(x,y) = (x-y)K_{t_1,t_2}(x,y)$. Mamy

$$R_{t_1,t_2}\circ K = R_{t_1,t_2} - K = K\circ R_{t_1,t_2}, \tag{42}$$

czyli

$$\int_{t_1}^{t_2} R_{t_1,t_2}(x,z)K(z,y)dz = R_{t_1,t_2}(x,y) - K(x,y) = \int_{t_1}^{t_2} K(x,z)R_{t_1,t_2}(z,y)dz.$$

Mnożąc pierwszą równość stronami przez $(x-y) = (x-z)+(z-y)$, otrzymujemy

$$\tilde{R}_{t_1,t_2}\circ K + R_{t_1,t_2}\circ\tilde{K} = \tilde{R}_{t_1,t_2} - \tilde{K}. \tag{43}$$

Składając teraz obie strony prawostronnie z R_{t_1,t_2} (poprzez interpretację $\circ$ jako złożenie operatorów jądrowych lub bezpośrednio z definicji całkowej łatwo wykazać, że $\circ$ jest łączna) i korzystając ponownie z (42), dostajemy

$$\tilde{R}_{t_1,t_2} \circ (R_{t_1,t_2} - K) + R_{t_1,t_2} \circ \tilde{K} \circ R_{t_1,t_2} = \tilde{R}_{t_1,t_2} \circ R_{t_1,t_2} - \tilde{K} \circ R_{t_1,t_2},$$

czyli (korzystając z rozdzielności $\circ$ względem dodawania)

$$-\tilde{R}_{t_1,t_2} \circ K + R_{t_1,t_2} \circ \tilde{K} \circ R_{t_1,t_2} = -\tilde{K} \circ R_{t_1,t_2},$$

co po dodaniu stronami do (43) daje

$$\tilde{R}_{t_1,t_2} = \tilde{K} + R_{t_1,t_2} \circ \tilde{K} \circ R_{t_1,t_2} + \tilde{K} \circ R_{t_1,t_2} + \tilde{R}_{t_1,t_2} \circ K \tag{44}$$

Mamy $K(x,y) = K(y,x)$, skąd wynika, że również $R_{t_1,t_2}(x,y) = R_{t_1,t_2}(y,x)$. Skorzystamy także z definicji funkcji K (zauważmy, że Lemat 20 korzystał jedynie z dobrych własności analitycznych funkcji K, nie z jej postaci), która implikuje, że

$$\tilde{K}(x,y) = \pi^{-1}(\sin x \cos y - \cos x \sin y) = f(x)f'(y) - f(y)f'(x),$$

Dzięki symetrii funkcji R_{t_1,t_2} mamy

$$\begin{aligned} R_{t_1,t_2} \circ \tilde{K} \circ R_{t_1,t_2}(x,y) &= \int_{t_1}^{t_2} \int_{t_1}^{t_2} R_{t_1,t_2}(x,s)(f(s)f'(z) - f(z)f'(s))R_{t_1,t_2}(y,z)dsdz \\ &= (Q_{t_1,t_2}(x) - f(x))(P_{t_1,t_2}(y) - f'(y)) - (Q_{t_1,t_2}(y) - f(y))(P_{t_1,t_2}(x) - f'(x)). \end{aligned}$$

Ponadto

$$\begin{aligned} \tilde{K} \circ R_{t_1,t_2}(x,y) &= f(x) \int_{t_1}^{t_2} f'(z)R_{t_1,t_2}(y,z)dz - f'(x) \int_{t_1}^{t_2} f(z)R_{t_1,t_2}(y,z)dz \\ &= f(x)(P_{t_1,t_2}(y) - f'(y)) - f'(x)(Q_{t_1,t_2}(y) - f(y)) \end{aligned}$$

oraz podobnie

$$\begin{aligned} R_{t_1,t_2}(x,y) \circ \tilde{K} &= f'(y) \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)f(z)dz - f(y) \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)f'(z)dz \\ &= f'(y)(Q_{t_1,t_2}(x) - f(x)) - f(y)(P_{t_1,t_2}(x) - f'(x)). \end{aligned}$$

Dodając stronami cztery ostatnie nierówności otrzymujemy (38). □

Potrafimy już różniczkować funkcje $P_{t_1,t_2}(x)$, $Q_{t_1,t_2}(x)$ względem parametrów t_i. W kolejnym lemacie wyprowadzimy wzory na pochodne cząstkowe względem argumentu x.

Lemat 22. *Dla dowolnych $t_1 < t_2$ oraz $x \in (t_1, t_2)$ mamy*

$$\frac{\partial Q_{t_1,t_2}(x)}{\partial x} = P_{t_1,t_2}(x) - R_{t_1,t_2}(x,t_2)Q_{t_1,t_2}(t_2) + R_{t_1,t_2}(x,t_1)Q_{t_1,t_2}(t_1)$$

oraz

$$\frac{\partial P_{t_1,t_2}(x)}{\partial x} = -Q_{t_1,t_2}(x) - R_{t_1,t_2}(x,t_2)P_{t_1,t_2}(t_2) + R_{t_1,t_2}(x,t_1)P_{t_1,t_2}(t_1)$$

Dowód. Zdefiniujmy operator różniczkowy T działający na funkcjach dwóch zmiennych wzorem

$$T = \frac{\partial}{\partial x} + \frac{\partial}{\partial y}$$

Aby obliczyć pochodną względem argumentu x, wyznaczymy TR_{t_1,t_2}, co pozwoli zamienić pochodną po x w definicji $P_{t_1,t_2}(x)$, $Q_{t_1,t_2}(x)$ na pochodną względem y i poprzez całkowanie przez części „przerzucić” tę pochodną na funkcję f, która jest dana jawnym wzorem.

Przypomnijmy, że z Lematu 19.

$$\int_{t_1}^{t_2} R_{t_1,t_2}(x,z)K(z,y)dz = R_{t_1,t_2}(x,y) - K(x,y) = \int_{t_1}^{t_2} K(x,z)R_{t_1,t_2}(z,y). \tag{45}$$

Stąd

$$T(R_{t_1,t_2} - K)(x,y) = \int_{t_1}^{t_2} \frac{\partial}{\partial x} R_{t_1,t_2}(x,z)K(z,y)dz + \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)\frac{\partial}{\partial y}K(z,y)dz,$$

czyli

$$TR_{t_1,t_2}(x,y) = \int_{t_1}^{t_2} \frac{\partial}{\partial x} R_{t_1,t_2}(x,z)K(z,y)dz + \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)\frac{\partial}{\partial y}K(z,y)dz, \tag{46}$$

gdyż jak łatwo sprawdzić

$$TK(x,y) = \frac{1}{\pi}\Big(\frac{(x-y)\cos(x-y) - \sin(x-y)}{(x-y)^2} + \frac{-(x-y)\cos(x-y) + \sin(x-y)}{(x-y)^2}\Big) = 0.$$

Zauważmy, że z wzoru na całkowanie przez części,

$$R_{t_1,t_2}(x,t_2)K(t_2,y) - R_{t_1,t_2}(x,t_1)K(t_1,y) = \int_{t_1}^{t_2} (\frac{\partial}{\partial z} R_{t_1,t_2}(x,z))K(z,y)dz + \int_{t_1}^{t_2} R_{t_1,t_2}(x,z)(\frac{\partial}{\partial z}K(z,y))dz.$$

Dodając ostatnią równość do (46), otrzymujemy

$$\begin{aligned} &TR_{t_1,t_2}(x,y) + R_{t_1,t_2}(x,t_2)K(t_2,y) - R_{t_1,t_2}(x,t_1)K(t_1,y) \\ &= (TR)\circ K(x,y) + R_{t_1,t_2}\circ(TK)(x,y) = (TR_{t_1,t_2})\circ K(x,y), \end{aligned} \tag{47}$$

