

Uroš Bojanić
2019/0077 (IR)

LABORATORIJSKA VEŽBA IZ VEROVATNOĆE I STATISTIKE

Maj 2021, Beograd

ZADATAK 7.

7 (glava 2-primer 42 i glava 3 primer 53) Bernulijevi opiti- pročitati tekst primera 42.

Zadatak: Napraviti program koji za dato n i p simulira n Bernulijevih opita sa verovatnoćom uspeha p , i ponavlja simulaciju sa istim parametrima N puta. Ulazne veličine su n, p, N a izlaz iz programa je vektor dužine $n + 1$ sa relativnim frekvencijama mogućih vrednosti broja uspeha (od 0 do n). Za $n = 12$, $p = 0.4$ i $N = 100$ nacrtati histogram i uporediti sa tačnim verovatnoćama. Ponoviti isto za $N = 1000$.

Odbrana: Prikazati program i rezultate i komentarisati odstupanja od binomne raspodele.

BERNULIJEVI OPITI

```
def BernoullisTrials(n, p, N):  
    f = np.zeros(n + 1)  
    for simulation in range(0, N):  
        countSuccessfulTrials = 0  
        for trial in range(0, n):  
            if np.random.rand() < p:  
                countSuccessfulTrials += 1  
        f[countSuccessfulTrials] += 1  
    return f / N
```

- Funkcija `BernoullisTrials` napisana na programskom jeziku *Python* izvršava N simulacija. Svaka simulacija se sastoji od n Bernulijevih opita – n puta se generiše nasumičan broj koji pripada segmentu $[0,1]$ i njegova vrednost se upoređuje sa zadatom verovatnoćom uspeha p .
- Ukoliko je nasumično generisan broj manji od p , opit smatramo uspešnim.
- Povratna vrednost funkcije je vektor dužine $n + 1$ koji sadrži relativne frekvencije mogućih vrednosti broja uspešnih opita prilikom simulacije (od 0 do n).
- Za generisanje slučajnih brojeva koristimo biblioteku `numpy`.

BINOMNA RASPODELA

```
def BinomialDistribution(n, p):  
    f = np.arange(n + 1) * 1.0  
    for x in f.astype(int):  
        f[x] = math.comb(n, x) * (p**(x))  
                * ((1-p)**(n-x))  
    return f
```

