STATISTICKÁ ANALÝZA PODVĚDOMÉHO JEDNÁNÍ

David Kordek, Pavel Kříž

Univerzita Hradec Králové

Abstrakt

Statistická analýza je hojně používána při studiu jednání lidí. Cílem bylo ukázat, že existuje matematická závislost mezi podvědomým jednáním dvou a více osob. Možností na nichž by se tato závislost dala zkoumat existuje určitě mnoho (např. zívání, mrkání,…). Navrhli jsme proto následující experiment, sledovat chování hostů v restauračních zařízeních při konzumaci nápojů.

1 Úvod

Statistická analýza je hojně používána při studiu jednání lidí. Setkáme se s ní v pracích zabývajících se velkou škálou situací, při nichž je jednání lidí studováno. Od úsměvných studií, např. analýzy změn v jednání osob ve Velké Británii v pátek třináctého, [1], až po studie mnohem vážnější popisující např. závislost mezi sebevražednými myšlenkami a pokusy o sebevraždy mezi čínskými prostitutkami, [2].

Cílem této práce bylo ukázat, že existuje závislost mezi podvědomým jednáním dvou a více osob. Možností na nichž by se tato závislost dala zkoumat existuje určitě mnoho (např. zívání, mrkání,…), tyto možnosti jsou však špatně měřitelné. Navrhli jsme následující experiment, sledovat chování hostů v restauračních zařízeních při konzumaci nápojů. Zajímala nás vždy skupina sedící u téhož stolu. Časy napití jednotlivých osob jsem zaznamenávali pomocí vytvořeného programu do notebooku.

2 Experiment

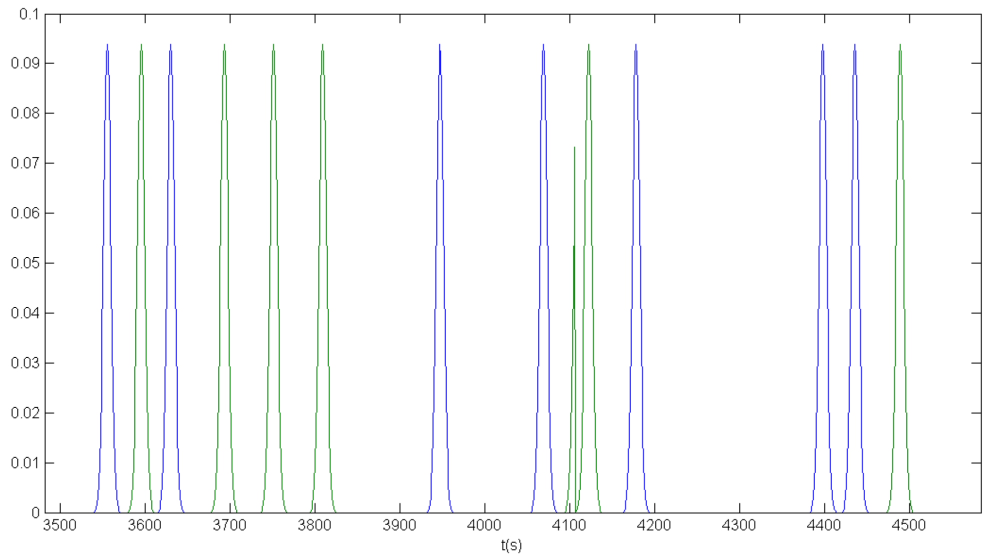
Provedli jsme tři měření v různých restauracích v Hradci Králové. Dvakrát jsme sledovali dvojici osob a jednou trojici osob. Výstupem měření byly konečné vícerozměrné posloupnosti časových údajů, konkrétně dvě dvourozměrné a jedna třírozměrná. Prvky posloupností reprezentují časy napití pozorovaných osob. Označíme-li čas *i*-tého napití *j*-té osoby $\tau_j(i)$, zmíněnou posloupnost tedy zapíšeme, $\{\tau_j(i)\}_{j=1,\dots,N}^{i=1,\dots,T_j}$, kde T_j je celkový počet napití *j*-té osoby a *N* je počet osob.

3 Zpracování dat

Uvažujme, že pokud je časový rozdíl napití dvou osob větší než *h*0, není mezi napitími obou osob žádná souvislost. Pro každou osobu ze skupiny, tj. , definujme na množině reálných čísel funkci

$$\widetilde{x}_j(t) = \sum_{i=1}^{T_j} \exp\left(-\frac{(t-\tau_j(i))^2}{2\sigma^2} \right)$$

kde $\sigma = \frac{h_0}{4\sqrt{2\ln 2}}$. Do každého okamžiku napití jsme tedy umístili vrchol (nenormované) Gaussovy křivky o pološířce $h=h_0/2$, viz obr. 1. K tomuto účelu jsme pomocí programu Matlab vytvořili následující funkci VytorRady na následující stránce.