Składając obie strony (47) prawostronnie z R i ponownie korzystając z Lematu 19, otrzymujemy

$$\begin{aligned} &(TR_{t_1,t_2})\circ R(x,y) + R_{t_1,t_2}(x,t_2)(R_{t_1,t_2}(t_2,y) - K(t_2,y)) - R_{t_1,t_2}(x,t_1)((R_{t_1,t_2}(t_1,y) - K(t_1,y)) \\ &= (TR_{t_1,t_2})\circ R(x,y) - (TR_{t_1,t_2})\circ K(x,y), \end{aligned}$$

czyli

$$R_{t_1,t_2}(x,t_2)(R_{t_1,t_2}(t_2,y) - K(t_2,y)) - R_{t_1,t_2}(x,t_1)((R_{t_1,t_2}(t_1,y) - K(t_1,y)) = -(TR_{t_1,t_2})\circ K(x,y),$$

co po dodaniu stronami do (47) daje

$$TR_{t_1,t_2}(x,y) = R_{t_1,t_2}(x,t_1)R_{t_1,t_2}(t_1,y) - R_{t_1,t_2}(x,t_2)R_{t_1,t_2}(t_2,y)$$

lub równoważnie

$$\frac{\partial}{\partial x}R_{t_1,t_2}(x,y) = R_{t_1,t_2}(x,t_1)R_{t_1,t_2}(t_1,y) - R_{t_1,t_2}(x,t_2)R_{t_1,t_2}(t_2,y) - \frac{\partial}{\partial y}R_{t_1,t_2}(x,y).$$

Zatem

$$\begin{aligned} \frac{\partial Q_{t_1,t_2}(x)}{\partial x} =& f'(x) + \int_{t_1}^{t_2} \frac{\partial}{\partial x}R_{t_1,t_2}(x,y)f(y)dy \\ =& f'(x) - \int_{t_1}^{t_2} \frac{\partial}{\partial y}R_{t_1,t_2}(x,y)f(y)dy \\ &+ \int_{t_1}^{t_2} (R_{t_1,t_2}(x,t_1)R_{t_1,t_2}(t_1,y) - R_{t_1,t_2}(x,t_2)R_{t_1,t_2}(t_2,y))f(y)dy \\ &= f'(x) + \int_{t_1}^{t_2} R_{t_1,t_2}(x,y)f'(y)dy - R_{t_1,t_2}(x,t_2)f(t_2) + R_{t_1,t_2}(x,t_1)f(t_1) \\ &= P_{t_1,t_2}(x) - R_{t_1,t_2}(x,t_2)Q_{t_1,t_2}(t_2) + R_{t_1,t_2}(x,t_1)Q_{t_1,t_2}(t_1), \end{aligned}$$

co kończy dowód pierwszej równości z tezy lematu. Dowód drugiej równości jest analogiczny (korzysta z postaci funkcji f, która gwarantuje, że $f'' = -f$). □

Dla uproszczenia notacji oznaczmy

$$p_i(t_1,t_2) = P_{t_1,t_2}(t_i)$$
$$q_i(t_1,t_2) = Q_{t_1,t_2}(t_i),$$

$i = 1,2$ oraz $R(t_1,t_2) = R_{t_1,t_2}(t_1,t_2) = R_{t_1,t_2}(t_2,t_1)$.

Lemat 23.

$$R(t_1,t_2) = \frac{q_1(t_1,t_2)p_2(t_1,t_2) - p_1(t_1,t_2)q_2(t_1,t_2)}{t_1 - t_2}, \tag{48}$$

Ponadto dla $i \neq j$,

$$\frac{\partial q_j(t_1,t_2)}{\partial t_i} = (-1)^i R(t_1,t_2) q_i(t_1,t_2), \tag{49}$$
$$\frac{\partial p_j(t_1,t_2)}{\partial t_i} = (-1)^i R(t_1,t_2) p_i(t_1,t_2), \tag{50}$$
$$\frac{\partial q_i(t_1,t_2)}{\partial t_i} = p_i(t_1,t_2) + (-1)^i R(t_1,t_2) q_j(t_1,t_2), \tag{51}$$
$$\frac{\partial p_i(t_1,t_2)}{\partial t_i} = -q_i(t_1,t_2) + (-1)^i R(t_1,t_2) p_j(t_1,t_2). \tag{52}$$
$$R_{t_1,t_2}(t_i,t_i) = p_i(t_1,t_2)\frac{\partial}{\partial t_i} q_i(t_1,t_2) - q_i(t_1,t_2)\frac{\partial}{\partial t_i} p_i(t_1,t_2) \tag{53}$$

Dowód. Przypomnijmy sobie Lemat 21 i zauważmy, że równość (48) wynika bezpośrednio z (38), zaś (49) i (50) z (40) i (41) (argument t_j występuje jedynie w jednym miejscu wśród argumentów i parametrów funkcji $Q_{t_1,t_2}(t_i)$, $P_{t_1,t_2}(t_i)$).

Z kolei ze wzoru na różniczkowanie funkcji złożonej mamy

$$\frac{\partial q_i(t_1,t_2)}{\partial t_i} = \frac{\partial Q_{t_1,t_2}(t_i)}{\partial t_i} = \frac{\partial}{\partial x} Q_{t_1,t_2}(x)\Big|_{x=t_i} + \frac{\partial}{\partial t_i} Q_{t_1,t_2}(x)\Big|_{x=t_i},$$

co z lematów 21 oraz 22 jest równe

$$\begin{aligned}&P_{t_1,t_2}(t_i) - R_{t_1,t_2}(t_i,t_2)Q_{t_1,t_2}(t_2) + R_{t_1,t_2}(t_i,t_1)Q_{t_1,t_2}(t_1) + (-1)^i R_{t_1,t_2}(t_i,t_i)Q_{t_1,t_2}(t_i)\\ &= p_i(t_1,t_2) + (-1)^i R(t_1,t_2) q_j(t_1,t_2),\end{aligned}$$

dowodząc (51). Dowód równości (52) jest analogiczny.

Pozostaje udowodnić (53). Z Lematu (21) mamy

$$\begin{aligned}
R_{t_1,t_2}(t_i,t_i) &= \frac{\partial Q_{t_1,t_2}(x)}{\partial x}\Big|_{x=t_i} P_{t_1,t_2}(t_i) - \frac{\partial P_{t_1,t_2}(x)}{\partial x}\Big|_{x=t_i} Q_{t_1,t_2}(t_i)\\
&= \frac{\partial Q_{t_1,t_2}(x)}{\partial x}\Big|_{x=t_i} p_i(t_1,t_2) - \frac{\partial P_{t_1,t_2}(x)}{\partial x}\Big|_{x=t_i} q_i(t_1,t_2)\\
&= \Big(\frac{\partial}{\partial t_i} q_i(t_1,t_2) - \frac{\partial}{\partial t_i} Q_{t_1,t_2}(x)\Big|_{x=t_i}\Big) p_i(t_1,t_2) - \Big(\frac{\partial}{\partial t_i} p_i(t_1,t_2) - \frac{\partial}{\partial t_i} P_{t_1,t_2}(x)\Big|_{x=t_i}\Big) q_i(t_1,t_2)\\
&= \Big(\frac{\partial}{\partial t_i} q_i(t_1,t_2) - (-1)^i R_{t_1,t_2}(t_i,t_i) Q_{t_1,t_2}(t_i)\Big) p_i(t_1,t_2) - \Big(\frac{\partial}{\partial t_i} p_i(t_1,t_2) - (-1)^i R_{t_1,t_2}(t_i,t_i) P_{t_1,t_2}(t_i)\Big) q_i(t_1,t_2)\\
&= \Big(\frac{\partial}{\partial t_i} q_i(t_1,t_2) - (-1)^i R_{t_1,t_2}(t_i,t_i) q_i(t_1,t_2)\Big) p_i(t_1,t_2) - \Big(\frac{\partial}{\partial t_i} p_i(t_1,t_2) - (-1)^i R_{t_1,t_2}(t_i,t_i) p_i(t_1,t_2)\Big) q_i(t_1,t_2)\\
&= p_i(t_1,t_2)\frac{\partial}{\partial t_i} q_i(t_1,t_2) - q_i(t_1,t_2)\frac{\partial}{\partial t_i} p_i(t_1,t_2),
\end{aligned}$$

dowodząc (53). □

Do dowodu Twierdzenia 29 potrzebny nam będzie jeszcze jeden prosty lemat, wynikający z definicji Δ_{t_1,t_2} oraz funkcji p_i, q_i. Jego dowód zostawiamy jako ćwiczenie.

