Problem urni

Solić, Daria

Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Diplomski studij matematike
Financijska matematika i statistika

Daria Solić

Problem urni

Diplomski rad

Osijek, 2020.

Sveučilište J.J. Strossmayera u Osijeku
Odjel za matematiku
Diplomski studij matematike
Financijska matematika i statistika

Daria Solić

Problem urni

Diplomski rad

Mentor: izv. prof. dr. sc. Nenad Šuvak
Komentor: dr. sc. Ivan Papić

Osijek, 2020.

Sadržaj

Uvod **1**

1 Osnovno o Markovljevim lancima **2**

2 Pozadina modela urni **7**

3 Ehrenfestov model urni **9**

3.1 Vrijeme zaustavljanja u Ehrenfestovom modelu urni 14

3.2 Primjene Ehrenfestovog modela urni 19

4 Bernoulli-Laplaceov model urni **20**

5 Urne u bioznanosti **27**

5.1 Wright-Fisherov model urni alela bez mutacija 27

5.2 Wright-Fisherov model urni alela s mutacijama 29

5.3 Pogrešno kopiranje gena 30

Literatura **33**

Sažetak **34**

Summary **35**

Životopis **36**

Uvod

Pojam modela urni općenito se odnosi na jednu ili više urni koje sadrže objekte različitih vrsta. Objekti se obično nazivaju kuglice i razlikuju se po bojama. Izvlačenje kuglica iz urne može se nazvati uzorkovanjem, a ono se može provesti na različite načine, međutim najčešći način je uzorkovanje sa zamjenom ili uzorkovanje bez zamjene. Uzorkovanje sa zamjenom znači da kuglicu koju smo izvukli iz urne vratimo natrag u urnu prije nego što izvučemo drugu, dok uzorkovanje bez zamjene znači da izvučenu kuglicu ne vraćamo u urnu. Polyine urne je zajednički naziv za modele koji uključuju jednu urnu koja sadrži kuglice jedne, dvije ili više različitih boja. Iz urne na slučajan način izvlačimo jednu kuglicu te, ovisno o boji izvučene kuglice, dodamo (ili oduzmemo) deterministički ili slučajan broj kuglica pojedine boje u urnu i taj postupak ponavljamo. Teško je točno odrediti prvi zapis u kojem se pojavio model urni. Upotreba pojma urni i slučajnih izvlačenja iz urni spominju se još u Starom zavjetu i židovskoj teologiji, a opisani su u Rabinovitchevoj studiji (1973). Prema nekim izvorima prva upotreba modela urni u problemima vjerojatnosti pojavljuje se u djelima Huygensa (1629.-1695.) u 14. izdanju njegovih skupljenih djela posvećenih teoriji vjerojatnosti. U svojim radovima ih spominje i puno matematičara kao što su Bernoulli, Laplace, Markov i de Moivre. Modeli urni imaju bezbroj primjena i služe kao racionalni modeli za različite uobičajene situacije u znanosti, ekonomiji, poslovanju, industriji i tehnologiji. Također, imaju primjenu u raznim granama bioznanosti kao što su evolucija vrste i zdravstvo poput epidemiologije, ekologije i kliničkih ispitivanja.

Cilj ovog rada je modelirati određene modele urni Markovljevim lancima. U prvom poglavlju ukratko ćemo opisati Markovljeve lance i navesti najvažnije tvrdnje vezane za njih. U drugom dijelu ukratko ćemo opisati pozadinu modela urni i njegove početke. U trećem poglavlju bavit ćemo se Ehrenfestovim modelom urni. Modelirat ćemo ga pomoću Markovljevih lanaca, odredit ćemo očekivano vrijeme čekanja koje je potrebno da urna dođe u ravnotežu, odnosno da u njoj bude jednak broj plavih i bijelih kuglica uz pretpostavku da su na početku bile sve bijele kuglice u urni. Četvrto poglavlje posvetit ćemo Bernoulli-Laplaceovom modelu urni te ćemo navesti osnovne tvrdnje vezanje uz njega. Na kraju, u petom poglavlju kroz primjer opisat ćemo Wright-Fisherov model za populaciju sa i bez selekcije, mutacije i spola.

1 Osnovno o Markovljevim lancima

Na početku ovog rada navest ćemo osnovne definicije i teoreme o Markovljevim lancima koji će nam biti potrebni u daljnjoj analizi modela urni. Detalji i dokazi navedenih rezultata mogu se pronaći u [6].
Markovljev proces je slučajni proces sa svojstvom da uz danu vrijednost X_t (sadašnjost) vrijednost X_s za $s > t$ (budućnost) ne ovisi o X_u, za $u < t$ (prošlost).

Definicija 1. *Slučajni proces* $(X_t, t \in T)$ *je Markovljev proces ako je*

$$P(a < X_t \leqslant b | X_{t_1} = x_1, \ldots X_{t_n} = x_n) = P(a < X_t \leqslant b | X_{t_n} = x_n) \tag{1}$$

za proizvoljne $t_1 < t_2 < \ldots < t_n < t$ *iz* T.

Svojstvo (1) nazivamo Markovljevo svojstvo. U našem radu bavit ćemo se Markovljevim lancima u diskretnom vremenu.

Definicija 2. *Neka je* S *diskretan skup. Slučajni proces* $(X_n, n \in \mathbb{N}_0)$ *na vjerojatnosnom prostoru* $(\Omega, \mathcal{F}, P)$ *je Markovljev lanac ako vrijedi*

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i), \tag{2}$$

za svaki $n \in \mathbb{N}_0$, *za proizvoljne* $i, j, i_{n-1}, \ldots, i_0 \in S$ *za koje su obje uvjetne vjerojatnosti dobro definirane.*

Svojstvo (2) naziva se Markovljevo svojstvo, a njegova interpretacija je sljedeća. Vjerojatnosno ponašanje Markovljevog lanca $(X_n, n \in \mathbb{N}_0)$ u neposrednoj budućnosti (X_{n+1}), uvjetno na sadašnjost (X_n) i prošlost $(X_{n-1}, \ldots, X_0)$ jednako je vjerojatnosnom ponašanju Markovljevog lanca u neposrednoj budućnosti uvjetno samo na sadašnjost.
Osnovu proučavanja Markovljevih lanaca čine funkcije prijelaznih vjerojatnosti.
Za stanja $i, j \in S$ i $0 \leqslant s < t$, $s, t \in \mathbb{N}_0$, definiramo funkciju prijelaznih vjerojatnosti Markovljevog lanca

$$p(i, s; j, t) = P(X_t = j | X_s = i).$$

Teorem 1. *Markovljev lanac je u potpunosti određen poznavanjem distribucije od* X_0 *i funkcije prijelaznih vjerojatnosti u jednom koraku, tj. funkcije*

$$p(i, n; j, n + 1) = P(X_{n+1} = j | X_n = i).$$

Ako prijelazne vjerojatnosti u jednom koraku ne ovise o n, tj. ako je za svaki $n \in \mathbb{N}_0$

$$p(i,n;j,n+1) = p(i,j) = p_{ij}, i,j \in S$$

onda govorimo o vremenski homogenom Markovljevom lancu. Tada funkcija prijelaznih vjerojatnosti ovisi samo o stanju i iz kojeg proces polazi i stanju j u kojeg proces dolazi, a $p_{ij}, i,j \in S$ organiziramo u matricu $M = \begin{bmatrix} p_{11} & p_{12} & \cdots \\ p_{21} & p_{22} & \cdots \\ \cdots & \cdots & \cdots \end{bmatrix} = [p_{ij}]_{i,j\in S}$.

M se naziva matrica 1-koračnih prijelaznih vjerojatnosti Markovljevog lanca i ima dva važna svojstva:

(i) $p_{ij} = P(X_{n+1} = j | X_n = i) \in [0,1], \forall i,j \in S$

(ii) $\sum_{j\in S} p_{ij} = \sum_{j\in S} P(X_{n+1} = j | X_n = i) = P(X_{n+1} \in S | X_n = i) = 1.$

Iz svojstva (ii) možemo zaključiti da ako je lanac u trenutku n bio u stanju i, vjerojatnost da će se u trenutku $n+1$ naći u skupu stanja S je 1. Također, sume elemenata matrice M u istom retku su 1, a takva matrica zove se stohastička matrica. Karakterizirati u potpunosti Markovljev lanac znači znati distribuciju od X_0 (oznaka λ) i matricu M pa u tu svrhu navodimo sljedeći teorem.

Teorem 2. *Neka je $(X_n, n \in \mathbb{N}_0)$ slučajni proces s diskretnim skupom stanja S i konačnodimenzionalnim distribucijama*

$$P(X_0 = i_0, \ldots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n},$$

gdje je $\lambda_i = P(X_0 = i)$ i $\lambda = (\lambda_i, i \in S)$ neka distribucija na S, a $M = [p_{ij}]_{i,j\in S}$ neka stohastička matrica na S. Tada je slučajani proces $(X_n, n \in \mathbb{N}_0)$ (λ, M)-Markovljev lanac.

Zbog homogenosti Markovljevog lanca za m-koračne prijelazne vjerojatnosti vrijedi

$$P(X_m = j | X_0 = i) = P(X_{n+m} = j | X_n = i) = (M^m)_{ij}$$

Teorem 3. *Prijelazne vjerojatnosti homogenog (λ, M)-Markovljevog lanca zadovoljavaju tzv. CHAPMAN-KOLMOGOROVLJEVE JEDNADŽBE, tj.*

$$p_{ij}^{(m)} = \sum_{k\in S} p_{ik}^{(r)} p_{kj}^{(m-r)}, \forall r \in \{1,2,\ldots,m-1\}, \forall i,j \in S, tj.$$

$$M^m = M^r \cdot M^{m-r}.$$

Sada se postavlja pitanje koja stanja Markovljev lanac može posjetiti ako je krenuo iz nekog zadanog stanja?
Neka je $(X_n, n \in \mathbb{N}_0)$ (λ, M)-Markovljev lanac sa skupom stanja S. Za podskup $B \subset S$ definiramo vrijeme prvog posjeta lanca skupu B:

$$T_B = min\{n \in \mathbb{N}_0 : X_n \in B\},$$

uz dogovor da je $min\emptyset = \infty$. Ako je B jednočlan, $B = \{j\}, j \in S$, onda ćemo umjesto $T_{\{j\}}$ pisati T_j.

Definicija 3. *Za stanja $i, j \in S$ kažemo da je j dostižno iz i, oznaka $i \to j$, ako vrijedi da je $P(T_j < \infty | X_0 = i) > 0$.*

To znači da Markovljev lanac s pozitivnom vjerojatnošću posjećuje $j \in S$ u konačnom vremenu, ako je krenuo iz $i \in S$.