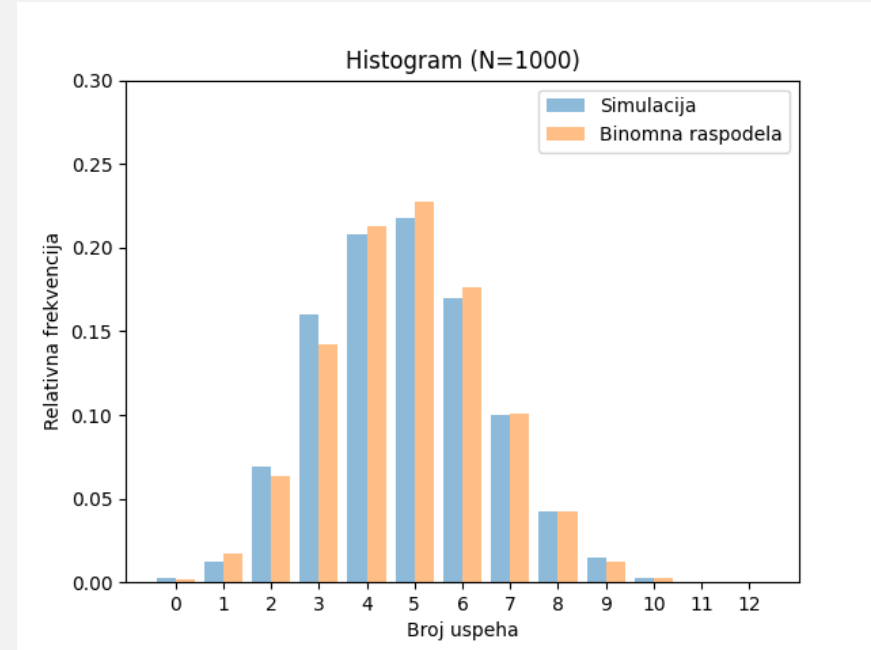
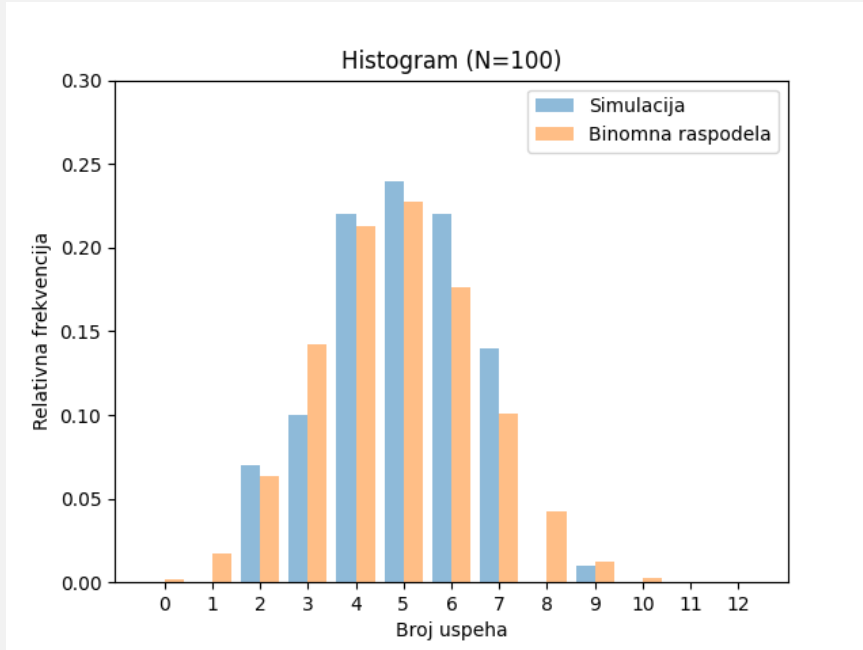
- Funkcija `BinomialDistribution` napisana na programskom jeziku *Python* izračunava tačne verovatnoće mogućih vrednosti broja uspešnih opita, na osnovu formule:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

za $k = 0, 1, \dots, n$.

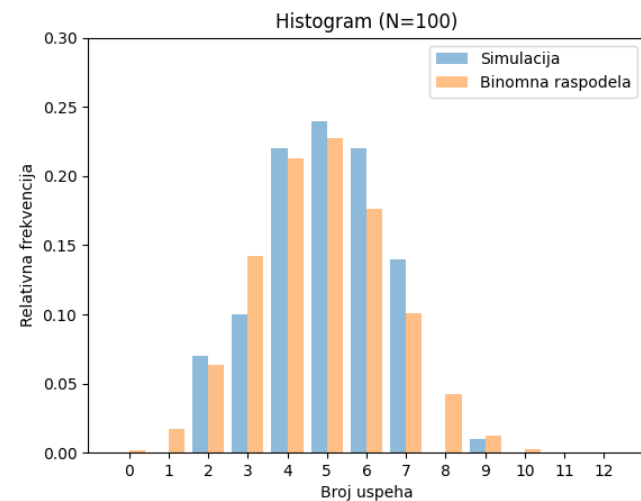
- Za računanje binomnih koeficijenata koristimo biblioteku `math`.