Lemat 24. *Dla dowolnych $t_1 < t_2$,*

$$\Delta_{t_1,t_2} = \Delta_{-t_2,-t_1}$$

oraz dla $i \neq j$, $i, j \in \{1, 2\}$.

$$p_i(t_1, t_2) = -p_j(-t_2, -t_1),\ q_i(t_1, t_2) = -q_j(-t_2, -t_1).$$

Dowód Twierdzenia 29. Oznaczmy

$$\Delta(t) = \Delta_{-t/2,t/2}.$$

Łącząc lematy 20 oraz 23, otrzymujemy dla $i, j \in \{1, 2\}$, $i \neq j$

$$\begin{aligned}\frac{\partial \log \Delta_{t_1,t_2}}{\partial t_i} =&(-1)^{i+1}\Big(p_i(t_1,t_2)\frac{\partial}{\partial t_i}q_i(t_1,t_2) - q_i(t_1,t_2)\frac{\partial}{\partial t_i}p_i(t_1,t_2)\Big)\\ =&(-1)^{i+1}\Big(p_i(t_1,t_2)(p_i(t_1,t_2) + (-1)^i R(t_1,t_2)q_j(t_1,t_2))\\ &- q_i(t_1,t_2)(-q_i(t_1,t_2) + (-1)^i R(t_1,t_2)p_j(t_1,t_2))\Big)\\ =&(-1)^{i+1}\Big(p_i(t_1,t_2)^2 + q_i(t_1,t_2)^2\Big) - R(t_1,t_2)p_i(t_1,t_2)q_j(t_1,t_2) + R(t_1,t_2)q_i(t_1,t_2)p_j(t_1,t_2)\end{aligned}$$

Stąd

$$\begin{aligned}&\frac{1}{2}(\frac{\partial}{\partial t_2} - \frac{\partial}{\partial t_1})\log \Delta_{t_1,t_2}\\ =&-\frac{1}{2}(p_1(t_1,t_2)^2 + p_2(t_1,t_2)^2 + q_1(t_1,t_2)^2 + q_2(t_1,t_2)^2) - R(t_1,t_2)p_2(t_1,t_2)q_1(t_1,t_2) + R(t_1,t_2)q_2(t_1,t_2)p_1(t_1,t_2)\\ =&-\frac{1}{2}(p_1(t_1,t_2)^2 + p_2(t_1,t_2)^2 + q_1(t_1,t_2)^2 + q_2(t_1,t_2)^2) + R(t_1,t_2)^2(t_2 - t_1),\end{aligned}$$

przy czym w ostatniej równości ponownie skorzystalismy z Lematu 23.

Zatem, uwzględniając symetrie podane w Lemacie 24, otrzymujemy

$$\begin{aligned}\frac{d}{dt}\log \Delta_{-t/2,t/2} &= \Big(\frac{\partial}{\partial t_2} - \frac{\partial}{\partial t_1}\Big)\log \Delta_{t_1,t_2}\Big|_{t_1=-t/2,t_2=t/2}\\ &= -p(t)^2 - q(t)^2 - tr(t)^2,\end{aligned}$$

gdzie

$$\begin{aligned}p(t) =&p_1(-t/2, t/2),\\ q(t) =&q_1(-t/2, t/2),\\ r(t) =&R_{-t/2,t/2}(-t/2, t/2).\end{aligned}$$

Korzystając ponownie z Lematu 23, mamy

$$r(t) = -2p(t)q(t)/t,$$

skąd otrzymujemy

$$\frac{d}{dt}\log \Delta_{-t/2,t/2} = -p(t)^2 - q(t)^2 - 4p(t)^2q(t)^2 \tag{54}$$

Ponadto z Lematu 23,

$$\begin{aligned}q'(t) &= \frac{1}{2}\Big(-\frac{\partial q_1(t_1,t_2)}{\partial t_1} + \frac{\partial q_1(t_1,t_2)}{\partial t_2}\Big)\Big|_{t_1=-t/2,t_2=t/2}\\ &= \frac{1}{2}\Big(-p_1(-t/2,t/2) + R_{-t/2,t/2}(t)q_2(-t/2,t/2) + R_{-t/2,t/2}q_2(-t/2,t/2)\Big)\\ &= -p(t)/2 + r(t)q(t)\\ &= -p(t)/2 + 2p(t)q(t)^2/t\end{aligned}$$

oraz podobnie

$$p'(t) = q(t)/2 - 2p(t)^2 q(t)/t.$$

Przypomnijmy, że

$$\sigma(t) = t\frac{d}{dt}\log \Delta_{-t/2,t/2} = -t(p(t)^2 + q(t)^2) + 4q(t)^2 p(t)^2$$

Używając ostatnich trzech równości możemy obliczyć

$$\begin{aligned}\sigma'(t) &= -p(t)^2 - q(t)^2 - 2tp(t)p'(t) - 2tq(t)q'(t) + 8q(t)q'(t)p(t)^2 + 8q(t)^2p(t)p'(t)\\ &= -p(t)^2 - q(t)^2.\end{aligned}$$

oraz

$$\begin{aligned}t\sigma''(t) &= -2tp(t)p'(t) - 2tq(t)q'(t)\\ &= -2tp(t)q(t) + 4p(t)^3q(t) - 2tq(t)p(t) - 4p(t)q(t)^3\\ &= 4(p(t)^3q(t) - q(t)^3p(t)) = 4p(t)q(t)(p(t)^2 - q(t)^2).\end{aligned}$$

W szczególności

$$\begin{aligned}t\sigma'(t) - \sigma(t) &= -4q(t)^2p(t)^2\\ t\sigma'(t) - \sigma(t) + (\sigma'(t))^2 &= (p(t)^2 - q(t)^2)^2\end{aligned}$$

oraz

$$(t\sigma''(t))^2 = 4(4p(t)q(t))^2(p(t)^2 - q(t)^2)^2 = 4(t\sigma'(t) - \sigma(t))(t\sigma'(t) - \sigma(t) + (\sigma'(t))^2),$$

co dowodzi tezy twierdzenia. □

6.4 Zachowanie wartości własnych przy brzegu nośnika

Na zakończenie wykładu omówimy krótko odpowiedniki Twierdzeń Gaudina-Mehty oraz Jimbo-Miwy-Mori-Sato, opisujące zachowanie wartości własnych macierzy typu GUE w małym otoczeniu punktów $-2\sqrt{N}$ oraz $\sqrt{N}$. Nie przedstawimy żadnych dowodów, ograniczymy się jedynie do sformułowania odpowiednich faktów. Skoncentrujemy się na zachowaniu wartości własnych w okolicy $2\sqrt{N}$, gdyż z symetrii zachowanie na przeciwnym krańcu nośnika miary Wignera jest analogiczne.