Definicija 4. *Stanja $i, j \in S$ komuniciraju (oznaka $i \leftrightarrow j$) ako je $i \to j$ i $j \to i$.*

Relacija komuniciranja K inducira particiju skupa S na tzv. klase komuniciranja.

Definicija 5. *Markovljev lanac je ireducibilan ako se skup stanja S sastoji od samo jedne klase komuniciranja, tj. ako za $\forall i, j \in S$ vrijedi $i \leftrightarrow j$.*

Definicija 6. *Za podskup $C \subset S$ skupa stanja kažemo da je zatvoren ako za svako stanje $i \in C$ vrijedi $P(T_{S \setminus C} = \infty | X_0 = i) = 1$.*

Stanje u koje Markovljev lanac jednom uđe te u njemu ostaje zauvijek naziva se apsorbirajuće stanje Markovljevog lanca. Iz zatvorenog podskupa skupa stanja Markovljev lanac ne može izaći, ali u njega može ući te kažemo da biva apsorbiran u njemu.

Neka je $T_i^{(1)} = min\{n \in \mathbb{N} : X_n = i\}$ vrijeme prvog povratka Markovljevog lanca u stanje $i \in S$, pri čemu je $\{X_0 = i\}$.

Definicija 7. *(i) Stanje $i \in S$ je povratno stanje Markovljevog lanca ako je*

$$P(T_i^{(1)} < \infty | X_0 = i) = 1$$

(ii) Stanje $i \in S$ je prolazno stanje Markovljevog lanca ako je

$$P(T_i^{(1)} < \infty | X_0 = i) < 1, \textit{tj. ako je}$$

$$P(T_i^{(1)} = \infty | X_0 = i) > 0.$$

Propozicija 1. *Pretpostavimo da je S konačan prostor stanja. Tada S sadrži barem jedno povratno stanje.*

Teorem 4. *Neka je $i \in S$ povratno stanje [prolazno] te neka $i \leftrightarrow j$. Tada je $j \in S$ također povratno [prolazno] stanje.*

Definicija 8. *Stanje $i \in S$ je pozitivno povratno ako je*

$$E_i[T_i^{(1)}] = E[T_i^{(1)}|X_0 = i] < \infty.$$

Definicija 9. *Slučajni proces $(X_t, t \in T)$ je strogo stacionaran ako su distribucije slučajnih vektora $(X_{t_1}, \dots, X_{t_n})$ i $(X_{t_{1+h}}, \dots, X_{t_{n+h}})$ jednake za proizvoljne $t_1, \dots t_n \in T$ t.d je $t_1 < t_2 < \dots < t_n$ i h t.d $t_{1+h}, t_{2+h}, \dots, t_{n+h} \in T$*

Definicija 10. *Neka je $(X_n, n \in \mathbb{N}_0)$ Markovljev lanac s prebrojivim skupom stanja S i matricom prijelaznih vjerojatnosti M. Vjerojatnosna distribucija $\pi = (\pi_i, i \in S)$ na S je stacionarna ili invarijantna distribucija ovog Markovljevog lanca ako vrijedi da je $\pi = \pi M$, tj. po komponentama $\pi_j = \sum\limits_{k \in S} \pi_k p_{kj}$ za sve $j \in S$.*

Teorem 5. *Neka je $(X_n, n \in \mathbb{N}_0)$ (π, M) Markovljev lanac sa stacionarnom distribucijom π. Tada je $(X_n, n \in \mathbb{N}_0)$ (strogo) stacionaran proces.*

Teorem 6. *Neka je S konačan skup stanja Markovljevog lanca te pretpostavimo da za neki $i \in S$ vrijedi*

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j, \forall j \in S.$$

Tada je $\pi = (\pi_j, j \in S)$ stacionarna distribucija tog Markovljevog lanca.

Napomena 1. *Ako je Marovljev lanac ireducibilan i pozitivno povratan, on ima jedinstvenu stacionarnu distribuciju.*

Definicija 11. *Neka je $X = (X_n, n \in \mathbb{N}_0)$ Markovljev lanac sa skupom stanja S i matricom prijelaznih vjerojatnosti M. Vjerojatnosna distribucija $\pi = (\pi_i, i \in S)$ naziva se graničnom distribucijom Markovljevog lanca X ako $\forall i, j \in S$ vrijedi*

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j.$$

Teorem 7. *Neka je π granična distribucija Markovljevog lanca X. Tada je ona i njegova stacionarna distribucija.*

Definicija 12. *Neka je X Markovljev lanac sa matricom prijelaznih vjerojatnosti M. Za stanje $i \in S$ s $d(i)$ označavamo najveći zajednički djelitelj skupa $\{n \in \mathbb{N} : p_{ii}^{(n)} > 0\}$, gdje je $d(i) = 1$ ako je taj skup prazan. Kažemo da je stanje $i \in S$ aperiodično ako je $d(i) = 1$. U suprotnom, kažemo da je stanje $i \in S$ periodično s periodom $d(i)$.*

Lema 1. *Periodičnost je svojstvo klase komuniciranja, tj. za proizvoljna stanja $i, j \in S$ t.d. $i \leftrightarrow j$ vrijedi $d(i) = d(j)$.*

Teorem 8. *Neka je λ proizvoljna vjerojatnosna distribucija na skupu stanja S. Pretpostavimo da je $X = (X_n, n \in S)$ (λ, M)-Markovljev lanac koji je ireducibilan i aperiodičan te neka ima stacionarnu distribuciju π. Tada je*

$$\lim_{n\to\infty} P(X_n = j) = \pi_j, \forall j \in S.$$

Specijalno, $\lim_{n\to\infty} p_{ij}^{(n)} = \pi_j, \forall i, j \in S$, tj. stacionarna distribucija je ujedno i granična.

Teorem 9. *Neka je $X = (X_n, n \in \mathbb{N}_0)$ (π, M)-Markovljev lanac gdje je π stacionarna distribucija za M. Tada je X stacionaran proces. Preciznije, X je stacionaran uz vjerojatnost $P_\pi = (P_{\pi_1}, \ldots, P_{\pi_n})$ gdje je $P_{\pi_i} = \sum_{i\in S} \pi_i p_{ij}$ za sve $j \in S$. Nadalje, za svaki $m \in \mathbb{N}$ je $(X_{m+n}, n \in \mathbb{N})$ ponovno (π, M)-Markovljev lanac.*

2 Pozadina modela urni

Model urni konstruira se zamišljajući određeni broj urni od kojih neke ili sve sadrže kuglice različitih boja. U specifičnim slučajevima izvode se nizovi pokusa u kojima se kuglice izvlače i potencijalno vraćaju u urnu po posebnim pravilima. Ova pravila mogu uključivati potrebu da se dodaju ili oduzimaju kuglice iz određenih urni u različitim fazama pokusa. Također se može dogoditi da određene kuglice promijene boju u skladu s propisanim pravilima. Ako u urni imamo s kuglica koje izvlačimo na slučajan način te ako svaka kuglica može biti izvučena s jednakom vjerojatnošću onda je ta vjerojatnost $\frac{1}{s}$. Pomoću ovog jednostavnog rezultata možemo (barem u pravilu) izračunati vjerojatnost bilo kojeg specifičnog ishoda bilo kojeg pokusa (ili niza pokusa) opisane vrste. U određenim slučajevima, naravno, računica može biti tehnički komplicirana, ali i tada ćemo samo koristiti, u suštini, jednostavne rezultate. Obično nas zanima:

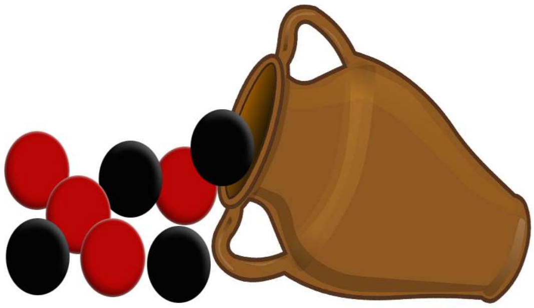
1. Distribucija određenog broja različitih kuglica u urnama
2. Distribucija vremena čekanja dok određeni uvjet ili uvjeti nisu zadovoljeni.

Možemo vidjeti da se preko modela urni mogu izvesti mnoge poznate parametarske diskretne distribucije. Primjer povijesne važnosti je [3] u kojoj je Wilhelm Lexis analizirao određene urne. Uzimao je u obzir skup urni koje sadrže različiti udio crnih i bijelih kuglica. Ako planiramo izvući određeni broj (n) kuglica iz urne, možemo:

1. Na slučajan način odabrati urnu i iz nje izvući n kuglica (bez vraćanja izvučene kuglice u urnu) ili
2. Izabrati isti broj kuglica iz svake urne (ako je to moguće).

Uzimajući u obzir ekstremne slučajeve kada imamo dvije urne, jednu koja sadrži samo crne i jednu koja sadrži samo bijele kuglice, vidljivo je da metoda 1 vodi do toga da je broj bijelih kuglica ili 0 ili n, s tim da urna mora sadržavati barem n bijelih kuglica, dok metoda 2 vodi do neobično stabilnih rezultata (ako je n paran onda je broj bijelih kuglica uvijek $\frac{n}{2}$). Slučaj 1 naziva se *supernormalna disperzija*, a slučaj 2 *subnormalna disperzija*, u usporedbi s rezultatima jednostavnog slučajnog uzimanja uzorka (bez vraćanja izvučene kuglice) iz jedne urne koja sadrži p bijelih, odnosno $1 - p$ crnih kuglica.

Modele urni u svojim radovima spominje puno matematičara kao što su Bernoulli, Laplace, Markov i de Moivre. Bernoulli je bio prvi koje je predložio model koji je sada poznat kao Ehrefestov model, kao i druge manje poznate modele o kojima ćemo nešto više reći u sljedećim poglavljima.

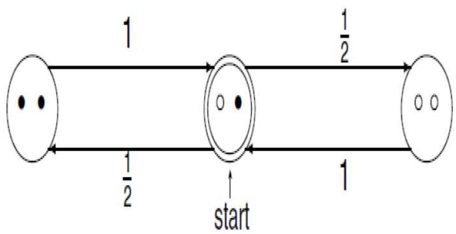


Slika 1: Urna s kuglicama

3 Ehrenfestov model urni

Kako bi lakše razumjeli Ehrenfestov model urni poglavlje ćemo započeti primjerom.

