Fluktuacija jednostavne simetrične slučajne šetnje

Radoš, Borna

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Borna Radoš

FLUKTUACIJE JEDNOSTAVNE SIMETRIČNE SLUČAJNE ŠETNJE

Diplomski rad

Voditelj rada:
doc. dr. sc. Hrvoje Planinić

Zagreb, 2022

Ovaj diplomski rad obranjen je dana ____________________ pred ispitnim povjerenstvom u sastavu:

1. ____________________________, predsjednik
2. ____________________________, član
3. ____________________________, član

Povjerenstvo je rad ocijenilo ocjenom _______________.

Potpisi članova povjerenstva:

1. ____________________________
2. ____________________________
3. ____________________________

Zahvaljujem mentoru doc. dr. sc. Hrvoju Planiniću na velikoj podršci i pomoći pri izradi ovog diplomskog rada. Posebno hvala mojoj majci i obitelji na pomoći pri studiranju.

Sadržaj

Uvod

U ovom diplomskom radu proučavati ćemo jednostavne simetrične slučajne šetnje, točnije njihove fluktuacije. Pokazat ćemo da se javljaju neki iznenađujući rezultati. Obratit ćemo pozornost na vjerojatnosti određenih događaja poput vjerojatnosti povratka u ishodište, dolaska u određenu poziciju te vjerojatnost stalnog vodstva. Kroz primjere ćemo bolje razviti intuiciju ponašanja ovih slučajnih šetnji. Struktura diplomskog rada je slijedeća:

U prvom poglavlju ćemo se prisjetiti nekih definicija u vjerojatnosti koje ćemo koristiti kroz rad u teoremima i primjerima. Posebno ćemo obratiti pozornost na uvjetnu vjerojatnost te slučajne varijable i njihove distribucije.

U drugom poglavlju definiramo jednostavnu slučajnu simetričnu šetnju. Princip refleksije će nam biti koristan u nastavku poglavlja pri dokazivanju nekih teorema. Istražit ćemo kako se računa vjerojatnost povratka u ishodište, te prvog povratka u ishodište. Proučavamo specifičan slučaj kockarove propasti vezan za temu rada. Nadalje pokazat ćemo da je najveća vjerojatnost da se zadnji povratak u ishodište dogodio ili blizu početka ili kraja te da je za očekivat da je čestica provela većinu vremena u jednom od kvadranata. Vidjet ćemo da te slučajne varijable imaju arkus sinusovu distribuciju reda n. Proučit ćemo neke primjere za bolje dočarati ove vjerojatnosti. U nastavku poglavlja pokazujemo da je očekivanje kako neće biti mnogo promjena vodstva.

U trećem poglavlju ćemo prikazati rezultate odrađenih simulacija koristeći programski jezik R. Grafovima ćemo vizualno prezentirati rezultate koje ćemo zatim usporediti sa ranije iskazanim u radu i vidjeti jesmo li dobili očekivane rezultate. U prvom dijelu prikazano je 10 igara duljine 10000 koraka te njihovi grafovi. U drugom dijelu analiziramo dobivene igre, a naposlijetku analiziramo veći broj igara (500) kako bi bolje uvidjeli dobivene rezultate.

Poglavlje 1

Osnovni pojmovi i rezultati teorije vjerojatnosti

1.1 Vjerojatnosni prostor

Pokus ili eksperiment u vjerojatnosti će biti dobro definirana procedura (ima svoj uzrok i posljedicu/ishode) te ga možemo ponoviti beskonačno mnogo puta. Ishode pokusa ćemo nazivati još i elementarnim događajima. Tako će na primjer pri bacanju igraće kocke biti 6 mogućih elementarnih događaja (kocka je pala na 1, 2, 3, 4, 5 ili 6). Skup svih mogućih elementarnih događaja ćemo označavati s Ω. Taj skup ćemo nazivati prostor elementarnih događaja. Njegove elemente ćemo označavati s ω ili ω_i gdje nam je i neki indeks.

Definicija 1.1.1. *Neka je Ω neprazan skup. Neka je $\mathcal{F}$ familija podskupova od Ω. Kažemo da je $\mathcal{F}$ σ-algebra ako vrijedi:*

1. $\Omega \in \mathcal{F}$;
2. *Ako je $A \in \mathcal{F}$ onda je i $A^c \in \mathcal{F}$;*
3. *Ako su $A_j \in \mathcal{F}$, $j \in \mathbb{N}$ onda je i $\bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$.*

Uređen par $(\Omega, \mathcal{F})$ nazivamo izmjeriv prostor.

Definicija 1.1.2. *Neka je Ω neprazan skup i $\mathcal{F}$ σ-algebra događaja. Funkciju $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ na $(\Omega, \mathcal{F})$ nazivamo vjerojatnost ako vrijedi:*

1. $\mathbb{P}(A) \geq 0$ *za svaki* $A \in \mathcal{F}$
2. $\mathbb{P}(\Omega) = 1$

3. *za $A_1, A_2, ...$ niz disjunktnih događaja u $\mathcal{F}$ ($A_i \cap A_j = \emptyset$ za $i \neq j$) vrijedi*

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j).$$

Definicija 1.1.3. *Neka je Ω prostor elementarnih događaja, $\mathcal{F}$ σ-algebra događaja na Ω i $\mathbb{P}$ vjerojatnost na $(\Omega, \mathcal{F})$. Uređenu trojku $(\Omega, \mathcal{F}, \mathbb{P})$ nazivamo vjerojatnosni prostor.*

Definicija 1.1.4. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor te $B \in \mathcal{F}$ događaj takav da je $\mathbb{P}(B) > 0$. Uvjetna vjerojatnost događaja A uz uvjet B definira se formulom*

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{1.1}$$

Iz (1.1) slijedi

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\mathbb{P}(A|B).$$

Definicija 1.1.5. *Neka je $(H_i)_{i \in I}$ ($I \subseteq \mathbb{N}$) konačna ili prebrojiva familija događaja iz $\mathcal{F}$ takva da je $\mathbb{P}(H_i) > 0$ za sve $i \in I$, $H_i \cap H_j = \emptyset$ za $i \neq j$ i $\bigcup_{i \in I}^{\infty} H_i = \Omega$. Takvu familiju $(H_i)_{i \in I}$ zovemo potpun sustav događaja.*

Definicija 1.1.6. *(Formula potpune vjerojatnosti) Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor i $(H_i)_{i \in I}$ potpun sustav događaja na njemu. Tada za proizvoljan $A \in \mathcal{F}$ vrijedi*

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A \mid H_i)\mathbb{P}(H_i). \tag{1.2}$$

Definicija 1.1.7. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor i neka su $A, B \in \mathcal{F}$, Kažemo da su događaji A i B nezavisni ako vrijedi*

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B). \tag{1.3}$$

1.2 Slučajna varijabla

Definicija 1.2.1. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor. Za funkciju $X : \Omega \to \mathbb{R}$ kažemo da je diskretna slučajna varijabla ako postoji prebrojiv ili konačan skup $A = \{a_1, a_2, ...\} \subseteq \mathbb{R}$ tako da vrijedi*

1. *$X(\omega) \in A$, $\forall \omega \in \Omega$;*

2. *$\{X = a_j\} = \{\omega : X(\omega) = a_j\} \in \mathcal{F}$, $\forall j \in \mathbb{N}$.*

Ako stavimo $p_i = \mathbb{P}(X = a_i)$, $i = 1, 2, ...$, brojeve $a_1, a_2...$ zajedno s pripadnim vjerojatnostima $p_1, p_2, ...$ nazivamo distribucijom slučajne varijable X, te pišemo

$$X \sim \begin{pmatrix} a_1 & a_2 & ... \\ p_1 & p_2 & ... \end{pmatrix} \tag{1.4}$$

Definicija 1.2.2. *Ako je $X : \Omega \to \{0, 1\}$, X se zove Bernoullijeva slučajna varijabla. Neka je $p = \mathbb{P}(X = 1)$ i $q = 1 - p = \mathbb{P}(X = 0)$. To jest*

$$X \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}. \tag{1.5}$$

U tom slučaju pišemo $X \sim B(1, p)$ te kažemo da je X Benoullijeva slučajna varijabla s parametrom $p \in [0, 1]$.

Posebno će nas zanimati slučajna varijabla u slijedećem primjeru.

Primjer 1.2.3. *Neka je $X : \Omega \to \{-1, 1\}$ slučajna varijabla za koju vrijedi $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = \frac{1}{2}$, to jest*

$$X \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{1.6}$$

Naprimjer, X može predstavljati bacanje simetričnog novčića tako da vrijedi

$$X = \begin{cases} 1, & \textit{pala je glava} \\ -1, & \textit{palo je pismo} \end{cases}.$$

Primijetimo da je to samo prilagođena verzija Bernoullijeve slučajne varijable kojoj je kodomena skup $\{-1, 1\}$ i $p = q = \frac{1}{2}$, a nekad se naziva i Radermacherova slučajna varijabla.

Definicija 1.2.4. *Za diskretne slučajne varijable na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, \mathbb{P})$ kažemo da su nezavisne ako vrijedi*

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \forall x, y. \tag{1.7}$$

Definicija 1.2.5. *Neka je $(\Omega, \mathcal{F}, \mathbb{P})$ vjerojatnosni prostor i X diskretna slučajna varijabla na njemu. Neka je $(x_i)_{i \in I}$ skup svih mogućih ishoda od X. Očekivanje varijable X definiramo ovako*

$$E[X] := \sum_{i \in I} x_i \mathbb{P}(X = x_i), \tag{1.8}$$

ukoliko je $\sum_{i \in I} |x_i| \, \mathbb{P}(X = x_i) < +\infty$, te u tom slučaju kažemo da očekivanje od X postoji.

U slučaju da je $x_i \geq 0$, $\forall i \in I$, uvijek definiramo $E[X]$ kao u (1.8) s tim da dopuštamo mogućnost $[X] = +\infty$.

Definicija 1.2.6. *Neka je* $(\Omega, \mathcal{F}, \mathbb{P})$ *vjerojatnosni prostor i* X *diskretna slučajna varijabla na njemu koja ima očekivanje. Varijancu definiramo kao*

$$Var[X] = E[X - E[X]]^2 \in [0, +\infty]. \tag{1.9}$$

Definicija 1.2.7. *(Općenitija definicija slučajne varijable) Neka je* $(\Omega, \mathcal{F}, \mathbb{P})$ *vjerojatnosni prostor. Funkcija* $X : \Omega \to \mathbb{R}$ *zove se slučajna varijabla ako vrijedi* $\{X \leq x\} \in \mathcal{F}$ *za sve* $x \in \mathbb{R}$ *gdje je* $\{X \leq x\} = \{\omega : X(\omega) \leq x\} \in \mathcal{F}$*,* $x \in \mathbb{R}$.

Definicija 1.2.8. *Slučajna varijabla* $X : \Omega \to \mathbb{R}$ *je apsolutno neprekidna ako postoji* $f : \mathbb{R} \to [0, \infty)$ *takva da sve* $x \in \mathbb{R}$ *vrijedi*

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f(t)\mathrm{d}t.$$

Funkciju f *zovemo funkcijom gustoće od* X.

Definicija 1.2.9. *Funkcija distribucije slučajne varijable* X *je funkcija* $F : \mathbb{R} \to [0, 1]$ *definirana formulom*

$$F(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Definicija 1.2.10. *Za neprekidnu slučajnu varijablu* $X : \Omega \to \mathbb{R}$ *kažemo da ima standardnu normalnu distribuciju s paramterima* 0 *i* 1 *i označavamo s* $X \sim N(0, 1)$ *ako ima funkciju gustoće*

$$\phi(x) = \frac{1}{2\pi} \cdot e^{\frac{-1}{2}x^2}, \quad x \in \mathbb{R}.$$

Funkcija distribucije slučajne varijable $X \sim N(0, 1)$ *je*

$$\Phi(x) = \frac{1}{2\pi} \int_{-\infty}^{x} e^{\frac{-1}{2}t^2}\mathrm{d}t, \quad x \in \mathbb{R}.$$

Definicija 1.2.11. *Funkciju* $\Gamma : \mathbb{R}_+ \to \mathbb{R}_+$ *zadanu formulom*

$$\Gamma(z) = \int_{0}^{\infty} t^{z-1}e^{-t}\mathrm{d}t, \quad z > 0.$$

zovemo gama funkcija.

