

KOLOKVIJ 2 IZ RAČUNARSKE STATISTIKE, GRUPA
1

PMF-Matematika, Školska godina 2015/2016

Predavač: Doc.dr.sc. Vesna Lužar-Stiffler

Datum: _____

Ime, Prezime: _____

Potpis: _____

NAPOMENA: Rješenja kolokvija (sas programe)
poslati na email adresu:

lubura.snjezana@gmail.com

1. Monte Carlo metodom procijenite slijedeći određeni integral:

$$I = \int_0^1 \tanh(x) dx$$

Koristite slijedeće Importance Sampling (IS) funkcije (tj. funkcije distribucija):

- a) Uniformnu na intervalu 0,1 (U(0,1))
- b) Beta na intervalu (0,1), uz vrijednosti parametara a=1, b=2
- c) Beta na intervalu (0,1), uz vrijednosti parametara a=1, b=1.2
- d) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.01
- e) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.001
- f) Podrezanu („truncated“) normalnu distribuciju sa vrijednostima parametara $\mu=0$, $\sigma=0.6$ (N(0,0.6)) , podrezanu („truncated“) na intervalu (0,1).

Za svaku importance sampling funkciju generirajte 4000 replikacija/ponavljanja.

Koristite SEED= 557338 (za sve importance sampling funkcije).

U svakom ponavljanju izračunajte kumulativni prosjek i kumulativnu standardnu pogrešku.

Odredite procjene standardne pogreške za metode a)-f) nakon 4000 replikacija, te efikasnosti ($\text{var}_a/\text{var}_i$, $i=b,c,\dots,f$) metoda b)-f) u odnosu na a).

UPUTA ZA f: Kod generiranja po podrezanoj normalnoj distribuciji zadržavaju se samo one generirane vrijednosti koje su u intervalu (0,1). Funkciju gustoće vjerojatnosti takve „truncated“ normalne distribucije na (0,1) intervalu potrebno je prilagoditi kako bi zadovoljavala svojstva funkcije gustoće vjerojatnosti. Koristite funkcije PROBNORM i PDF.

UPUTA ZA IZRAČUNAVANJE PODINTEGRALNE FUNKCIJE: Koristite SAS funkcije TANH i SQRT

Upišite rezultate zaokružene na 3 decimalna mjesta:

Dio	Procjena integral	Procjena standardne pogreške	Procjena efikasnosti
a)			
b)			
c)			
d)			
e)			
f)			

2. U data setu UCENICI su zapisane prosječne ocjene učenika dvaju razreda (A i B).

```
data UCENICI;
input RAZRED $ OCJENA @@;
datalines;
A 4.4 A 4.8 A 3.8 A 4.2 A 4.7 A 4.6
B 5.0 B 4.7 B 4.8 B 4.9 B 5.0 B 4.5
;
```

- a) Testirajte hipotezu $H_0: \mu_A = \mu_B$ (da je prosječna OCJENA jednake bez obzira o kojem se razredu radi (A ili B)) nasuprot $H_1: \mu_A < \mu_B$ (da je prosječna OCJENA u razredu A manja nego u razredu B na razini statističke značajnosti $\alpha = 0.05$ primjenom t testa za 2 nezavisna uzorka.

UPUTA: Koristite proceduru TTEST. Prvo testirajte hipotezu o jednakosti varijanci na razini statističke značajnosti $\alpha=0.05$, te u zavisnosti od ishoda, primijenite ili „pooled“ ili „Satterthwaite“ test za hipotezu $H_0: \mu_A = \mu_B$.

Može li se hipoteza H_0 odbaciti na razini statističke značajnosti $\alpha = 0.05$?

Upišite rezultate zaokružene na 2 decimalna mjesta:

t-vrijednost (za H_0)	DF (br.stupnjeva slobode)	p-vrijednost za dvostranu hipotezu	p-vrijednost za jednostranu hipotezu

- b) Testirajte hipotezu $H_0: \mu_A = \mu_B$ nasuprot $H_1: \mu_A < \mu_B$ na razini statističke značajnosti $\alpha = 0.05$ primjenom bootstrap testa.