Primjer 1. *Ehrenfestova urna započinje s jednom bijelom i jednom plavom kuglicom. Kuglice se uzorkuju iz urne. Kad god se odabere kuglica određene boje, izbacimo je iz urne i zamijenimo kuglicom druge boje. Urna se uvijek vraća u početno stanje nakon parnog broja izvlačenja, a uvijek je izvan tog stanja nakon neparnog broja izvlačenja. Kada izađe iz početnog stanja, može se nalaziti u jednom od dva podjednako vjerojatna stanja: stanje s dvije bijele kuglice ili stanje s dvije plave kuglice. Na Slici 2 mali ispunjeni krugovi predstavljaju plave kuglice, mali neispunjeni krugovi predstavljaju bijele kuglice, veliki krugovi su stanja Markovljevog lanca, veliki dvostruki krug označava početno stanje, dok su brojevi sa strelicama vjerojatnosti prijelaza. Neka je W_n broj bijelih kuglica u urni nakon n izvlačenja. Tada,*



Slika 2: Markovljev lanac za Ehrenfestov model urni

$$W_n = \begin{cases} 1, & \textit{ako je n paran;} \\ 2X, & \textit{ako je n neparan,} \end{cases}$$

pri čemu X ima Bernoullijevu distribuciju s parametrom $p = \frac{1}{2}$ koja je dana slijedećom tablicom distribucije $\begin{pmatrix} 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$. Ovaj proces je Markovljev lanac. Stanja se mogu predstaviti brojem bijelih kuglica u urni. Imamo Markovljev lanac s tri stanja koja se mogu označiti s 0,1 i 2 i s matricom prijelaznih vjerojatnosti

$$M = \begin{bmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$$

Takav lanac ima stacionarnu distribuciju danu vektorom retkom $\pi = (\pi_0 \ \ \pi_1 \ \ \pi_2)$ *koji je rješenje sustava jednadžbi*

$$\pi M = \pi$$

$$\pi_0 + \pi_1 + \pi_2 = 1.$$

Naime, imamo

$$\pi M = (\pi_0 \ \ \pi_1 \ \ \pi_2) \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} = \left(\frac{1}{2}\pi_1 \ \ \pi_0 + \pi_2 \ \ \frac{1}{2}\pi_1 \right).$$

Izjednačavajući dobiveno s $(\pi_0 \ \ \pi_1 \ \ \pi_2)$ *vidimo da je stacionarna distribucija binomna distribucija s parametrima* $n = 2$ *i* $p = \frac{1}{2}$*, odnosno* $\pi = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$*. Štoviše, ovaj Markovljev lanac nema graničnu distribuciju, s obzirom na periodičnost*

$$\begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} = M = M^3 = M^5, \ldots,$$

i

$$\begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} = M^2 = M^4 = M^6, \ldots,$$

to jest, lanac je periodičan s periodom 2, a M^n *izmjenjuje dvije vrijednosti ne dostižući graničnu vrijednost. Ako se nalazimo u bilo kojem od tri stanja, potrebno je barem dva prijelaza da se vratimo u to stanje.*

Općenito promatrajući modele urni, neka urna može sadržavati kuglice k različitih boja ($k \in \mathbb{N}$). Promjene u sadržaju urne događaju se u diskretnim trenucima. U svakom koraku prvo dobro promiješamo našu urnu i nakon toga izvlačimo jednu kuglicu iz nje (pri čemu su vjerojatnosti izvlačenja za sve kuglice jednake). Kada izvučemo kuglicu, pogledamo koje je ona boje i stavimo ju natrag u urnu. Ako je boja izvučene kuglice jednaka i, gdje je $i \in \{1, \ldots, k\}$ onda u urnu stavljamo A_{ij} kuglica boje j, $j \in \{1, \ldots, k\}$. Općenito, A_{ij} može biti deterministička ili slučajna vrijednost te može biti pozitivna ili negativna. Uobičajeno je pravila izmjene sadržaja urni, ovisno o boji izvučene kuglice, zapisati matrično. Tu matricu nazivamo shema izmjena. U ovom općenitom slučaju shema izmjena izgleda ovako

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1k} \\ A_{21} & A_{22} & \ldots & A_{2k} \\ \ldots & \ldots & \ldots & \ldots \\ A_{k1} & A_{k2} & \ldots & A_{kk} \end{pmatrix},$$

pri čemu elementi u i-tom retku predstavljaju koliko se kuglica pojedine boje stavlja u urnu ako je boja izvučene kuglice i, a elementi u j-tom stupcu predstavljaju broj kuglica j-te boje koji stavljamo u urnu ovisno o tom koje je boje izvučena kuglica.

U slučaju Ehrenfestovog modela urne iz prethodnog primjera kada imamo dvije kuglice koje mogu biti bijele ili plave boje shema izmjena izgleda ovako

$$A = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

što znači da kada god odaberemo kuglicu određene boje iz urne, izvadimo ju i zamijenimo kuglicom druge boje.

U Ehrenfestovom modelu urni broj kuglica je konstantna veličina, ali dolazi do promjene broja bijelih i plavih kuglica. Drugim riječima, ukupan broj kuglica u urni je fiksan, recimo τ, ali broj kuglica jedne određene boje ima diskretnu distribuciju sa slikom koja sadrži nenegativne cijele brojeve iz skupa $\{0, 1, \ldots, \tau\}$. Pretpostavimo da u početnoj urni imamo nula bijelih i τ plavih kuglica. U prvom izvlačenju odaberemo jednu kuglicu iz urne, a kako su sve kuglice plave boje odabrana kuglica je plave boje te ju zamijenimo s kuglicom bijele boje. Sada u urni imamo jednu kuglicu bijele boje i $\tau - 1$ kuglica plave boje. U drugom izvlačenju možemo odabrati ili plavu ili bijelu kuglicu. Ako odaberemo plavu kuglicu zamijenimo je s bijelom pa u urni imamo dvije bijele i $\tau - 2$ plavih kuglica. Ako odaberemo bijelu kuglicu zamijenimo je s plavom pa u urni imamo nula bijelih i τ plavih kuglica. Postupak dalje nastavljamo na analogan način dok ne dođemo do željenog broja izvlačenja ili do željenog broja bijelih i plavih kuglica u urni. Pretpostavljamo da započinjemo s nula bijelih i τ plavih kuglica, a mogli smo započeti i s različitim brojem bijelih i plavih kuglica. Sada ćemo reći nešto o stacionarnoj distribuciji ovog Markovljevog lanca.

Teorem 10. *Binomna distribucija $B(\tau, \frac{1}{2})$ je stacionarna distribucija Ehrenfestovog Markovljevog lanca.*

Dokaz: Provjerimo zadovoljava li $\pi = B(\tau, \frac{1}{2})$ Definiciju 10, odnosno vrijedi li:

$$\pi M = \pi$$

$$\pi_0 + \ldots + \pi_\tau = 1.$$

Matricu prijelaznih vjerojatnosti dobijemo iz

$$p_{xy} = P(W_{n+1} = y | W_n = x) = \begin{cases} \frac{\tau - x}{\tau}, & y - x = 1, \\ \frac{x}{\tau}, & x - y = 1, \\ 0, & \textit{inače} \end{cases}$$

pri čemu su $x, y \in \{0, 1, \ldots, \tau\}$. Dakle, matrica prijelaznih vjerojatnosti dana je s

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ \frac{1}{\tau} & 0 & \frac{\tau-1}{\tau} & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & \frac{2}{\tau} & 0 & \frac{\tau-2}{\tau} & 0 & 0 & \ldots & 0 & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & \frac{\tau-1}{\tau} & 0 & \frac{1}{\tau} \\ 0 & 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 1 & 0 \end{pmatrix},$$

a distribucija s

$$\pi = \left(\frac{1}{2^\tau}\binom{\tau}{0}, \frac{1}{2^\tau}\binom{\tau}{1}, \ldots, \frac{1}{2^\tau}\binom{\tau}{\tau} \right).$$

Množenjem π i matrice M dobivamo da je $\pi M = \pi$. Provjerimo još vrijedi li drugi uvjet, odnosno $\pi_0 + \ldots + \pi_\tau = 1$. Dobivamo

$$\begin{aligned} \pi_0 + \ldots + \pi_\tau &= \frac{1}{2^\tau}\binom{\tau}{0} + \frac{1}{2^\tau}\binom{\tau}{1} + \ldots \frac{1}{2^\tau}\binom{\tau}{\tau} \\ &= \frac{1}{2^\tau}\left(\binom{\tau}{0} + \binom{\tau}{1} + \cdots + \binom{\tau}{\tau} \right) \\ &= \frac{1}{2^\tau}\dot{2}^\tau \\ &= 1. \end{aligned}$$

Dakle, $B(\tau, \frac{1}{2})$ je stacionarna distribucija. □

Teorem 11. *Pretpostavimo da Ehrenfestova urna počinje s* $W_0 = B(\tau, \frac{1}{2})$ *bijelih kuglica i* $B_0 = \tau - W_0$ *plavih kuglica. Neka je* W_n *broj bijelih kuglica nakon* n *izvlačenja. Tada,* $W_n \xrightarrow{D} B(\tau, \frac{1}{2}), n \to \infty$.

Dokaz: Stroga stacionarnost kod slučajnih procesa znači da su konačnodimenzionalne distribucije procesa invarijantne na vremenske pomake (Definicija 9). S obzirom da je ovaj proces Markovljev lanac sa stacionarnom distribucijom π, distribucija broja bijelih kuglica W_n u urni ne mijenja se kroz vrijeme (Teorem 9). Nakon n izvlačenja, u Ehrenfestovoj urni pojavljuje se W_n bijelih i B_n plavih kuglica. Da bi W_n bio k nakon n izvlačenja, u prethodnom koraku moramo imati ili $k+1$ bijelih kuglica, smanjujući na taj način bijele kuglice za jednu, ili $k-1$ bijelih kuglica te moramo izvući plavu kuglicu, povećavajući na taj način bijele kuglice:

$$P(W_n = k) = \frac{k+1}{\tau} P(W_{n-1} = k+1) + \frac{\tau - (k-1)}{\tau} P(W_{n-1} = k-1).$$

Neka je

$$\lim_{n\to\infty} P(W_n = k) = p(k).$$

S obzirom da se distribucija ne mijenja kroz vrijeme $p(k)$ zadovoljava

$$p(k) = \frac{k+1}{\tau} p(k+1) + \frac{\tau - (k-1)}{\tau} p(k-1).$$

Ako to raspišemo dobivamo:

$$\tau p(k) = (k+1)p(k+1) + \tau p(k-1) - (k-1)p(k-1)$$

$$\tau(p(k) - p(k-1)) = (k+1)p(k+1) - (k-1)p(k-1).$$

Napišimo sada sustav jednadžbi za $k, k-1, \ldots, 0$ (s prirodnom interpretacijom da je $p(-1) = 0$)

$$\begin{gathered}
\tau(p(k) - p(k-1)) = (k+1)p(k+1) - (k-1)p(k-1) \\
\tau(p(k-1) - p(k-2)) = kp(k) - (k-2)p(k-2) \\
\cdots \\
\tau(p(0) - p(-1)) = p(1) - (-1)p(-1).
\end{gathered}$$

Zbrajanjem ovih jednadžbi dobivamo:

$$p(k+1) = \frac{\tau - k}{k+1} p(k).$$

Raspisivanjem $p(k)$ dobivamo

$$\begin{aligned}
p(k) &= \frac{(\tau - (k-1))(\tau - (k-2)) \cdots \tau}{k(k-1) \cdots 1} p(0) \\
&= \binom{\tau}{k} P(W_0 = 0) \\
&= \frac{1}{2^\tau} \binom{\tau}{k},
\end{aligned}$$

za svaki $k = 0, \ldots, \tau$.

