Milí studenti,

projděte si nejprve soubor geost_7_odpovedi.pdf. Výsledky nových úkolů vložte do elaborátu.

S pozdravem,

JJ

Statistický popis prostorové veličiny

V předchozí lekci jsme se seznámili s nástroji pro popis prostorové veličiny, s variogramem a kovariancí. Náš přístup byl intuitivní a nevycházel z žádné teorie.

Statistické metody, mezi než geostatistika patří, mají svoji teorii, která je založena na počtu pravděpodobnosti. Abychom alespoň částečně geostatistiku pochopili, musíme se seznámit s některými základními pojmy a představami, jako je náhodná funkce nebo stacionarita. Opřeme se při tom o znalosti, které byste měli mít ze základního kurzu statistiky. Půjdeme podle textu GPI s ukázkami simulací v Matlabu.

V textu GPI si nejprve přečtěte začátek kapitoly Prostorová proměnná, stránky 71-73 až po obrázek 6.1.

K ilustraci poslouží následující ukázky, které si vyrobíte sami v Matlabu.

Úkol U1:

Příkazem `x=rand(1,100)` lze vygenerovat 100 hodnot náhodné proměnné, která má *rovnoměrné rozdělení* v intervalu (0,1). Je to její „realizace“ (náhodný vzorek). Příkaz `hist(x)` vykreslí její histogram.

Vyrobte funkci, kde se budou v cyklu opakovat tyto dva kroky a pozorujte tvar výsledného histogramu.

Spusťte stejnou simulaci pro n=1000 a 10000.

Pozn.: Před konec cyklu (end) je vhodné dát příkaz pause(0.1), jak jsme to již použili vícekrát. Matlab se na chvilku pozdrží a ukáže aktuální výsledky a obrázek.

Funkci a případně některou ukázku simulace mi vložte do elaborátu. Napište, co znamená, že náhodná veličina má rovnoměrné rozdělení.

U2:

Příkazem `x=randn(1,100)` lze vygenerovat 100 hodnot náhodné proměnné, která má *normální rozdělení* s nulovou střední hodnotou ($\mu = 0$) a jednotkovou směrodatnou odchylkou ($\sigma = 1$). Označujeme ho $\mathrm{N}(0,1)$.

Vyrobte podobnou funkci jako v úkolu 1, kde budete simulovat náhodný výběr z normálního rozdělení (n=100, 1000, 10000).

Patrně již víte, že příkazy mean(x) , var(x) a std(x) spočtou průměr, rozptyl a směrodatnou odchylku.

Jelikož jsou spočteny z náhodného výběru, říká se těmto charakteristikám výběrové. Pozorujte, jak při simulaci vycházejí blízko kolem zadané střední hodnoty, rozptylu a směrodatné odchylky (0 a 1).

Změňte funkci tak, že bude generovat hodnoty výšky náhodně vybraných dospělých lidí (například v=m+s*randn(1,100), kde m=175, s=7) .

Pošlete mi výsledný histogram některé simulace. Vysvětlete, proč se v literatuře označuje normální rozdělení jako $\mathrm{N}(\mu, \sigma^2)$.

U3:

Kovarianci 2 proměnných x a y spočteme tak, že použijeme matlabovskou funkci cov, která počítá kovarianční matici C a z ní si vezmeme člen mimo diagonálu: `C=cov(x,y)`, `c=C(1,2)`.Malé písmeno c je tedy hodnota kovariance (na diagonále matice C jsou rozptyly proměnných x a y).

Ze základního kurzu statistiky víme, že kovariance 2 nezávislých náhodných proměnných by měla být rovna nule. Ověříme si to simulací. V cyklu generujte x=randn(1,100) a y=randn(1,100), vykreslujte bodový graf (příkaz `plot(x,y,'.')`) a vypisujte hodnotu kovariance.

Do elaborátu popište svoje pozorování a vysvětlete, proč nevychází „čistá“ 0.

