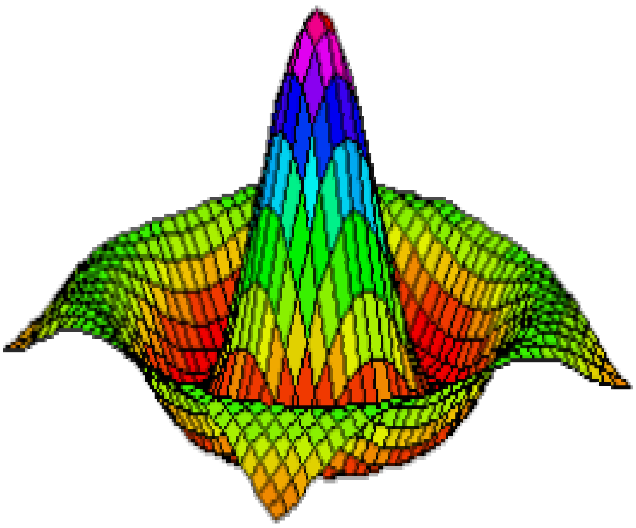
SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA
ZAVOD ZA TELEKOMUNIKACIJE

Željko Ilić, Smiljan Pilipović



LABORATORIJSKA VJEŽBA

Teorija informacije

STATISTIČKA SVOJSTVA MARKOVLJEVIH IZVORA

Priroda informacijskih izvorišta koja se susreću u praksi vrlo je različita. Već na prvi pogled moguće je ustanoviti da TV signal po prestanku TV programa i TV slika imaju posve različita informacijska svojstva. Slične razlike moguće je naslutiti i kod usporedbe sadržaja binarne datoteke s tekstualnom. U slučaju binarne datoteke podaci (okteti) dolaze u slučajnom slijedu, ne postoji određena veza između susjednih okteta. Za razliku od toga, u tekstualnoj datoteci vjerojatnost pojave pojedinog okteta (slova) ovisi o njemu susjednom slovu. (U hrvatskom jeziku vrlo je mala vjerojatnost pojave primjerice para "kp").

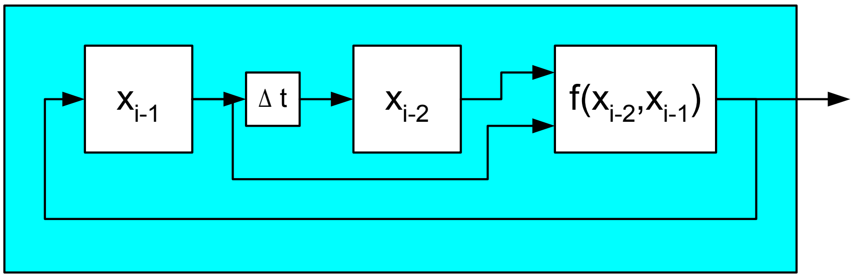
Sukladno tome postoje i različiti modeli kojima se modelira izlaz iz određenog izvorišta. Kao najrašireniji koriste se Markovljevi modeli.

Kako bi se kvantitativno opisao informacijski izvor, odnosno uočile razlike među njima koriste se veličine iz teorije vjerojatnosti: srednja vrijednost, standardna devijacija, korelacija, momenti višeg reda, …

Zadatak ove laboratorijske vježbe je da se prouče Markovljevi izvori nultog, prvog i drugog reda te da se uoči svrhovitost primjene kvantitativnih veličina u opisivanju informacijskih izvora.

Markovljev izvor

Markovljev izvor ima po volji "dugotrajno pamćenje", tj. trenutačni izlaz koji daje Markovljev izvor ovisi o proizvoljnom broju prethodnih izlaza. Stupanj izvora definira broj prethodnih izlaza koji se koriste kao argumenti funkcije za izračunavanje novog izlaza. Na slici 1 prikazan je Markovljev izvor drugog stupnja. Kod njega je trenutačni izlaz ovisan o dva prethodna izlaza.



