

Využití teorie náhodných matic

Hana Srbová

Fakulta jaderná a fyzikálně inženýrská, ČVUT Praha

17. 4. 2012

Obsah

- 1 Úvod
- 2 Chaos
- 3 Approach to financial cross-correlations
- 4 Analysis of Human EEG Data
- 5 Control of HIV infection

Využití teorie náhodných matic

- Statistika více proměnných
- Numerická analýza
- Teorie grafů
- Finanční korelace při dokonalé konkurenci
- Modelování dopravy a pohybu chodců
- Jaderná a kvantová fyzika
- Bezdrátová komunikace
- Šíření zvuku
- Studium lidského EEG
- Studium HIV

CHAOS IN ONE-DIMENSIONAL INTEGRABLE QUANTUM SYSTEM

Šeba P., Vašata D.. Physics Letters A, Vol. 373 (2009), pp. 826–831.

Model

- Kvantová částice rozmístěna na kruhu s n bodovými interakcemi rozmístěnými na pozicích

$$0 < x_1 < \dots < x_n < 2\pi$$

- Bodová interakce na počátku popsána hraničními podmínkami

$$\frac{1}{\alpha} f(0_-) = f(0_+), \quad f'(0_-) = f'(0_+),$$

kde $\alpha \in (0, +\infty)$ je parametr

Model

- Hamiltonián systému je reprezentován jednorozměrným Laplaceovým operátorem

$$T_\alpha = -\frac{d^2}{dx^2}$$

a hraničními podmínkami:

$$f(2\pi_-) = f(0_+), \quad f'(2\pi_-) = f'(0_+), \quad (1)$$

$$\frac{1}{\alpha} f(x_{k-}) = f(x_{k+}), \quad f'(x_{k-}) = f'(x_{k+}). \quad (2)$$

- Pozice bodových interakcí jsou $x_k = \frac{2\pi\sqrt{p_k}}{\sqrt{p_{n+1}}}$, $k = 1, 2, \dots, n$, kde p_k označuje k -té prvočíslo

Výsledky pro případ $\alpha = 1$

- Místo vlastních čísel $E_l = k_l^2$, které jsou dvakrát degenerované, uvažujeme hodnoty $e_l = 2|k_l|$
- Řešení: $e_{2l-1} = e_{2l} = 2l, l = 1, 2, \dots$
- Definujeme-li rozestupy $s_l = e_{l+1} - e_l$, je potom hustota pravděpodobnosti tvaru

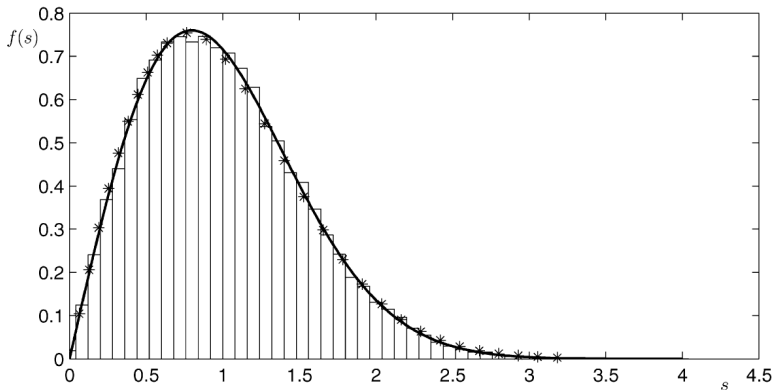
$$P(s) = \frac{(\delta(s) + \delta(s - 2))}{2}$$