U4:

Jinak je tomu u kovariance 2 závislých proměnných. Abychom to ilustrovali, vyrobíme si proměnnou z=x+y. Díky této volbě je proměnná z závislá na x i y. Tedy kovariance z a x i kovariance z a y by měly vycházet výrazně nenulové. Ověřte si to pomocí simulace podobné předchozímu úkolu.

Do elaborátu napište v jakých jednotkách bude kovariance dvou veličin x a y, když jejich jednotky jsou metry a kilogramy, když se například jedná o výšku a váhu člověka. Jak se změní hodnota kovariance, když výšku udáme v cm?

Nyní si znovu prohlédněte obr. 6.1 a dočtěte až na stránku 79 nahoře.

Náhodné funkce

Dále postupně projdeme příklady náhodných funkcí v textu, str. 79-85. Využijeme 2 matlabovské funkce z minulé lekce, `variogram` a `kovariance`, které umějí zpracovat 1D ekvidistantní data.

Příklad 1

Přečtěte si odpovídající text a v Matlabu spusťte příkazy

```
z=randn(1,100);
g=variogram(z,30);
c=kovariance(z,10);
```

Příklad 2

Čtěte dále a prohlédněte si vzorec (6.10) pro vážený průměr. Pak funkci `wma`, kterou vám zasílám (název je zkratkou weighted moving average) a zjistěte, jak je v ní vzorec naprogramován.

Datový vektor, jehož hodnoty jsou vzájemně nekorelované vyrobí odpovídající kovarianci

```
z=randn(1,100);  c=kovariance(z,30);
```

Pomocí funkce wma vyrobíme korelovaná data

```
zs=wma(z,[1 1 1]/3);
```

Zadali jsme váhy 1/3,1/3,1/3 a výsledek zapsali do proměnné zs. Nyní spočteme její kovarianci

```
c=kovariance(zs,30);
```

Opakovaně simulujte data z, počítejte kovarianci c a porovnávejte s obr. 6.7 nahoře.

U5:

Kontrolní otázka 1: Data zs jsme vyrobili jako vážený průměr z náhodné sekvence. Jak souvisí délka klesání kovariance u počátku s počtem průměrovaných datových hodnot (s počtem vah)?

Kontrolní otázka 2: Proč je na 9. řádku wma příkaz `zs=mean(z)*ones(1,n);`?

Kontrolní otázka 3: Jak lze pomocí vhodné volby vah napodobit obr. 6.7 dole?

Proveďte podobné simulace s variogramem.

z=randn(1,100); g=variogram(z,30); zs=wma(z,[1 1 1]/3); g=variogram(zs,30);

U6:

Kontrolní otázka 4: Data zs jsme vyrobili jako vážený průměr z náhodné sekvence. Jak souvisí délka stoupání variogramu u počátku s počtem průměrovaných datových hodnot (s počtem vah)?

Příklad 3

Spusťte zaslanou funkci `randwalk`, spusťte ji, spočtěte variogram z výstupního datového vektoru a porovnejte s obr. 6.8. Tím jsme vyrobili jednu složku v levém grafu obr. 6.9.

U7:

Zkuste funkci doplnit tak, aby vygenerovala ten pravý graf, tj. náhodnou procházku ve 2D. Pokud se to podaří, kód funkce mi vložte do elaborátu. Pokud ne, odpovězte alespoň na následující otázku.

Kontrolní otázka 5: Jaká je hodnota parametru sill (česky práh) variogramu proměnné, kterou lze modelovat jako náhodnou procházku?

Nyní už můžeme dočíst kapitolu do str. 87.

Po krátkém oddechu (kafe :)) si projděte kapitolu Modely variogramu, 89-105.

Tato pasáž je náročná a bez předběžných teoretických znalostí není možno ji pochopit do hloubky. Zůstaneme tak trochu „na povrchu“. Zdá se mi vhodné, abyste tuto kapitolu prošli teď hned.

Je možné, že v součtu s předchozí prací tím trochu přeženeme časovou náročnost. V tom případě mi dejte vědět a já to zohledním v příští přednášce.

Jistě budou nějaké dotazy. Zašlete mi je spolu s odpověďmi na kontrolní otázky.

Další lekce opět příští pondělí.

Hezký den,

JJ