Obrázek 1: Okamžikům napití odpovídají vrcholy Gaussových křivek

Funkce VytvorRady:

```
function x = VytvorRady(hwidth,varargin);
n = length(varargin);
for j=1:n;
    y0(1:length(varargin{j}),j) = varargin{j};
    y = round(y0*24*60*60);
end
[m n] = size(y);
zacatek = floor(min([y(1,1) y(1,2)])-3*hwidth);
konec = floor(max(max(y))+3*hwidth+1);
x = zeros(konec - zacatek +1,n);
sigma = hwidth/sqrt(2*log(2));
time = floor((-3*hwidth:3*hwidth)+0.5);
for j=1:n;
    for k=1:m;
        if y(k,j) > zacatek
            x(time + y(k,j) - zacatek + 1,j) = normpdf(time,0,sigma);
        end
    end;
end;
```

Oříznutím a diskretizací (s krokem 1s) těchto funkcí získáme N-rozměrnou časovou řadu $\{x_j(i)\}_{j=1,\dots,N}^{i=0,\dots,T}$, kde diskrétní čas $i=0$ odpovídá např. času $t=\min\limits_{j\in\{1,\dots,N\}}\tau_j(1)-3h$ a $i=T$ odpovídá např. času $t=\max\limits_{j\in\{1,\dots,N\}}\tau_j(1)-3h$, jeli h celočíselné (ukládané hodnoty byly celočíselné v sekundách).

Za míru závislosti podvědomého chování j-té a k-té osoby uvnitř sledované skupiny jsme zvolili korelační koeficient funkcí $\widetilde{x}_j$ a $\widetilde{x}_k$, resp. časových řad x_j a x_k,

$$\widetilde{C}_{jk} = \frac{\int_{-\infty}^{\infty} \widetilde{x}_j(t)\,\widetilde{x}_k(t)\,dt}{\sqrt{\int_{-\infty}^{\infty} \left(\widetilde{x}_j(t)\right)^2 dt \int_{-\infty}^{\infty} \left(\widetilde{x}_k(t)\right)^2 dt}} \quad \text{resp. } C_{jk} = \frac{\sum_{i=0}^{T} x_j(i)\, x_k(i)}{\sqrt{\sum_{i=0}^{T} \left(x_j(i)\right)^2 \sum_{i=0}^{T} \left(x_k(i)\right)^2}}$$

Pokud by obě sledované osoby pily synchronně, tj. $\widetilde{x}_j \equiv \widetilde{x}_k$, dosáhl by korelační koeficient $\widetilde{C}_{jk}$ jedné. Uvažujeme-li např. pouze jedno napití každé osoby ze sledované dvojice, vzájemně posunuté o pološířku Gaussovy křivky h, dostaneme přímým výpočtem (s využitím substituce $t = x + h/2$) hodnotu $\widetilde{C}_{jk} = 1/4$.

Nechť naopak platí $\left|\tau_j(i) - \tau_k(l)\right| \geq h_0$. Podle předpokladu není pak pití jedné osoby vůbec ovlivněno pitím osoby druhé a naopak. Odhadněme hodnotu $\widetilde{C}_{jk}$ shora. Označme

$$I = \int_{-\infty}^{\infty} \left(\exp\left(-\frac{t^2}{2\sigma^2} \right) \right)^2 dt .$$

Jmenovatele korelačního koeficientu odhadneme zdola pomocí nerovnosti,

$$\int_{-\infty}^{\infty} \left(\sum_{i=1}^{T_j} \exp\left(-\frac{\left(t - \tau_j(i)\right)^2}{2\sigma^2} \right) \right)^2 dt > \sum_{i=1}^{T_j} \int_{-\infty}^{\infty} \left(\exp\left(-\frac{\left(t - \tau_j(i)\right)^2}{2\sigma^2} \right) \right)^2 dt = T_j I .$$

