# LABORATORIJSKA VEŽBA IZ VEROVATNOĆE I STATISTIKE

Ime i prezime: Rastko Lazarević Datum: 10.6.2024

Indeks: 2022/0247 prof. dr Bojana Mihailović

**Zadatak broj 7**

**7.** Kutija sadrži 90 ispravnih i 10 neispravnih proizvoda. Na slučajan način biraju se odjednom dva proizvoda. Neka je slučajna promenljiva $X$ – broj neispravnih proizvoda u dva izabrana. Naći zakon raspodele slučajne promenljive $X$ i očekivanje $EX$.

Napisati program koji generiše $n$ ($n > 30$) slučajnih promenljivih $X_1, \ldots, X_n$ sa ovom raspodelom. Dobijeni niz $X_1, \ldots, X_n$ posmatrati kao slučajni uzorak iz nepoznate raspodele čiji je zakon raspodele $Bin(2, p)$, $0 < p < 1$ raspodele sa nepoznatim parametrom *p*. Na osnovu ovog uzorka:
1) oceniti nepoznati parametar *p* metodom maksimalne verodostojnosti, pa na osnovu te ocene, naći ocenu za $EX$
2) primenom centralne granične teoreme naći aproksimativni 95% interval poverenja za nepoznato *p*, pa na osnovu tog interval naći interval za $EX$.
Uporediti dobijene rezultate sa početno izračunatim očekivanjem.

**Rešenje:**

a) 90 – ispravnih proizvoda, 10 – neispravnih proizvoda

**X –** broj neispravnih proizvoda u dva izabrana

Jednostavnim rezonovanjem vidimo da **X** može uzimati vrednosti iz skupa {0,1,2}, dakle diskretna slučajna promenjiva. Zakon raspodele slučajne promenjive **X** dat je sa:

$$\boldsymbol{X} \sim \begin{pmatrix} 0 & 1 & 2 \\ p_1 & p_2 & p_3 \end{pmatrix}$$

Gde imamo da važi: $p_1 = \frac{\binom{90}{2}}{\binom{100}{2}} \approx 0.81$, $p_2 = \frac{\binom{90}{1}\binom{10}{1}}{\binom{100}{2}} \approx 0.18$, $p_3 = \frac{\binom{10}{2}}{\binom{100}{2}} \approx 0.01$.

Odredimo očekivanje: $\boldsymbol{EX} = \sum_{i=1}^{3} p_i x_i = 0.2$

b) Vršimo generisanje slučajne sekvence brojeva, čija raspodela odgovara goredobijenoj. Prvo ćemo generisati sekvencu brojeva sa uniformnom raspodelom i iz nje ivesti našu raspodelu. (koristeći C++ programski jezik)

```
vector<double> genUniform(int numOfNumbers) {
    random_device rd;
    mt19937 gen(rd());
    uniform_real_distribution<> distrib(0.0, 1.0);
    vector<double> random_numbers;
    random_numbers.reserve(numOfNumbers);
    for (size_t i = 0; i < numOfNumbers; ++i) {
        random_numbers.push_back(distrib(gen));
    }
    return random_numbers;
}
```

*Generator uniformne raspodele*

Dalje vršimo generisanje po zadotoj raspodeli, a prema datom kodu.

```
//generator nase raspodele
vector<int> genSpecificDistribution(int numOfNumbers) {
    vector<double>uniformSequence = genUniform(numOfNumbers);
    vector<int> finalSequence;
    for (int i = 0;i < numOfNumbers;i++) {
        double uniNum =uniformSequence[i];
        //cout << uniNum<<endl;
        if (0 <= uniNum && uniNum < 0.81) { finalSequence.push_back(0);}
        else if (0.81 <= uniNum && uniNum < 0.99) { finalSequence.push_back(1);}
        else if (0.99 <= uniNum && uniNum <= 1) { finalSequence.push_back(2);}
        else { throw exception("Greska u generator uniforomne raspodele!"); }
    }
    return finalSequence;
}
```

*Generator naše raspodele*

Generisaćemo slučajni niz za $n=50$. Dobijeni slučajna sekvenca $\boldsymbol{X_1,\ldots,X_{50}}$ (prema raspodeli slučajne promenjive $\boldsymbol{X}$).

```
Microsoft Visual Studio Debug
0 1 0 1 1 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 1 1 2 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0
C:\Users\Korisnik\source\repos\vis_projekat\x64\Debug\vis_projekat.exe (process 24888) exited with code 0.
Press any key to close this window . . .
```

0 1 0 1 1 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 1 1 2 0 0 0 0 0 0 1 1 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0

Dalje smatramo da je data sekvenca brojeva slučajni uzorak nepoznate raspodele, zakona raspodele $\boldsymbol{X} \sim Bin(2,\boldsymbol{p}),\ \ 0<p<1$. Želimo da izvršimo ocenu parametra $\boldsymbol{p}$. Još znamo E$\boldsymbol{X}$$=2p$.