Výsledky pro případ $\alpha \neq 1$

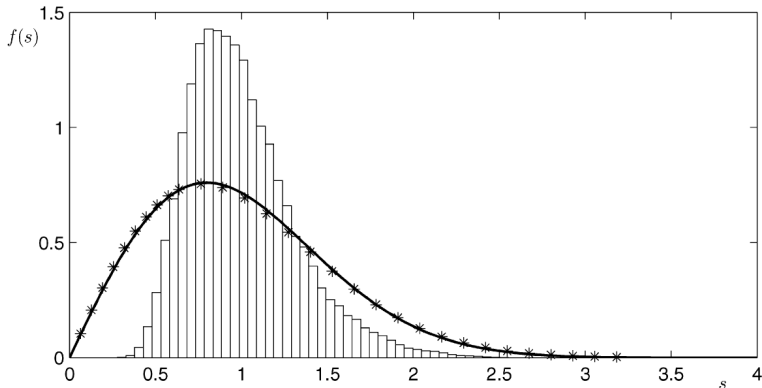
- Popis systému se liší pro sudá a lichá vlastní čísla
- Rozestupy lichých vlastních čísel podléhají Wignerově distribuci blíží se distribuci P_{GOE}

$$P_W(s) = \frac{\pi}{2} s e^{-\pi s^2/4}$$

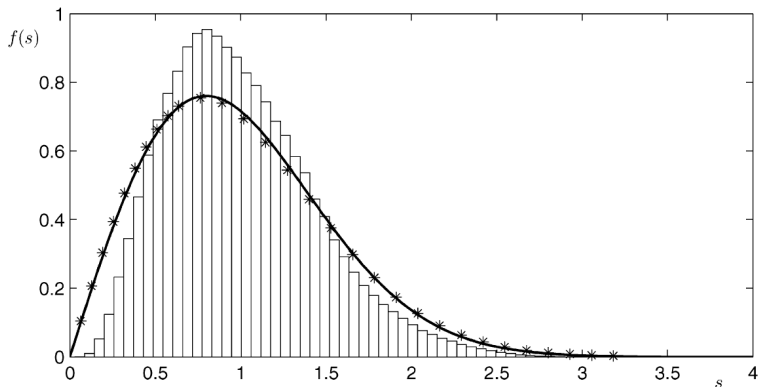
- Distribuce rozestupů sudých vlastních čísel se nikdy nepřiblíží teorii náhodných matic



Obrázek: Srovnání hustoty pravděpodobnosti rozestupů s_{2j-1} lichých vlastních čísel pro 47 bodových interakci, $\alpha = 1,001$ a počet použitých rozestupů roven $1 \cdot 10^5$ s Wignerovou distribucí (plná čára) a přesným řešením GOE (hvězdy).



Obrázek: Srovnání hustoty pravděpodobnosti rozestupů sudých vlastních čísel pro 9 bodových interakci, $\alpha = 1,9$ a počet použitých rozestupů roven $1 \cdot 10^5$ s Wignerovou distribucí (plná čára) a přesným řešením GOE (hvězdy).



Obrázek: Srovnání hustoty pravděpodobnosti rozestupů vlastních čísel pro model na přímce pro 9 bodových interakcí, $\alpha = 1,8$ a počet použitých rozestupů $1 \cdot 10^5$ s Wignerovou distribucí (plná čára) a přesným řešením GOE (hvězdy).

A RANDOM MATRIX THEORY APPROACH TO FINANCIAL CROSS-CORRELATIONS

Plerou V., Gopikrishnan P., Rosenow B., Amaral L.A.N., Stanley H.E.. J. Phys. A:
Statistical Mechanics and its Applications, Vol. 278 (2000), pp. 374–382.

Úvod do problematiky

- Analyzujeme 30-minutový výnos 1 000 největších US akcií v rozmezí dvou let 1994–1995
- 20 z největších vlastních čísel (2%) vykazují odchylky od RMT \Rightarrow ukazují na vlastnosti specifické pro daný systém
- U zbylých určujeme korelační matici a testujeme rozestupy vl. čísel, abychom určili, zda jsou opravdu náhodné
- Rozestupy získáme ze vztahu $s \equiv \lambda_{i+1} - \lambda_i$, kde posl. λ_i označuje rozbalená vlastní čísla seřazená podle velikosti

Výsledky

- Pro korelační matici \mathbf{C} vypočtenou ze vztahu

$$C_{ij} = \frac{\langle G_i G_j \rangle - \langle G_i \rangle \langle G_j \rangle}{\sigma_i \sigma_j},$$

kde G_i je časový sled cenových fluktuací pro i -tou akci
 $i = 1, \dots, N$ a $\sigma_i = \sqrt{\langle G_i^2 \rangle - \langle G_i \rangle^2}$ byla prokázána shoda
s reálnou symetrickou náhodnou maticí (či GOE maticí)