Čitatele odhadneme shora,

$$\int_{-\infty}^{\infty} \left(\sum_{i=1}^{T_j} \exp\left(-\frac{\left(t - \tau_j(i)\right)^2}{2\sigma^2} \right) \sum_{l=1}^{T_k} \exp\left(-\frac{\left(t - \tau_k(l)\right)^2}{2\sigma^2} \right) \right) dt =$$

$$= \sum_{i,l} \int_{-\infty}^{\infty} \exp\left(-\frac{\left(t - \tau_j(i)\right)^2 + \left(t - \tau_k(l)\right)^2}{2\sigma^2} \right) dt = I \sum_{i,l} \exp\left(-\frac{\left(\tau_j(i) - \tau_k(l)\right)^2}{4\sigma^2} \right) <$$

$$< I \sum_{i,l} \exp\left(-\frac{h_0^2}{4\sigma^2} \right) = I T_j T_k 2^{-8},$$

kde jsme v první rovnosti užili substituci $t = x + \left(\tau_j(i) + \tau_k(l)\right)/2$. Celkově tedy máme odhad $\widetilde{C}_{jk} < \sqrt{T_j T_k}\, 2^{-8}$. Tedy, je-li hodnota korelačního koeficientu větší než tento odhad, existuje závislost mezi pitím obou osob.

Poznamenejme, že výše uvedený horní odhad korelačního koeficientu je velice hrubý. Ve skutečnosti bude časový rozdíl pouze dvou nejbližších napití přibližně h_0, v odhadu počítáme tuto mez pro všechny dvojice napití.

Prokážeme-li výše uvedeným způsobem vzájemnou závislost pití dvou osob, může nás dále zajímat průměrné zpoždění napití osoby pijící pod vlivem napití osoby jiné. Toto můžeme zjistit pomocí normované křížové korelační funkce (viz např. [3], odstavec 9.1),

$$\widetilde{K}_{jk}(t') = \frac{\int_{-\infty}^{\infty} \widetilde{x}_j(t)\, \widetilde{x}_k(t-t')\, dt}{\sqrt{\int_{-\infty}^{\infty} (\widetilde{x}_j(t))^2 dt \int_{-\infty}^{\infty} (\widetilde{x}_k(t))^2 dt}}, \text{ resp. } K_{jk}(l) = \frac{\sum_{i=0}^{T} x_j(i)\, x_k(i-l)}{\sqrt{\sum_{i=0}^{T} (x_j(i))^2 \sum_{i=0}^{T} (x_k(i))^2}}.$$

Je zřejmé, že platí, $\widetilde{K}_{jk}(0) = \widetilde{C}_{jk}$, resp. $K_{jk}(0) = C_{jk}$. Polohy lokálních maxim funkce $\widetilde{K}_{jk}$ odpovídají průměrným zpožděním napití osoby, která pije pod vlivem napití osoby jiné. Konkrétně, lokální maximum odpovídající kladnému času *t'* udává zpoždění pití *k*-té osoby za *j*-tou a lokální maximum odpovídající zápornému času *t'* udává zpoždění pití *j*-té osoby za *k*-tou.

