

## KOLOKVIJ 2 IZ RAČUNARSKE STATISTIKE

PMF-Matematika, Školska godina 2011/2012

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

Datum: \_\_\_\_\_

Ime, Prezime: \_\_\_\_\_

Potpis: \_\_\_\_\_

1. Monte Carlo metodom procijenite slijedeći određeni integral:

$$I = \int_0^1 x^{2.5} e^{-x/4} dx$$

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 vrijednost parametara a=2.5, b=1
- c) Beta na intervalu (0,1), uz vrijednost parametara a=2, b=2
- d) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.5
- e) Triangularnu na intervalu (0,1), uz vrijednost parametra h=0.9
- f) Gamma distribuciju („truncated“) na intervalu (0,1), uz vrijednost parametra a=2

Za svaku importance sampling funkciju generirajte 10000 replikacija/ponavljanja.

Koristite SEED= 4772565 (za sve importance sampling funkcije).

U svakom ponavljanju izračunajte kumulativni prosjek i kumulativnu standardnu pogrešku.

Odredite procjene standardne pogreške za metode a)-f) nakon 10000 replikacija, te efikasnosti metoda b)-f) u odnosu na a).

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

2. Izmjerena je količina šećera (u %) sadržana u različitim brandovima žitarica za djecu i za odrasle. Podaci za žitarice za djecu su zapisani u datasetu `zitarice_za_djecu`, a za odrasle u datasetu `zitarice_za_odrasle`.

```
data zitarice_za_djecu;  
  input secer @@;  
  datalines;  
40.3 55 45.7 43.3 50.3 45.9 23.5 43 44.2 30.3  
;  
data zitarice_za_odrasle;  
  input secer @@;  
  datalines;  
20 30.3 25.2 74.5 4.4 22.2 16.6 14.5 21.4 33.3 6.6 37.8 10.6 16.2 44.1  
;  
run;
```

- a) Testirajte hipotezu  $H_0: \mu_D = \mu_O$  (da je količina šećera (varijabla SECER) u žitaricama za djecu jednaka količini šećera u žitaricama za odrasle) nasuprot  $H_1: \mu_D > \mu_O$  (da je količina šećera (varijabla SECER) u žitaricama za djecu veća od količine šećera u žitaricama za odrasle) na razini statističke značajnosti  $\alpha = 0.01$  primjenom t testa za 2 nezavisna uzorka.

UPUTA: Koristite proceduru TTEST. Prvo testirajte hipotezu o jednakosti varijanci na razini statističke značajnosti  $\alpha=0.10$ , te u zavisnosti od ishoda, primijenite ili „pooled“ ili „Satterthwaite“ test za hipotezu  $H_0: \mu_D = \mu_O$ .

Može li se hipoteza  $H_0$  odbaciti na razini statističke značajnosti  $\alpha = 0.01$ ?

- b) Testirajte hipotezu  $H_0: \mu_D = \mu_O$  nasuprot  $H_1: \mu_D > \mu_O$  na razini statističke značajnosti  $\alpha = 0.01$  primjenom bootstrap testa.

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

```
%let seed=87654; *za grupu 1 (Djeca);  
%let seed= 58667; *za grupu 2 (Odrasli);
```

Za svaku grupu izvedite po 1000 bootstrap ponavljanja (replikacija).

Može li se hipoteza  $H_0$  odbaciti na razini statističke značajnosti  $\alpha = 0.01$ ?

3. Koristite podatke iz zadatka 2.

a) Parametarskom bootstrap metodom odredite bootstrap standardnu pogrešku i 90% bootstrap interval pouzdanosti za apsolutnu vrijednost razlike između prosječnih vrijednosti varijable SECER za žitarice za djecu i prosječnih vrijednosti varijable SECER za žitarice za odrasle.

b) Parametarskom bootstrap metodom odredite bootstrap standardnu pogrešku i 90% bootstrap interval pouzdanosti za apsolutnu vrijednost razlike između 10-postotne podrezane sredine (engl. 10% trimmed mean) varijable SECER za žitarice za djecu i 10-postotne podrezane sredine varijable SECER za žitarice za odrasle.

U a) i b):

Primijenite slijedeće vrijednosti makro varijabli SEED:

%let seed= 58849; \*za grupu 1 (djeca);

%let seed= 28566; \*za grupu 2 (odrasli);

Pretpostavite da varijabla SECER slijedi normalnu distribuciju.

Za svaku grupu izvedite 1000 bootstrap ponavljanja (replikacija).

UPUTA za b):

Umjesto procedure MEANS, koristite procedure UNIVARIATE i SORT:

```
proc sort data= imeulaznogdataseta;  
  by rep imegrupnevarijable;  
run;  
  
ods select trimmedmeans;  
ods output TrimmedMeans=imeizlaznogdataseta;  
proc univariate data=imeulaznogdataseta trimmed=0.10;  
  var imevarijable;  
  class imegrupnevarijable;  
  by rep;  
run;
```

NAPOMENA: Nakon izvođenja procedure UNIVARIATE će u izlaznom datasetu („imeizlaznogdataseta“) 10-postotna podrezana sredina biti spremljena kao varijabla sa imenom MEAN.