Slika: 1

Na primjer *izvor s pamćenjem nultog stupnja* generira izlazne simbole nezavisno od prethodno generiranih. *Izvor s pamćenjem prvog stupnja* generirat će simbol ovisno o jednom prethodnom simbolu. *Izvor s pamćenjem j-tog stupnja* koristi prethodnih j izlaznih simbola kako bi se generirao novi simbol na izlazu.

Stanje

Pojam stanja usko je vezan za Markovljev izvor. Svako stanje definirano je izlaznim simbolom i vjerojatnostima prijelaza u sva slijedeća stanja. Za izvor nultog reda koji na izlazu generira simbole iz skupa od q znakova postoji q stanja kojima su pridjeljena q izlazna znaka. Izvor prvog reda također ima q stanja dok u slučaju izvora drugog reda postoji q^2 stanja jer je svako stanje definirano uređenim parom izlaznih simbola (ima ukupno q^2 takvih uređenih parova). U općem slučaju izvora k-tog reda postoji q^k stanja.

Pojam matrice vjerojatnosti prijelaza

Prijelazi između pojedinih stanja daju se uvjetnim vjerojatnostima:

$$p[x(t)=x_j|x(t-1)=x_i]=p(x_j|x_i)=p_{i,j}$$

Vjerojatnosti moraju zadovoljavati slijedeća svojstva:

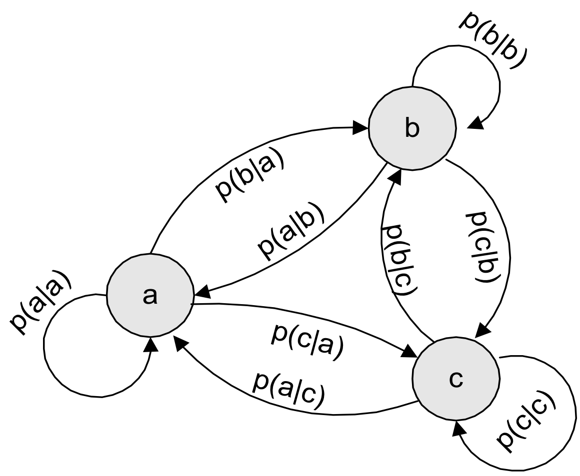
$$0<p(x_j|x_i)<1 \tag{1}$$

$$\sum_j p\left(x_j \middle| x_i\right) = 1 \tag{2}$$

$$\sum_i p\left(x_i\right) p\left(x_j \middle| x_i\right) = p\left(x_j\right) \tag{3}$$

Vjerojatnosti $p(x_i)$ u jednadžbi (3) nazivaju se apriorne vjerojatnosti i predstavljaju vjerojatnost pojavljivanja simbola *i* u ukupnom slijedu simbola.

Prijelazne vjerojatnosti mogu se prikazati pomoću tablice prijelaznih vjerojatnosti $\mathbf{P} = \left[p_{i,j}\right]$ ili pomoću grafa prijelaza. Jedan graf prijelaza prikazan je na slici 1 za slučaj 3 stanja. Svaka usmjerena linija označava prijelaz iz jednog u drugo stanje i uz nju je vezana pripadna uvjetna vjerojatnost.



Slika 1: Graf prijelaza

Pojmove stanja i vjerojatnosti prijelaza zornije ćemo pojasniti na primjeru Markovljevog izvora s tri izlazna simbola (*a, b, c*). U tablici 1 definirana su stanja i izlazni simboli uz stanja za takav Markovljev izvor nultog, prvog i drugog reda.