Zaczniemy od zdefiniowania pewnej funkcji specjalnej, występującej dość często w różnych działach matematyki

Definicja 13. *Funkcją Airy'ego nazywamy funkcję* $\mathrm{Ai}\colon \mathbb{R} \to \mathbb{R}$*, daną wzorem*

$$\mathrm{Ai}(x) = \frac{1}{2\pi i}\int_C e^{z^3/3 - xz}dz,$$

gdzie C jest konturem w płaszczyźnie zespolonej, złożonym z półprostych $\{e^{-\pi i/3}t\colon t \in \mathbb{R}_+\}$ oraz $\{e^{\pi i/3}t\colon t \in \mathbb{R}_+\}$, zorientowanym zgodnie z ruchem wskazówek zegara.

Definicja 14. *Jądrem Airy'ego nazwiemy jedyną funkcję ciągłą $K_{\mathrm{Ai}}\colon \mathbb{R}^2$ to$\mathbb{R}$, taką że dla $x \neq y$,*

$$K_{\mathrm{Ai}}(x,y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x-y}.$$

Odpowiednikiem Twierdzenia Gaudina-Mehty jest

Twierdzenie 30. *Dla dowolnych $-\infty < t_1 < t_2 \leqslant \infty$,*

$$\lim_{N\to\infty} \mathbb{P}(N^{1/6}(\lambda_1^N - 2\sqrt{N}), \ldots, N^{1/6}(\lambda_1^N - 2\sqrt{N}) \notin [-t_1, t_2]) = \Delta(K_{\mathrm{Ai}}, [t_1, t_2]).$$

Zauważmy, że czynnik skalujący wynosi $N^{1/6}$, czyli jest mniejszy niż w Twierdzeniu Gaudina-Mehty (gdzie był równy $N^{1/2}$). Intuicyjnie odzwierciedla to fakt, że gęstość miary Wignera znika w 2 i dlatego możemy spodziewać się mniejszej gęstości wartości własnych (po przeskalowaniu do przedziału $[-2, 2]$, jak w Twierdzeniu Wignera, z powyższego twierdzenia otrzymujemy, że odstęp między 2 i najbliższą wartością własną jest równy $N^{-1/6-1/2} = N^{-2/3} > N^{-1}$). Cechą wspólną powyższego faktu i Twierdzenia Gaudina-Mehty jest, że graniczne zachowanie prawdopodobieństw opisane jest przez wyznaczniki Fredholma. Również schemat dowodu jest podobny, korzystając z reprezentacji prawdopodobieństw przez wyznaczniki Fredholma z jądrem K_N (Twierdzenie 28), sprowadza się do analizy odpowiednio przeskalowanych funkcji. Szczegóły analityczne są oczywiście nietrywialne (jak można się spodziewać już na podstawie raczej skomplikowanej definicji funkcji K_{Ai}).

Oznaczając przez λ_N^{max} i podstawiając w powyższym twierdzeniu $t_1 = t, t_2 = \infty$, otrzymujemy, że

$$\lim_{N\to\infty} \mathbb{P}\Big(N^{2/3}\Big(\frac{\lambda_N^{max}}{\sqrt{N}} - 2\Big) \leqslant t\Big) = G(t),$$

gdzie

$$G(t) = \Delta(K_{\mathrm{Ai}}, [t,\infty)) = 1 + \sum_{n=1}^{\infty} \frac{(-1)^n}{n!} \int_t^\infty \cdots \int_t^\infty \det[K_{\mathrm{Ai}}(x_i, x_j)]_{i,j\leqslant n} dx_1 \cdots dx_n.$$

Podobnie jak w Twierdzeniu Jimbo-Miwy-Mori-Sato mamy więc zbieżność dystrybuant przeskalowanych wartości własnych do pewnej funkcji granicznej, o której a priori nie wiemy czy jest dystrybuantą. Okazuje się, że jest, a zatem największa wartość własna macierzy GUE jest po przeskalowaniu słabo zbieżna. Dokładniej, zachodzi następujące

Twierdzenie 31 (Tracy-Widom)**.** *Funkcja G jest dystrybuantą pewnej zmiennej losowej. Co więcej, zachodzi równość*

$$G(t) = \exp\Big(-\int_t^\infty (x-t)q(x)^2 dx\Big),$$

gdzie q jest rozwiązaniem tzw. równanie Painlevé II:

$$q''(t) = tq(t) + 2q^3,$$

takim że dla $t \to \infty$, $q(t) \simeq \mathrm{Ai}(t)$.

Literatura

[AGZ] G. Anderson, A. Guionnet, O. Zeitouni. *Lecture notes on random matrices.* Książka jeszcze się nie ukazała, ale wybrane fragmenty można znaleźć pod adresem `http://www.mathematik.uni-muenchen.de/~lerdos/SS09/Random/randommatrix.pdf`

[BS] Z. Bai, J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices*

[F] W. Feller. *Wstęp do rachunku prawdopodobieństwa*

[Fer]

[G] A. Guionnet. *Large Random Matrices: Lectures on Macroscopic Asymptotics*, École d'Été de Probabilités de Saint-Flour XXXVI – 2006 Dostępne pod adresem `http://www.umpa.ens-lyon.fr/~aguionne/cours.pdf`

[M] M. Mehta. *Random matrices*

[P] G. Pisier. Volumes of convex bodies and geometry of Banach spaces.

Elementy teorii macierzy losowych - seria 1:

przypomnienie podstawowych faktów dotyczących słabej zbieżności miar, tw. Wignera.

1. Niech $X_1, X_2, \ldots$ będzie ciągiem i.i.d. rzeczywistych zmiennych losowych. Rozważmy ciąg macierzy losowych $A_N = \mathrm{diag}(X_1, \ldots, X_N)$ (tzn. macierz A_N jest diagonalna, wymiarów $N \times N$ i elementy na przekątnej to $X_1, \ldots, X_N$). Zbadać zbieżność miary spektralnej macierzy A_N.

Zadania 2–4 są ze sobą ściśle powiązane i można je robić w różnej kolejności, wykorzystując już udowodnione fakty do wykazania kolejnych.

2. Niech C_k oznacza k-tą liczbę Catalana. Wykazać, że C_k jest
   a) liczbą poprawnych nawiasowań złożonych z k par nawiasów,
   b) liczbą ścieżek Dycka długości $2k$,
   c) liczbą drzew zorientowanych o $k+1$ wierzchołkach.

   Wskazać bijekcje między powyższymi zbiorami.
3. Wykazać, że $C_k = C_0C_{k-1} + C_1C_{k-2} + \ldots + C_{k-1}C_0$.
4. Wyznaczyć funkcję tworzącą ciągu C_k.
5. Wykazać, że
$$C_k = \int_{\mathbb{R}} x^{2k} d\sigma(x),$$
   gdzie σ jest miarą Wignera.
6. Wyznaczyć transformatę Stieltjesa σ, zdefiniowaną jako
$$S(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\sigma(x),$$
   dla $z \in \mathbb{C} \backslash \mathbb{R}$.
7. Wykazać, że jeżeli μ, μ_n, $n \geqslant 1$, są miarami probabilistycznymi na prostej, posiadającymi skończone momenty dowolnego rzędu, μ jest wyznaczona jednoznacznie przez swoje momenty oraz $\int x^k d\mu_n \to \int x^k d\mu$ dla $k \in \mathbb{N}$, to $\mu_n \to \mu$ słabo.
8. Wykazać, że dowolna miara probabilistyczna na prostej o zwartym nośniku jest wyznaczona przez swoje momenty.
9. Podać przykład miary probabilistycznej na prostej, która nie jest wyznaczona przez swoje momenty.
10. Wykazać, że jeśli $G = (V, E)$ jest grafem spójnym, to $|V| \leqslant |E| + 1$ oraz że równość zachodzi wtedy i tylko wtedy, gdy graf G jest drzewem.
11. Wykazać, że punktami ekstremalnymi zbioru macierzy bistochastycznych są macierze permutacji (Tw. Birkhoffa).