Koristimo metod maksimalne verodostojnosti, sa funkcijom verodostojnosti oblika:

$$L(p) = \prod_{\mathrm{i}} \mathrm{p}(X_i = \mathrm{x_i})$$

Kao i pomoćnu funkciju:

$$l(p)=ln(L(p))$$

Takođe, znamo da je $Bin(2,p)$ oblika $\begin{pmatrix} 0 & 1 & 2 \\ (1-p)^2 & 2p(1-p) & p^2 \end{pmatrix}$. Te onda imamo:

$$L(p)=2^{13}\, p^{15}(1\text{-}p)^{85}$$

$$l(p)=13ln2+85ln(1\text{-}p)+15ln(p)$$

Tražimo maksimum date funkcije, tj. $p$ za koje

$$l'(p) = \frac{-85}{1-p} + \frac{15}{p} = 0 \quad (1)$$

Rešavanjem (1) imamo $\boldsymbol{p} = \frac{\mathbf{3}}{\mathbf{20}}$, a iz znaka drugog izvoda zaista vidimo da to jeste tačka maksimuma.

Odatle dobijamo procenu za parameter $\boldsymbol{p} = \frac{\mathbf{3}}{\mathbf{20}} = \mathbf{0.15}.$

Ocena za očekivanje **EX=*2p=0.30.***

---

U prethodnom delu smo se bavili tačkastom ocenom parametra *p*, a sada radimo intervalnu ocenu (intervalima poverenja, 95%); $X \sim Bin(2, \boldsymbol{p}),\ 0<p<1.$

Tražimo A i B tako da $P(A \le p \le B) = 0{,}95$ *gde su A,B realni brojevi.* (2)

Koristićemo centralnu graničnu teoremu. Za niz nezavisnih slučajnih promenjivih $X_1, \ldots, X_n$, gde $EX_i = np = 2p$, kao i $VarX_i = np(1-p) = 2p(1-p)$, za svako $i = 1, \ldots, n$. Tada prema CGT za slučajnu promenjivu definisanu sa $\boldsymbol{Y} = \frac{1}{n}\sum_{i=1}^{n} X_i$, gde $E\mathbf{Y} = 2p$ i $Var\mathbf{Y} = \frac{2p(1-p)}{n}$ važi:

$$\mathbf{Y} \sim \mathcal{N}\left(2p, \frac{2p(1-p)}{n}\right)$$

Odnosno imaćemo:

$$\boldsymbol{Z} = \frac{\mathbf{Y} - 2p}{\sqrt{\frac{2p(1-p)}{n}}} \sim \mathcal{N}(0,1) \quad (3)$$

Sada ćemo (2) napisati u ekvivalentnom obliku:

$$p\left(\frac{\mathbf{Y}-2B}{\sqrt{\frac{2p(1-p)}{n}}} \le \frac{\boldsymbol{Y}-2p}{\sqrt{\frac{2p(1-p)}{n}}} \le \frac{\mathbf{Y}-2A}{\sqrt{\frac{2p(1-p)}{n}}}\right) = 0{,}95$$

Na osnovu generisanog uzorka daćemo ocenu za **Y** i *p*. Odnosno, iz standardne ocene aritmetičkom sredinom: $\widehat{\boldsymbol{Y}} = \frac{1}{50}\sum_{i=1}^{50} X_i = \frac{15}{50} = 0.3.$ Metodom momenata ocenićemo parameter *p*, imajući u vidu $E\mathbf{Y} = 2p$ i prethodno dobijene vrednosti imamo $\hat{p} = \frac{\hat{Y}}{2} = 0.15$.(što odgovara oceni metodom maksimalne verodosnojnosti iz prethodnog dela).

Koristeći dobijene ocene imamo:

$$p\left(\frac{\widehat{\boldsymbol{Y}}-2B}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}} \le \boldsymbol{Z} \le \frac{\widehat{\boldsymbol{Y}}-2A}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}}\right) = 0{,}95$$

Formiraćemo dvostrani interval poverenja, pri čemu koristimo parnost (simetriju) normalne raspodele (u cilju da sam interval bude što kraći).

Neka je $t = \frac{\hat{Y}-2A}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}}$ ,odnosno -$t = \frac{\hat{Y}-2B}{\sqrt{\frac{2\hat{p}(1-\hat{p})}{n}}}$ (odgovarajući kvantili).

Iz Gausove raspodele imamo $\phi(t) = 0{,}975$. Odakle se dobija $t = 1.96$.

Uzimajući činjenicu da je $n = 50$, imaćemo:

$$A=0{,}08$$

$$B=0.22$$

Tj. 95% interval poverenja za ***p*** **[0.08 , 0.22].**

Iz činjenice da je *E**X***=*2p*, tada je odgovarajući interval poverenja za *E**X*** [0.16, 0.44].

Komentar:

Prvo treba naglasiti da ocene dobijane metodom maksimalne verodostojnosti potpuno odgovaraju vrednostima koje smo dobili pri proceni odgovarajućih parametara za intervale poverenja. Sami intervali poverenja 2) su u skladu su sa onim što smo već izračunali u delu pod 1). Kada poredimo očekivanje koje smo na početku izračunali, u odnosu na ocenu dobijenu u 1) imamo odstupanje od približno 0.1, ali ono što treba naglasiti jeste da sama početna vrednost očekivanja pripada intervalu poverenja dobijenom u 2).