Odatle se može zaključiti da $W_n \xrightarrow{D} B(\tau, \frac{1}{2}), n \to \infty$. □

Do sada je naglasak bio na sastavu urne nakon niza izvlačenja, odnosno na broju pojava određene boje u nizu. U sljedećem poglavlju raspravljat ćemo o problemu čekanja ili vremenu zaustavljanja.

3.1 Vrijeme zaustavljanja u Ehrenfestovom modelu urni

Neka Ehrenfestov model urni s bijelim i plavim kuglicama započinje s urnom koja sadrži sve bijele kuglice. Na početku, broj bijelih kuglica u urni je $2M$. Zanima nas koliko je očekivano vrijeme čekanja $\gamma(2M)$ dok se ravnoteža ne pojavi prvi put. Ravnoteža je definirana kao stanje jednakog boja kuglica svake boje u urni. Za početak opišimo stanja Markovljevog lanca ovog modela. Neka je A_i stanje u kojem je i kuglica plave boje u urni. Na primjer, urna počinje u stanju A_0, bez plavih kuglica, ali prelazi u stanje A_1 odmah nakon prvog izvlačenja (u kojem morate odabrati bijelu kuglicu); bijela kuglica je u prvom izvlačenju zamijenjena plavom kuglicom. Neka je X_i vrijeme koje je potrebno da urna prijeđe iz stanja A_i u stanje A_{i+1}. Vrijeme čekanja $\gamma(2M)$ je

$$\gamma(2M) = X_0 + X_1 + \ldots + X_{M-1}. \tag{3}$$

Ako je urna u stanju A_i (s i plavih i $2M-i$ bijelih kuglica), može odmah prijeći u stanje A_{i+1} u jednom koraku, ako je izvučena bijela kuglica (s vjerojatnošću $\frac{2M-i}{2M}$). Inače, dolazi do zastoja u usponu prema ravnoteži, izvlači se plava kuglica (s vjerojatnošću $\frac{i}{2M}$), broj plavih kuglica se smanjuje i urna dolazi u stanje A_{i-1}. Vrijeme čekanja da se urna vrati u stanje A_i distribuirano je kao X_{i-1}, dok je vrijeme čekanja da urna dođe u stanje A_{i+1} distribuirano kao X_i. Imamo

$$X_i \overset{D}{=} \begin{cases} 1, & \text{s vjerojatnošću } \frac{2M-i}{2M} \\ 1 + X_{i-1} + X_i, & \text{s vjerojatnošću } \frac{i}{2M}. \end{cases}$$

Očekivanje od X_i je

$$E[X_i] = \frac{2M-i}{2M} + (1 + E[X_{i-1}] + E[X_i]) \cdot \frac{i}{2M},$$

a ako to malo sredimo dobivamo

$$\begin{aligned} E[X_i] &= \frac{2M-i}{2M} + \frac{i}{2M} + \frac{i}{2M} \cdot E[X_{i-1}] + \frac{i}{2M} \cdot E[X_i] \\ 2ME[X_i] &= 2M - i + i + i \cdot E[X_{i-1}] + i \cdot E[X_i] \\ (2M-i)E[X_i] &= 2M + i \cdot E[X_{i-1}] \\ E[X_i] &= \frac{2M}{2M-i} + \frac{i}{2M-i} E[X_{i-1}] \end{aligned} \tag{4}$$

Nizovi $E[X_{i-1}]$ i $2M-i$ idu u "suprotnim smjerovima", odnosno $E[X_{i-1}]$ je rastući niz, a $2M-i$ padajući pa je jedna od mogućnosti da $E[X_i]$ prikažemo pomoću integrala. Pokazat ćemo matematičkom indukcijom da je

$$E[X_i] = 2MI_i,$$

gdje je

$$I_i := 2M \int_0^1 x^{2M-i-1}(2-x)^i dx.$$

Baza indukcije, formula vrijedi za $k = 0$:

$$E[X_0] = 2M \int_0^1 x^{2M-0-1}(2-x)^0 dx = 2M \int_0^1 x^{2M-1} dx = 1.$$

Znamo da to vrijedi za $E[X_0]$ jer je potreban samo jedan prijelaz da urna sadrži jednu plavu kuglicu.
Pretpostavimo da za neki prirodan broj $k > 0$ vrijedi:

$$E[X_k] = 2M \int_0^1 x^{2M-k-1}(2-x)^k dx = 2MI_k.$$

U koraku indukcije pokažimo da tvrdnja vrijedi za prirodan broj $k+1$. Prema (4)

$$\begin{aligned} E[X_{k+1}] &= \frac{2M}{2M-(k+1)} + \frac{k+1}{2M-(k+1)} E[X_k] \\ &= \frac{2M}{2M-k-1} + 2M\frac{k+1}{2M-k-1} I_k. \end{aligned}$$

Kada integriramo:

$$I_{k+1} = \int_0^1 x^{2M-k-1-1}(2-x)^{k+1} dx = \left| \begin{array}{cc} u = (2-x)^{k+1} & dv = \int x^{2M-k-2} dx \\ du = -(k+1)(2-x)^k dx & v = \dfrac{x^{2M-k-1}}{2M-k-1} \end{array} \right|$$

$$= (2-x)^{k+1} \frac{x^{2M-k-1}}{2M-k-1} \Bigg|_{x=0}^{1} + \int_0^1 (k+1)(2-x)^k \frac{x^{2M-k-1}}{2M-k-1} dx. \tag{5}$$

Uvrštavanjem granica u (5) dobivamo:

$$I_{k+1} = \frac{1}{2M-k-1} + \frac{k+1}{2M-k-1} I_k.$$

Indukcijom smo dokazali da je $E[X_{k+1}] = 2MI_k$.
Očekivano vrijeme čekanja do ravnoteže iznosi

$$\begin{aligned}E[\gamma(2M)] &= E[X_0] + E[X_1] + \ldots + E[X_{M-1}]\\ &= 2M\sum_{i=0}^{M-1}\int_0^1 x^{2M-i-1}(2-x)^i dx\\ &= 2M\sum_{i=0}^{M-1}\int_0^1 x^{2M-1}\Big(\frac{2-x}{x}\Big)^i dx.\end{aligned}$$

Zbog svojstva linearnosti integrala suma i integral zamijene mjesta pa imamo da je

$$E[\gamma(2M)] = 2M\int_0^1 \sum_{i=0}^{M-1} x^{2M-1}\Big(\frac{2-x}{x}\Big)^i dx.$$

Sada ćemo izračunati sumu prvih M članova geometrijskog niza

$$\begin{aligned}E[\gamma(2M)] &= 2M\int_0^1 x^{2M-1}\frac{(\frac{2-x}{x})^M - 1}{\frac{2-x}{x} - 1}dx\\ &= 2M\int_0^1 x^{2M-1}\frac{\frac{(2-x)^M - x^M}{x^M}}{\frac{2-2x}{x}}dx\\ &= M\int_0^1 x^M\frac{(2-x)^M - x^M}{1-x}dx.\end{aligned}$$

Zbog lakšeg izračuna koristit ćemo supstituciju $x = 1-y$, što daje

$$\begin{aligned}E[\gamma(2M)] &= M\int_0^1 \frac{1}{y}(1-y)^M((1+y)^M - (1-y)^M)dy\\ &= M\int_0^1 \frac{1}{y}((1-y^2)^M - (1-y)^{2M})dy \qquad (6)\\ &= M\int_0^1 \frac{1}{y}(1-(1-y)^{2M})dy - M\int_0^1 \frac{1}{y}(1-(1-y^2)^M)dy.\end{aligned}$$

Podintegralne funkcije možemo protumačiti kao sumu prvih n članova geometrijskog niza

$$\sum_{k=0}^{n-1}(1-z)^k = \frac{1-(1-z)^n}{1-(1-z)}.$$

Koristeći to dolazimo do

$$\begin{aligned}
E[\gamma(2M)] &= M\int_0^1 \sum_{k=0}^{2M-1}(1-y)^k dy - M\int_0^1 y\sum_{j=0}^{M-1}(1-y^2)^j dy \\
&= M\sum_{k=0}^{2M-1}\int_0^1 (1-y)^k dy - M\sum_{j=0}^{M-1}\int_0^1 y(1-y^2)^j dy \\
&= -M\sum_{k=0}^{2M-1}\Big(0-\frac{1}{k+1}\Big) + \frac{M}{2}\sum_{j=0}^{M-1}\Big(0-\frac{1}{j+1}\Big) \\
&= M\sum_{k=0}^{2M-1}\frac{1}{k+1} - \frac{M}{2}\sum_{j=0}^{M-1}\frac{1}{j+1} \\
&= MH_{2M} - \frac{M}{2}H_M,
\end{aligned}$$

gdje je H_r r-ti harmonijski broj $1+\frac{1}{2}+\cdots+\frac{1}{r}$. Harmonijski brojevi imaju dobro poznatu aproksimaciju

$$H_r = \ln(r) + \Gamma + O\Big(\frac{1}{r}\Big), \quad \text{kada} \quad r\to\infty,$$

gdje je $\Gamma \approx 0.5772$ Eulerova konstanta. Za velike M očekivano vrijeme do ravnoteže je približno

$$\begin{aligned}
E[\gamma(2M)] &\approx M\Big(\ln(2M)+\Gamma+O\Big(\frac{1}{M}\Big)\Big) - \frac{M}{2}\Big(\ln(M)+\Gamma+O\Big(\frac{1}{M}\Big)\Big) \\
&\approx M\ln(2)+M\ln(M)+M\Gamma+MO\Big(\frac{1}{M}\Big) - \frac{M}{2}\ln(M) - \frac{M}{2}\Gamma - \frac{M}{2}O\Big(\frac{1}{M}\Big) \\
&\approx M\Big(\ln(2)+\frac{\Gamma}{2}\Big) + \frac{M}{2}\ln(M) + O(1).
\end{aligned}$$

Nema ništa posebno u tome što se kreće s $2M$ kuglica i nema nikakvog posebnog razloga da se definira da je ravnotežno stanje (stanje A_m) cilj. Mogli smo započeti s N bijelih kuglica i pitati se koliko vremena treba da se dođe u stanje A_k, za proizvoljni $k = 1, 2, \dots, N$, argument bi imao samo manje prilagodbe parametara. Pogledajmo sada primjer.