Elementy teorii macierzy losowych - seria 2:

macierze GOE i GUE, niezmienniczość, gęstość łączna wartości własnych

1. Wykazać, że rozkład macierzy GOE (GUE) nie zmienia się przy sprzężeniu macierzą ortogonalną (unitarną), czyli nie zależy od wyboru bazy ortogonalnej, w której reprezentujemy macierz.
2. Wykażać, że zbieżność zmiennych gaussowskich jest równoważna zbieżności ich parametrów.
3. Wykazać, że

$$\det \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \dots & \lambda_1^{N-1} \\ 1 & \lambda_2 & \lambda_2^2 & \dots & \lambda_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \lambda_N & \lambda_N^2 & \dots & \lambda_N^{N-1} \end{bmatrix} = \prod_{i<j}(\lambda_j - \lambda_i).$$

Kolejne zadania dotyczą istnienia miary Haara na grupach zwartych i ich przestrzeniach jednorodnych. Przyjmujemy w nich, że G jest grupą, która działa na pewnej zwartej przestrzeni metrycznej (M, d) poprzez izometrie.

4. Niech $\{x_1, \dots, x_n\}$, $\{y_1, \dots, y_n\}$ będą minimalnymi ε-sieciami w M. Wykazać, że istnieje permutacja π zbioru $\{1, \dots, n\}$, taka że $d(x_i, y_{\pi(i)}) \leqslant 2\varepsilon$ dla dowolnego $i \leqslant n$ (wsk. użyć twierdzenia Halla o małżeństwach).
5. Niech A_n będzie minimalną n^{-1}-siecią w M, zaś μ_n unormowaną miarą liczącą skupioną na A_n. Wykazać, że z ciągu μ_n można wybrać podciąg zbieżny, a jego granica μ jest niezmiennicza ze względu na działanie grupy G, tzn. $\mu = \mu \circ g$ dla dowolnego $g \in G$ (taką miarę nazywamy miarą Haara na M).
6. Wykazać, że jeżeli dla dowolnych $x, y \in M$ istnieje $g \in G$, t. że $g(x) = y$, to miara Haara jest wyznaczona jednoznacznie (zakładamy, że jest to miara probabilistyczna).
7. Niech U będzie macierzą o kolumnach będących wektorami własnymi macierzy typu GOE, odpowiadającymi kolejnym wartościom własnym (wybieramy wektor, którego pierwsza niezerowa współrzędna jest dodatnia). Wykazać, że U jest niezależna od wektora wartości własnych macierzy oraz jest rozłożona jak $[v_{ij}\mathrm{sgn} v_{1j}]_{i,j \leqslant N}$, gdzie $[v_{ij}]_{ij \leqslant N}$ jest macierzą losową rozłożoną względem miary Haara na O_N.
8. Wykazać, że jeżeli wielomian rzeczywisty d zmiennych $x_1, \dots, x_d$ znika gdy $x_i = x_j$ dla pewnych $i \neq j$, to jest podzielny przez $\prod_{i<j}(x_i - x_j)$.

Elementy teorii macierzy losowych - seria 3:

macierze GOE i GUE, gęstość łączna wartości własnych – c.d., twierdzenie Marczenki-Pastura, transformata Stieltjesa,

1. Niech P będzie niezerowym wielomianem n zmiennych. Wykazać, że $\{x \in \mathbb{R}^N : P(x) = 0\}$ ma miarę Lebesgue'a zero.
2. Wykazać, że jeżeli P jest wielomianem stopnia n, to $\Delta(P) = (-1)^{n(n-1)/2} r(P, P')$.
3. Udowodnić, że rugownik dwóch wielomianów jest równy wyznacznikowi odpowiedniej macierzy Sylvestera, zbudowanej z ich współczynników.
4. Wykazać, że dla dowolnej macierzy A, niezerowe wartości własne macierzy AA^T oraz $A^T A$ pokrywają się i mają równe krotności.
5. Rozważmy dowolną macierz A wymiarów $N \times n$, gdzie $N/n = y$. Niech s oznacza transformatę Stieltjesa miary spektralnej AA^T, zaś $\tilde{s}$ transformatę Stieltjesa miary spektralnej $A^T A$. Wykazać, że wówczas
$$s(z) = y^{-1}\tilde{s}(z) - \frac{1 - y^{-1}}{z}.$$
6. Niech $A = [a_{ij}]_{i,j \leqslant N}$ będzie macierzą $N \times N$. Załóżmy, że A są oraz $A^{(k)}$ $(k \leqslant N)$ są odwracalne oraz oznaczmy $H = [h_{ij}]_{i,j \leqslant N} = A^{-1}$. Wykazać, że
$$h_{kk} = \frac{1}{a_{kk} - \alpha_k^T (A^{(k)})^{-1} \beta_k},$$
gdzie α_k^T oraz β_k są wektorami otrzymanymi odpowiednio z k-tego wiersza i k-tej kolumny A przez usunięcie k-tej współrzędnej.
7. Wykazać, że jeśli macierz A jest symetryczna i odwracalna oraz $A^{(k)}$ jest odwracalna, to
$$\operatorname{tr} A^{-1} - \operatorname{tr} (A^{(k)})^{-1} = \frac{1 + \alpha_k^T (A^{(k)})^{-2} \alpha_k}{a_{kk} - \alpha_k^T (A^{(k)})^{-1} \alpha_k}$$
8. Udowodnić, że dla dowolnej miary μ i dowolnego $z \in \mathbb{C}^+$, $\operatorname{Im} s_\mu(z) > 0$.
9. Udowodnić, że jeśli μ jest miarą probabilistyczną skupioną na $[0, \infty)$, to dla każdego $z \in \mathbb{C}^+$ oraz $y > 0$,
$$\operatorname{Im} (y + z - 1 + yzs_\mu(z)) > 0.$$

Elementy teorii macierzy losowych - seria 4:

twierdzenie Marczenki-Pastura, transformata Stieltjesa, oszacowania normy operatorowej macierzy losowych

1. Wykazać, że transformata Stieltjesa jest funkcją analityczną w $\mathbb{C}^+$.
2. Wykazać, że jeżeli μ jest miarą probabilistyczną, to $\lim_{y\to\infty} y\mathrm{Im}\, s_\mu(iy) = 1$.
3. Wykazać, że dla dowolnej miary skończonej μ, $\lim_{\mathrm{Im}\, z\to\infty} s_\mu(z) = 0$.
4. (Uzupełnienie dowodu twierdzenia Marczenki-Pastura). Przy oznaczeniach z notatek wykazać, że $\mathrm{tr}\, CC^* \leqslant (n+N)K(z)$, gdzie $K(z)$ jest stałą zależną tylko od z.
5. (związek transformaty Stieltjesa z transformatą Laplace'a) Niech μ będzie miarą skończoną na $[0,\infty)$. Wykazać, że
$$s_\mu(z) = \int_0^\infty e^{sz} \int_0^\infty e^{-ts} d\mu(t) ds.$$
6. Dla dowolnej rzeczywistej zmiennej losowej X zdefiniujmy normę
$$\|X\|_{\psi_2} = \inf \Big\{ C\colon \mathbb{E}\exp\Big(\Big(\frac{|X|}{C}\Big)^2\Big) \leqslant 2\},$$
przy czym przyjmujemy, że infimum zbioru pustego to ∞. Wykazać, że dla $t \geqslant 0$,
$$\mathbb{P}(|X| \geqslant t) \leqslant 2\exp(-(t/\|X\|_{\psi_2})^2).$$
7. Niech $X_1,\ldots,X_n$ będą niezależnymi zmiennymi losowymi o średniej 0. Wykazać, że
$$\|X_1+\ldots+X_n\|_{\psi_2} \leqslant C\Big(\sum_{i=1}^n \|X_i\|_{\psi_2}^2\Big)^{1/2},$$
gdzie C jest pewną stałą uniwersalną. Wywnioskować stąd, że
$$\mathbb{P}(|X_1+\ldots+X_n| \geqslant t) \leqslant 2\exp\Big(-c\frac{t^2}{\sum_{i=1}^n \|X_i\|_{\psi_2}^2}\Big)$$
dla pewnej stałej uniwersalnej $c > 0$.
8. Niech $X_1,\ldots,X_N$ będą niezależnymi wektorami losowymi w $\mathbb{R}^n$ o średniej zero i macierzy kowariancji Id. Załóżmy, że istnieje C, takie że
$$\sup_{y\in S^{n-1}} \|\langle X_i, y\rangle\|_{\psi_2} \leqslant C.$$
Niech M będzie macierzą losową o wierszach $X_1,\ldots,X_N$. Wykazać, że istnieje stała K, zależna tylko od C, taka że dla dowolnego $\varepsilon\in(0,1)$, jeżeli $N > C\varepsilon^{-2}n$ to $N^{-1/2}M$ jest z dużym prawdopodobieństwem $(1+\varepsilon)$-izometrią między $\mathbb{R}^n$ i $\mathbb{R}^N$, tzn.
$$(1-\varepsilon)|x| \leqslant N^{-1/2}|Mx| \leqslant (1+\varepsilon)|x|$$
dla wszystkich $x\in\mathbb{R}^n$. Zastanowić się jaki jest związek między tym faktem, a empiryczną aproksymacją macierzy kowariancji.