Definicija 1.2.12. *Funkciju* $B : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$ *zadanu formulom*

$$B(z_1, z_2) = \int_{0}^{1} t^{z_1-1}(1 - t)^{z_2-1}\mathrm{d}t, \quad z_1, z_2 > 0.$$

zovemo beta funkcija.

Vrijedi

$$B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1 + z_2)}.$$

Definicija 1.2.13. *Slučajna varijabla X ima beta distribuciju s parametrima α i β ako je neprekidna s funkcijom gustoće*

$$f(x) = \frac{x^{(\alpha-1)}(1-x)^{(\beta-1)}}{B(\alpha,\beta)} \tag{1.10}$$

i pišemo $X \sim Beta(\alpha,\beta)$.

Vrijedi

$$E[X] = \frac{\alpha}{\alpha+\beta}$$

i

$$Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Teorem 1.2.14. ***(Jaki zakon velikih brojeva)*** *Neka je $X_1, X_2, \ldots$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s očekivanjem $E[X] = \mu$. Neka je*

$$\overline{X_n} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Jaki zakon velikih brojeva tvrdi da $\overline{X_n}$ gotovo sigurno konvergira prema očekivanoj vrijednosti, to jest

$$\mathbb{P}(\lim_{n\to\infty} \overline{X_n} = \mu) = 1.$$

Teorem 1.2.15. ***(Centralni granični teorem)*** *Neka je $X_1, X_2, \ldots$ niz nezavisnih jednako distribuiranih slučajnih varijabli. Neka je $\mu = E[X_k]$ i neka je $\sigma^2 = Var[X_k] < +\infty$ za sve $k \in \mathbb{N}$. Neka je $S_n = X_1 + \ldots + X_n$. Tada za svaki fiksni β vrijedi*

$$\mathbb{P}\left\{\frac{S_n - n\mu}{\sigma\sqrt{n}} < \beta\right\} \to \Phi(\beta)$$

gdje je $\Phi(x)$ funkcija standardne normalne distribucije.

Poglavlje 2

Jednostavna simetrična slučajna šetnja

U prirodi postoji mnoštvo procesa koji se mogu dobro aproksimirati slučajnom šetnjom poput kretanja molekula u plinu ili tekućini, kretanja cijena dionica ili praćenje financijskog stanja kockara.

Definicija 2.0.1. *Neka je $X_1, X_2, ...$ niz nezavisnih i jednako distribuiranih slučajnih varijabli s distribucijom*

$$X_i \sim \begin{pmatrix} -1 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Neka je

$$S_n = \sum_{i=1}^{n} X_i,\ n \geq 1, \qquad i\ \ S_0 = 0.$$

Niz slučajnih varijabli $(S_n)_{n\in\mathbb{N}}$ nazivamo jednostavnom simetričnom slučajnom šetnjom.

Uočimo da vrijedi $S_n = S_{n-1} + X_n$ za sve $n \geq 1$.

2.1 Princip refleksije i teorem o glasanju

Uvedimo oznake. Uređena n-torka $(s_1, ..., s_n)$ predstavljat će mogući ishod slučajne šetnje $(S_1, ..., S_n)$, $(x_1, .., x_n)$ ishod slučajnih varijabli $(X_1, ..., X_n)$. Dakle vrijedi

$$x_i \in \{-1, 1\},\ i = 1, ..., n\ , \quad s_j - s_{j-1} = x_j,\ j = 1, ..., n\ , \quad (s_0 := 0). \tag{2.1}$$

Koristit ćemo geometrijsku reprezentaciju kretanja niza. Horizontalnu os označavati ćemo s x, a vertikalnu s y. Uređenu n-torku $(s_1, ..., s_n)$ prikazivat ćemo izlomljenom linijom koja se sastoji od n dužina čija k-ta točka ima koordinate (k, s_k). Takve linije zvat ćemo *putovima*.

Definicija 2.1.1. *Neka je $n > 0$ i $n, x \in \mathbb{Z}$. Put $(s_1, s_2..., s_n)$ od ishodišta $(0,0)$ do točke (n, x) je izlomljena linija čiji vrhovi imaju apscise $0, 1, ..., n$, i ordinate $s_1, s_2, ..., s_n$ koje zadovoljavaju* (2.1) *te $s_n = x$.*

Za svaki put $(s_1, ..., s_n)$ [dakle mogući ishod slučajne šetnje do s_n] zbog nezavisnosti slučajnih varijabli $X_1, X_2, ...$ vrijedi

$$\mathbb{P}(S_1 = s_1, ..., S_n = s_n) = \mathbb{P}(X_1 = x_1, ..., X_n = x_n) = \mathbb{P}(X_1 = x_1)...\mathbb{P}(X_n = x_n) = \left(\frac{1}{2}\right)^n = \frac{1}{2^n}.$$

Ako je B neki skup putova pri čemu su svi duljine n vrijedi

$$\mathbb{P}((S_1, ..., S_n) \in B) = \sum_{(s_1,...,s_n)\in B} \mathbb{P}((S_1, ..., S_n) = (s_1, ..., s_n)) = \frac{|B|}{2^n},$$

pri čemu je $|B|$ broj putova (članova) u skupu B.

Ako u putu $(s_1, ...s_n)$ od $(0,0)$ do (n, x) imamo p pozitivnih, a q negativnih x_i-ova tada vrijedi

$$n = p + q, \quad x = p - q, \quad p, q \in \mathbb{N}_0. \tag{2.2}$$

Nadalje za proizvoljne $n \in \mathbb{N}$, $x \in \mathbb{N}_o$ put od ishodišta do točke (n, x) postoji ako i samo ako su n i x oblika (2.2). Pozicije za p pozitivnih jedinica možemo odabrati na

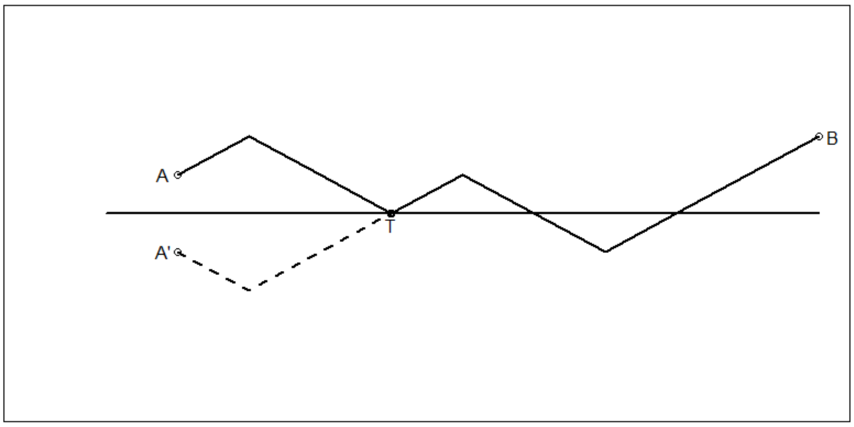
$$N_{n,x} := \binom{p+q}{p} = \binom{p+q}{q}$$

načina. Za n i x koji nisu oblika (2.2) definiramo $N_{n,x} = 0$. Tada $\forall n, x$ postoji točno $N_{n,x}$ putova od ishodišta do točke (n, x).

Postavimo teren za slijedeću lemu. Neka su $A = (a, \alpha)$ i $B = (b, \beta)$ točke u pozitivnom kvadrantu te neka vrijedi $b > a \geq 0$, $\alpha > 0$, $\beta > 0$, $a, b, \alpha, \beta \in \mathbb{N}$. Osnom simetrijom oko osi x dobijemo točku $A' = (a, -\alpha)$. Put od A do B je definiran na isti način kao ranije.

Lema 2.1.2. ***(Princip refleksije)*** *Broj putova od točke A do točke B koji diraju ili prelaze os apscisu jednak je broju putova od točke A' do točke B.*

Dokaz. Pretpostavimo da je $(s_a = \alpha, s_{a+1}, ..., s_b = \beta)$ neki put od točke A do točke B koji barem jednom dira ili siječe x-os. Neka je $T = \inf\{k \ : \ s_k = 0\}$. Neka je $s'_k = -s_k$ za sve $k \leq T$, $s'_k = s_k$ za $T < k \leq b$. Onda je s'_k, $k = a, a+1, ..., b$, put od točke A' do točke B (vidi sliku 2.1). S druge strane, neka je $(t_a = -\alpha, t_{a+1}, ..., t_b = \beta)$ put od točke A' do točke B (koji očito mora sijeći $x-$os barem jednom). Neka je $T = \inf\{k \ : \ t_k = 0\}$. Neka je $t'_k = -t_k$ za sve $k \leq T$, $t'_k = t_k$ za $T < k \leq b$. Onda je $t'_k, k = a, a+1, ...b$, put od točke A' do točke B koji ima vrijednost 0 u T. Ovo pokazuje da između ove dvije klase putova postoji bijekcija te je njihov broj jednak. □



Slika 2.1: Ilustracija principa refleksije

Teorem 2.1.3. ***(Teorem o glasanju)*** *Neka su n i x prirodni brojevi. Postoji točno $\frac{x}{n}N_{n,x}$ putova $(s_1, s_2, ..., s_n = x)$ od ishodišta do točke (n, x) tako da vrijedi $s_1 > 0, s_2 > 0, ..., s_n > 0$. Specijalno, ako su n, x oblika (2.2) i $(S_i)_{i\geq 0}$ jednostavna simetrična slučajna šetnja vrijedi*

$$\mathbb{P}(S_1 > 0, ..., S_n > 0 \mid S_n = x) = \frac{p-q}{p+q}.$$

Dokaz. Broj putova od $(0, 0)$ do (n, x) koji ne prelaze i ne diraju os x isti je kao broj putova od točke $(1, 1)$ do (n, x) koji ne prelaze i ne diraju os x. Primijetimo, $N_{n-1,x+1}$ predstavlja broj putova od točke $(1, -1)$ do točke (n, x), a prema principu refleksije to je i broj svih putova od točke $(1, 1)$ do (n, x) koji diraju ili prelaze os x. Sada preostaje od svih putova od točke $(1, 1)$ oduzeti one koji diraju ili prelaze os x.

$$N_{n-1,x-1} - N_{n-1,x+1} = \binom{p+q-1}{p-1} - \binom{p+q-1}{p}.$$

No to je upravo $\frac{x}{n}N_{n,x}$. Provjerimo:

$$\begin{aligned}
&\binom{p+q-1}{p-1} - \binom{p+q-1}{p} = \\
&= \frac{(p+q-1)...(p+q-1-(p-1)+1)}{(p-1)!} - \frac{(p+q-1)...(p+q-1-p+1)}{p!} \\
&= (p-q)\frac{(p+q-1)(p+q-2)...(q+1)}{p!} \\
&= \frac{p-q}{p+q}\frac{(p+q)(p+q-1)(p+q-2)...(q+1)}{p!} \\
&= \frac{x}{n}\binom{p+q}{p} = \frac{x}{n}N_{n,x}.
\end{aligned}$$

Pokažimo sada i drugu tvrdnju. Broj putova do točke $(n, x) = (p + q, p - q)$ je očito $N_{n,x}$. Vrijedi

$$\mathbb{P}(S_1 > 0, ..., S_n > 0, S_n = x) = \frac{\frac{x}{n}N_{n.x}}{2^n}; \tag{2.3}$$

$$\mathbb{P}(S_n = x) = \frac{N_{n,x}}{2^n}. \tag{2.4}$$

Pa imamo

$$\begin{aligned}
\mathbb{P}(S_1 > 0, ..., S_n > 0 \mid S_n = x) = \frac{\mathbb{P}(S_1 > 0, ..., S_n > 0, S_n = x)}{\mathbb{P}(S_n = x)} &= \Big[(2.3), (2.4)\Big] = \frac{\frac{x}{n}N_{n,x}}{N_{n,x}} \\
&= \frac{x}{n} = \frac{p-q}{p+q}.
\end{aligned}$$

□

Primjer 2.1.4. *Pretpostavimo da na izborima imamo 2 kandidata, nazovimo ih P i Q te n glasača. Neka vrijedi*

$$X_i := \begin{cases} 1, & \textit{i-ti glasač dao glas za kandidata P} \\ -1, & \textit{i-ti glasač dao glas za kandidata Q} \end{cases} \quad i \tag{2.5}$$

$$S_k = \sum_{i=1}^{k} X_i = \textit{razlika u glasovima nakon k prebrojanih listića.}$$

Ako pretpostavimo da je $X_1, X_2, ..., X_n$ niz slučajnih varijabli s distribucijom (1.6)*, to jest svaki glasač s jednakom vjerojatnošću glasa za bilo kojeg od kandidata, i to nezavisno od* *drugih glasača. Teorem o glasanju tvrdi*

$$\mathbb{P}\Big(P \text{ cijelo vrijeme bio u vodstvu} \;\Big|\; \begin{matrix} P \text{ dobio } p \text{ glasova ,} \\ Q \text{ dobio } q \text{ glasova} \end{matrix}\Big) =$$
$$= \mathbb{P}(S_1 > 0, ..., S_n > 0 \mid S_n = x) = \frac{p-q}{p+q}.$$

p	q	$\frac{p-q}{p+q}$
18	0	100%
17	1	88.89%
16	2	77.78%
15	3	66.67%
14	4	55.56%
13	5	44.44%
12	6	33.33%
11	7	22.22%
10	8	11.11%

Tablica 2.1: Ilustracija teorema o glasovanju za $n = 18$.

