

Seminář Matematické metody v praxi

LS 2021/2022

Volitelný zápočtový úkol: Bootstrap

(Submission Deadline: 30.06.2022 | Koniec skúškového obdobia)

Všeobecné informácie

- ☐ Vypracovanie aspoň jedného úkolu (výber konkrétneho úkolu z niektorého tematického okruhu je na princípe individuálnej voľby) je v prípade nekompletnej účasti na seminároch nutnou podmienkou k získaniu zápočtu.
- ☐ V prípade úkolu z teoretického okruhu *bootstrap* je potrebné tento úkol vyracovať pomocou ľubovolnej mince (prvá časť) a následne pomocou vhodného (ľubovlného) softwaru (druhá časť týkajúca sa resamplingu). Prvú časť úkolu (fyzické hádzanie mincou) je možné nahradiť simuláciou pomocou vhodného generátora pseudo-háhodných čísel.
- ☐ V prípade, že je pre vypracovanie úkolu zvolený štatistický program R, je možné celý úkol vypracovať pomocou príkazov, ktoré sú súčasťou štandardnej inštalácie. Program je možné stiahnuť (GNU free of charge public license) a nainštalovať podľa inštrukcii na stránke
<https://cran.r-project.org/index.html>
- ☐ Vypracovanie úkolu je potrebné zaslať emailom doc.Michalovi Peštovi, t.j. na email
pesta[AT]karlin.mff.cuni.cz

Zadanie úkolu

1. V prvej časti úkolu využijte ľubovolnú mincu (nie nutne symetrickú, resp. spravodlivú) a 100-krát ju po sebe (nezávisle) vyhoďte a nechajte dopadnúť—pričom zakaždým sledujete/zaznamenáte výsledok. Celkovým výstupom bude teda napr. postupnosť čísel 0 a 1 s celkovou dĺžkou 100 znakov.
2. Spočítajte výberový priemer (resp. relatívny počet jednotiek v postupnosti o dĺžke 100 znakov), ktorý je možné interpretovať aj ako odhadnutú pravdepodobnosť, že minca spadne po vyhodení na stranu, ktorú označujeme/kódujeme hodnotou jedna.
3. V druhej časti úkolu urobíme tzv. bootstrapový resampling—t.j., z pôvodnej postupnosti núl a jednotiek (o celkovej dĺžke 100 znakov) vytvoríme novú postupnosť (tzv. bootstrapovú verziu) a to postupným (náhodným a nezávislým) vyberaním núl a jednotiek (s vracaním), až kým nedostaneme novú postupnosť o celkovej dĺžke 100. V novej (bootstrapovej) postupnosti opäť spočítame výberový priemer a hodnotu zaznamenáme/uložíme.

Celý proces bootstrapového resamplingu (t.j. krok 3) je nutné opakovať dostatočne veľa krát (napr. 200, alebo až 1000 bootstrapových postupností, každá o dĺžke 100 znakov, pričom z každej postupnosti spočítame výberový priemer). Práve z dôvodu veľkého počtu bootstrapových výberov (t.j., nových bootstrapových postupností) je nutné využiť výpočetné možnosti počítačov a vhodného softwaru.
4. Výberové priemery spočítané na veľkom množstve bootstrapových postupností následne využijeme k tomu, aby sme odhadli rozptyl (resp. rozdelenie) pôvodného výberového priemeru, ktorý sme získali hádzaním mince.

Hodnoty bootstrapových priemerov môžeme napríklad zoradiť od najmenšieho po najväčší a následne z takto zoradenej postupnosti (200, prípadne až 1000 hodnôt) vybrať tzv. výberové kvantily, ktoré určujú príslušný interval spoľahlivosti pre neznámu pravdepodobnosť toho, že po vyhodení spadne minca na stranu označenú hodnotou 1.
5. V prípade, že hodnota 0.5 (to znamená, že minca je spravodlivá) je uvnitř získaného intervalu, môžeme udeľovať záver, že získané data nenaznačujú problém s férovosťou danej mince. V prípade, že získaný interval hodnotu 0.5 nepokrýva, lze tvrdiť, že máme štatistický preukázateľný dôvod tvrdiť, že mince není spravodlivá.