Elementy teorii macierzy losowych - seria 5:

Lemat Slepiana, koncentracja gaussowska

1. Wykazać, że w dowodzie lematu Slepiana-Fernique'a możemy zakładać, że wektory losowe X, Y mają rozkład absolutnie ciągły.
2. Niech X_t, Y_t będą scentrowanymi procesami gaussowskimi na pewnym zbiorze przeliczalnym t. Załóżmy, że
$$\|X_t - X_s\|^2 \leqslant \|Y_t - Y_s\|^2$$
dla dowolnych $t, s \in T$ oraz $\|X_t\|^2 = \|Y_t\|^2$ dla dowolnego $t \in T$. Wykazać, że wówczas, dla dowolnych $t_1, \ldots, t_n$ oraz liczb rzeczywistych $r_1, \ldots, r_n$ zachodzi nierówność
$$\mathbb{P}(\exists_{i \leqslant n} X_{t_i} \geqslant r_i) \leqslant \mathbb{P}(\exists_{i \leqslant n} Y_{t_i} \geqslant r_i).$$
W szczególności $\mathbb{P}(\sup_t X_t \geqslant r) \leqslant \mathbb{P}(\sup_t Y_t \geqslant r)$ dla dowolnego $r \in \mathbb{R}$.
3. (Zasada minoryzacyjna Sudakowa) Dla dowolnej przestrzeni metrycznej (T, d) oraz $\varepsilon > 0$, niech $N(T, \varepsilon)$ oznacza moc minimalnej ε-sieci w T. Rozważmy dowolny scentrowany proces gaussowski na zbiorze przeliczalnym T. Niech $d(t, s) = \|X_t - X_s\|_2$ będzie metryką pochodzącą od tego procesu. Wykazać, że istnieje stała uniwersalna $c > 0$, taka że
$$\sup_{\varepsilon > 0} \varepsilon \sqrt{\log N(T, \varepsilon)} \leqslant \mathbb{E} \sup_{t \in T} X_t.$$
4. (Twierdzenie Dudleya) Niech X_t będzie scentrowanym procesem gaussowskim określonym na zbiorze przeliczalnym T. Wykazać, że istnieje stała uniwersalna K, taka że
$$\mathbb{E} \sup_t |X_t| \leqslant |X_{t_0}| + K \int_0^\infty \sqrt{\log N(T, \varepsilon)} d\varepsilon,$$
gdzie t_0 jest dowolnym elementem zbioru T.
5. Niech $M_n = [X_{ij}]$ będzie ciągiem macierzy losowych wymiarów n na N_n (gdzie $N_n/n \to y$), których współczynniki są niezależne i mają wspólny rozkład $\mathcal{N}(a, \sigma^2)$, gdzie $a > 0$. Znaleźć normalizację c_n, taką że
$$\lim_{n \to \infty} c_n^{-1} \|M_n\|$$
istnieje p.n.? Jaka jest wartość tej granicy?

Elementy teorii macierzy losowych - seria 6:

Wielomiany Hermite'a, Twierdzenie Gaudina-Mehty, Twierdzenie JMMS

1. Niech μ będzie miarą na $\mathbb{R}$ posiadającą wszystkie momenty, zaś P_n ciągiem wielomianów ortogonalnych dla tej miary. Udowodnić, że istnieją liczby a_n, b_n, c_n, takie że
$$P_{n+1}(x) = (a_n x + b_n)P_n(x) - cP_{n-1}(x).$$

2. Wykazać, że
   a) H_n jest stopnia n i ma wiodący współczynnik równy 1,
   b) $\{H_k\}_{k\leqslant n}$ jest bazą w przestrzeni wielomianów stopnia co najwyżej n,
   c) $\{H_k/\sqrt{k!}\}_{k\geqslant 0}$ jest bazą ortonormalną przestrzeni $L_2(\mathbb{R},\gamma)$, gdzie γ jest standardową miarą gaussowską, równoważnie $\{\psi_k\}_{k\geqslant 0}$ jest bazą ortonormalną w $L_2(\mathbb{R}, dx)$.
   d) $H_{n+1}(x) = xH_n(x) - nH_{n-1}(x)$.
   e) Dla dowolnych $x \neq y$,
$$\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(n-1)!(x-y)}$$
   oraz
$$\sum_{k=0}^{n-1} \psi_k(x)\psi_k(y) = \sqrt{n}\frac{\psi_n(x)\psi_{n-1}(y) - \psi_{n-1}(x)\psi_n(y)}{x-y}.$$

3. Udowodnić, że dla dowolnych funkcji $f_1, \ldots, f_N$ oraz $g_1, \ldots, g_N$, całkowalnych z kwadratem, zachodzi wzór
$$\begin{aligned}
&\frac{1}{N!}\int_{\mathbb{R}^n} \det\Big[\sum_{k=1}^{N} f_k(x_i)g_k(x_j)\Big]_{i,j\leqslant N} dx_1\cdots dx_N \\
=&\frac{1}{N!}\int_{\mathbb{R}^N} \det\Big[f_i(x_j)\Big]_{i,j\leqslant N} \det\Big[g_i(x_j)\Big]_{i,j\leqslant N} dx_1\cdots dx_N \\
=&\det\Big[\int_{\mathbb{R}} f_i(x)g_j(x)dx\Big]_{i,j\leqslant N}.
\end{aligned}$$

4. Wykazać, że dla dowolnych ograniczonych funkcji dwóch zmiennych F, G, zachodzi
$$\Big|\det[F(x_i,x_j)]_{i,j\leqslant n} - \det[G(x_i,x_j)]_{i,j\leqslant n}\Big| \leqslant n^{1+n/2}\|F-G\|_\infty \max(\|F\|_\infty, \|G\|_\infty)^{n-1}.$$

5. Wykazać, że operacja złożenia zdefiniowana na funkcjach dwóch zmiennych odpowiada składaniu operatorów jądrowych wyznaczonych przez te funkcje.

Elementy teorii macierzy losowych - seria 7:

Twierdzenie JMMS, Twierdzenie Tracy'ego-Widoma, wartości własne macierzy niehermitowskich

1. Wykazać, że funkcja $t \mapsto \Delta_{-t/2,t/2}$ rozszerza się do funkcji analitycznej w całej płaszczyźnie zespolonej.