Primjer 2. *Imamo Ehrenfestovu urnu koja može sadržavati bijele i plave kuglice. Na početku u urni su samo bijele kuglice kojih ima M. Zanima nas koliko je očekivano vrijeme čekanja do potpune zamjene boja, tj. koliki je očekivani broj izvlačenja potreban da bi sve kuglice u urni po prvi puta bile plave.*

Rješenje: *Neka je A_i stanje u kojem je i kuglica plave boje u urni, a X_i neka bude vrijeme koje je potrebno da urna prijeđe iz stanja A_i u stanje A_{i+1}. Kao što smo naveli u prethodnom razmatranju, općenito vrijeme čekanja $\gamma_n(M)$ koje je potrebno da prijeđemo iz stanja A_0 u stanje A_n možemo predstaviti kao sumu slučajnih varijabli*

$$\gamma_n(M) = X_0 + X_1 + \ldots + X_n.$$

U prethodnom razmatranju bavili smo se slučajem $\gamma_M(2M)$ kojeg smo jednostavno označili s $\gamma(2M)$, a ovdje ćemo se baviti s $\gamma_M(M)$. Slično kao u (6) za općeniti slučaj imamo

$$E[\gamma_M(2M)] = \frac{M}{2}\int_0^1 \frac{1}{y}(1-y)^{M-n}((1+y)^n - (1-y)^n)dy.$$

U slučaju našeg konkretnog primjera imamo:

$$a_M = E[\gamma_M(2M)] = \frac{M}{2}\int_0^1 \frac{1}{y}((1+y)^M - (1-y)^M)dy.$$

Radi lakšeg izračuna imamo sljedeći postupak:

$$\begin{aligned}
\frac{2a_M}{M} - \frac{2a_{M-1}}{M-1} &= \int_0^1 \frac{1}{y}((1+y)^M - (1-y)^M)dy - \int_0^1 \frac{1}{y}((1+y)^{M-1} - (1-y)^{M-1})dy \\
&= \int_0^1 \frac{1}{y}((1+y)^M - (1+y)^{M-1})dy - \int_0^1 \frac{1}{y}((1-y)^M - (1-y)^{M-1})dy \\
&= \int_0^1 (1+y)^{M-1}dy + \int_0^1 (1-y)^{M-1}dy \\
&= \left.\frac{(1+y)^M}{M}\right|_{y=0}^1 - \left.\frac{(1-y)^M}{M}\right|_{y=0}^1 \\
&= \frac{2^M}{M}.
\end{aligned}$$

Dakle, imamo sljedeće:

$$\begin{aligned}\frac{2a_M}{M} &= \frac{2a_{M-1}}{M-1} + \frac{2^M}{M} \\ &= \frac{2a_M - 2}{M-2} + \frac{2^{M-1}}{M-1} + \frac{2^M}{M} \\ &= \sum_{i=1}^{M} \frac{2^i}{i}.\end{aligned}$$

Na kraju dobivamo da je očekivano vrijeme čekanja da u urni budu sve plave kuglice jednako

$$E[\gamma_M(M)] = \frac{M}{2}\sum_{i=1}^{M} \frac{2^i}{i}.$$

3.2 Primjene Ehrenfestovog modela urni

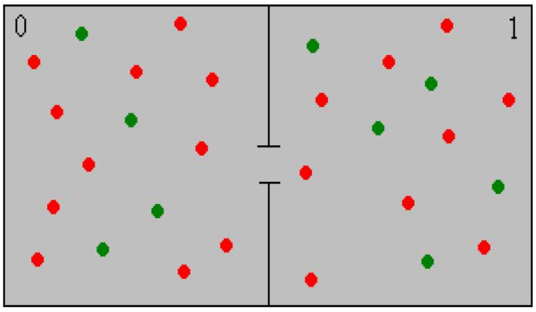
Na kraju ovog poglavlja navest ćemo gdje se Ehrenfestov model urni primjenjuje. Jedna od primjena Ehrenfestovog modela urni je model za miješanje čestica dvije povezane plinske komore, A i B. U svakoj komori postoji niz čestica te se u svakom koraku pokusa na slučajan način bira čestica iz populacije dviju komora. Odabrana čestica iz plinske komore A prelazi u komoru B i obrnuto. Situaciju je moguće usporediti s urnom koja sadrži dvije vrste kuglica, bijele koje predstavljaju čestice u komori A i plave koje predstavljaju čestice u komori B. Promjenu pozicije čestice iz jedne u drugu komoru modeliramo kuglicom u urni koja dobiva suprotnu boju. Ako kod modela miješanja plinova govorimo o problemu čekanja ili vremenu zaustavljanja, postavlja se pitanje je li u početku sav plin u jednoj komori te koliko je očekivano vrijeme čekanja da dobijemo takvu podjelu u kojoj polovica od ukupne količine plina u svakoj komori.

Ehrenfestov model urni koristi se i kao model za izmjenu topline između dva izolirana tijela nejednakih temperatura. U tom slučaju temperature tijela su simbolizirane brojem kuglica u urnama, a izmjena topline je slučajna kao u kinetičkoj teoriji plinova.

Također, iz ovog modela može se izvesti Newtonov zakon hlađenja.

4 Bernoulli-Laplaceov model urni

U ovom poglavlju reći ćemo nešto više o Bernoulli-Laplaceovom modelu urni koji se također može formulirati kao Ehrenfestov model pomoću urni i kuglica. Ovo je jednostavan diskretan model za difuziju dva nestlačljiva plina između dva spremnika. Pretpostavimo da imamo dvije urne s oznakama 0 i 1. Urna 0 sadrži j kuglica, a urna 1 sadrži k kuglica gdje su $j, k \in \mathbb{N}$. Od $j+k$ kuglica, r je crvenih, a preostalih $j+k-r$ su zelene. Stoga vrijedi da je $r \in \mathbb{N}$ i $0 < r < j+k$. U svakom diskretnom trenutku, neovisno o prošlom izvlačenju na slučajan način bira se kuglica iz svake urne te se odabrane kuglice stavljaju u suprotne urne (mijenjaju mjesta). Kuglice različitih boja odgovaraju molekulama različitih vrsta, urne predstavljaju spremnike, a nestlačljivo svojstvo očituje se u činjenici da broj kuglica u svakoj urni s vremenom postaje stabilan.



Slika 3: Bernoulli-Laplaceov model urni

Napomena 2. *Ako s X_n označimo broj crvenih kuglica u urni 1 u trenutku $n \in \mathbb{N}$ onda je:*

(*i*) $k - X_n$ *broj zelenih kuglica u urni 1 u trenutku n*

(*ii*) $r - X_n$ *broj crvenih kuglica u urni 0 u trenutku n*

(*iii*) $j - r + X_n$ *broj zelenih kuglica u urni 0 u trenutku n.*

Propozicija 2. $(X_n, n \in \mathbb{N})$ *je Bernoulli-Laplaceov Markovljev lanac u diskretnom vremenu sa skupom stanja $S = \{max\{0, r-j\}, \dots, min\{k, r\}\}$ i matricom prijelaznih vjerojatnosti P danom s*

$$P(x, x-1) = \frac{(j-r+x)x}{jk},$$

$$P(x,x) = \frac{(r-x)x + (j-r+x)(k-x)}{jk},$$

$$P(x,x+1) = \frac{(r-x)(k-x)}{jk}; \quad x \in S,$$

$$P(x_1,x_2) = 0, \quad |x_1 - x_2| > 1, \quad x_1, x_2 \in S.$$

Dokaz: Ako pogledamo skup stanja S prema prethodnoj napomeni jasno je da broj crvenih kuglica x u urni 1 mora zadovoljavati nejednakosti $x \geqslant 0$, $x \leqslant k$, $x \leqslant r$ i $x \geqslant r - j$.
Markovljevo svojstvo jasno je iz modela.
Za vjerojatnost prijelaza iz stanja x u stanje $x-1$ moramo odabrati zelenu kuglicu iz urne 0 i crvenu iz urne 1. Vjerojatnosti ovih događaja su $\frac{j-r+x}{j}$ i $\frac{x}{k}$ za $x, x-1 \in S$, a događaji su nezavisni pa dobivamo

$$P(x,x-1) = \frac{(j-r+x)x}{jk}.$$

Kako bi broj crvenih kuglica u urni 1 ostao nepromijenjen moramo odabrati crvenu kuglicu iz obje urne ili zelenu kuglicu iz obje urne, pa je vjerojatnost da smo odabrali crvenu kuglicu iz obje urne jednaka $\frac{x(r-x)}{kj}$ za $x \in S$, dok je vjerojatnost da smo odabrali zelenu kuglicu iz obje urne jednaka $\frac{(k-x)(j-r+x)}{kj}$ za $x \in S$, a događaji su nezavisni. U konačnici imamo da je

$$P(x,x) = \frac{(r-x)x + (j-r+x)(k-x)}{jk}.$$

Na kraju, da bi smo prešli iz stanja x u stanje $x+1$ moramo odabrati crvenu kuglicu iz urne 0 i zelenu iz urne 1. Vjerojatnosti tih događaja su $\frac{r-x}{j}$ i $\frac{k-x}{k}$, za $x, x+1 \in S$, a događaji su nezavisni pa na kraju imamo da je

$$P(x,x+1) = \frac{(r-x)(k-x)}{jk}.$$

Također vrijedi da je $P(x,x) = 1 - P(x,x-1) - P(x,x+1)$. $\square$
Zbog broja parametara ovo je prilično kompliciran model. Razmotrimo neke specifične slučajeve.

Napomena 3. *(i) Ako je $j = k$, odnosno ako svaka urna ima isti broj kuglica onda je skup stanja $S = \{max\{0, r-k\}, \ldots, min\{k, r\}\}$, a matrica prijelaznih* *vjerojatnosti dana je s*

$$P(x, x-1) = \frac{(k-r+x)x}{k^2},$$
$$P(x, x) = \frac{(r-x)x + (k-r+x)(k-x)}{k^2},$$
$$P(x, x+1) = \frac{(r-x)(k-x)}{k^2}, x \in S,$$
$$P(x_1, x_2) = 0, \ |x_1 - x_2| > 1, \ x_1, x_2 \in S.$$

(ii) *Ako je $r = j$, odnosno ako je ukupan broj crvenih kuglica jednak broju kuglica u urni 0 onda je skup stanja $S = \{0, \dots, min\{j, k\}\}$, a matrica prijelaznih vjerojatnosti dana je s*

$$P(x, x-1) = \frac{x^2}{jk},$$
$$P(x, x) = \frac{x(j+k-2x)}{jk},$$
$$P(x, x+1) = \frac{(j-x)(k-x)}{jk}, x \in S,$$
$$P(x_1, x_2) = 0, \ |x_1 - x_2| > 1, \ x_1, x_2 \in S.$$

(iii) *Ako je $r = k$, odnosno ako je ukupan broj crvenih kuglica jednak broju kuglica u urni 1 onda je skup stanja $S = \{max\{0, k-j\}, \dots, k\}$, a matrica prijelaznih vjerojatnosti dana je s*

$$P(x, x-1) = \frac{(j-k+x)x}{jk},$$
$$P(x, x) = \frac{(k-x)(j-k+2x)}{jk},$$
$$P(x, x+1) = \frac{(k-x)^2}{jk}, x \in S,$$
$$P(x_1, x_2) = 0, \ |x_1 - x_2| > 1, \ x_1, x_2 \in S.$$

(iv) *Ako je $j = k = r$, odnosno ako svaka urna ima isti broj kuglica i on je jednak ukupnom broju crvenih kuglica onda je skup stanja $S = \{0, \dots, k\}$, a matrica* *prijelaznih vjerojatnosti dana je s*

$$\begin{aligned} P(x, x-1) &= \frac{x^2}{k^2}, \\ P(x, x) &= \frac{2x(k-x)}{k^2}, \\ P(x, x+1) &= \frac{(k-x)^2}{k^2}, x \in S, \\ P(x_1, x_2) &= 0, \quad |x_1 - x_2| > 1, \quad x_1, x_2 \in S. \end{aligned}$$

Riješimo sada jedan konkretan primjer.