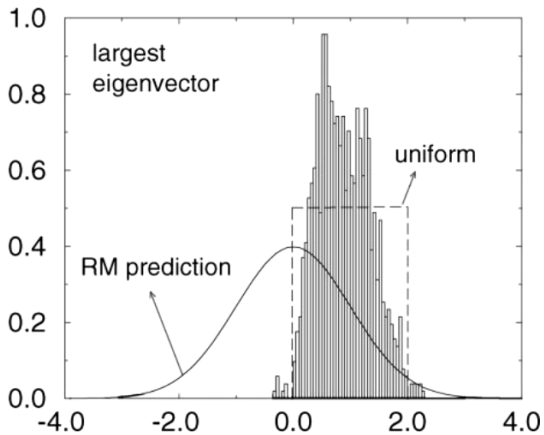
- Aplikace mnohem citlivějšího testu - rozptyl počtu vlastních čísel potvrdila shodu \mathbf{C} s RMT

Problémy vyčíslení korelací

- Chybí algoritmus pro určení interakční síly mezi dvěma společnostmi (přestože by každá ze společností měla interagovat přímo či nepřímo, přesná povaha interakcí je neznámá)
- Korelace by neměly zahrnovat jen dvě společnosti, ale celý shluk
- Korelace C_{ij} se mění v čase
- Pro každou akcii máme pouze konečný počet interakcí G_{ij} , ze kterých tvoříme průměrnou korelaci mezi firmami i a j

Analýza vlastních vektorů

- Složky vlastních vektorů náhodné korelační matice jsou normálně rozloženy se střední hodnotou rovnou 0 a rozptylem 1
- Největší vlastní číslo je silně negaussovské, „má tendenci být z rovnoměrného rozdělení“, což signalizuje, že všechny společnosti jsou navzájem propojeny
- Stabilita vlastních vektorů klesá směrem k hranici předpovědi RMT



Obrázek: Rozdělení složek vlastních vektorů.

Analýza výnosů a S&P 500 indexu

- $G_{\text{SP500}} \equiv \sum_{i=1}^{500} w_i G_i$, kde G_i jsou výnosy jednotlivých firem, $w_i \equiv S_i / \sum_{j=1}^N S_j$, S_i označuje tržové financování i -té společnosti
- Očekávání konvergence G_{SP500} ke Gaussovu rozdělení
- G_i i G_{SP500} mají stejné asymptotické chování blížíící se k $\frac{1}{x^{1+\alpha}}$, kde $1 + \alpha \approx 4$
- V případě promíchání času pro $G_i(t)$, pozorované chování mezi výnosy S&P 500 indexů již neplatí, což naznačuje existenci netriviálních korelací

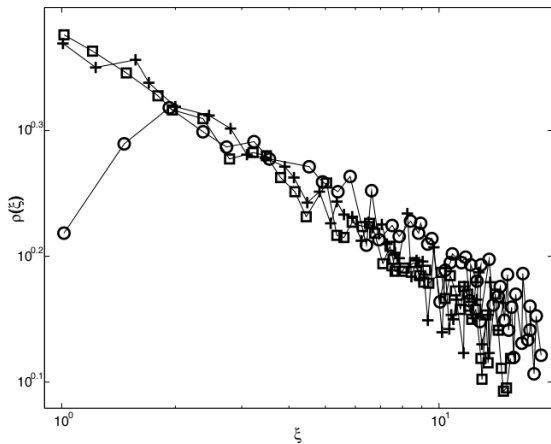
Další zajímavosti

- Distribuce růstových sazeb firemních tržeb nezávisí na odvětví průmyslu ani na trhovém financování
- Šíře σ těchto distribucí klesá se zvyšujícím se S ,
 $\sigma(S) \sim S^{-\beta}$ pro $\beta \approx 1/6$
- Podobné β objeveny také např. v případě, že za S byl zvolen počet zaměstnanců
- Obdobné statistické vlastnosti nalezeny i pro GDP (gross domestic product) či pro financování univerzitních výzkumů

RANDOM MATRIX ANALYSIS OF HUMAN EEG DATA

Šeba P.. Phys. Rev. Lett., Vol. 91 (2003), 198104.