2. Przy oznaczeniach z wykładu, wykazać że dla dowolnych $t_1 < t_2$,

$$\Delta_{t_1,t_2} = \Delta_{-t_2,-t_1}$$

oraz dla $i \neq j$, $i, j \in \{1, 2\}$.

$$p_i(t_1, t_2) = -p_j(-t_2, -t_1),\ q_i(t_1, t_2) = -q_j(-t_2, -t_1).$$

3. Funkcją Airy'ego nazywamy funkcję $\mathrm{Ai}\colon \mathbb{R} \to \mathbb{R}$, daną wzorem

$$\mathrm{Ai}(x) = \frac{1}{2\pi i} \int_C e^{z^3/3 - xz} dz,$$

gdzie C jest konturem w płaszczyźnie zespolonej, złożonym z półprostych $\{e^{-\pi i/3}t\colon t \in \mathbb{R}_+\}$ oraz $\{e^{\pi i/3}t\colon t \in \mathbb{R}_+\}$, zorientowanym zgodnie z ruchem wskazówek zegara. Z kolei jądrem Airy'ego nazwiemy jedyną funkcję ciągłą $K_{\mathrm{Ai}}\colon \mathbb{R}^2\ to\mathbb{R}$, taką że dla $x \neq y$,

$$K_{\mathrm{Ai}}(x, y) = \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}(y)\mathrm{Ai}'(x)}{x - y}.$$

Wykazać, że powyższe definicje są poprawne.

4. Rozpatrzmy macierze n na n,

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix},\ B = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \\ \varepsilon_n & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

gdzie $\varepsilon_n = n^{-\alpha}$ dla pewnego $\alpha > 0$. Obliczyć widma tych macierzy i wywnioskować stąd, że dla macierzy niesymetrycznych widmo nie jest asymptotycznie stabilne ze względu na niewielkie zaburzenie macierzą o niskim rzędzie.

5. Niech M_n będzie macierzą $n \times n$ (niekoniecznie hermitowską), zaś μ jej miarą spektralną. Oznaczmy przez $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ funkcję charakterystyczną miary μ (utożsamiamy tu płaszczyznę zespoloną z $\mathbb{R}^2$). Wykazać, że dla $uv \neq 0$,

$$\varphi(u, v) = \frac{u^2 + v^2}{4iu\pi} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{\partial}{\partial s} \Big[\int_0^\infty \log x \nu_n(dx, s + it) \Big] e^{ius + ivt} dt ds,$$

gdzie dla $z \in \mathbb{C}$, $\nu(\cdots, z)$ oznacza miarę spektralną symetrycznej macierzy $(M - z\mathrm{Id})(M - z\mathrm{Id})^*$.

Wsk. Skorzystać ze wzoru na funkcję charakterystyczną rozkładu Cauchy'ego.

Elementy teorii macierzy losowych - egzamin

Proszę o wybranie sześciu z poniższych zadań.

1. Niech X będzie wektorem losowym w $\mathbb{R}^n$ ($n \geqslant 2$), o niezależnych współrzędnych i rozkładzie niezmienniczym ze względu na obroty (tzn. dla każdego $U \in O_n$, $UX \sim X$). Wykazać, że X ma rozkład gaussowski o średniej zero i macierzy kowariancji równej $\lambda \mathrm{Id}$ dla pewnego $\lambda \geqslant 0$.
2. Niech N_n będzie ciągiem liczb naturalnych, takim że $\lim_{n\to\infty} N_n/n = y \in (0,\infty)$. Niech $X_1^n, \ldots, X_n^n$ będą niezależnymi wektorami losowymi, rozłożonymi jednostajnie na sferze S^{N_n-1}, zaś M_n niech będzie macierzą losową $N_n \times n$ o kolumnach X_i^n, $i = 1, \ldots, n$. Zdefiniujmy macierz
$$S_n = M_n M_n^T$$
i oznaczmy przez L_n jej miarę spektralną. Wykazać, że dla dowolnej funkcji ciągłej, ograniczonej $f\colon \mathbb{R} \to \mathbb{R}$,
$$\int f(x) dL_n(x) \to \int f(x) d\nu^y(x),$$
wg. prawdopodobieństwa (ν^y jest miarą Marchenki-Pastura z parametrem y).
Wsk. Wyrazić miarę jednostajną na sferze przy pomocy niezależnych zmiennych gaussowskich oraz użyć minimaksowej charakteryzacji wartości własnych (Lemat Couranta-Fishera), aby porównać L_N z miarą spektralną odpowiedniej macierzy generowanej przez te zmienne.
3. Wyprowadzić następujący wzór na momenty miary Marchenki-Pastura z parametrem $y \in (0,\infty)$:
$$\int_{\mathbb{R}} x^k d\nu^y(x) = \sum_{r=0}^{k-1} \frac{1}{r+1} \binom{k}{r} \binom{k-1}{r} y^r.$$
4. Niech H_n będzie ciągiem macierzy GOE, wymiarów $n \times n$ (unormowanym przez $\sqrt{n}$). Wykazać, że największa (odp. najmniejsza) wartość własna H_n zbiega p.n. do 2 (odp. -2).
5. Niech $g(x)$ oznacza gęstość wektora „nieuporządkowanych" wartości własnych macierzy GUE o wymiarach $N \times N$, czyli
$$g(x) = C_N \prod_{i<j} |x_i - x_j|^2 \exp\Big(-\sum_{i=1}^n x_i^2/2\Big).$$
Niech X będzie wektorem losowym o gęstości g oraz dla $p = 1, \ldots, n$, niech g_p będzie gęstością p-wymiarowego rozkładu brzegowego X (g jest permutacyjnie niezmiennicza, więc wszystkie rozkłady brzegowe sa równe). Wykazać, że
$$g_p(x_1, \ldots, x_p) = \frac{(N-p)!}{N!} \det[K_N(x_i, x_j)]_{i,j\leqslant p},$$
gdzie $K_N(x,y) = \sum_{i=0}^{N-1} \psi_i(x)\psi_i(y)$ (ψ_i oznacza i-tą funkcję falową oscylatora).
6. Niech $X_1, X_2, \ldots$ będą zmiennymi i.i.d. o średniej 0, wariancji 1 i wszystkich momentach skończonych. Wykazać, że dla dowolnego $k \in \mathbb{N}$,
$$\lim_{n\to\infty} \mathbb{E}\Big(\frac{X_1 + \ldots + X_n}{\sqrt{n}}\Big)^k = \mathbb{E} g^k,$$
gdzie g jest standardową zmienną gaussowską. Wywnioskować stąd centralne twierdzenie graniczne (dla dowolnych zmiennych o średniej 0 i skończonym drugim momencie).
7. Niech $X, X_{ij}, i,j = 1,2,\ldots$ będą zmiennymi i.i.d., zaś $M_n = [X_{ij}]_{i,j\leqslant n}$. Wykazać, że jeżeli $n^{-1/2}\|M_n\|$ jest zbieżne wg prawdopodobieństwa, to dla każdego $0 \leqslant p < 4$, $\mathbb{E}|X|^p < \infty$ ($\|\cdot\|$ oznacza normę operatorową macierzy).
Wsk. Zbadać zachowanie największego współczynnika macierzy.

8. Niech M będzie macierzą $n \times n$, której współczynniki są niezależnymi zmiennymi $\mathcal{N}(0,1)$. Zdefiniujmy $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant 0$, jako wartości własne macierzy MM^T. Wykazać, że wektor losowy $(\lambda_1, \ldots, \lambda_n)$ ma gęstość postaci

$$g(\lambda_1, \ldots, \lambda_n) = K_n \exp\Big(-\frac{1}{2} \sum_{i=1}^{n} \lambda_i \Big) \prod_{i=1}^{n} \lambda_i^{-1/2} \prod_{i<j} (\lambda_i - \lambda_j) \mathbf{1}_{\{\lambda_1 \geqslant \ldots \geqslant \lambda_n \geqslant 0\}},$$

gdzie K_n jest pewną stałą (zależną od wymiaru n).
Uwaga: Nie trzeba wyliczać dokładnej wartości stałej K_n.