Pogledajmo još primjer na izborima u Sjedinjenim američkim državama. Joe Biden je osvojio ukupno 81,282,896 glasova, a Donald Trump 74,222,484. Ukoliko pretpostavimo da se pretpostavke o glasovanju s početka primjera (barem približno) zadovoljene, vjerojatnost da je Biden vodio od početka do kraja je

$$\frac{81\,282\,896 - 74\,222\,484}{81\,282\,896 + 74\,222\,484} = \frac{7\,060\,412}{155\,505\,380} \approx 0.0454 = 4.54\%.$$

2.2 Posjeti ishodištu

Definiramo $\forall n \geq 1,\; r \in \mathbb{N}$

$$p_{n,r} = \mathbb{P}(S_n = r) = \binom{n}{\frac{n+r}{2}} 2^{-n},$$

gdje se podrazumijeva da je binomni koeficijent jednak nuli ukoliko $(n+r)/2 \notin \{0, 1, 2, ..., n\}$. Kažemo da se povratak u ishodište pojavljuje u trenutku k ukoliko je $S_k = 0$ gdje je k nužno paran broj. Uočimo vrijednost parcijalne sume S_k će uvijek biti jednake parnosti kao i trenutak k. To jest:

$$2|k \Leftrightarrow 2|S_k.$$

Neka je $k = 2v$. Vjerojatnost povratka u ishodište u trenutku $2v$ (to jest $\mathbb{P}(S_{2v} = 0) = p_{2v,0}$) označavat ćemo s u_{2v}. Dakle,

$$u_{2v} = \binom{2v}{v} 2^{-2v}, \quad v \geq 1. \tag{2.6}$$

Iskažimo sada Stirlingovu formulu koja će nam služiti za aproksimaciju računanja vjerojatnosti povratka u ishodište. Kada $n \to \infty$ vrijedi

$$n! \sim n^n e^{-n} \sqrt{2\pi n},$$

pri čemu za nizove $(a_n)_{n\in\mathbb{N}}$ i $(b_n)_{n\in\mathbb{N}}$ pišemo $a_n \sim b_n$ za $n \to \infty$, ako $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$.

Lema 2.2.1. *Kada $n \to \infty$,*

$$u_{2n} \sim \frac{1}{\sqrt{n\pi}}. \tag{2.7}$$

Dokaz.

$$\begin{aligned}
u_{2v} &= \binom{2v}{v} 2^{-2v} = \frac{(2v)!}{v! \cdot v!} \\
&\sim \frac{\sqrt{2\pi}(2v)^{2v+\frac{1}{2}} e^{-2v}}{(\sqrt{2\pi} v^{v+\frac{1}{2}} e^{-v})^2} \\
&= \frac{\cancel{\sqrt{2\pi}} \cdot \cancel{v^{2v}} \cdot \sqrt{2} \cdot 2^{2v} \cdot \sqrt{v} \cdot \cancel{e^{-2v}}}{(\sqrt{2\pi})^{\cancel{2}} \cdot \cancel{v^{2v}} \cdot v \cdot \cancel{e^{-2v}}} \cdot 2^{-2v} \\
&= \frac{\sqrt{2} \cdot \sqrt{v} \cdot \cancel{2^{2v}}}{\sqrt{2\pi} \cdot \cancel{2^{2v}}} \\
&= \frac{1}{\sqrt{\pi v}}.
\end{aligned}$$

□

U Tablici 2.2 navedene su neke vjerojatnosti povratka u ishodište zajedno s aproksimacijom (2.7) dobivenom pomoću Stirlingove formule. Uočimo da već za $2v = 26$ relativna razlika između prave vrijednosti i aproksimacije pada ispod 0.01 (to jest 1%).

Posebno će nas zanimati slučajevi kada se prvi povratak u ishodište dogodi u trenutku $2v$. Neka je

$$R := \inf\{n \geq 1 \ : S_n = 0\} = \text{vrijeme prvog povratka u ishodište};$$

v	$2v$	$x = u_{2v}$	$y = \frac{1}{\sqrt{v\pi}}$	$\lvert x-y\rvert$	$\frac{\lvert x-y\rvert}{x}$
1	2	0.500000	0.56419	0.064189584	0.128379
2	4	0.375000	0.398942	0.02394228	0.063846
3	6	0.312500	0.325735	0.013235008	0.042352
4	8	0.273438	0.282095	0.008657292	0.031661
5	10	0.246094	0.252313	0.006219502	0.025273
6	12	0.225586	0.230329	0.004743495	0.021027
7	14	0.209473	0.213244	0.003770962	0.018002
8	16	0.196381	0.199471	0.003090525	0.015737
9	18	0.185471	0.188063	0.002592613	0.013979
10	20	0.176197	0.178412	0.00221536	0.012573
11	22	0.168188	0.17011	0.001921465	0.011424
12	24	0.161180	0.162868	0.001687246	0.010468
13	26	0.154981	0.156478	0.001497019	0.009659
14	28	0.149446	0.150786	0.001340028	0.008967
15	30	0.144464	0.145673	0.001208676	0.008367
100	200	0.056348	0.056419	$7.05 \cdot 10^{-5}$	0.001251

Tablica 2.2: Usporedba u_{2v} sa aproksimacijom dobivenom pomoću Stirlingove formule

pa vrijedi

$$\{R = 2v\} = \{S_1 \neq 0, ..., S_{2v-1} \neq 0, S_{2v} = 0\}, \quad v \geq 1;$$

i neka je

$$f_{2v} := \mathbb{P}(R = 2v), \quad v \geq 1;$$
$$f_0 := 0.$$

Vjerojatnosti f_{2n} i u_{2n} su povezane. Povratak u ishodište u trenutku 2n može biti prvi povratak ili se prvi povratak dogodio u nekom ranijem trenutku $2k < 2n$ i popraćen je povratkom $2n - 2k$ vremenskih jedinica kasnije.

Po formuli potpune vjerojatnosti znamo da je

$$u_{2n} = \mathbb{P}(S_{2n} = 0) = \sum_{k=1}^{n} \mathbb{P}(S_{2n} = 0 \mid R = 2k)\mathbb{P}(R = 2k).$$

Vrijedi

$$\begin{aligned}
&\mathbb{P}(S_{2n} = 0 \mid R = 2k) = \mathbb{P}(S_{2n} = 0 | S_{2k} = 0, S_{2k-2} \neq 0, ..., S_2 \neq 0) \\
&= [S_{2n} = S_{2k} + X_{2k+1} + ... + X_{2n} = \sum_{i=2k+1}^{2n} x_i \\
&= \mathbb{P}\left(\sum_{i=2k+1}^{2n} X_i = 0 \mid S_{2k} = 0, S_{2k-2} \neq 0, ...S_2 \neq 0 \right) \\
&= \mathbb{P}\left(\sum_{i=2k+1}^{2n} X_i = 0 \right) = \mathbb{P}(S_{2n-2k} = 0) = u_{2n-2k},
\end{aligned}$$

pa imamo

$$u_{2n} = f_2 u_{2n-2} + ... + f_{2n} u_0 \qquad n \geq 1. \tag{2.8}$$

Proučimo kako izračunati vjerojatnost prvog povratka u trenutku $2n$.

Lema 2.2.2. *Vrijedi*

$$\mathbb{P}(S_1 \neq 0, S_2 \neq 0, ..., S_{2n} \neq 0) = \mathbb{P}(S_{2n} = 0) = u_{2n}, \quad n > 0,$$

to jest

$$\mathbb{P}(R > 2n) = u_{2n}, \quad n > 0.$$

Uočimo svi S_i $(i = 1, 2, ..., 2n)$ moraju biti ili isključivo pozitivni ili isključivo negativni. Budući da su to jednako vjerojatni događaji možemo pisati

$$\mathbb{P}(S_1 > 0, S_2 > 0, ..., S_{2n} > 0) = \frac{1}{2} u_{2n}$$

Dokaz. Promatrajući sve moguće vrijednosti od S_{2n} $(2, ..., 2n)$ očigledno vrijedi:

$$\mathbb{P}(S_1 > 0, S_2 > 0, ..., S_{2n} > 0) = \sum_{r=1}^{n} \mathbb{P}(S_1 > 0, S_2 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r).$$

Kao u dokazu teorema o glasanju broj putova koji zadovoljavaju uvjet zdesna je jednak $N_{2n-1,2r-1} - N_{2n-1,2r+1}$ pa je r-ti sumand jednak

$$\begin{aligned}
&\mathbb{P}\{S_1 > 0, S_2 > 0, ..., S_{2n-1} > 0, S_{2n} = 2r\} = \frac{N_{2n-1,2r-1} - N_{2n-1,2r+1}}{2^n} \\
&= \frac{\binom{2n-1}{n+r-1} - \binom{2n-1}{n+r}}{2^n} \\
&= \frac{1}{2}\left(\frac{\binom{2n-1}{n+r-1}}{2^{n-1}} - \frac{\binom{2n-1}{n+r}}{2^{n-1}}\right) \\
&= \frac{1}{2}(p_{2n-1,2r-1} - p_{2n-1,2r+1}).
\end{aligned}$$

Dakle imamo

$$\begin{aligned}
&\mathbb{P}(S_1 > 0, ..., S_{2n} = 0) \\
&= \frac{1}{2}(p_{2n-1,1} - p_{2n-1,3}) + \frac{1}{2}(p_{2n-1,3} - p_{2n-1,5}) + ... + \frac{1}{2}(p_{2n-1,2n-1} - p_{2n-1,2n+1}).
\end{aligned}$$

Primijetimo da se drugi član prvog i prvi član sljedećeg sumanda uvijek krate pa na kraju budući da je $p_{2n-1,2n+1} = 0$ ostane samo $\frac{1}{2}p_{2n-1,1}$. Provjerimo još da vrijedi $p_{2n-1,1} = u_{2n}$.

$$\begin{aligned}
p_{2n-1,1} &= \binom{2n-1}{\frac{2n-1+1}{2}} 2^{(-2n-1)} = \binom{2n-1}{n} 2^{-2n+1} \\
&= \frac{(2n-1)...(2n-1-n+1)}{n!} 2^{-2n} \cdot 2 = \frac{(2n-1)...n}{n!} 2^{-2n} \cdot 2 \\
&= \frac{2n(2n-1)...(n+1)}{n!} 2^{-2n} = \binom{2n}{n} 2^{-2n} = u_{2n}.
\end{aligned}$$

□

Ova se lema može iskazati na nekoliko načina. Na primjer,

$$\mathbb{P}(S_1 \geq 0, S_2 \geq 0, ...S_{2n} \geq) = u_{2n}. \tag{2.9}$$

Zaista, put duljine $2n$ sa svim vrhovima iznad osi x nužno prolazi kroz $(1, 1)$. Tada dobivamo put duljine $2n - 1$ s ishodištem u $(1, 1)$ čiji su svi vrhovi iznad ili na novoj x-osi podignutoj za 1. Slijedi

$$\begin{aligned}\frac{1}{2}u_{2n} = \mathbb{P}(S_1 > 0, ..., S_{2n} > 0) &= \mathbb{P}(S_1 = 1)\mathbb{P}(S_2 > 0, ..., S_{2n} > 0 \mid S_1 = 1)\\ &= \frac{1}{2}\mathbb{P}(S_1 \geq 0, ..., S_{2n-1} \geq 0)\\ &= \frac{1}{2}\mathbb{P}(S_1 \geq 0, ..., S_{2n} \geq 0).\end{aligned}$$

Budući da je S_{2n-1} neparan broj i $S_{2n-1} \geq 0$ to povlači i $S_{2n} \geq 0$.