- Záznamy 90 lidí nejprve v klidu, poté stimulovaných
- Měření elektrodami umístěnými na pokožku hlavy - některé měření pomocí 19 elektrod, některé pomocí 15 ⇒ sestavení korelační matice
- Potvrzení očekávání, že spektrální analýza je závislá na jedinci, avšak některé rysy jsou společné
- Pro malá vl. čísla závisí od jedince, pro velká pozorován algebraický chvost ⇒ zdánlivě individuální mozková aktivita obsahuje určitou hladinu synchronizace

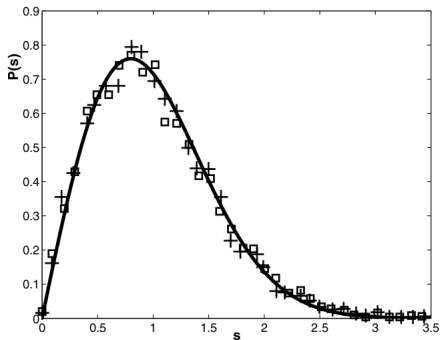


Obrázek: Hustota vlastních čísel pro data získaná od 3 osob v klidu.

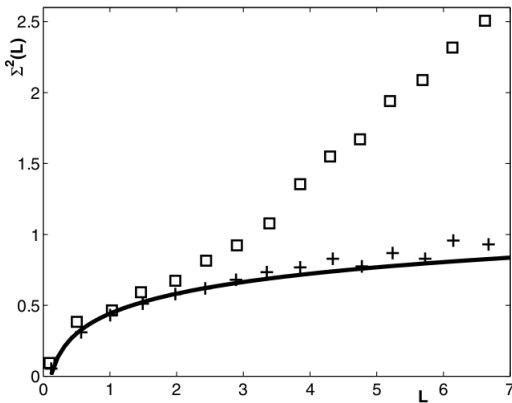
- Seřazení vlastních čísel ξ_n podle velikosti a definování rozestupů $s_n = \xi_{n+1} - \xi_n$
- Rozdělení rozestupů shodné s Wignerovou distribucí $P(s)$
- Rozptyl počtu $\Sigma^2(L)$, kde L je počet po sobě jdoucích prvočísel, lze přibližně vyjádřit jako

$$\Sigma^2(L) \approx \frac{2}{\pi^2} \left(\log(2\pi L) + 1,5772 - \frac{\pi^2}{8} \right)$$

- Při analýze vynechaná data s vlivy typu mrkání, pohyb jedince atd.



Obrázek: Hustota rozestupů pro jedince v klidu (křížky) a pro stimulovaní jedince (čtverce) srovnaná s Wignerovou formulí (plná čára).



Obrázek: Rozptyl počtu pro jedince v klidu (křížky) a pro stimulované jedince (čtverce) srovnaný s výše uvedenou aproximací.

EFFECTS OF THYMIC SELECTION OF THE T-CELL REPERTOIRE ON HLA CLASS I-ASSOCIATED CONTROL OF HIV INFECTION

Košmrlj A., Read E.L., Qi Y., Allen T.M., Altfeld M., Deeks S.G., Pereyra F., Carrington M., Walker B.D., Chakraborty A.K.. Nature, Vol. 465 (2010), pp. 350–354.