Tablica 1: Broj i vrsta stanja za pojedine izvora

Stupanj izvora:	Broj stanja:	Stanja	Izlazni simbol pridjeljen stanju
0	3	*(a), (b), (c)*	*(a), (b), (c)*
1	3	*(a), (b), (c)*	*(a), (b), (c)*
2	9	*(a,a), (a,b), (a,c), (b,a), (b,b), (b,c), (c,a), (c,b), (c,c)*	*(a), (b), (c), (a), (b), (c), (a), (b), (c)*

Markovljev izvor nultog reda naziva se još i izvor bez pamćenja jer mu izlazni simbol ne ovisi o prethodnim simbolima, odnosno slijedeće stanje ne ovisi o prethodnim stanjima. Stoga se njegova matrica prijelaznih vjerojatnosti (za primjer s tri stanja) može napisati na ovaj način:

$$\mathbf{P} = \left[p_{i,j}\right] = \begin{bmatrix} p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \\ p_1 & p_2 & p_3 \end{bmatrix} \tag{4}$$

Pri tome su vjerojatnosti p_i apriorne vjerojatnosti pojave simbola *i* na izlazu iz izvora. Kako su svi redovi matrice u (4) jednaki, vjerojatnost generiranja novog simbola ne ovisi o prethodnom simbolu.

Markovljev izvor prvog reda generira izlazni simbol j u ovisnosti o trenutačnom simbolu i s vjerojatnošću $p_{i,j}$. Matrica prijelaza u slučaju 3 izlazna simbola (tri stanja) može se prikazati na slijedeći način:

$$\mathbf{P} = \left[p_{i,j}\right] = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \tag{5}$$

U slijedećem poglavlju opisat ćemo način izračunavanja apriornih vjerojatnosti pojavljivanja pojedinih simbola na izlazu iz matrice (5).

U slučaju Markovljevog izvora drugog reda vjerojatnost pojave slijedećeg simbola k ovisi o trenutačnom simbolu j i prethodnom simbolu i, odnosno možemo ju pisati $p_{i,j,k}$. Matrica prijelaza u ovom slučaju bi bila trodimenzionalna matrica s dimenzijama za trenutačni, prethodni i slijedeći izlazni simbol.

Ovu matricu moguće je napisati i u dvodimenzionalnom obliku ukoliko se promijeni veza između stanja i simbola. U prethodnim slučajevima su stanje i simbol bili direktno vezani i imali smo jednak broj stanja i simbola. U slučaju Markovljevog izvora drugog reda treba pojedinom stanju pridjeliti **uređeni par izlaznih simbola**. Prema tome u općem slučaju s q izlaznih simbola imali bi $q \times q = q^2$ stanja. U primjeru s tri izlazna simbola ukupan broj stanja iznosi 9 i prikazan je u tablici 1. Uređeni par simbola (a,b) označava da je izvor generirao na izlazu simbol b a prije toga simbol a. Generiranjem novog izlaznog simbola x prelazimo u stanje (b, x). Bitno je uočiti da iz stanja (a,b) nije moguće prijeći u bilo koje stanja koje kao prvi član uređenog para nema b. Na taj način imamo veći broj stanja, ali neki od prijelaza između stanja su zabranjeni. U našem primjeru s 3 stanja, matrica prijelaza može se napisati na slijedeći način:

$$\mathbf{P} = \left[p_{i,j}\right] = \begin{bmatrix} p_{1,1,1} & p_{1,1,2} & p_{1,1,3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{1,2,1} & p_{1,2,2} & p_{1,2,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{1,3,1} & p_{1,3,2} & p_{1,3,3} \\ p_{2,1,1} & p_{2,1,2} & p_{2,1,3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{2,2,1} & p_{2,2,2} & p_{2,2,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{2,3,1} & p_{2,3,2} & p_{2,3,3} \\ p_{3,1,1} & p_{3,1,2} & p_{3,1,3} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{3,2,1} & p_{3,2,2} & p_{3,2,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{3,3,1} & p_{3,3,2} & p_{3,3,3} \end{bmatrix} \tag{6}$$

Daljnim poopćenjem moguće je svaki Markovljev izvor k-tog stupnja s q izlaznih simbola svesti na izvor prvog stupnja s q^k stanja.

Izračunavanje apriornih vjerojatnosti izlaznih simbola

U dosadašnjem tekstu pojašnjen je pojam matrice vjerojatnosti prijelaza i pojam stanja. Osim vjerojatnosti prijelaza (uvjetne vjerojatnosti) postoje i apriorna vjerojatnost pojavljivanja simbola te trenutačna vjerojatnost pojavljivanja simbola. **Apriorna vjerojatnost pojavljivanja** odnosi se na frekvenciju simbola u dužem slijedu izlaznih simbola, dok se **trenutačna vjerojatnost pojavljivanja** odnosi na točno određen ishod.