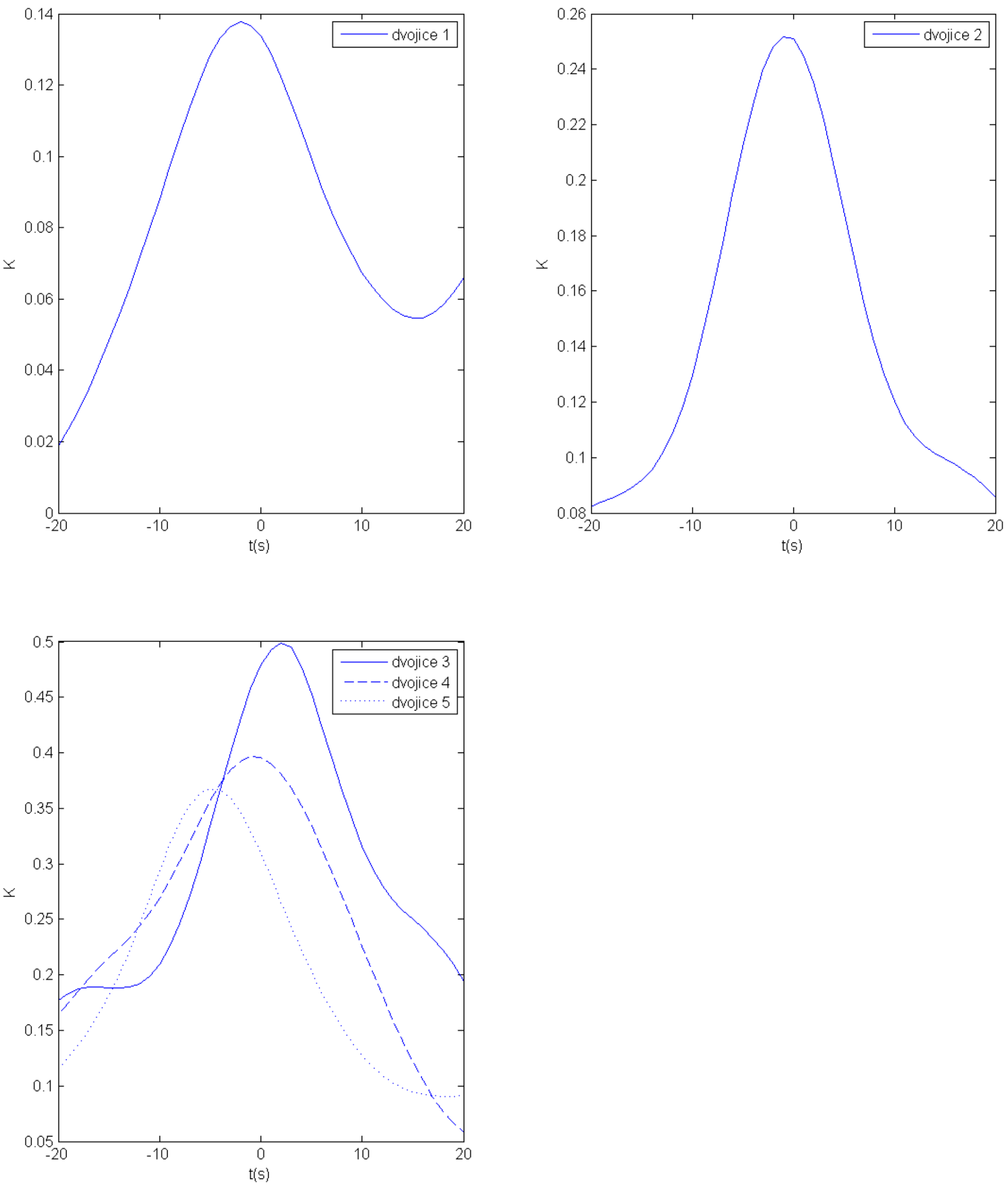
4 Výsledky

Mezní časový rozdíl, kdy už nedochází k ovlivnění napití, jsme zvolili h_0=20 s. Hodnoty korelačního koeficientu C_{jk} uvádíme v 6. sloupci tabulky 1. V pátém sloupci této tabulky je pro porovnání výše odvozená mez závislosti pití $\sqrt{T_j T_k}\, 2^{-8}$. Hodnoty korelačního koeficientu C_{jk} lze také odečíst z grafů na obr. 2, kde je zakreslena závislost křížové korelační funkce K_{jk} na čase (jak bylo uvedeno výše, $K_{jk}(0) = C_{jk}$).

Všechny grafy na obr. 2 mají výrazné lokální maximum v blízkosti času t=0 s. Poloha tohoto maxima udává nejpravděpodobnější zpoždění napití osoby, která podvědomě pije v důsledku napití osoby druhé. V některých případech, dojde k rozpadu zmíněného maxima na dvě, použijeme-li v definici $\widetilde{x}_j$ Gaussovy křivky s menší pološířkou (např. h=2 s), viz obr. 3. Pokud k rozpadu maxima nedojde, lze konstatovat, že jedna z osob pila většinou dříve (*j*-tá osoba ze skupiny pila většinou dříve než *k*-tá osoba, je-li časová souřadnice maxima záporná a naopak, *k*-tá osoba ze skupiny pila většinou dříve než *j*-tá osoba, je-li časová souřadnice maxima kladná). Pokud k rozpadu maxima dojde, získáme nejpravděpodobnější zpoždění napití *j*-té osoby za napitím *k*-té i nejpravděpodobnější zpoždění napití *k*-té osoby za napitím osoby *j*-té. Polohy maxim (ve dvou případech po rozštěpení) jsou v tabulce 1 v 7. sloupci.

Tabulka 1: HODNOTY KORELAČNÍCH KOEFICIENTŮ VŠECH MĚŘENÝCH DVOJIC V POROVNÁNÍ S MEZNÍ HODNOTOU, POLOHY MAXIM KŘÍŽOVÉ KORELACE.