REZULTATI

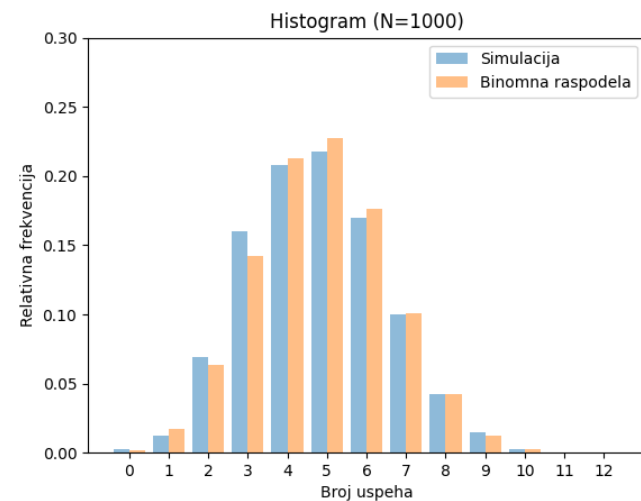


- Glavna funkcija našeg programa poziva simulaciju dva puta – jednom sa 100 iteracija (histogram levo), i jednom sa 1000 iteracija (histogram desno). Na histogramima su prikazane relativne frekvencije dobijene simulacijom Bernulijevih opita (plavo), kao i tačne verovatnoće (narandžasto). Kao što je očekivano, rezultat simulacije liči na binomnu raspodelu u oba slučaja, pri čemu je sličnost veća za veći broj iteracija (1000).
- Histogrami su iscrtani pomoću biblioteke [matplotlib](#).

GREŠKA



Broj uspešnih opita	Relativna frekvencija	Tačna verovatnoća	Apsolutna greška
0	0.00000	0.00218	0.00218
1	0.00000	0.01741	0.01741
2	0.07000	0.06385	0.00615
3	0.10000	0.14189	0.04189
4	0.22000	0.21284	0.00716
5	0.24000	0.22703	0.01297
6	0.22000	0.17658	0.04342
7	0.14000	0.10090	0.03910
8	0.00000	0.04204	0.04204
9	0.01000	0.01246	0.00246
10	0.00000	0.00249	0.00249
11	0.00000	0.00030	0.00030
12	0.00000	0.00002	0.00002



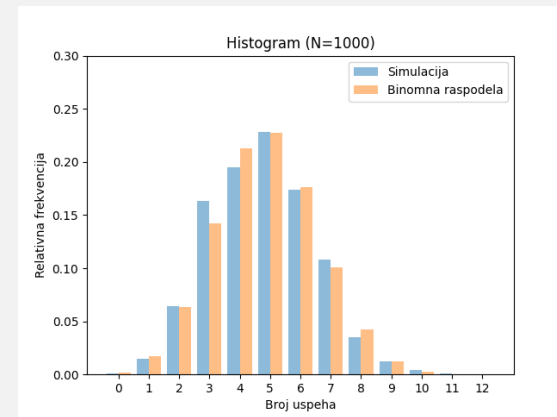
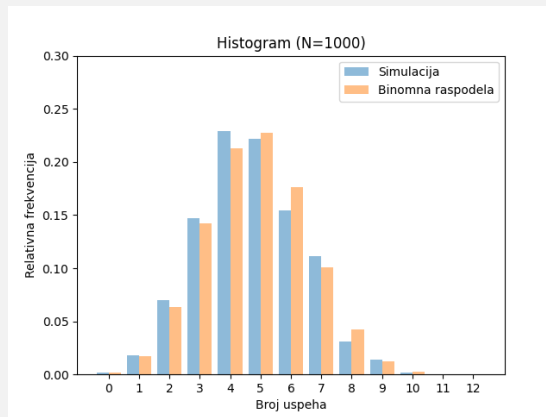
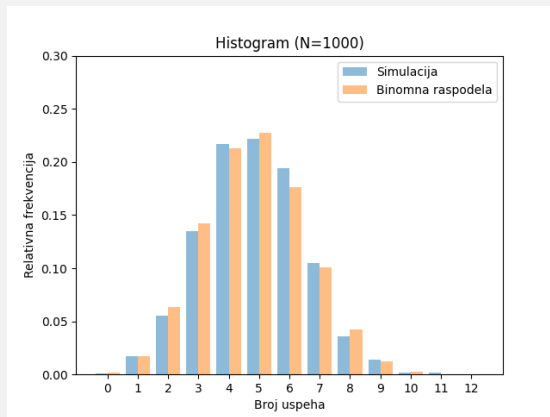
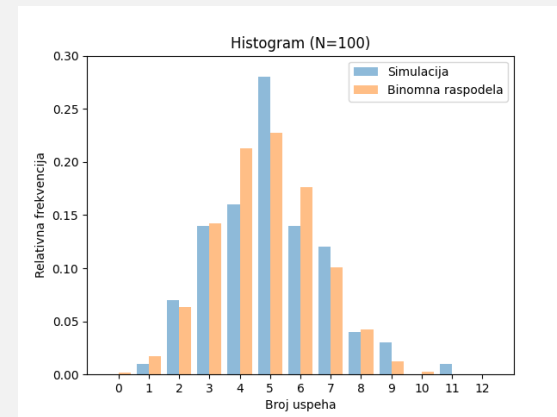
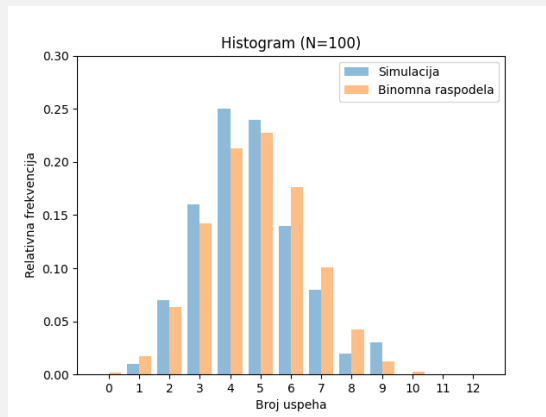
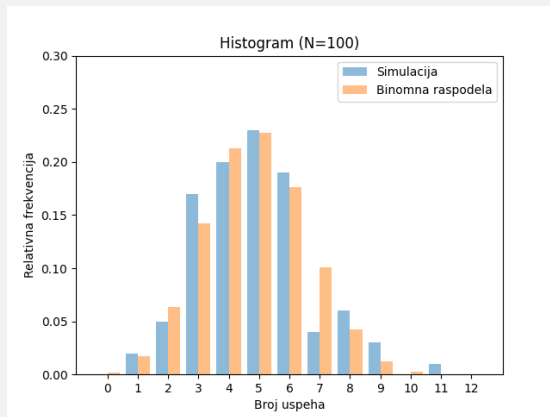
Broj uspešnih opita	Relativna frekvencija	Tačna verovatnoća	Apsolutna greška
0	0.00300	0.00218	0.00082
1	0.01200	0.01741	0.00541
2	0.06900	0.06385	0.00515
3	0.16000	0.14189	0.01811
4	0.20800	0.21284	0.00484
5	0.21800	0.22703	0.00903
6	0.17000	0.17658	0.00658
7	0.10000	0.10090	0.00090
8	0.04200	0.04204	0.00004
9	0.01500	0.01246	0.00254
10	0.00300	0.00249	0.00051
11	0.00000	0.00030	0.00030
12	0.00000	0.00002	0.00002