Pojasnimo ovu razliku na slijedećem primjeru:

> Promatrajmo Markovljev izvor koji ima matricu prijelaza prikazanu izrazom (5). Neka je u početku promatranja na njegovom izlazu simbol a. U slijedećem koraku može se pojaviti simbol a s vjerojatnošću $p_{1,1}$, simbol b s vjerojatnošću $p_{1,2}$ i simbol c s vjerojatnošću $p_{1,3}$. To su trenutačne vjerojatnosti pojavljivanja. Ukoliko nakon n koraka razdioba trenutačnih

vjerojatnosti ne ovisi o početnom koraku, onda se takva razdioba naziva apriornim vjerojatnostima pojavljivanja.

Označimo početne vjerojatnosti triju stanja s $p = [p_1 \quad p_2 \quad p_3]$. Tada će nakon prvog koraka nove vjerojatnosti iznositi:

$$p' = [p_1' \quad p_2' \quad p_3'] = [p_1 \quad p_2 \quad p_3] \times \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \tag{7}$$

Izraz (7) predstavlja množenje vektora s matricom.

Vjerojatnosti nakon drugog koraka iznose

$$p'' = [p_1'' \quad p_2'' \quad p_3''] = [p_1' \quad p_2' \quad p_3'] \times \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \tag{8}$$

$$p'' = [p_1 \quad p_2 \quad p_3] \times \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \times \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \tag{9}$$

$$p'' = [p_1 \quad p_2 \quad p_3] \times \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix}^2 = p\mathbf{P}^2 \tag{10}$$

Matrica $\mathbf{P^2}$ predstavlja novu matricu prijelaznih vjerojatnosti kojom bi se opisivala vjerojatnost prijelaza iz stanja u stanje kad bi se promatrao svaki drugi ishod procesa. Ukoliko se žele promatrati vjerojatnosti nakon k koraka procesa formira se prijelazna matrica $\mathbf{P^k}$.

Ukoliko matrica $\mathbf{P^k}$ ima sve retke jednake konačna razdioba vjerojatnosti ne ovisi o početnoj razdiobi tj. nakon nekog vremena proces poprimi stacionarne vrijednosti razdiobe i tada je riječ o apriornim vjerojatnostima pojavljivanja.

Stacionarne vjerojatnosti moguće je dobiti i analitički slijedećim razmatranjem. Kod stacionarnog stanja se vjerojatnosti pojavljivanja p ne mijenjaju između prijelaza, odnosno vrijedi:

$$\begin{gathered} p\mathbf{P} = p \\ \sum_i p_i = 1 \end{gathered} \tag{11}$$

Na osnovu izraza (11) konstruira se sustav jednadžbi od n nepoznanica (n je broj stanja procesa) čijim se rješavanjem dobiva stacionarno rješenje.

Kvantitativne veličine za opisivanje informacijskih izvorišta

Za opisivanje svih slučajnih procesa, pa tako i informacijskih izvorišta koristimo slijedeće mjere:

- srednja vrijednost

  ukoliko izlazni simboli informacijskog izvorišta imaju numeričke vrijednosti ili im se mogu pridružiti numeričke analogne veličine, moguće je izračunati srednju vrijednost izlaznog simbola u vremenskom nizu:

$$\mu = \frac{1}{N}\sum_{n=1}^{N} x(n)$$

- standardna devijacija

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\mu\right)^2$$

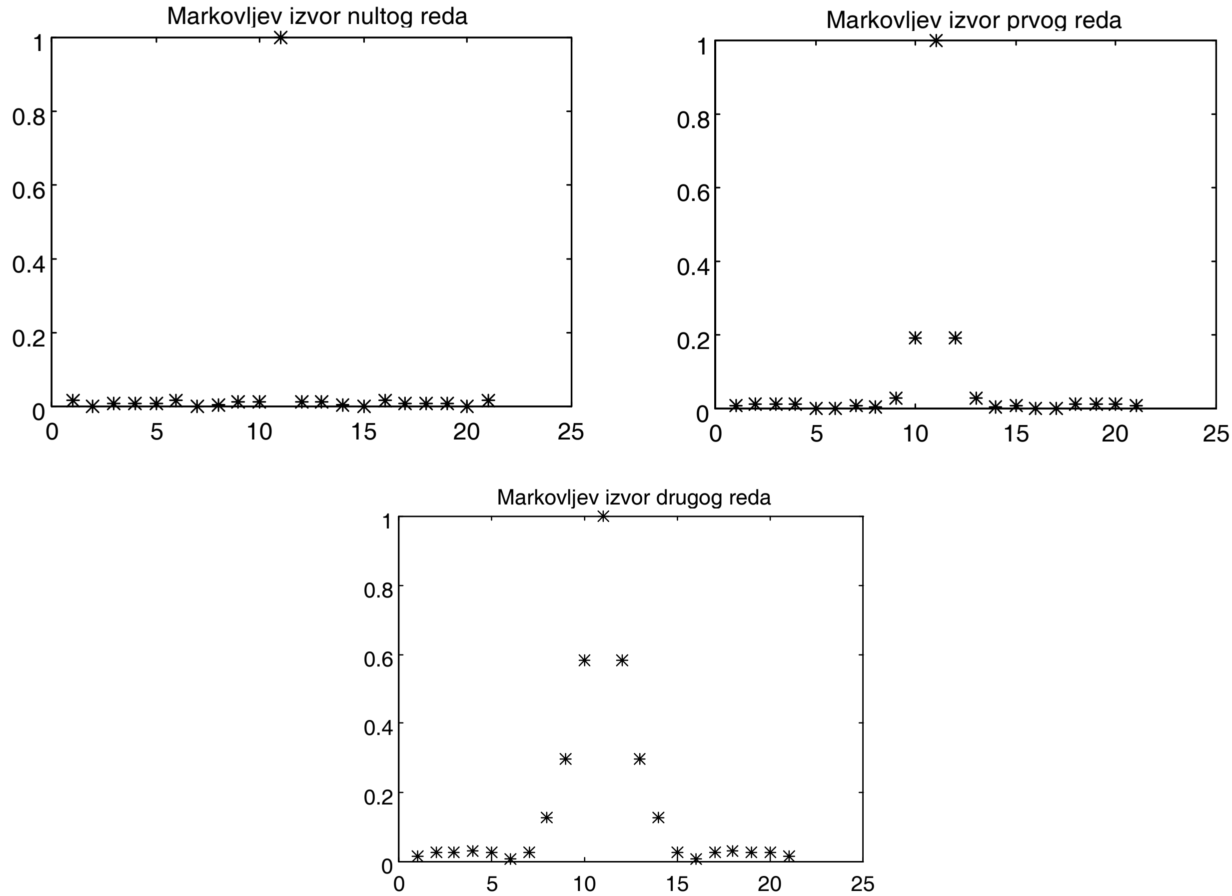
- autokorelacijska funkcija

$$R(\tau) = \frac{1}{N}\sum_{n=1}^{N} x(n)\times x(n+\tau)$$

- normirana autokorelacijska funkcija

$$r(\tau) = \frac{R(\tau)}{R(0)}$$

Na slici 2 prikazane su normirane autokorelacijske funkcije za Markovljeve izvore nultog, prvog i drugog reda.