Měření	*Číslo dvojice (obr. 2)*	*Počet napití T_j*	*Počet napití T_k*	*Mezní hodn.* $\sqrt{T_j T_k}\, 2^{-8}$	*Korelační koeficient C_{jk}*	*Poloha maxima $t_{max}(s)$*	
1	1	22	29	0,0987	0,1337	-2	
2	2	24	16	0,0765	0,2509	-3	1
3	3	19	13	0,0614	0,4793	2	
	4	19	19	0,0742	0,3955	-3	5
	5	13	19	0,0614	0,3097	-5	

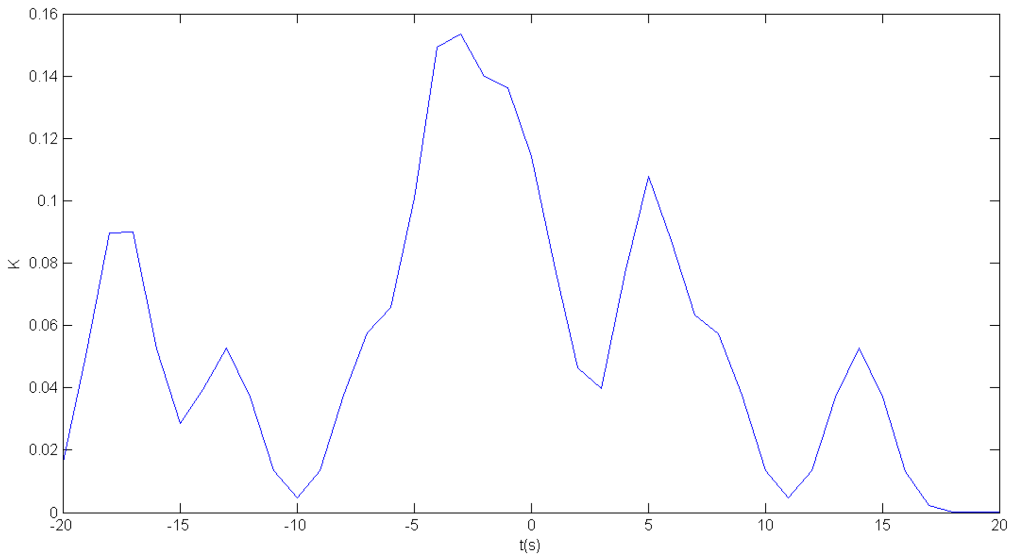


Obrázek 2: Křížová korelační funkce pro všechny měřené dvojice

5 Diskuze a závěr

Jak ukazují výsledky prezentované v tab. 1, nalezli jsme hodnotu korelačního koeficientu vždy vyšší, než byla teoreticky vypočtená mezní hodnota (kdyby měl korelační koeficient právě tuto mezní hodnotu nebo hodnotu nižší, nebyla by závislost pití prokazatelná), kromě jednoho případu (dvojice 1) dokonce několikanásobně vyšší. Lze tedy tvrdit, že lidé ve skupině pijí synchronně, tedy existuje závislost jejich podvědomého chování. Nejpravděpodobnější hodnota zpoždění napití se pohybovala (v absolutní hodnotě) mezi 1 s a 5 s. Volba h_0=20 s, tedy časového rozdílu, při kterém už nelze hovořit o ovlivnění napití, byla tedy přiměřená.

K větší přesvědčivosti našich závěrů bychom potřebovali větší množství experimentálních dat. Jak ale vyplývá z popisu našeho experimentu, získávání dat je velmi časově náročné.



Obrázek 3: Rozpad maxima křížové korelační funkce na dvě (dvojice 4 z obr. 2) při h=2s

Literatura

[1] Scanlon, T. J., Luben, R. N., Scanlon, F. L. and Singleton, N., *Is Friday 13th bad for your health*?, *BMJ* **307** (1993), 1584-1586

[2] Hong, Y., Li, X., Fang, X. and Zhao, R., Correlates of Suicidal Ideation and Attempt Among Female Sex Workers in China, *Health Care Women Int.* **28** (2007), 490–505

[3] Priestly, M. B., *Spectral Analysis and Time Series*, Elsevier Academic Press, London, 1981

Mgr. David Kordek
Katedra fyziky a iformatiky, Pedagogická fakulta
Univerzita Hradec Králové
Rokitanského 62, Hradec Králové, 500 03
david.kordek@uhk.cz

Ing. Pavel Kříž
Katedra informatiky a kvantitativních metod, Fakulta informatiky a managementu
Univerzita Hradec Králové
Rokitanského 62, Hradec Králové, 500 03
pavel.kriz@uhk.cz