9. a) Wyznaczyć transformatę Stieltjesa miary o gęstości

$$g(x) = \frac{1}{\pi\sqrt{(x-a)(b-x)}} \mathbf{1}_{[a,b]}(x),$$

gdzie $a, b \in \mathbb{R}$, $a < b$.

b) Wykazać, że transformata Stieltjesa miary μ o nośniku skończonym jest analityczna w otoczeniu nieskończoności i posiada rozwinięcie

$$s_\mu(z) = -\sum_{n=0}^{\infty} m_n z^{-(n+1)},$$

dla $|z|$ odpowiednio dużego, gdzie m_n to n-ty moment miary μ.

Elementy teorii macierzy losowych - termin II

Proszę o wybranie pięciu z poniższych zadań.

1. Niech N_n będzie ciągiem liczb naturalnych, takim że $\lim_{n\to\infty} N_n/n = y \in (0,\infty)$. Niech $X_1^n, \ldots, X_n^n$ będą niezależnymi wektorami losowymi, rozłożonymi jednostajnie na sferze S^{N_n-1}, zaś M_n niech będzie macierzą losową $N_n \times n$ o kolumnach X_i^n, $i = 1, \ldots, n$. Zdefiniujmy macierz
$$S_n = M_n M_n^T$$
i oznaczmy przez L_n jej miarę spektralną. Wykazać, że dla dowolnej funkcji ciągłej, ograniczonej $f\colon \mathbb{R} \to \mathbb{R}$,
$$\int f(x)dL_n(x) \to \int f(x)d\nu^y(x),$$
wg. prawdopodobieństwa (ν^y jest miarą Marchenki-Pastura z parametrem y).
Wsk. Wyrazić miarę jednostajną na sferze przy pomocy niezależnych zmiennych gaussowskich oraz użyć minimaksowej charakteryzacji wartości własnych (Lemat Couranta-Fishera), aby porównać L_N z miarą spektralną odpowiedniej macierzy generowanej przez te zmienne.
2. Niech H_n będzie ciągiem macierzy GOE, wymiarów $n \times n$ (unormowanym przez $\sqrt{n}$). Wykazać, że największa (odp. najmniejsza) wartość własna H_n zbiega p.n. do 2 (odp. -2).
3. Dla $n \in \mathbb{N}$ zdefiniujmy wielomiany Czebyszewa drugiego stopnia wzorami rekurencyjnymi:
$$U_0(x) = 1, U_1(x) = 2x,$$
$$U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x),\ n \geqslant 1.$$
   a) Wykazać, że wielomiany U_n spełniają równość
   $$U_n(\cos\theta) = \frac{\sin((n+1)\theta)}{\sin\theta}$$
   b) Niech σ będzie miarą Wignera, przeskalowaną do odcinka $[-1,1]$, czyli miarą z gęstością $g(x) = 2\pi^{-1}(1-x^2)_+^{1/2}$. Wykazać, że wielomiany te stanowią ciąg wielomianów ortogonalnych dla miary Wignera (czyli, że $\int U_i U_j d\sigma = \delta_{ij}$).
4. Miarą Kestena-McKaya z parametrem $d > 1$ nazywamy miarę μ_d na $\mathbb{R}$ o gęstości
$$g_d(x) = \frac{d}{2\pi}\frac{\sqrt{4(d-1)-x^2}}{d^2-x^2}\mathbf{1}_{[-2\sqrt{d-1},2\sqrt{d-1}]}$$
   a) Znaleźć transformatę Stieltjesa miary μ_d (ozn. $s_d(z)$).
   b) Niech dla $n > 1$, X_n będzie zmienną losową o rozkładzie μ_n. Znaleźć ciąg liczb a_n, t. że X_n/a_n jest słabo zbieżny do niezdegenerowanej granicy i wyznaczyć tę granicę.
5. Niech $g(x)$ oznacza gęstość wektora „nieuporządkowanych” wartości własnych macierzy GUE o wymiarach $N \times N$, czyli
$$g(x) = C_N \prod_{i<j} |x_i - x_j|^2 \exp\Big(-\sum_{i=1}^n x_i^2/2\Big).$$
Niech X będzie wektorem losowym o gęstości g oraz dla $p = 1, \ldots, n$, niech g_p będzie gęstością p-wymiarowego rozkładu brzegowego X (g jest permutacyjnie niezmiennicza, więc wszystkie rozkłady brzegowe sa równe). Wykazać, że
$$g_p(x_1, \ldots, x_p) = \frac{(N-p)!}{N!}\det[K_N(x_i,x_j)]_{i,j\leqslant p},$$
gdzie $K_N(x,y) = \sum_{i=0}^{N-1}\psi_i(x)\psi_i(y)$ (ψ_i oznacza i-tą funkcję falową oscylatora).

6. Niech $(g_{ij})_{i,j\geqslant 1, i\neq j}$ będzie tablicą niezależnych zmiennych losowych o rozkładzie $\mathcal{N}(0,1)$. Rozważmy macierze losowe

$$M_n = \begin{pmatrix} -\sum_{\substack{j=1\\ j\neq 1}}^n g_{1j} & g_{12} & g_{13} & \dots & g_{1n} \\ g_{21} & -\sum_{\substack{j=1\\ j\neq 2}}^n g_{2j} & g_{23} & \dots & g_{2n} \\ \vdots & \dots & \ddots & \dots & \vdots \\ g_{n-1,1} & g_{(n-1)2} & \dots & -\sum_{\substack{j=1\\ j\neq n-1}}^n g_{(n-1),j} & g_{n-1,n} \\ g_{n1} & g_{n2} & \dots & g_{n,n-1} & -\sum_{\substack{j=1\\ j\neq n}}^n g_{nj} \end{pmatrix}.$$

Wykazać, że

$$0 < \liminf_{n\to\infty} \frac{\mathbb{E}\|M_n\|}{\sqrt{n\log n}} \leqslant \limsup_{n\to\infty} \frac{\mathbb{E}\|M_n\|}{\sqrt{n\log n}} < \infty.$$

7. Niech $X_1, X_2, \dots$ będą ograniczonymi zmiennymi i.i.d. o średniej zero i wariancji 1. Zdefiniujmy zmienne

$$S_n = \frac{X_1X_2 + X_2X_3 + \ldots + X_{n-1}X_n}{\sqrt{n}}.$$

Korzystając z metody momentów zbadać słabą zbieżność ciągu S_n.

8. Niech $X, X_{ij}, i,j = 1,2,\dots$ będą zmiennymi i.i.d., zaś $M_n = [X_{ij}]_{i,j\leqslant n}$. Wykazać, że jeżeli $n^{-1/2}\|M_n\|$ jest zbieżne wg prawdopodobieństwa, to dla każdego $0 \leqslant p < 4$, $\mathbb{E}|X|^p < \infty$ ($\|\cdot\|$ oznacza normę operatorową macierzy).

   Wsk. Zbadać zachowanie największego współczynnika macierzy.

9. Niech M będzie macierzą $n\times n$, której współczynniki są niezależnymi zmiennymi $\mathcal{N}(0,1)$. Zdefiniujmy $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n \geqslant 0$, jako wartości własne macierzy MM^T. Wykazać, że wektor losowy $(\lambda_1,\dots,\lambda_n)$ ma gęstość postaci

$$g(\lambda_1,\dots,\lambda_n) = K_n \exp\Big(-\frac{1}{2}\sum_{i=1}^n \lambda_i\Big)\prod_{i=1}^n \lambda_i^{-1/2}\prod_{i<j}(\lambda_i-\lambda_j)\mathbf{1}_{\{\lambda_1\geqslant\ldots\geqslant\lambda_n\geqslant 0\}},$$

   gdzie K_n jest pewną stałą (zależną od wymiaru n).

   Uwaga: Nie trzeba wyliczać dokładnej wartości stałej K_n.