Slika 2: Normirane autokorelacijske funkcije

MATLAB

U ovoj laboratorijskoj vježbi koristit će se programski paket Matlab. Za potrebe ove vježbe potrebne su slijedeće komande:

Sintaksa komande	Objašnjenje	Argumenti i napomene
z=ti(x,y,[q]);	generira niz simbola Markovljeva izvora duljine x iz izlaznog skupa od y simbola, ukoliko nije navedena matrica prijelaza q, funkcija se generira tijekom izvođenja	x - duljina generirane sekvencije y - veličina abecede izlaznih simbola q - matrica prijelaza
size(z,[i])	veličina vektora z, ukoliko i nije specificirano daju se obje dimenzije	i - 1 - broj redaka matrice - 2 - broj stupaca matrica
mean(z)	srednja vrijednost elemenata u matrici z	
std(z)	standardna devijacija elemenata u matrici z	
find(*izraz*)	daje sve redne brojeve elemenata koji zadovoljavaju *izraz*	find(z(find(z==1)+1)==1) izlistat će sva mjesta na kojima se nalaze uzastopno dvije jedinice
hist(z,a)	crta histogram na osnovu podataka iz z, a koji ima a razreda	

Rad u laboratoriju:

1. Funkcija *ti*

1. U Matlabu otvoriti prozor za upisivanje funkcije **File/New/M-File** i upisati slijedeću funkciju:

```
function z=ti(x,y,q)
if (nargin<3)
   q=rand(y);
end
for i=1:y
   q(i,:)=cumsum(q(i,:));
   q(i,:)=q(i,:)/q(i,y);
end
z=zeros(x,1);
z(1)=1;
for i=2:x
   s=rand(1);
   s1=find(q(z(i-1),:)>s);
   z(i)=s1(1);
end
```

2. Funkciju snimiti pod nazivom **ti.m** u neki od direktorija u koji je to dozvoljeno.
3. U Matlabu podesiti put (**File/Set Path...**) do direktorija u kojem se nalazi funkcija *ti*.

   Drugi način postavljanja puta: u radnom polju Matlaba otipkati:

   >> addpath *disk*:*direktorij*\\...*ime_finkcije*

2. Markovljev izvor nultog reda

1. Generirajte slučajnu matricu q veličine 1×3.

   $q =$

2. Normirajte matricu q tako da bude matrica vjerojatnosti.

$q =$

3. Generirajte *w*, slučajni slijed simbola "*1*", "*2*", "*3*", duljine 1000 u kojem će se simboli pojavljivati s vjerojatnostima definiranim matricom *q*.

   Napomena: Kod zadavanja ulaznih parametara funkcije *ti* pripazite na dimenzije matrice *q*, tj. radi se o Markovljevom izvorištu nultog reda.

4. Izračunajte srednju vrijednost i standardnu devijaciju elemenata u nizu.

   $\mu =$ $\sigma^2 =$

5. Nacrtajte histogram slijeda *w*.

   Odredite vjerojatnosti pojavljivanja simbola u slijedu *w*. $p(1)=$ _______; $p(2)=$ _______; $p(3)=$_______. **Napomena:** koristite Matlab funkciju *find*.

6. Izračunajte entropiju izvorišta.

   H(izv.)=________ [bita/simbolu]

2. Markovljev izvor prvog reda

1. Pomoću funkcije *ti* generirajte slijed od 1000 simbola iz skupa od 4 simbola (matricu prijelaza ne zadavati).
2. Pomoću histograma odredite vjerojatnost pojave pojedinog simbola. Vjerojatnosti pohraniti u niz *p*.

   $p =$ [ ]

3. Odredite uvjetne vjerojatnosti pojave pojedinih simbola i pohraniti ih u tablicu *q*.

$q=p(y_j/x_i) =$

4. Provjeriti ispravnost jednakosti $p=p\times q$.

   $p\times q =$

5. Izračunajte srednju vrijednost i standardnu devijaciju elemenata u nizu.

   $\mu =$ $\sigma^2 =$

6. Pretpostavimo da izvor u danom trenutku ima vrijednost "1". Izračunajte vjerojatnosti pojave simbola u nekoliko slijedećih koraka:

   $p' =$

   $p'' =$

7. Nacrtajte graf prijelaza za Markovljevo izvorište i naznačite vjerojatnosti prijelaza na granama prijelaza.

8. Izračunajte entropiju izvorišta uzimajući u obzir ovisnosti među simbolima.

   H(izv.)=________ [bita/simbolu]