Lema 2.2.2 vodi direktno do eksplicitne jednadžbe za prvi povratak u ishodište:

$$\begin{aligned}f_{2n} &= \mathbb{P}(R = 2n)\\ &= \mathbb{P}(\{R > 2n-2\} \backslash \{R > 2n\}) = \mathbb{P}(\{R > 2n-2\}) - \mathbb{P}(\{R > 2n\})\\ &= u_{2n-2} - u_{2n},\end{aligned}$$

dakle

$$f_{2n} = u_{2n-2} - u_{2n}, \quad n \geq 1. \tag{2.10}$$

Propozicija 2.2.3. *Vrijedi*

$$f_{2n} = \frac{1}{2n-1}u_{2n}, \quad n \geq 1.$$

Dokaz. Imamo

$$\begin{aligned}f_{2n} &= u_{2n-2} - u_{2n}\\ &= \binom{2n-2}{n-1}2^{(2n-2)} - \binom{2n}{n}2^{-2n}\\ &= \frac{(2n-2)(2n-3)...(2n-2-n+1+1)}{(n-1)!}2^{(2n-2)} - \frac{2n(2n-1)...(n+1)}{n!}2^{-2n}\\ &= \frac{(2n-2)(2n-3)...n}{(n-1)!} \cdot \frac{4}{2^{2n}} - \frac{2n(2n-1)...(n+1)}{n!}2^{-2n}\\ &= \frac{n^2}{2n(2n-1)}\binom{2n}{n}\frac{4}{2^{2n}} - \binom{2n}{n}\frac{1}{2^{2n}}\\ &= \left(\frac{4n^2}{2n(2n-1)} - 1\right)\binom{2n}{n}\frac{1}{2^{2n}}\\ &= \left(\frac{2n}{2n(2n-1)}\right)\binom{2n}{n}2^{-2n} = \frac{1}{2n-1}u_{2n}.\end{aligned}$$

□

2.3 Kockarova propast

Problem kockarove propasti dobro je poznat problem u povijesti vjerojatnosti. Riječ je o igri između dva igrača, A i B, koji se natječu međusobno. Igrač A započinje s nekom svotom novaca a, a igrač B ima $N - a$ novaca ($a, N \in \mathbb{N}$, $N > a$). Bacamo simetrični idealni novčić koji ima jednaku šansu pasti na obe strane. Kada padne glava igrač A dobiva jednu kunu, a kada padne pismo igrač B dobiva kunu. Igra se nastavlja dok neki od igrača ne ostane bez novaca tj. jedan od igrača ne skupi N kuna. Stanje ćemo mjeriti tako da pratimo koliko A ima kuna. Šetnja će umjesto u 0 početi u a i završiti u 0 ili N koje ćemo zvati upijajućim granicama. Reći ćemo da je A pobijedio u igri ukoliko šetnja završi u N, a B ako završi u 0. Zanima nas koja je vjerojatnost da igrač A pobijedi. Odgovor slijedi u teoremu u nastavku.

Teorem 2.3.1. ***(Kockarova propast)*** *Promotrimo jednostavnu simetričnu slučajnu šetnju na* $0, 1, ..., N$ *s upijajućim granicama 0 i N. Ako šetnja počinje u a, gdje je* $0 \leq a \leq N$*, onda je vjerojatnost* $v(a)$ *da se šetnja upije u N dana sa:*

$$v(a) = \frac{a}{N}.$$

Dokaz. Uočimo,

$$v(a) = \mathbb{P}(A \text{ pobjeđuje}) = \frac{1}{2}\mathbb{P}(A \text{ pobjeđuje} \mid \text{pala je glava}) + \frac{1}{2}\mathbb{P}(\text{A pobjeđuje} \mid \text{palo je pismo})$$

Ovo možemo zamijeniti sa:

$$v(a) = \frac{1}{2}v(a+1) + \frac{1}{2}v(a-1) \qquad \text{za } 0 < a < N. \tag{2.11}$$

Po uvjetima igre vrijedi

$$v(0) = 0, \quad v(N) = 1,$$

jer je tada igra već izgubljena odnosno dobivena. Sada riješimo ove diferencijalne jednadžbe. Primjetimo (2.11) je linearna homogena rekurzija reda 2. Linearna jer se javljaju samo prve potencije od a, a homogena jer nitijedan od 2 člana nije konstantan. Jednadžba (2.11)

$$t = \frac{1}{2}t^2 + \frac{1}{2}.$$

je njena karakteristična jednažba. Nultočke su

$$t_{1,2} = \frac{2 \pm \sqrt{4-4}}{2} = 1.$$

Pa imamo da je opće rješenje dano s $v(a) = \alpha \cdot 1^a + \beta a \cdot 1^a$. Koristimo početni uvjet $v(0) = 0$

$$\begin{aligned} v(0) = 0 &= \alpha \cdot 1^0 + \beta \cdot 0 \cdot 1^0 \\ \alpha &= 0 \end{aligned}$$

Iz drugog uvjeta $V(N) = 1$ sada imamo

$$\begin{aligned} 1 &= \alpha \cdot 1^N + \beta \cdot N \cdot 1^N \\ 1 &= 0 \cdot 1 + \beta N \\ \beta &= \frac{1}{N}. \end{aligned}$$

Imamo $\alpha = 0$, i $\beta = 1/N$ pa vrijedi

$$v(a) = \frac{a}{N} \qquad \text{za sve } 0 \leq a \leq N.$$

□

Primjer 2.3.2. *Dva igrača igraju igru objašnjenu u teoremu . Neka igrač A započne igru s a novaca , a igrač B s N novaca. Pogledajmo kako se kreću vjerojatnosti za pobjedu igrača A u ovisnosti o početnom novcu.*

a	$N - a$	N	$v(a)$
1	49	50	2%
10	40	50	20%
20	30	50	40%
25	25	50	50%
30	20	50	60%
1	99	100	1%
32	68	100	32%
1	2020	2021	0.0495%

Tablica 2.3: Primjer šansi za pobjedu

2.4 Zadnji posjeti i duga vodstva

Možemo krenuti u dublju analizu prirode vjerojatnosti fluktuacija u slučajnoj šetnji. Mogli bismo očekivati će u jednoj dugoj igri bacanja novčića oba igrača voditi otprilike pola vremena i da će se vodstvo često izmjenjivati. Zamislimo veliki uzorak igara bacanja idealnog novčića duljine $2n$ te odaberimo jedan primjer nasumično. Promatrat ćemo trenutke posljednjeg izjednačenja.

Definicija 2.4.1. *Neka je I_{2n} slučajna varijabla za $n \geq 0$ takva da vrijedi:*

$$I_{2n} = \sup\{m : 0 \leq m \leq 2n, S_m = 0\} = \textit{zadnji posjet šetnje ishodištu do trenutka } 2n$$

pri čemu je $(S_n)_{n\geq 0}$ jednostavna simetrična slučajna šetnja.

Slučajna varijabla I_{2n} očito poprima samo parne vrijednosti, to jest vrijednosti $2k$ za $0 \leq k \leq n$. Česte promjene vodstva implicirale bi da I_{2n} s velikom vjerojatnosti poprima vrijednosti $2k$ za k blizu n. Ipak, teorem 2.4.3 u nastavku pokazuje da to nije slučaj. Na primjer distribucija od I_{2n} je simetrična u smislu da je

$$\mathbb{P}(I_{2n} = 2k) = \mathbb{P}(I_{2n} = 2n - 2k), \quad k \in \{0, 1, ..., n\}.$$

Posebno vrijedi

$$\mathbb{P}(I_{2n} < n) = \mathbb{P}(I_{2n} > n)$$

to jest vjerojatnost da se niti jedno izjednačenje nije dogodilo u drugoj polovici igre je jednako vjerojatnosti da se posljednje izjednačenje dogodilo u drugoj polovici igre.

Primjer 2.4.2. *Pretpostavimo da se provodi veliki broj igara bacanja novčića te da se novčić baca svake sekunde, cijeli dan, cijele godine. To jest u godini novčić bacimo 31,536,000 ($2n = 31536000$) puta. Idući teorem (2.4.3) povlači da će se u prosjeku u jednoj od 10 igara zadnje izjednačenje dogoditi prije 9. dana (9 dana = 777600 sekundi) i vodstvo se neće mijenjati idućih 356 dana, to jest*

$$\mathbb{P}(I_{2n} \leq 777600) \approx 0.1.$$

Pogledajmo još vjerojatnosti da se zadnja promjena vodstva dogodi u prvih 54h odnosno u prvih 130 minuta.

t	$\mathbb{P}(I_{2n} \leq t)$
9 dana	0.1
54 sata	0.05
130 minuta	0.01

Tablica 2.4: t = vrijeme zadnje promjene vodstva

Distribucija od I_{2n} dana je sljedećim teoremom.

Teorem 2.4.3. ***(Zakon arkus sinusa za zadnje posjete ishodištu)*** *Za sve $n \geq 0$ vrijedi*

$$\mathbb{P}(I_{2n} = 2k) = u_{2k}u_{2n-2k} =: \alpha_{2k,2n}, \qquad k = 0, 1, ..., n. \tag{2.12}$$

Dokaz. Promatrat ćemo puteve koji zadovoljavaju sljedeće uvjete: $S_{2k} = 0$ i $S_{2k+1} \neq 0, ..., S_{2n} \neq 0$. Imamo

$$\mathbb{P}(I_{2n} = 2k) = \mathbb{P}(S_{2k} = 0, S_{2k+1} \neq 0, ..., S_{2n} \neq 0),$$

a budući da su bacanja nezavisna vrijedi

$$\begin{aligned}
\mathbb{P}(S_{2k} = 0, S_{2k+1} \neq 0, ..., S_{2n} \neq 0) &= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2k+1} \neq 0, ..., S_{2n} \neq 0 \mid S_{2k} = 0) \\
&= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_{2k} + X_{2k+1} \neq 0, ..., S_{2k} + \sum_{i=2k+1}^{2n} X_i \neq 0 \mid S_{2k} = 0) \\
&= [X_1,\ i \in \mathbb{N}, \text{ imaju istu distribuciju}] \\
&= \mathbb{P}(S_{2k} = 0)\mathbb{P}(X_1 \neq 0, ..., \sum_{i=1}^{2n-2k} X_i \neq 0) \\
&= \mathbb{P}(S_{2k} = 0)\mathbb{P}(S_1 \neq 0, .., S_{2n-2k} \neq 0) \\
&= [\text{koristimo lemu 2.2.2 }] = u_{2k}u_{2n-2k}.
\end{aligned}$$

□

Distribucija koja će pridavati vrijednost $\alpha_{2k,2n}$ točci $2k$ za $k = 0, ..., n$ će se zvati diskretna arkus sinus distribucija reda n, jer inverz funkcije sinus daje dobru numeričku aproksimaciju. Dakle I_{2n} ima diskretnu arkus sinus distribuciju reda n. Distribucija je simetrična u smislu da je $\alpha_{2k,2n} = \alpha_{2n-2k,2n}$. Za $n = 2$ vrijednosti su $\frac{2}{8}, \frac{3}{8}, \frac{2}{8}$ (za primjer za $n = 10$ pogledati prvi i drugi redak tablice 2.5). Primijetimo, srednji član je uvijek najmanji, to jest najmanja je šansa da se posljednja promjena vodstva dogodi točno na pola igre.