Primjer 3. *Imamo Bernoulli-Laplaceov Markovljev lanac gdje je $j = 10$, $k = 5$ i $r = 4$. Odredimo skup stanja S i matricu prijelaznih vjerojatnosti M.*
Rješenje: *Jasno je da je skup stanja $S = \{0, 1, 2, 3, 4\}$. Kako je $j = 10$ ukupan broj kuglica u urni 0, $k = 5$ ukupan broj kuglica u urni 1, a $r = 5$ ukupan broj crvenih kuglica onda iz $j + k - r$ dobivamo da je ukupan broj zelenih kuglica 11. Odredimo sada matricu prijelaznih vjerojatnosti.*

$$\begin{aligned} p_{00} &= P(X_{n+1} = 0 | X_n = 0) = \frac{\binom{5}{1}}{\binom{5}{1}} \frac{\binom{6}{1}}{\binom{10}{1}} = \frac{30}{50} \\ p_{01} &= P(X_{n+1} = 1 | X_n = 0) = \frac{\binom{5}{1}}{\binom{5}{1}} \frac{\binom{4}{1}}{\binom{10}{1}} = \frac{20}{50} \\ p_{02} &= P(X_{n+1} = 2 | X_n = 0) = 0 \\ p_{03} &= P(X_{n+1} = 3 | X_n = 0) = 0 \\ p_{04} &= P(X_{n+1} = 4 | X_n = 0) = 0 \\ p_{10} &= P(X_{n+1} = 0 | X_n = 1) = \frac{\binom{1}{1}}{\binom{5}{1}} \frac{\binom{7}{1}}{\binom{10}{1}} = \frac{7}{50} \\ p_{11} &= P(X_{n+1} = 1 | X_n = 1) = \frac{\binom{1}{1}}{\binom{5}{1}} \frac{\binom{3}{1}}{\binom{10}{1}} + \frac{\binom{4}{1}}{\binom{5}{1}} \frac{\binom{7}{1}}{\binom{10}{1}} = \frac{31}{50} \\ p_{12} &= P(X_{n+1} = 2 | X_n = 1) = \frac{\binom{4}{1}}{\binom{5}{1}} \frac{\binom{3}{1}}{\binom{10}{1}} = \frac{12}{50} \\ p_{13} &= P(X_{n+1} = 3 | X_n = 0) = 0 \\ p_{14} &= P(X_{n+1} = 4 | X_n = 0) = 0 \\ p_{20} &= P(X_{n+1} = 0 | X_n = 2) = 0 \\ p_{21} &= P(X_{n+1} = 1 | X_n = 2) = \frac{\binom{2}{1}}{\binom{5}{1}} \frac{\binom{8}{1}}{\binom{10}{1}} = \frac{16}{50} \end{aligned}$$

$$p_{22} = P(X_{n+1} = 2|X_n = 2) = \frac{\binom{2}{1}}{\binom{5}{1}}\frac{\binom{2}{1}}{\binom{10}{1}} + \frac{\binom{3}{1}}{\binom{5}{1}}\frac{\binom{8}{1}}{\binom{10}{1}} = \frac{28}{50}$$
$$p_{23} = P(X_{n+1} = 3|X_n = 2) = \frac{\binom{3}{1}}{\binom{5}{1}}\frac{\binom{2}{1}}{\binom{10}{1}} = \frac{6}{50}$$
$$p_{24} = P(X_{n+1} = 4|X_n = 2) = 0$$
$$p_{30} = P(X_{n+1} = 0|X_n = 3) = 0$$
$$p_{31} = P(X_{n+1} = 1|X_n = 3) = 0$$
$$p_{32} = P(X_{n+1} = 2|X_n = 3) = \frac{\binom{3}{1}}{\binom{5}{1}}\frac{\binom{9}{1}}{\binom{10}{1}} = \frac{27}{50}$$
$$p_{33} = P(X_{n+1} = 3|X_n = 3) = \frac{\binom{3}{1}}{\binom{5}{1}}\frac{\binom{1}{1}}{\binom{10}{1}} + \frac{\binom{2}{1}}{\binom{5}{1}}\frac{\binom{9}{1}}{\binom{10}{1}} = \frac{21}{50}$$
$$p_{34} = P(X_{n+1} = 4|X_n = 3) = \frac{\binom{2}{1}}{\binom{5}{1}}\frac{\binom{1}{1}}{\binom{10}{1}} = \frac{2}{10}$$
$$p_{40} = P(X_{n+1} = 0|X_n = 4) = 0$$
$$p_{41} = P(X_{n+1} = 1|X_n = 4) = 0$$
$$p_{42} = P(X_{n+1} = 2|X_n = 4) = 0$$
$$p_{43} = P(X_{n+1} = 3|X_n = 4) = \frac{\binom{4}{1}}{\binom{5}{1}}\frac{\binom{10}{1}}{\binom{10}{1}} = \frac{40}{50}$$
$$p_{44} = P(X_{n+1} = 4|X_n = 4) = \frac{\binom{4}{1}}{\binom{5}{1}}\frac{\binom{0}{1}}{\binom{10}{1}} + \frac{\binom{1}{1}}{\binom{5}{1}}\frac{\binom{10}{1}}{\binom{10}{1}} = \frac{10}{50}$$

U konačnici, matrica prijelaznih vjerojatnosti je oblika

$$M = \frac{1}{50}\begin{bmatrix} 30 & 20 & 0 & 0 & 0 \\ 7 & 31 & 12 & 0 & 0 \\ 0 & 16 & 28 & 6 & 0 \\ 0 & 0 & 27 & 21 & 2 \\ 0 & 0 & 0 & 40 & 10 \end{bmatrix}.$$

Na kraju poglavlja reći ćemo nešto o graničnoj i stacionarnoj distribuciji ovog Markovljevog lanca.

Propozicija 3. *Bernoulli-Laplaceov Markovljev lanac je ireducibilan.*

Dokaz: U našem slučaju imamo da je $P(x, x-1) > 0, \forall x, x-1 \in S$ i $P(x, x+1) > 0, \forall x, x+1 \in S$. S obzirom da iz svakog stanja možemo doći u bilo koje drugo stanje lanac je ireducibilan. $\square$

Propozicija 4. *Osim u trivijalnom slučaju kada je $j = k = r = 1$, Bernoulli-Laplaceov Markovljev lanac je aperiodičan.*

Dokaz: Markovljev lanac koji se nalazi u stanju x ostaje u stanju x s pozitivnom vjerojatnošću, odnosno imamo da je $P(x,x) > 0$, pa osim kada je $j = k = r = 1$, stanje x je aperiodično. Budući da je lanac ireducibilan, prema Lemi 1, sva stanja su aperiodična. □

S obzirom na to da je stacionarna distribucija ovog Markovljevog lanca hipergeometrijska, navedimo najprije njezinu općenitu definiciju.

Definicija 13. *Diskretna slučajna varijabla X ima hipergeometrijsku distribuciju s parametrima N, M i n, $N, M, n \in \mathbb{N}$, ako prima vrijednosti iz skupa $\mathcal{R}(X) = \{k \in \mathbb{N}_0 : max(0, n - N + M) \leqslant k \leqslant min(n, M)\}$ s vjerojatnostima*

$$p_k = P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}.$$

Očekivanje hipergeometrijske distribucije je

$$E[X] = \frac{nM}{N}, \tag{7}$$

a njezina varijanca je

$$VarX = \frac{Mn(N-M)(N-n)}{N^2(N-1)}. \tag{8}$$

Napomena 4. *Stacionarna distribucija Bernoulli-Laplaceovog Markovljevog lanca je hipergeometrijska s parametrima $j + k$,r i k.*

(*i*) *Gustoća je jednaka*

$$f(x) = \begin{cases} \dfrac{\binom{r}{x}\binom{j+k-r}{k-x}}{\binom{j+k}{k}}, & x \in \mathcal{R}(X) \\ 0, & inače \end{cases}$$

gdje je $\mathcal{R}(X) = \{x \in \mathbb{N}_0 : max(0, n - N + M) \leqslant x \leqslant min(n, M)\}$.

Objašnjenje: *Skup iz kojeg vršimo odabir elemenata je konačan i sastoji se od točno $j+k$ elemenata (kuglica) od kojih je r tipa 1 (crvene boje), a $j+k-r$ tipa*

2 (zelene boje). Pretpostavimo da smo na slučajan način iz tog skupa odabrali $k < k + j$ *elemenata i to bez vraćanja prethodno izvučenih elemenata. Tada broj izvučenih elemenata tipa 1 modeliramo hipergeometrijskom distribucijom s parametrima* $j + k$, r *i* k *pa je gustoća jednaka upravo gore navedenoj što zaključujemo iz prethodne definicije hipergeometrijske distribucije.*

(ii) *Očekivanje je* $\mu = k\dfrac{r}{j+k}$
Objašnjenje: *Očekivanje dobijemo tako da parametre ove hipergeometrijske distribucije uvrstimo u (7).*

(iii) *Varijanca je* $\sigma^2 = \dfrac{jkr(j+k-r)}{(j+k)^2(j+k-1)}$
Objašnjenje: *Varijancu dobijemo tako da parametre ove hipergeometrijske distribucije uvrstimo u (8).*

5 Urne u bioznanosti

Modeli urni također pronalaze upotrebu i primjenu u raznim granama bioznanosti kao što su evolucija vrste i zdravstvo poput epidemiologije, ekologije i kliničkih ispitivanja. U sljedećim potpoglavljima prikazat ćemo nekoliko primjera urni preko Markovljevih lanaca koje se koriste kao modeli za evoluciju vrste. Za razliku od standardnih shema Polya urni gdje postoji fiksan skup dopuštenih boja koje se mogu pojaviti u urni, u modelima za evoluciju vrste moramo dopustiti da se broj boja povećava s vremenom te moramo razmotriti beskonačni skup dopuštenih boja. Razlog tomu je taj što su u modelima za evoluciju vrste, životinje određene vrste predstavljene pomoću kuglica iste boje, a s obzirom na evoluciju, u nekom se trenutku može pojaviti nova vrsta (nova boja). Ovi modeli se smatraju i razumnim prikazom genetskih alela, gdje se s vremena na vrijeme pojavljuje novi alelni tip (nova boja) kao rezultat mutacije. Upravo ćemo u sljedećem potpoglavlju navesti jedan takav model.