UPUTA: Koristite metodu B bootstrap uzorkovanja za testiranje hipoteza. Primijenite slijedeće vrijednosti makro varijabli SEED:

```
%let seed=34567; *za grupu 1 (razred="A");
%let seed= 76543; *za grupu 2 (razred="B");
```

Za svaku grupu izvedite po 500 bootstrap ponavljanja (replikacija). Procijenite p-vrijednosti Bootstrap testa za dvostranu i jednostranu hipotezu, te procijenite 95% interval pouzdanosti za p-vrijednosti (UPUTA: koristite proceduru FREQ i odaberite u rezultatima „Exact Conf Limits“.)

Može li se hipoteza H_0 odbaciti na razini statističke značajnosti $\alpha = 0.05$ (uzimajući u obzir 95% interval pouzdanosti za p-vrijednost)?

Upišite rezultate zaokružene na 3 decimalna mjesta:

	Bootstrap test za dvostranu hipotezu	Bootstrap test za jednostranu hipotezu H_0 : $\mu_A = \mu_B$ nasuprot H_1 : $\mu_A < \mu_B$
p-vrijednost		
95% interval pouzdanosti za p- vrijednost (donosi dodatne bodove)		

3. Koristite podatke iz zadatka 2.

a) Parametarskom bootstrap metodom odredite bootstrap standardnu pogrešku i 90% bootstrap interval pouzdanosti za **t vrijednost** testa za testiranje razlike između prosječnih vrijednosti varijable OCJENA za razred A i prosječnih vrijednosti varijable OCJENA za razred B.

$$(t=(\text{meanA}-\text{meanB})/\sqrt{(\text{stdeA}^2 + \text{stdeB}^2)},$$

gdje su meanA i stdeA prosječna vrijednost i standardna pogreška za razred A, a meanB i stdeB prosječna vrijednost i standardna pogreška za razred B).

b) Parametarskom bootstrap metodom odredite bootstrap standardnu pogrešku i 90% bootstrap interval pouzdanosti za apsolutnu vrijednost razlike između medijana varijable OCJENA za razred A i medijana varijable OCJENA za razred B.

U a) i b):

Primijenite slijedeće vrijednosti makro varijabli SEED:

```
%let seed1= 23456; *za grupu 1 (razred="A");
```

```
%let seed2= 65432; *za grupu 2 (razred="B");
```

Koristite proceduru RAND(„NORMAL“) i naredbu call streaminit(&seed)

Pretpostavite da varijabla OCJENA slijedi normalnu distribuciju sa sredinama i standardnim devijacijama procijenjenim iz podataka.

Za svaku grupu izvedite 500 bootstrap ponavljanja (replikacija).

UPUTA za a):

Koristite proceduru MEANS :

```
proc means data=imeulaznogdataseta nway noprint;
  var imevarijable;
  class rep imegrupnevarijable;
  output out=imeizlaznogdataseta mean=mean stderr=stderr;
run;
```

NAPOMENA: Nakon izvođenja procedure MEANS će u izlaznom datasetu („imeizlaznogdataseta“) prosječna vrijednost će biti spremljena kao varijabla sa imenom MEAN, a standardna pogreška kao varijabla sa imenom STDERR.

UPUTA za b):

Koristite proceduru MEANS :

```
proc means data=imeulaznogdataseta nway noprint;
  var imevarijable;
  class rep imegrupnevarijable;
  output out=imeizlaznogdataseta median=medijan;
run;
```

NAPOMENA: Nakon izvođenja procedure MEANS će u izlaznom datasetu („imeizlaznogdataseta“) medijan biti spremljen kao varijabla sa imenom MEDIJAN.

Upišite rezultate zaokružene na 3 decimalna mjesta:

Dio	Bootstrap procjena sredine	Bootstrap procjena standardne pogreške	Bootstrap procjena 90% intervala pouzdanosti
a)			
b)			