Úvod do biologických pojmů

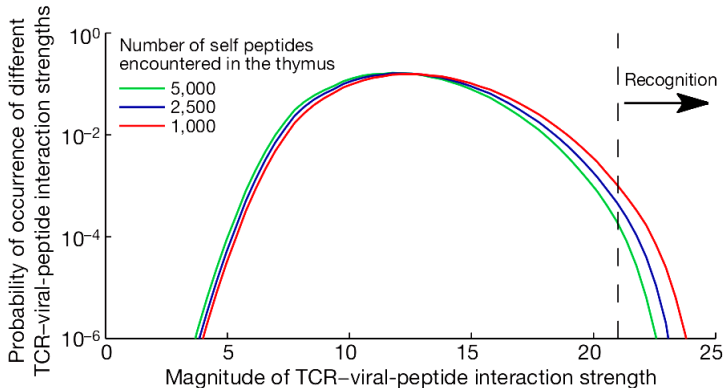
- T-buňka – imunitní buňka patřící ke skupině bílých krvinek
- TCR – receptor T-buňky
- MHC (major histocompatibility complex) – komplex, který vystavuje části proteinů T-buňce \Rightarrow v případě, že se jedná o „cizí“ proteiny, nastává imunitní reakce
- HLA (human leukocyte antigen) – lidský MHC
- Mamu – MHC makaků
- pMHC – peptid MHC

Článek se skládá z těchto 4 částí:

- HLA-peptide binding predictions
- Thymic selection model and antigen recognition
- Host-pathogen interaction dynamics
- HLA-allele association with ability to control HIV

HLA-peptide binding predictions

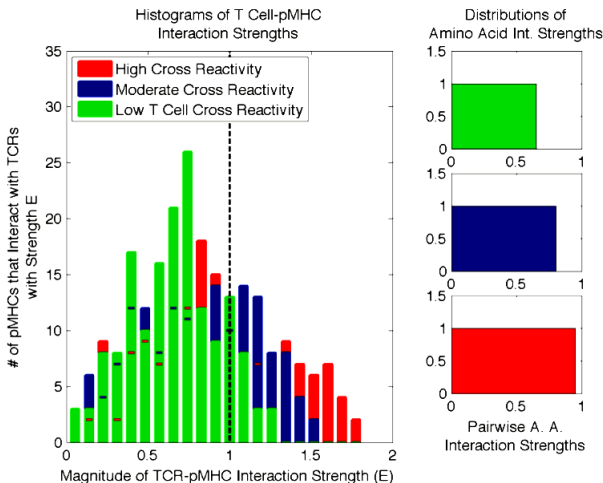
- Peptidy odvozené od lidského proteomu dávaly v podstatě stejné výsledky jako náhodně generované peptidy
- Pro cca 10^7 jedinečných peptidů odvozených od lidského proteomu máme tyto výsledky:
 - $7 \cdot 10^4$ se naváže na HLA-B*5701 \Rightarrow imunitní reakce
 - $18 \cdot 10^4$ se spojí s HLA-B*0701 \Rightarrow imunitní reakce nenastává
- Využité metody:
 - Artificial neural networks (ANN)
 - Stabilized matrix method (SMM)



Obrázek: Pravděpodobnost, s kterou se repertoár T-buňky spojí určitou interakční silou s peptidy viru. Konkrétní repertoár T-buňky rozpozná peptidu viru, když vazebná síla překročí poznávací práh (přerušovaná čára).

Host-pathogen interaction dynamics

- Koeficient šíření řetězce viru pochází z rovnoměrného rozdělení mezi 0–2 000 nových buněk viru za den.
- Interakce volné energie mezi T-buňkou daného klonotypu a zbytkem viru na určeném epitopu je náhodná veličina z rovnoměrného rozdělení, přičemž šířka tohoto rozdělení určuje, jaká je pravděpodobnost, že daná T-buňka rozpozná peptid viru.
- Rozdělení repertoárů T-buněk se blíží k normálnímu rozdělení interagujících energií.
- Rozdílnosti ve vzájemném působení mezi antigenem a imunitním receptorem jsou kromě virové šířitelnosti zachycené ve veličině σ_{ij} , jejíž rozdělení je zobrazené viz následující obrázek.



Obrázek: Model generování prvků σ_{ij} matice popisující rozpoznání pMHC T-buňkou. Stupeň reaktivity v simulacích závisí na rovnoměrném rozdělení, ze kterého jsou náhodně vybrány interakční síly mezi jednotlivými pozůstatky epitopů a TRC (vpravo).

Děkuji za pozornost.