Definicija 2.4.4. *Slučajna varijabla X ima arkus sinus distribuciju ako je neprekidna s funkcijom gustoće*

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}, \qquad x \in (0,1) \qquad [f(x) = 0,\ \textit{inače}]. \tag{2.13}$$

Uočimo, $\forall t \in [0, 1]$

$$\begin{aligned}
\int_0^t f(x)\,\mathrm{d}x &= \frac{1}{\pi}\int_0^t \frac{1}{\sqrt{x}}\frac{1}{\sqrt{1-x}} = \left[u = \sqrt{x},\ \mathrm{d}x = 2\sqrt{x}\,\mathrm{d}u\right] \\
&= \frac{1}{\pi}\int_0^t \frac{1}{\sqrt{x}}\frac{1}{\sqrt{1-u^2}}2\sqrt{x}\,\mathrm{d}u \\
&= \frac{2}{\pi}\int_0^t \frac{1}{\sqrt{1-u^2}}\,\mathrm{d}u \\
&= \frac{2}{\pi}(\arcsin(\sqrt{t}) - \arcsin(0)) = \frac{2}{\pi}\arcsin\sqrt{t}.
\end{aligned}$$

Specijalno,

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^1 f(x)\mathrm{d}x = \frac{2}{\pi}\frac{\pi}{2} = 1 \Rightarrow f \text{ je zaista } f\text{-ja gustoće te}$$

$$\mathbb{P}(X \le t) = \frac{2}{\pi}\arcsin(\sqrt{t}), \qquad \forall t \in [0,1].$$

Koristeći Stirlingovu formulu (2.7) vidjeli smo da je u_{2n} blizu $1/\sqrt{\pi n}$ osim za jako male n. Iz toga slijedi aproksimacija (za velike k i n)

$$\begin{aligned}\alpha_{2k,2n} = u_{2k}u_{2n-2k} &\approx \frac{1}{\sqrt{\pi k}}\frac{1}{\sqrt{\pi(n-k)}}\\ &= \frac{1}{\pi\sqrt{k(n-k)}}\\ &= \frac{1}{\pi n\sqrt{(k/n)(1-k/n)}} = \frac{1}{n}f(k/n)\end{aligned}$$

Imamo

$$\alpha_{2k,2n} \approx \frac{1}{n}f(x_k), \qquad \text{za } x_k = \frac{k}{n}. \tag{2.14}$$

Greška je zanemariva osim za k-ove jako blizu 0 ili n.

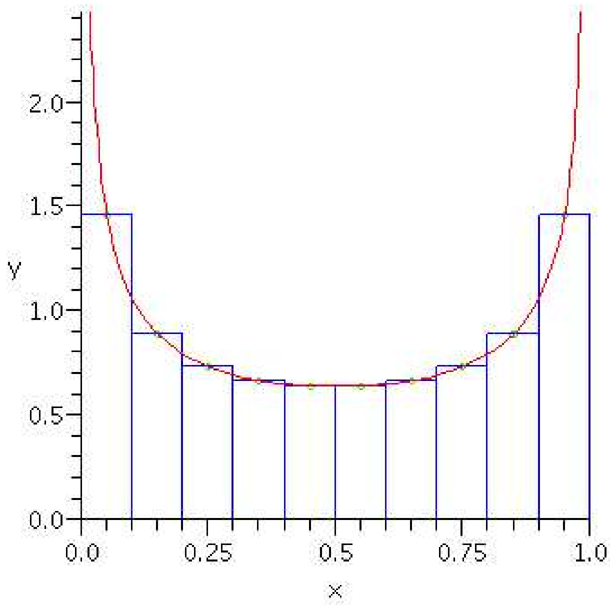
	$k=0$ $k=10$	$k=1$ $k=9$	$k=2$ $k=8$	$k=3$ $k=7$	$k=4$ $k=6$	$k=5$
$\alpha_{2k,20}$	0.1762	0.0927	0.0736	0.0655	0.0617	0.0606
$\frac{1}{n}f(x_k)$	/	0.1061	0.0796	0.0695	0.0650	0.6366

Tablica 2.5: Usporedba $\alpha_{2k,20}$ s aproksimacijom dobivenom pomoću 2.14

Desna strana u (2.14) je jednaka površini pravokuntika čija je visina $f(x_k)$, a širina interval duljine $1/n$ s polovištem u x_k (pogledaj sliku 2.2). Za $0 < p < q < 1$ i velike n suma vjerojatnosti $\alpha_{2k,2n}$ za $pn < k < qn$ je stoga približno jednako području ispod grafa funkcije f i iznad intervala $p < x < q$. Ovo vrijedi i za $p = 0$ i $q = 1$ jer je područje pod grafom jednako 1, kao i suma svih vjerojatnosti $\alpha_{2k,2n}$. Pokažimo precizno navedene tvrdnje.

Lema 2.4.5. *Za sve $p, q \in [0,1]$ vrijedi*

$$\sum_{pn\le k\le qn} \alpha_{2k,2n} \xrightarrow{n\to\infty} \int_p^q \frac{\mathrm{d}x}{\pi\sqrt{x(1-x)}}. \tag{2.15}$$



Slika 2.2: $f(x) = \frac{1}{\pi\sqrt{x(1-x)}}$.

Dokaz. Prisjetimo se

- vrijedi $u_{2n} \sim \dfrac{1}{\sqrt{n\pi}}$ kada $n \to \infty$ (vidi lemu 2.2.1),

  tj. $\forall \epsilon > 0, \quad \exists n_0 = n_0(\epsilon)$ t.d.

$$\frac{1}{\sqrt{\pi n}}(1-\epsilon) \le u_{2n} \le \frac{1}{\sqrt{\pi n}}(1+\epsilon), \quad \forall n \ge n_0. \tag{2.16}$$

- Pretpostavimo da je $p, q \in (0,1)$. Ako je k takav da $np \le k \le nq$, vrijedi $k \ge np$, te $n-k \ge n(1-q)$. Budući da je $p > 0$ i $1-q > 0$, za $n \ge n_0 \cdot \max\{\frac{1}{p}, \frac{1}{1-q}\}$, te za sve k t.d. $np \le k \le nq$, vrijedi

  1. $k \ge n_0(\epsilon)$
  2. $n-k \ge n_0(\epsilon), \quad \forall k$ t.d. $np \le k \le nq$.

  Dakle, za n dovoljno velik, (2.16) povlači:

$$\begin{aligned}\sum_{np\le k\le nq} u_{2k}u_{2n-2k} &\ge (1-\epsilon)^2 \sum_{np\le k\le nq} \frac{1}{\sqrt{\pi k}}\frac{1}{\sqrt{\pi(n-k)}}\\ &= (1-\epsilon)^2 \sum_{p\le \frac{k}{n}\le q} \frac{1}{n} f(\frac{k}{n}) \xrightarrow{n\to\infty} (1-\epsilon)^2 \int_p^q f(x)\mathrm{d}x.\end{aligned}$$

Dakle,

$$\liminf_{n\to\infty} \sum_{np\le k\le nq} u_{2k}u_{2n-2k} \ge (1-\epsilon)^2 \int_p^q f(x)\mathrm{d}x, \quad \forall \epsilon > 0.$$

Budući da

$$\text{za } \epsilon \to 0, \quad 1-\epsilon \to 1,$$

iz gornje ograde slijedi da je

$$\liminf_{n\to\infty} \sum_{np\le k\le nq} u_{2k}u_{2n-2k} \ge \int_p^q f(x)\mathrm{d}x.$$

Slično se pokaže i:

$$\limsup_{n\to\infty} \sum_{np\le k\le nq} u_{2k}u_{2n-2k} \le \int_p^q f(x)\mathrm{d}x$$

pa slijedi

$$\lim_{n\to\infty} \sum_{np\le k\le nq} u_{2k}u_{2n-2k} = \int_p^q f(x)\mathrm{d}x \quad p, q \in (0,1). \tag{2.17}$$

- Imamo: kada $n \to \infty$,

$$\begin{aligned}\sum_{0\le k<pn} \alpha_{2k,2n} + \sum_{n\ge k>qn} \alpha_{2k,2n} = 1 - \sum_{np\le k\le nq} \alpha_{2k,2n} &\to \int_0^1 f(x)\mathrm{d}x - \int_p^q f(x)\mathrm{d}x\\ &= \int_0^p f(x)\mathrm{d}x + \int_q^1 f(x)\mathrm{d}x, \quad \forall p, q \in (0,1).\end{aligned}$$

  - $\forall \epsilon > 0$,

$$\sum_{0\le k<pn} \alpha_{2k,2n} \ge \sum_{n\epsilon\le k\le pn} \alpha_{2k,2n} \xrightarrow{n\to\infty} \int_\epsilon^p f(x)\mathrm{d}x.$$

Dakle,

$$\liminf_{n\to\infty} \sum_{0\le k<pn} \alpha_{2k,2n} \ge \int_\epsilon^p f(x)\mathrm{d}x, \quad \forall \epsilon > 0.$$

Puštanjem $\epsilon \to 0$ slijedi

$$\liminf_{n\to\infty} \sum_{0\le k<pn} \alpha_{2k,2n} \ge \int_0^p f(x)\mathrm{d}x.$$

■ $\forall p > 0$,

$$\sum_{0\le k<pn} \alpha_{2k,2n} \le \sum_{0\le k<pn} \alpha_{2k,2n} + \sum_{n(1-\epsilon)\le k<n} \alpha_{2k,2n} \xrightarrow{n\to\infty} \int_0^p f(x)\mathrm{d}x + \int_{1-\epsilon}^1 f(x)\mathrm{d}x$$

$$\Rightarrow \limsup_{n\to\infty} \sum_{0\le k<pn} \alpha_{2k,2n} \le \int_0^p f(x)\mathrm{d}x + \int_{1-\epsilon}^1 f(x)\mathrm{d}x, \quad \forall \epsilon > 0$$

Budući da $\int_{1-\epsilon}^1 f(x)\mathrm{d}x \to 0$ kada $\epsilon \to 0$,

$$\limsup_{n\to\infty} \sum_{0\le k<pn} \alpha_{2k,2n} \le \int_0^p f(x)\mathrm{d}x, \quad \forall p \in (0,1).$$

Sve skupa, imamo

$$\lim_{n\to\infty} \sum_{0\le k<pn} \alpha_{2k,2n} = \int_0^p f(x)\mathrm{d}x, \quad \forall p \in (0,1). \tag{2.18}$$

Na isti način se pokaže:

$$\lim_{n\to\infty} \sum_{qn\le k\le n} \alpha_{2k,2n} = \int_q^1 f(x)\mathrm{d}x, \quad \forall q \in (0,1). \tag{2.19}$$

Iz (2.17), (2.18) i (2.19) imamo da (2.15) vrijedi za sve : $p, q \in [0,1]$. □

Korolar 2.4.6. *Vrijedi*

$$\mathbb{P}\left(\frac{I_{2n}}{2n} \le x\right) \xrightarrow{n\to\infty} \frac{2}{\pi} \arcsin \sqrt{x}, \quad \forall x \in [0,1].$$

Specijalno,

$$\mathbb{P}\left(\frac{I_{2n}}{2n} \le \frac{1}{2}\right) \xrightarrow{n\to\infty} \frac{1}{2}.$$

x	$\frac{2}{\pi}\arcsin\sqrt{x}$
0	0
0.05	0.1436
0.1	0.2048
0.15	0.2532
0.2	0.2952
0.25	0.3333
0.3	0.3690
0.35	0.4030
0.4	0.4359
0.45	0.4681
0.5	0.5

Tablica 2.6: Vrijednosti od $\frac{2}{\pi}\arcsin\sqrt{x}$

Dokaz. Slijedi iz (2.12) i (2.15). Kada $n \to \infty$ imamo

$$\begin{aligned}\mathbb{P}\left(\frac{I_{2n}}{2n} \leq x\right) &= [\text{jer } (2.12)] = \sum_{k \leq xn} \alpha_{2n,2k} \\ [\text{pa iz } (2.15) \text{ slijedi (uz } p = 0, q = x)] &\to \frac{1}{\pi}\int_0^x \frac{1}{\sqrt{t(1-t)}}\mathrm{d}t \\ &= \frac{2}{\pi}\arcsin\sqrt{x}.\end{aligned}$$

□

U tablici 2.6 možemo vidjeti neke vrijednosti arkus sinus distribucije.