5.1 Wright-Fisherov model urni alela bez mutacija

Model koji ćemo navesti u ovom poglavlju nazvan je po ranim pionirima teorije populacijske genetike, Sewallu Wrightu u čijim djelima se spominje još 1922. te Ronaldu A. Fisheru koji ga spominje u svom radu u 1930.-ima. U modernoj literaturi naziva se jednostavnim Wright-Fisherovim modelom te je jedan od najjednostavnijih genetskih modela za fiksnu populaciju alela (dvije vrste) bez selekcije, mutacije ili spola. Neka **A** i **a** označavaju dva alela (dva alternativna gena koja određuju istu osobinu, a u stanicama uvijek dolaze u paru) koja se zbog ponašanja kromosoma tijekom mejoze odvajaju jedan od drugoga i smještaju u različite gamete (spolne stanice). Neka je X_n slučajna varijabla kojom modeliramo broj alela **A** u n-toj populaciji s N diploidnih pojedinaca (može imati samo dva alela istog gena) koja ima skup stanja $S = \{0, 1, 2, \ldots, 2N\}$. Slučajna varijabla X_{n+1}, kojom modeliramo broj alela **A** u $(n + 1)$-oj populaciji, ovisi samo o broju alela **A** u n-toj populaciji, znači ispunjeno je Markovljevo svojstvo pa znamo da je slučajani proces $\{X_n, n \in \mathbb{N}_0\}$ Markovljev lanac u diskretnom vremenu. Generacija $(n+1)$ sastoji se od $2N$ alela i nastaje provođenjem $2N$ nezavisnih slučajnih odabira od po jednog alela iz početne generacije n, s tim da se nakon svakog odabira izvučeni alel "vraća" u roditeljsku populaciju. Ovaj proces opisuje binomno uzorkovanje alela svake generacije što nam omogućava da matricu prijelaznih vjerojatnosti Markovljevog lanca zapišemo kao

$$p_{ij} = \binom{2N}{j}\Big(\frac{i}{2N}\Big)^j\Big(1-\frac{i}{2N}\Big)^{2N-j}, i,j \in \{0,1,\ldots,2N\}.$$

Drugim riječima, vjerojatnost da imamo j alela $\mathbf{A}$ u trenutku $(n+1)$ kada imamo i alela $\mathbf{A}$ u trenutku n dana je s p_{ij}. Na kraju, broj alela $\mathbf{A}$ u populaciji, postat će 0 (što interpretiramo kao nestajanje alela $\mathbf{A}$) ili $2N$ (što interpretiramo kao nestajanje alela $\mathbf{a}$). Jednom izgubljen alel iz populacije nikada se ne vraća (jer pretpostavljamo da se mutacije u populaciji ne mogu dogoditi), tako da su stanja 0 i $2N$ apsorbirajuća stanja ovog Markovljevog lanca. Nakon što lanac uđe u jedno od tih stanja ne može ih više napustiti. Ovaj model također možemo prikazati pomoću kuglica i urne. Započinjemo s dvije boje, koje predstavljaju dva alela. Pretpostavljamo da u urni ima m kuglica od kojih je i bijelih (s jednim genskim karakteristikama) i $m-i$ plavih (s drugim genskim karakteristikama). Kuglice se uzorkuju zamjenom, m puta, kako bi se pružila prilika da očekivano pojavljivanje svake kuglice bude jedan. Kad god se u uzorku pojavi kuglica, u novu urnu koja predstavlja novu generaciju jedinki postavlja se kuglica iste boje. Ako je uzorkovana kuglica bijela, bijelu kuglicu stavljamo u novu urnu, a ako je uzorkovana kuglicu plava, u novu urnu stavimo plavu kuglicu. Broj bijelih kuglica koje se pojavljuju u uzorku veličine m je $B(m, \frac{i}{m})$, a očekivani broj bijelih kuglica u novoj urni je $m \times \frac{i}{m} = i$ što znači da nema promjene u omjeru bijelih kuglica uzastopnih urni (generacija). Nakon što se nova urna ispuni, postupak se ponavlja kako bi se proizvela treća urna i tako dalje. Jasno je da postoji pozitivna vjerojatnost da nova urna postane jednobojna. Jednom kada se postigne jednobojnost, urne će zauvijek ostati iste. Pogledajmo sada to na primjeru.

Primjer 4. *Imamo urnu koja može sadržavati bijele i plave kuglice. Pretpostavimo da na početku u urni imamo jednu bijelu i jednu plavu kuglicu. Evoluciju predstavljamo pomoću Markovljevog lanca s tri stanja gdje je svako stanje opisano brojem bijelih kuglica u urni. Pretpostavimo da krećemo iz Stanja 1. Iz Stanja 1 možemo prijeći u Stanje 0 ako najprije iz urne odaberemo plavu kuglicu (vjerojatnost tog događaja je $\frac{1}{2}$) te u novu urnu postavimo njezinu kopiju, a odabranu kuglicu vratimo u početnu urnu te nakon toga iz početne urne odaberemo opet plavu kuglicu (vjerojatnost $\frac{1}{2}$) te njezinu kopiju također postavimo u urnu u kojoj je prva kopija. Tako je vjerojatnost tog događaja $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Analogno, iz Stanja 1 možemo prijeći u Stanje 2 ako se u uzorku pojave dvije bijele kuglice (vjerojatnost tog događaja je*

$\frac{1}{4}$), *te možemo ostati u Stanju 1 ako se u uzorku pojavi samo jedna bijela kuglica (vjerojatnost tog događaja je* $\frac{1}{2}$). *Matrica prijelaznih vjerojatnosti dana je s*

$$M = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 1 \end{pmatrix}.$$

Jasno je da je izravnom indukcijom matrica n-koračnih prijelaznih vjerojatnosti dana s

$$M^n = \begin{pmatrix} 1 & 0 & 0 \\ \frac{2^n-1}{2^{n+1}} & \frac{2}{2^{n+1}} & \frac{2^n-1}{2^{n+1}} \\ 0 & 0 & 1 \end{pmatrix}.$$

Kako je početna distribucija Markovljevog lanca $\lambda = (0, 1, 0)$, *iz*

$$\pi_n = \lambda M^n$$

dobivamo da je (marginalna) distribucija lanca nakon n koraka jednaka

$$\begin{aligned} \pi_n &= (0 \quad 1 \quad 0) \begin{pmatrix} 1 & 0 & 0 \\ \frac{2^n-1}{2^{n+1}} & \frac{2}{2^{n+1}} & \frac{2^n-1}{2^{n+1}} \\ 0 & 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} \frac{2^n-1}{2^{n+1}} & \frac{2}{2^{n+1}} & \frac{2^n-1}{2^{n+1}} \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} + \begin{pmatrix} -\frac{1}{2^{n+1}} & \frac{1}{2^n} & -\frac{1}{2^{n+1}} \end{pmatrix}. \end{aligned}$$

Vidimo da ovaj vektor redak konvergira po komponentama prema $(\frac{1}{2} \quad 0 \quad \frac{1}{2})$. *U konačnici, gotovo sigurno se postiže samo jedna boja u urni te će s jednakom vjerojatnošću urna sadržavati samo bijele ili samo plave kuglice. Također je jasno da je* $(p \quad 0 \quad 1-p)$, *za bilo koji* $p \in [0, 1]$ *stacionarna distribucija, jer je ovaj vektor redak rješenje sustava* $\pi M = \pi$, *za* $\pi = (\pi_0 \quad \pi_1 \quad \pi_2)$ *s* $\pi_0 + \pi_1 + \pi_2 = 1$.

Prethodni primjer može se generalizirati na ukupno m kuglica od kojih je i bijelih, i gotovo sigurno dobivamo jednobojnu urnu. Vjerojatnost da su sve kuglice u urni bijele je $\frac{i}{m}$, što je jednako početnom udjelu bijelih kuglica u urni, a $(p \quad 0 \ldots 0 \quad 1-p)$ je stacionarna distribucija promatranog lanca.

5.2 Wright-Fisherov model urni alela s mutacijama

U ovom potpoglavlju kratko ćemo se osvrnuti na Wright-Fisherov model s mutacijama. Pretpostavimo, da nakon što odaberemo alel iz generacije n i prije nego što ga iskoristimo za formiranje nove populacije u generaciji $n+1$, **a** postaje **A** s vjerojatnošću u i **A** postaje **a** s vjerojatnošću v, tj. događaju se mutacije s vjerojatnostima u i v, $u, v \in (0,1)$. Vjerojatnost da postoji j alela **A** u trenutku $n+1$ kada ih je i u trenutku n dana je formulom

$$p_{ij} = \binom{2N}{j} p_i^j (1-p_i)^{2N-j},$$

gdje je p_i vjerojatnost odabira alela **A** (kada imamo i alela **A**) dana s

$$p_i = \frac{i}{2N}(1-v) + \frac{2N-i}{2N}u.$$

Dakle, ili izvučemo alel **A** i on ne mutira ili izvučemo alel **a** i on mutira u alel **A**. Posljedica postojanja mutacija je nestajanje apsorbirajućih stanja 0 i $2N$, tako da se u ovom modelu prostor stanja sastoji samo od jedne klase komuniciranja pa je lanac ireducibilan (Definicija 5). Prijelazna vjerojatnost za model s mutacijom je $p_{ij} > 0$, jer više nema apsorbirajućih stanja, za svaki i, j pa je ovaj Markovljev lanac aperiodičan (Definicija 12). Kako je skup stanja konačan i lanac je ireducibilan, slijedi da je povratan, tj. svako njegovo stanje je povratno (Propozicija 1 i Teorem 4). Lanac je i pozitivno povratan jer je očekivano vrijeme potrebno Markovljevom lancu da iz stanja i opet dođe do i konačno (Definicija 8). Budući je lanac ireducibilan i pozitivno povratan, slijedi da ima jedinstvenu stacionarnu distribuciju (Napomena 1). A budući da je lanac ireducibilan, aperiodičan i ima stacionarnu distribuciju, slijedi da je ta stacionarna distribucija upravo njegova granična distribucija (Teorem 8).