GREŠKA

Broj uspešnih opita	Apsolutna greška (N=100)	Apsolutna greška (N=1000)
0	0.00218	0.00082
1	0.01741	0.00541
2	0.00615	0.00515
3	0.04189	0.01811
4	0.00716	0.00484
5	0.01297	0.00903
6	0.04342	0.00658
7	0.03910	0.00090
8	0.04204	0.00004
9	0.00246	0.00254
10	0.00249	0.00051
11	0.00030	0.00030
12	0.00002	0.00002
Koren srednje vrednosti kvadrata greške (RMSE)	0.02404	0.00644

- Na prethodnom slajdu su prikazane relativne frekvencije dobijene simulacijom i tačne verovatnoće (binomna raspodela) za sve moguće vrednosti broja uspešnih opita (0 do 12), kao i apsolutna greška za $N = 100$ (gornja tabela) i $N = 1000$ (donja tabela).
- Greška se smanjuje sa povećanjem broja iteracija, što je i bilo očekivano.

PONOVLJENE SIMULACIJE



- Program je pokrenut još tri puta kako bismo utvrdili prethodna zapažanja.
- Naravno, rezultati u sva tri pokretanja su različiti, ali primećujemo slično ponašanje kao i u prethodnim simulacijama.

PONOVLJENE SIMULACIJE - GREŠKA

Simulacija br.	RMSE (N=100)	RMSE (N=1000)
0	0.02404	0.00645
1	0.02109	0.00639
2	0.01871	0.00914
3	0.02452	0.00823

- Histogrami relativnih frekvencija mogućih vrednosti broja uspeha liče na binomnu raspodelu pri svakom pokretanju programa, a greška opada sa povećanjem broja iteracija N .