Vidjeli smo da je velika vjerojatnost da u dugoj igri bacanja novčića jedan od igrača provede gotovo cijelo vrijeme na pobjedničkoj strani, a drugi na gubitničkoj. Sada će nas zanimati koliko će vremena vodstvo biti na jednoj strani. Možda će nas intuicija ponovno zavarati i očekivat ćemo da je najveća vjerojatnost da svaki igrač vodi točno pola vremena. No vrijedi suprotno od toga; najveća je vjerojatnost da je cijelo vrijeme jedan igrač u vodstvu. Neka je

$$L'_n := |\{k : 1 \leq k \leq n, S_k > 0 \text{ ili } (S_k = 0 \text{ i } S_{k-1} > 0)\}|.$$

Uočimo L'_n predstavlja ukupno vrijeme koje je čestica provela u pozitivnom kvadrantu.

Teorem 2.4.7. $\forall n \geq 0$ *Vrijedi*

$$\mathbb{P}(L'_{2n} = 2k) = \alpha_{2k,2n}, \quad k = 0, ..., n$$

tj. L'_{2n} *ima diskretnu arkus sinus razdiobu reda n.*

To znači da je vjerojatnost da je čestica provela $2k$ trenutaka u pozitivnom kvadrantu i $2n - 2k$ u negativnom jednaka kao i vjerojatnost da se zadnji povratak dogodio u $2k$, to jest $\mathbb{P}(L'_{2n} = 2k) = \mathbb{P}(I_{2n} = 2k)$.

Dokaz. Moramo dokazati da vrijedi:

$$\mathbb{P}(L'_{2n} = 2k) = \alpha_{2k,2n}, \quad k = 0, ..., n.$$

Pogledajmo rubne slučajeve. Vrijedi

$$\mathbb{P}(L'_{2n} = 2n) = \mathbb{P}(S_1 \geq 0, ..., S_{2n} \geq 0) = [\text{slijedi iz leme 2.2.2}] = u_{2n}.$$

Očito onda vrijedi i $\mathbb{P}(L'_{2n} = 0) = u_{2n}$. Sada nam preostaje dokazati tvrdnju još za $1 \leq k \leq n - 1$. Budući da je čestica u ovim slučajevima provodila vrijeme na pozitivnoj i negativnoj strani morao se dogoditi bar jedan povratak u ishodište. Prisjetimo se:

- $R = \inf\{n \geq 1 : S_n = 0\}$ = prvi povratak šetnje u ishodište
- $f_{2r} := \mathbb{P}(R = 2r), r \geq 1$

Neka je $R = 2r,\ 2r < 2n$. Imamo dva slučaja.

- Prvi slučaj:

  $(S_1 > 0, ...S_{2r-1} > 0, S_{2r} = 0)$ i vrijedi $|\{j : 2r < j \leq 2n, S_j > 0$ ili $S_j = 0$ i $S_{j-1} > 0\}| = 2k-2r$. To jest gledamo putove koji imaju prvi povratak u ishodište u $2r$ i vrijedi $S_1 > 0$, a od ostalih $2n-2r$ koraka je njih $2k-2r$ čestica provela na pozitivnoj strani. Vjerojatnost ovog događaja je

  $$\frac{1}{2} f_{2r} \mathbb{P}(L'_{2n-2r} = 2k - 2r).$$

  Uočimo da mora vrijediti $r \leq k$ (inače je vjerojatnost jednaka nuli).

- Drugi slučaj:

  $(S_1 < 0, ...S_{2r-1} < 0, S_{2r} = 0)$ i vrijedi $|\{j : 2r < j \leq 2n, S_j > 0$ ili $S_j = 0$ i $S_{j-1} > 0\}| = 2k$. Slično kao u prvom slučaju vjerojatnost ovog događaja je

  $$\frac{1}{2} f_{2r} \mathbb{P}(L'_{2n-2r} = 2k).$$

  Uočimo mora vrijediti $r \leq n - k$.

Za $1 \leq k \leq n-1$ dakle vrijedi

$$\mathbb{P}(L'_{2n} = 2k) = \frac{1}{2}\sum_{r=1}^{k} f_{2r}\mathbb{P}(L'_{2n-2r} = 2k-2r) + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r}\mathbb{P}(L'_{2n-2r} = 2k). \tag{2.20}$$

Koristit ćemo matematičku indukciju. Očito vrijedi

$$\mathbb{P}(L'_2 = 2k) = \frac{1}{2} = \alpha_{2k,2k}, \quad k = 0, 1.$$

Pretpostavimo sada da vrijedi

$$\mathbb{P}(L'_{2v} = 2l) = \alpha_{2l,2v} = u_{2l}u_{2v-2l}, \quad \forall v \leq n-1, \quad \forall l = 0, 1, ...v.$$

Dokazat ćemo da vrijedi i za n. Primijenit ćemo pretpostavku na sumande u (2.20). Primijetimo u prvoj sumi imamo $\mathbb{P}(L'_{2n-2r} = 2k-2r)$, a budući da smo pretpostavili da tvrdnja vrijedi za sve $v \leq n-1$, vrijedi $\mathbb{P}(L'_{2n-2r} = 2k-2r) = \alpha_{2k-2r,2n-2r} = u_{2k-2r}u_{2n-2k}$. Slično vrijedi i za drugu sumu, pa zbog (2.8) imamo

$$\begin{aligned}
\mathbb{P}(L'_{2n} = 2k) &= \frac{1}{2}\sum_{r=1}^{k} f_{2r}u_{2k-2r}u_{2n-2k} + \frac{1}{2}\sum_{r=1}^{n-k} f_{2r}u_{2k}u_{2n-2r-2k} \\
&= u_{2n-2k}\frac{1}{2}\sum_{r=1}^{k} f_{2r}u_{2k-2r} + u_{2k}\frac{1}{2}\sum_{r=1}^{n-k} f_{2r}u_{2n-2k-2r} \\
&= \frac{1}{2}u_{2n-2k}u_{2k} + \frac{1}{2}u_{2k}u_{2n-2k} = u_{2k}u_{2n-2k} = \alpha_{2k,2n}.
\end{aligned}$$

Stoga tvrdnja vrijedi i za n. □

Korolar 2.4.8. *Vrijedi za $n \to \infty$*

$$\mathbb{P}\left(\frac{L'_{2n}}{2n} \leq x\right) \to \frac{2}{\pi} \arcsin \sqrt{x}, \quad x \in [0, 1].$$

Primjer 2.4.9. *Ako promatramo igru koja je trajala 20 bacanja vjerojatnost da je jedan od igrača vodio 16 ili više puta je* 0.685.
Primijetimo: Vjerojatnost da je sretniji igrač u vodstvu 16 ili više trenutaka povlači to da je manje sretan igrač u vodstvu 4 ili manje trenutaka. Stoga, dovoljno nam je računati vjerojatnost da je jedan od igrača bio u vodstvu 16 ili više te 4 ili manje trenutaka. To jest za $2n = 20$

$$\mathbb{P}(L'_{20} \geq 16) + \mathbb{P}(L'_{20} \leq 4) = 0.685.$$

Provjerimo

$$\begin{aligned}\mathbb{P}(L'_{20} \geq 16) + \mathbb{P}(L'_{20} \leq 4) &= 2\mathbb{P}(L'_{20} \geq 16) = [\,jer\ \mathbb{P}(L'_{20} \geq 16) = \mathbb{P}(L'_{20} \leq 4)] \\ &= 2 \cdot (\alpha_{20,20} + \alpha_{18,20} + \alpha_{16,20}) \\ &\approx 0.1762 + 0.0927 + 0.0655 + 0.1762 + 0.0927 + 0.0655 = 0.685.\end{aligned}$$

Vjerojatnost da su oba igrača vodila pola vremena ili da je sretniji od njih vodio najviše 12 puta je 0.184.

To jest

$$\mathbb{P}(8 \leq L'_{20} \leq 12) = 0.184.$$

Provjerimo:

$$\begin{aligned}\mathbb{P}(8 \leq L'_{20} \leq 12) &= \alpha_{12,20} + \alpha_{10,20} + \alpha_{8,20} \\ &\approx 0.0617 + 0.0606 + 0.0617 = 0.184.\end{aligned}$$

Primjer 2.4.10. *Zamislimo da se svake sekunde baca simetrični novčić 365 dana zaredom, to jest da je broj bacanja* $2n = 31536000$*. Dakle, u prosjeku, 9 od 10 puta će manje sretni igrač biti u vodstvu kraće od 13301280 sekundi to jest* 153.95 *dana. Za razne vjerojatnosti* p *navedena su vremena* t_p *koja predstavljaju vremena trajanja vodstva za manje sretne igrače. U tablici 2.7 možemo vidjeti još neke vjerojatnosti za trajanje vodstva manje sretnih igrača.*

p	t_p
0.9	153.95 dana
0.8	126.1 dana
0.7	99.65 dana
0.6	75.23 dana
0.5	53.45 dana
0.4	34.85 dana
0.3	19.89 dana
0.2	8.93 dana
0.1	2.24 dana
0.01	32.4 minuta
0.001	2.6 minuta

Tablica 2.7: Vjerojatnosti za manje sretne igrače

2.5 Promjene vodstva/predznaka

U ovom dijelu ćemo se kratko osvrnuti na očekivani broj vodstva te koliko se povećava očekivani broj promjena vodstva sa rastom duljine igre.

Promjena vodstva se dogodila u trenutku n ako vrijedi

$$S_n = 0, S_{n-1} \neq S_{n+1}.$$

Definicija 2.5.1. *Neka je, za sve $n \geq 1$,*

$$Y_n := |\{1 \leq k \leq n : \ S_k = 0, \ S_{k-1} \neq S_{k+1}\} = [\textit{broj promjena vodstva do trenutka } n].$$

Teorem 2.5.2. *Vjerojatnost da se do trenutka $2n+1$ dogodilo točno r promjena vodstva je*

$$\xi_{r,2n+1} := \mathbb{P}(Y_{2n+1} = r) = 2\mathbb{P}(S_{2n+1} = 2r+1) = 2p_{2r+1,2n+1}.$$

(dokaz možemo pronaći u [2] poglavlje III, potpoglavlje 5)

Primijetimo da se vjerojatnost za n izjednačenja smanjuje kako n raste, odnosno

$$\xi_{0,n} > \xi_{1,n} > \xi_{2,n} > ...$$

jer

$$2p_{n,1} > 2p_{n,3} > ....$$

To jest ako fiksiramo neki proizvoljni $r > 0$, vjerojatnost da se dogodi r promjena vodstva manja je od vjerojatnosti da se uopće ne dogodi promjena vodstva. Provjerimo

Dovoljno je dokazati $p_{2n+1,1} > p_{2n+1,3}, \quad \forall n > 0$

$$\begin{aligned}
\binom{2n+1}{\frac{2n+2}{2}} = \binom{2n+1}{n+1} &= \frac{(2n+1)2n...(2n+1-(n+1)+1)}{(n+1)!} \\
&= \frac{(2n+1)2n...(n+1)}{(n+1)!} \\
= \frac{(2n+1)2n...(n+1)}{(n+2)!} \cdot (n+2) > [\text{za sve } n > 0] > &\frac{(2n+1)2n...(n+1)}{(n+2)!} \cdot n \\
&= \binom{2n+1}{n+2} = \binom{n}{\frac{2n+4}{2}} = p_{2n+1,3}.
\end{aligned}$$

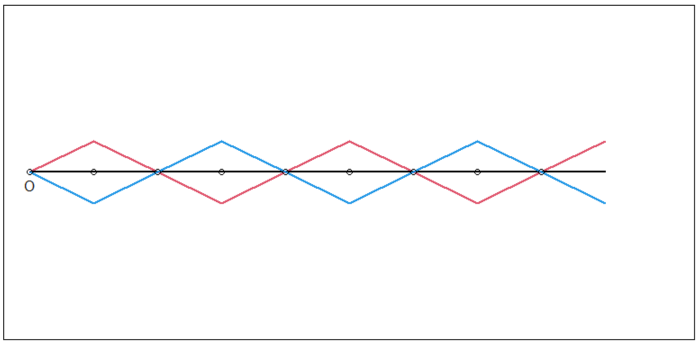
Za primjer pogledati tablicu 2.8.