Napomena 5. *Pretpostavimo da je veličina populacije N velika i neka je $q = 4Nu$, $r = 4Nv$. Tada se stacionarna distribucija za Wright-Fisherov model, kada je reskalirana unutar intervala $[0,1]$, može dobro aproksimirati Beta(q,r) distribucijom s funkcijom gustoće*

$$f(x) = c_{q,r} x^{q-1}(1-x)^{r-1}$$

pri čemu je $c_{q,r}$ konstanta koja osigurava da vrijedi $\int_0^1 f(x)dx = 1$.

5.3 Pogrešno kopiranje gena

U jednostavnom Wright-Fisherovom modelu iz prethodnog potpoglavlja, u urni postoji m kuglica koje mogu biti u jednoj od dvije boje i m kuglica su uzorkovane zamjenom. U međuvremenu se druga urna ispunjava točnom kopijom uzorka. Kad god se u uzorku pojavi kuglica, u novu urnu se postavlja kuglica iste boje. Kada se nova urna ispuni, postupak se ponavlja kako bi se ispunila još jedna urna. Kuglice predstavljaju gene. U stvarnosti kopiranje svojstava gena nije savršena operacija i postoji šansa za pogrešno kopiranje. U pojednostavljenom modelu dopuštamo pogrešno kopiranje boja. Kad se kuglica uzorkuje, postoji mala vjerojatnost da će "prijeći" u drugu boju. Kad se u uzorku pojavi bijela kuglica, u novu urnu stavljamo bijelu kuglicu s vjerojatnošću $1-\alpha$, inače u novu urnu stavljamo plavu kuglicu s malom vjerojatnošću α. I obrnuto, kada se u uzorku pojavi plava kuglica, u novu urnu stavljamo plavu kuglicu s vjerojatnošću $1-\beta$, inače u novu urnu stavljamo bijelu kuglicu s malom vjerojatnošću β. Operacija se zatim ponavlja s budućim generacijama novijih urni. Sada ćemo ponovo pogledati primjer iz prethodnog poglavlja, ali ćemo u njemu dopustiti pogrešno kopiranje.

Primjer 5. *Neka Markovljev lanac ima tri stanja koja označavaju broj bijelih kuglica u urni i neka se u početnoj urni nalaze dvije plave kuglice. U novu urnu možemo postaviti dvije kuglice plave boje (s vjerojatnošću* $(1-\beta)^2$*), obje plave kuglice možemo pretvoriti u bijele (s vjerojatnošću* β^2*) ili možemo postaviti jednu plavu i jednu bijelu kuglicu (s vjerojatnošću* $2\beta(1-\beta)$*). Ako se nalazimo u stanju 2 (u urni imamo dvije bijele kuglice) možemo prijeći u stanja 0, 1, 2 s vjerojatnostima* $\alpha^2, 2\alpha(1-\alpha), (1-\alpha)^2$*, redom. U stanju 1 imamo jednu bijelu i jednu plavu kuglicu. Možemo prijeći u stanje s dvije plave (nula bijelih) kuglica na četiri različita načina, ovisno o slijedu boja u uzorku i o tome jesu li pogrešno kopirani ili ne. Bijelu i plavu boju u uzorku predstavljamo s* W *i* B*, a postupak pogrešnog kopiranja s Da*(Y) *ili Ne*(N)*, i primijetimo da možemo prijeći iz stanja 1 u stanje 0 ako imamo jedan od četiri događaja:*

$$WWYY, \qquad WBYN, \qquad BWNY, \qquad BBNN,$$

čije su vjerojatnosti

$$\frac{1}{4}\alpha^2, \qquad \frac{1}{4}\alpha(1-\beta), \qquad \frac{1}{4}(1-\beta)\alpha, \qquad \frac{1}{4}(1-\beta)^2.$$

Zbroj ovih vjerojatnosti je $M_{1,0}$ *element u matrici prijelaznih vjerojatnosti* M*. Element* $M_{1,2}$ *je simetričan s* $M_{1,0}$*, a* $M_{1,1} = 1 - M_{1,0} - M_{1,2}$*. Matrica prijelaznih vjerojatnosti dana je s*

$$M = \begin{pmatrix} (1-\beta)^2 & 2\beta(1-\beta) & \beta^2 \\ \frac{1}{4}(1+\alpha-\beta)^2 & \frac{1}{2}-\frac{1}{2}(\alpha-\beta)^2 & \frac{1}{4}(1-\alpha+\beta)^2 \\ \alpha^2 & 2\alpha(1-\alpha) & (1-\alpha)^2 \end{pmatrix}.$$

Stacionarna distribucija rješenje je sustava $\pi M = \pi$, s komponentama koje zbrojene daju 1. Iz

$$\pi M = \pi,$$
$$(M - I)^T \pi^T = 0$$

dobivamo:

$$\pi_0 = \frac{\alpha((\alpha-\beta)(\alpha+\beta-2)-1)}{(\alpha+\beta)((\alpha+\beta)(\alpha+\beta-2)-1)},$$

$$\pi_1 = \frac{\beta((\beta-\alpha)(\alpha+\beta-2)-1)}{(\alpha+\beta)((\alpha+\beta)(\alpha+\beta-2)-1)},$$

$$\pi_2 = \frac{4\alpha\beta(\alpha+\beta-2)}{(\alpha+\beta)((\alpha+\beta)(\alpha+\beta-2)-1)}.$$

Očekivani broj bijelih kuglica u novoj urni u stanju ravnoteže možemo pronaći izravnim razmatranjem. Naime, u stanju ravnoteže očekivani broj pojavljivanja svake kuglice u uzorku je jedan. Ukoliko izvučemo bijelu kuglicu vjerojatnost da ćemo u novoj urni opet imati bijelu kuglicu je $1 - \alpha$. Ukoliko izvučemo plavu kuglicu vjerojatnost da ćemo u novoj urni nakon pogrešnog kopiranja imati bijelu kuglicu je β. S obzirom da u urni možemo imati dvije kuglice kako bi dobili broj plavih kuglica od te dvije oduzmemo očekivani broj bijelih kuglica te zatim taj dobiveni broj plavih kuglica pomnožimo s vjerojatnošću pogrešnog kopiranja β pa imamo da je očekivani broj bijelih kuglica u novoj urni:

$$E[W] = (1-\alpha)E[W] + \beta(2 - E[W])$$
$$E[W](1 - 1 + \alpha + \beta) = 2\beta$$
$$E[W] = \frac{2\beta}{\alpha+\beta}.$$

Ovo se može generalizirati i na veći primjer s m kuglica u početnoj urni, ali matrica prijelaznih vjerojatnosti postaje prilično složena. Analogno, očekivani broj bijelih kuglica u novoj urni je

$$E[W] = \frac{m\beta}{\alpha+\beta}.$$

Tipične vrijednosti vjerojatnosti pogrešnog kopiranja α i β su reda veličine 10^{-5}. Na primjer, s $\alpha = 10^{-5}$ i $\beta = 3 \times 10^{-5}$ očekivani udio bijelih kuglica je $\frac{3}{4}$.
Više informacija o ovakvim modelima može se pronaći u [4].

Literatura

[1] N. L. Johnson, S. Kotz, Urn Models and Their Application, John Wiley and Sons, 1977.

[2] S. Karlin, H. M. Taylor, A Second Course in Stochastic Processes, Academic Press, 1981.

[3] W.Lexis, Treatises on Population and Social Statistics, Izdavačka kuća Gustava Fichera u Jeni, 1903.

[4] H.M. Mahmoud, Polya Urn Models,Chapman and Hall book/CRC, 2008.

[5] S. Šebek, Modeli urni i martingalne metode, Diplomski rad, Prirodoslovno-matematički fakultet sveučilište u Zagrebu, 2014.

[6] Z.Vondraček, Slučajni pocesi, Prirodoslovno-matematički fakultet Sveučilište u Zagrebu, skripta
https://web.math.pmf.unizg.hr/ vondra/sp18-predavanja.html

Sažetak

U ovom radu bavili smo se modelima urni koji su vrlo stari te se spominju još u Starom zavjetu i židovskoj teologiji. Imaju bezbroj primjena u ekonomiji, industriji, tehnologiji, bioznanostima i zdravstvu. Kako je cilj rada bio modelirati određene modele urni Markovljevim lancima najprije smo naveli osnovne tvrdnje vezane za njih. Detaljno smo opisali Ehrenfestov i Bernoulli-Laplaceov model urni preko Markovljevih lanaca, naveli njihove primjene i dali povijesni pregled. Za kraj rekli smo nešto o urnama u bioznanosti, konkretno u evoluciji vrsta opisujući Wright-Fisherov model sa i bez selekcije, spola i mutacije preko Markovljevih lanaca.

Ključne riječi: Markovljevi lanci, urna, kuglice, izvlačenje, Ehrenfestov model urni, Bernoulli-Laplaceov model urni, Wright-Fisherov model, vrijeme zaustavljanja

Summary

In this diploma thesis we are dealing with urn models, very old models mentioned in Old Testament and Jewish theology. Urn models have countless ways of applications in economy, industry, technology, biosciences and various disciplines regarding health. The aim of the papper is to model some urn models via Markov chains so we first stated the basic results regarding the theory of Markov chains. Ehrenfest urn model and Bernoulli-Laplaces urn model are described in details in this paper through Markov chains, and also we have described their applications and historical overview. Finally, we introduced urn models in biosciences, specifically in the evolution of species by describing Wright-Fisher model with and without selection, sex and mutations through Markov chains.

Keywords: Markov chains, urn models, balls, drawing, Ehrenfests urn model, Bernoulli-Laplaces urn model, Wright-Fisher model, stopping time

Životopis

Rođena sam 5. rujna 1995. godine u Novoj Gradiški. Pohađala sam Osnovnu školu Ljudevita Gaja u Novoj Gradiški koju sam završila 2010. godine nakon čega upisujem Opću gimnaziju u Novoj Gradiški. 2014. godine upisala sam preddiplomski studij Odjela za matematiku u Osijeku. Preddiplomski studij završila sam 2017. godine s temom završnog rada Centralni granični teorem pod mentorstvom doc.dr.sc. Slobodana Jelića, a nakon toga upisala sam diplomski studij na Odjelu za matematiku, smjer Financijska matematika i statistika. Tijekom diplomskog studija bila sam demonstrator iz kolegija Uvod u vjerojatnost i statistiku. Od 2017. godine do 2019. bila sam član Studentskog zbora Odjela za matematiku. 2018. godine sudjelujem na Zimskoj školi matematike s temom Geometrijski i statistički pristup određivanju vjerojatnosti istih događaja. Stručnu praksu obavljala sam u Privrednoj banci Zagreb na odjelu Customer Relationship Management u Zagrebu.