Najveći mogući broj izjednačenja u igri duljine n dan je s $r_{max} := \left\lfloor \frac{n-1}{2} \right\rfloor$.

r	$\xi_{r,21}$
0	0.33637619
1	0.28031349
2	0.19406319
3	0.11089325
4	0.05175018
5	0.01940632
6	0.00570774
7	0.00126839
8	0.00020027
9	0.00002003
10	0.00000095

Tablica 2.8: Tablica vjerojatnosti za $2n + 1 = 21$

Uočimo to će se dogoditi samo u slučajevima kada se šetnja vraća u ishodište svaka 2 koraka i ponavlja se isti uzorak X_i. Odnosno kada $X_1 = 1, X_2 = -1, X_3 = -1, X_4 = 1, X_5 = 1, X_6 = 1, \ldots$. ili obratno ako se počinje s $X_1 = -1$. Pogledajmo na slici (2.3) (crveno je prikazan prvi, a plavo drugi primjer). Budući da se neovisno o duljini igre ovaj slučaj javlja samo 2 puta vjerojatnost je očekivano mala.



Slika 2.3: Maksimalni broj promjena vodstva

Napomena 2.5.3. *Pretpostavimo da je n dovoljno velik i fiksiramo $x > 0$. Vjerojatnost* *da se manje od $x\sqrt{n}$ promjena vodstva dogodi prije trenutka n je približno $2\mathbb{P}(0 < S_n < 2x\sqrt{n})$. Za $\mu = 0, \sigma = 1$ po(1.2.15) vrijedi* ***(normalna aproksimacija)*** *da vjerojatnost da se dogodilo manje od $x\sqrt{n}$ promjena vodstva prije trenutka n teži u $2\Phi(2x) - 1$ za $n \to \infty$. To jest*

$$
\begin{aligned}
&\mathbb{P}(0 < S_n < 2x\sqrt{n})\\
&=\mathbb{P}\left(0 < \frac{S_n}{\sqrt{n}} < 2x\right)\\
[\textit{Po 1.2.15}] \longrightarrow &\int_0^{2x} \frac{1}{\sqrt{2\pi}} e^{\frac{-1}{2}t^2}\,\mathrm{d}t\\
&= \Phi(2x) - \Phi(0) = \Phi(2x) - \frac{1}{2}.
\end{aligned}
$$

Slijedi da je medijan za broj promjena vodstva približno jednak $0.337\sqrt{n}$ (jer za $x = 0.337,\ 2\Phi(2x) - 1 \approx \frac{1}{2}$) za velike n.

Poglavlje 3

Simulacija i analiza bacanja novčića

U ovom poglavlju pogledat ćemo rezultate nekih simulacija bacanja 10000 simetričnih novčića te analizirati ih. U prvom dijelu ćemo analizirati 10 igara dugih 10000 koraka. Prvo primijetimo da je arkus sinus distribucija poseban slučaj beta distribucije za $\alpha = \beta = \frac{1}{2}$ to jest za slučajnu varijablu Z sa arkus sinus distribucijom možemo pisati $Z \sim Beta(\frac{1}{2}, \frac{1}{2})$. Onda za slučajnu varijablu Z vrijedi

$$E[Z] = \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{2}} = \frac{1}{2}$$

pa za I_{10000} i L'_{10000} vrijedi

$$E[I_{10000}] = E[L'_{10000}] \approx 5000$$

i budući da je arkus sinusova distribucija simetrična medijan je jednak $\frac{1}{2}$ odnosno približno 5000 za I_{10000} i L'_{10000}. To znači da očekujemo da će podjednaki broj igara imati povratak prije odnosno poslije 5000-itog koraka. Slično, očekujemo da će otprilike pola igara provesti više od 5000 trenutaka u pozitivnom kvadrantu, a pola manje. Medijan za broj promjena vodstva u igri dugoj 10000 koraka je približno 33.7.

Uočimo očekivanje slučajne varijabla X_k jednako je

$$E[X_k] = \frac{1}{2} \cdot (-1) + \frac{1}{2} \cdot 1 = 0, \quad \forall k.$$

Budući da je to prilagođena verzija Bernoullijeve slučajne varijable možemo izračunati i varijancu

$$Var[X_k] = E[X_k^2] - (E[X_k])^2 = E[X_k^2] = (-1)^2 \cdot \frac{1}{2} + 1^2 \cdot \frac{1}{2} = 1, \quad \forall k.$$

Uočimo da je standardna devijacija jednaka $\sigma = 1$. Primjenimo i jaki zakon velikih brojeva na jednostavnu simetričnu slučajnu šetnju. Vrijedi

$$\mathbb{P}(\lim_{n\to\infty} \overline{X_n} = 0) = 1.$$

Napomena 3.0.1. *Kako bi dobili bolju sliku o kretanju vjerojatnosti $\mathbb{P}(S_n = r)$ koristimo aproksimacijsku formulu dobivenu aproksimacijskim teoremom koji predstavlja posebnu verziju centralnog graničnog teorema 1.2.15 (za više detalja pogledati [2] poglavlje III i poglavlje VII). Vjerojatnost $\mathbb{P}(a < S_n = r < b)$ dobivena je sumiranjem vjerojatnosti $p_{n,r}$ za sve r, takve da $a < r < b$. Pogledajmo aprokismaciju za $\mathbb{P}(S_n > a)$ za $n \to \infty$*

$$\mathbb{P}(S_n > x\sqrt{n}) \to 1 - \Phi(x) = \frac{1}{2\pi}\int_x^\infty e^{\frac{-1}{2}t^2}\mathrm{d}t.$$

To jest $\frac{S_n}{\sqrt{n}}$ ima približno normalnu distribuciju kada $n \to \infty$.

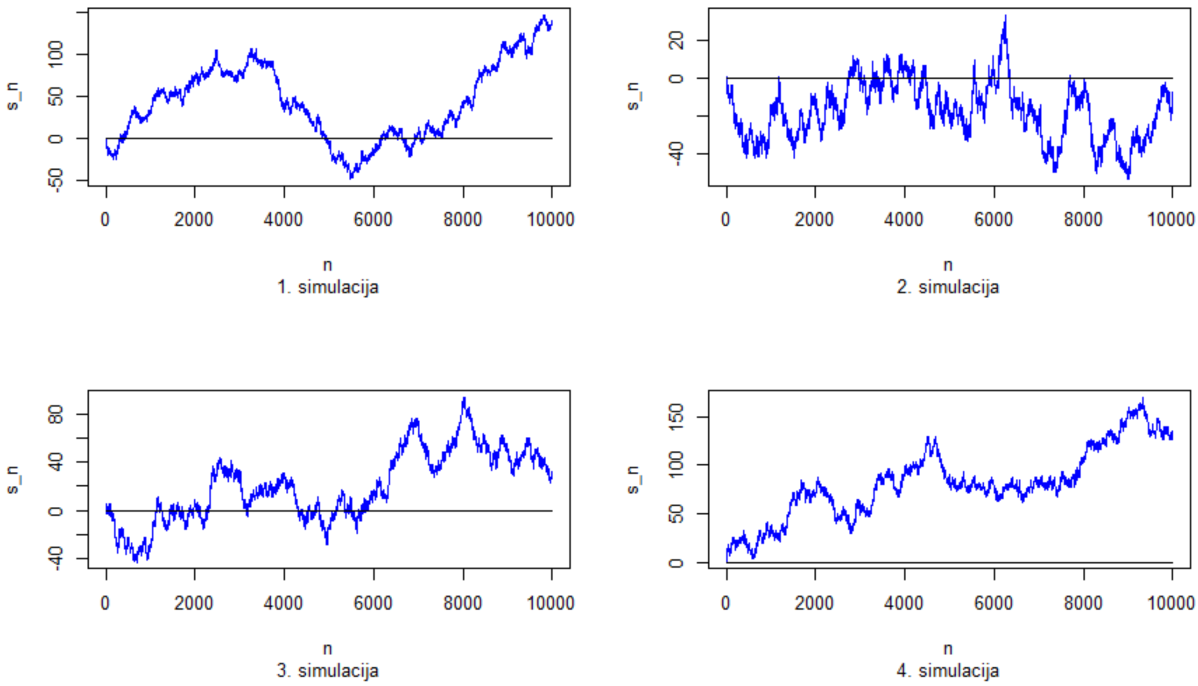
x	0.5	1.0	1.5	2.0	2.5	3.0
$\mathbb{P}(S_n > x\sqrt{n})$	0.309	0.159	0.067	0.023	0.006	0.001

Tablica 3.1: Primjer nekih vrijednosti za $\mathbb{P}(S_n > x\sqrt{n})$

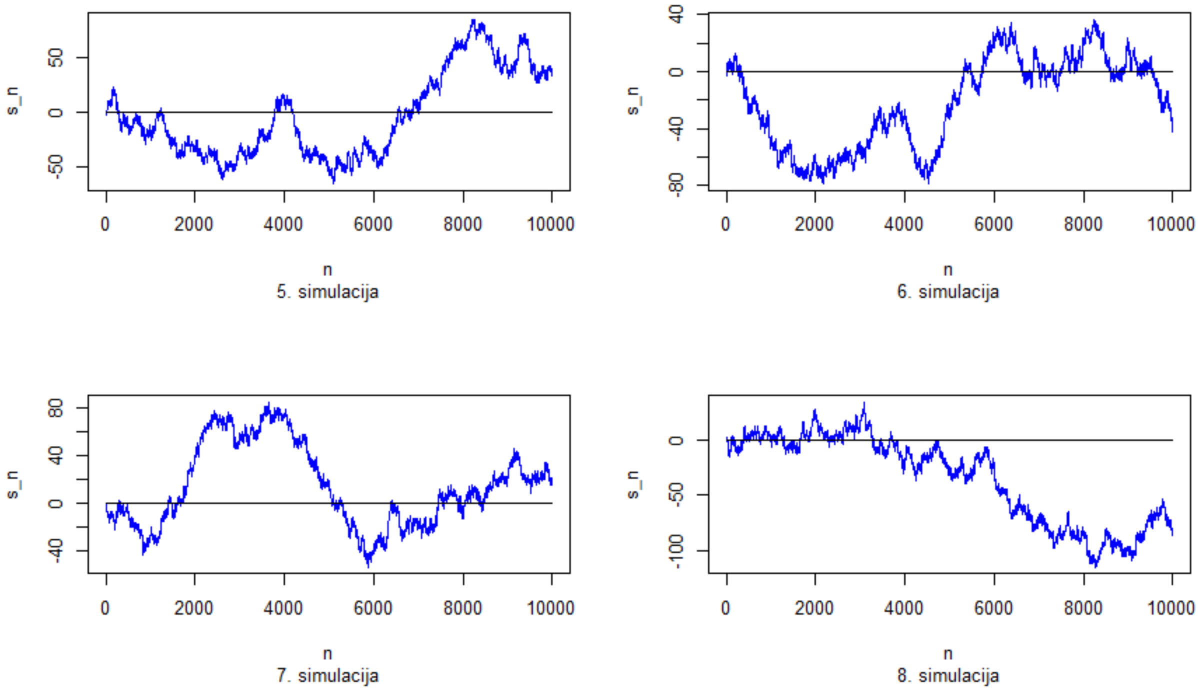
U idućem potpoglavlju se nalaze grafovi simulacija 10 igara duljine 10000 koraka.

3.1 Simulacije i prikaz 10 trajektorija

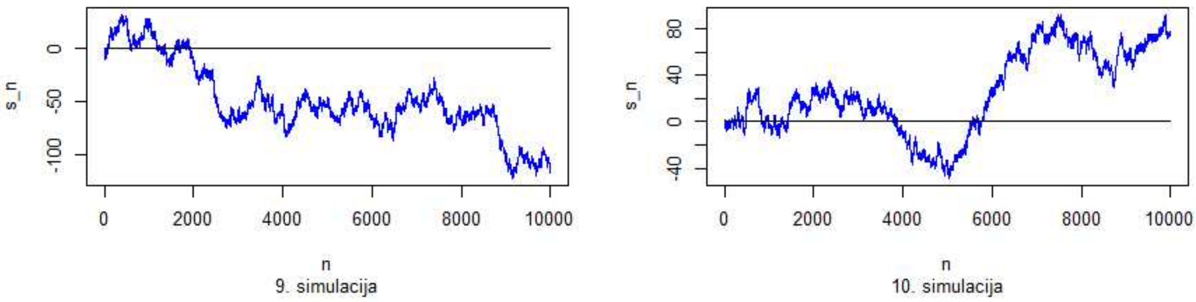
U ovom potpoglavlju nalaze se grafovi simulacija odrađenih u programskom jeziku R. Simulirano je 10 igara duljine 10000 koraka.



Slika 3.1: Simulacije 1-4



Slika 3.2: Simulacije 5-8



Slika 3.3: Simulacije 9-10

3.2 Analiza rezultata

Pogledajmo sada tablicu vjerojatnosti za posljednje povratke u igri dugoj 10000 koraka.

p	$t_p = 2nx$
0.9	4218
0.8	3455
0.7	2730
0.6	2061
0.5	1464
0.4	955
0.3	545
0.2	245
0.1	62
0.05	15

Tablica 3.2: tablica analogna 2.7 za $2n = 10000$

U idućoj tablici nalaze se rezultati simuliranih igara u prošlom potpoglavlju.

Igra	I_{2n}	L'_{2n}	Y_{2n}	S_{10000}
1	7537	7948	39	136
2	8019	1364	55	-10
3	5867	7162	99	32
4	3	10000	0	134
5	7009	3956	38	36
6	9571	3386	84	-42
7	8457	6044	66	16
8	3825	2374	82	-84
9	1931	1432	42	-118
10	5771	7394	70	74
Prosjek	5799	5106	57.5	17.4

Tablica 3.3: $I_{2n}, L_{2n}, Y_{2n}, S_{10000}$, 10 testiranih igara

Uspoređujući s tablicom 3.2 dobili smo očekivane rezultate. U 9 od 10 igara je manje sretni igrač bio u vodstvu kraće od 4218 trenutaka, 7 puta kraće od 3455, a u jednoj igri se nije ni dogodila promjena vodstva. Posljednji povratak se u prosjeku dogodio u 5799-om koraku što nije veliko odstupanje od očekivanog prosjeka (5000) s obzirom na mali broj igara, a prosječan provod vremena u pozitivnom kvadrantu bio je još bliži očekivanju.Kada promatramo broj promjena vodstva rezultati su neočekivani jer imamo samo jednu igru koja je imala manje pd 33.7 (medijan) promjena vodstva.

y	50	100	150	200	250	300
$\mathbb{P}(S_{10000} > y)$	0.309	0.159	0.067	0.023	0.006	0.001

Tablica 3.4: Tablica analogna s 3.1 prilagođena za $2n = 10000$

Najveće odstupanje S_{2n} od 0 koje smo dobili iznosi 136 te smo imali 3 igre koje su završavale s većim odmakom od 100. U teoriji ta je vjerojatnost

$$2\mathbb{P}(S_{10000} > 100) \approx 0.3174$$

što znači da i očekujemo da će u prosjeku 3 od 10 igara imati takav odmak.

3.3 Simulacija 500 trajektorija

U ovom dijelu rada proučit ćemo primjenu zakona velikih brojeva i centralnog graničnog teorema na jednostavnu slučajnu simetričnu šetnju te pogledati dodatne simulacije napravljene u R-u.

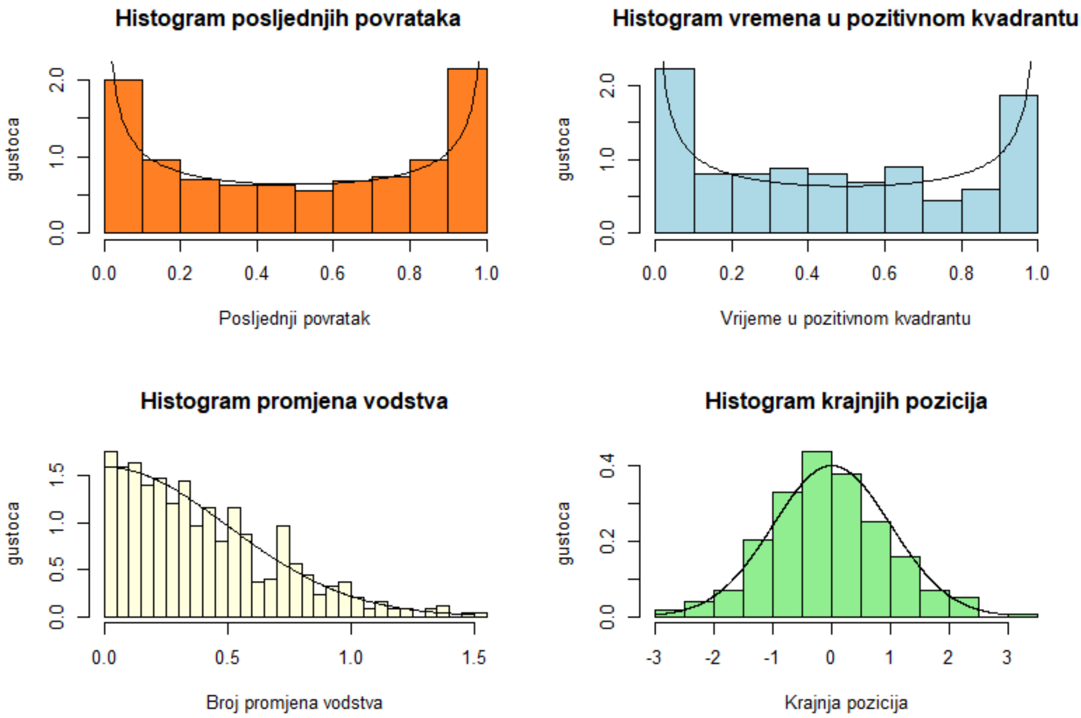
U testiranju imamo 500 igara u kojima su zabilježene vrijednosti u tablici ispod (samo ukupni prosjek).

	I_{2n}	L'_{2n}	Y_{2n}	S_{10000}
Prosjek	5092.564	4656.222	40.442	-4.584

Tablica 3.5: Rezultati 500 igara duljine 10000 koraka

Pogledajmo za početak kako izgledaju histogrami uzoraka za ranije spomenute vrijednosti.

Iz histograma promjena vodstva možemo primjetiti da su rezultati očekivani to jest da je najviše bilo igara s malim brojem promjena vodstva i da opada. U 248 igara bilo je više od 33.7 (medijan) promjena vodstva što je vrlo blizu očekivanja (250). Histogram krajnjih pozicija pokazuje da najveći broj igara završava blizu ishodišta. Histogrami posljednjih povrataka i pozitivnog vremena nalikuju na arkus sinus distribuciju. Najveći broj igara je imao povratak ili relativno blizu početka ili kraja igre, odnosno, u čak 41% slučajeva manje sretni igrač je bio u vodstvu kraće od 1000 poteza. A po korolaru 2.4.6 znamo da je $\mathbb{P}(I_{2n} < 1000) \approx \frac{2}{\pi} \arcsin \sqrt{x} = 0.205$.



Slika 3.4: Histogrami
Histogrami od I_{2n} (histogram posljednjih povrataka), L'_{2n} (histogram vremena u pozitivnom kvadrantu), Y_{2n} (histogram promjena vodstva), S_{2n} (histogram krajnjih pozicija) redom s asimptotskim distribucijama

Bibliografija

[1] E. Ackelsberg, *What is the Arcsine Law?*, (2018), https://math.osu.edu/sites/math.osu.edu/files/What_is_2018_Arcsine_Law.pdf.

[2] W. Feller, *An Introduction to Probability Theory and Its Applications, Volume I*, John Wiley & Sons, Inc., 1968.

[3] Welsh D. Grimmett G., *Probability, an introduction, Second edition*, Oxford university press, 1968.

[4] Vondraček Z. Sandrić N., *Vjerojatnost, skripta s predavanja*, (2019), https://www.pmf.unizg.hr/images/50023697/vjer_predavanja.pdf.

Sažetak

U ovom diplomskom radu proučavamo jednostavne simetrične slučajne šetnje i uočavamo neke naizgled neočekivane rezultate. Kroz primjere uviđamo da neki naizgled neočekivani rezultati imaju veću vjerojatnost nego što smo pretpostavili.

Pokazuje se da posjeti ishodištu nisu toliko učestali te da je vjerojatnost da se čestica vratila u ishodište u nekom trenutku jednaka vjerojatnosti da se do tog trenutka uopće nije vratila. Pokazujemo da intiucija zavarava te da prethodni dio šetnje ne utječe na nastavak (kockarova propast). Proučavamo slučajne varijable koje imaju arkus sinusovu distribuciju te pratimo svojstva iste. To nam pokazuje neke zanimljive rezultate koji su iznenađujući. Pokazuje se da ne očekujemo veliki broj promjena vodstva u igri i da postoji velika vjerojatnost da jedna strana provede većinu vremena u vodstvu.

Nalazimo dobru aproksimaciju za vjerojatnost u kojoj točki će slučajna šetnja završiti pomoću centralnog graničnog teorema.

Preko simulacija i grafova izrađenih simulacija bolje shvaćamo ponašanja jednostavnih simetričnih slučajnih šetnji i uspoređujemo s očekivanim rezultatima.

Summary

In this paper, we study simple symmetrical random walks and observe some seemingly unexpected results. Through some examples, we see that some seemingly unexpected results have a higher probability than we assumed.

It is shown that the visits to the origin are not so frequent and that the probability that the particle has returned to the origin at some point in time is equal to the probability that it has not returned at all until that moment. We show that intuition is deceiving and that the previous part of the walk does not affect the continuation (gambler's ruin). We study random variables that have an arcsine distribution and monitor its properties. This shows us some interesting results that are surprising. It turns out that we don't expect a lot of lead changes in the game and that there is a big chance one side spends most of the time in the lead.

An interesting result is that if we fix that our walk ends at the origin, the distribution of spending time in the positive quadrant ceases to be an arcsine distribution and becomes unfornm.

We find a good approximation for the probability at which point the random walk will end using the central limit theorem.

Using simulations and graphs of the created simulations, we better understand the behavior of simple symmetric random walks and compare them with the expected results.

Životopis

Rođen sam 28. siječnja 1996. u Zadru. Tamo sam završio gimnaziju Franje Petrića te sam 2014. upisao Preddiplomski sveučilišni studij Matematika na Prirodoslovno-Matematičkom fakultetu Sveučilišta u Zagrebu. 2016 godine prebacio sam se na smjer Matematika: Nastavnički te 2018. upisujem Diplomski sveučilišni studij Matematika i informatika; nastavnički smjer na istom fakultetu. 2020. godine zaposlio sam se u IT odjelu Plive gdje trenutno radim.