Vjerojatnosne razdiobe u nogometu

Brkić, Filip

SVEUČILIŠTE U ZAGREBU

PRIRODOSLOVNO–MATEMATIČKI FAKULTET

MATEMATIČKI ODSJEK

Filip Brkić

VJEROJATNOSNE RAZDIOBE U NOGOMETU

Diplomski rad

Voditelj rada:
doc. dr. sc. Franka Miriam Brückler

Zagreb, ožujak 2021.

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. ____________________________

Ovaj rad posvećujem svojoj supruzi Maji, roditeljima te sestri koji su mi uvijek bili podrška i nisu mi dali da odustanem.

Sadržaj

Uvod

Počeci vjerojatnosti vezani su uz igre na sreću u XVI. i XVII. stoljeću. Veliki doprinos u samim početcima dali su Pascal i Fermat dok se do značajnijih rezultata dolazi u XVIII. i početkom XIX. stoljeća. Za te su rezultate posebno značajni Jacob Bernoulli, De Moivre i Laplace. Danas se teorija vjerojatnosti, osim u matematici, primjenjuje u mnogim područjima kao što su fizika, biologija i tehnika. Među ostalim, vjerojatnost se primjenjuje i u analizi nogometnih utakmica i prvenstava i predviđanju njihovih ishoda, što ćemo opisati u ovom radu.

U prvom poglavlju ovog rada naveli smo ključne pojmove koje koristimo u glavnom dijelu rada. Drugo poglavlje opisuje primjene binomne razdiobe u opisivanju vjerojatnosti pojedinih rezultata utakmica. U trećem poglavlju opisujemo upotrebu Poissonove razdiobe u predviđanju ishoda i analizi nogomentnih utakmica, a u četvrtom poglavlju primjenjujemo normalnu razdiobu na analizu nogometnih prvenstava. Cilj je ovog rada dati pregled osnovnih primjena tih triju razdioba u analizi i predviđanju ishoda nogometnih utakmica i prvenstava, do razine razumljive studentima prirodnih i tehničkih znanosti.

Poglavlje 1

Osnovni pojmovi

1.1 Elementarni događaj. Vjerojatnosni prostor

Pogledajmo pokus bacanja jednog simetričnog novčića ili tzv. igru „pismo - glava”. Pokus se izvodi na način da se novčić baci uvis. Nakon što novčić padne na ravnu podlogu, na njegovoj se gornjoj strani nalazi pismo (P) ili glava (G). Dakle, možemo očekivati da su jedina dva ishoda bacanja novčića (možemo pretpostaviti da novčić neće pasti na rub ili da se neće dogoditi neki sličan događaj):

- novčić pokazuje pismo (P),
- novčić pokazuje glavu (G).

Ne znamo hoće li novčić pokazati pismo ili glavu, ali znamo da će ishod bacanja novčića biti jedan od elemenata dvočlanog skupa

$$\Omega = \{P, G\}.$$

Analogno, bacanje (simetrične) igraće kocke je pokus kod kojeg je svaki ishod jedan od elemenata skupa

$$\Omega = \{1, 2, 3, 4, 5, 6\},$$

no unaprijed ne možemo znati koji.

Navedeni pokusi primjeri su slučajnih pokusa. Ishodi slučajnih pokusa nisu unaprijed određeni. Teorija vjerojatnosti je matematička disciplina čiji je zadatak formiranje i proučavanje matematičkog modela slučajnog pokusa. Rezultate slučajnih pokusa zvat ćemo **događajima**. Razlikujemo složene (razložive) i elementarne (nerazložive) događaje. Složeni događaji međusobno se ne isključuju, tj. načelno je moguće da se istovremeno dogode dva složena događaja (primjerice, kad bacamo kocku moguće je da istovremeno padne paran broj i broj veći od 3). Elementarni događaji međusobno se isključuju, tj. ne mogu se istovremeno dogoditi dva različita elementarna događaja (primjerice, nemoguće je da pri bacanju novčića istovremeno padnu pismo i glava) [15].

Osnovni polazni objekt u teoriji vjerojatnosti je neprazan skup Ω, kojeg nazivamo **prostorom elementarnih događaja**, tj. to je skup svih elementarnih događaja slučajnog pokusa.

Prije definiranja vjerojatnosti, primijetimo da nije uvijek moguće sve podskupove od Ω uzeti za događaje. Naime, može se dogoditi da za neki podskup A i element ω od Ω ne možemo dati odgovor „Je li $\omega \in A$?". Zbog navedenog problema, vjerojatnost se formalno definira aksiomatski preko familija skupova i σ-algebri, te je vjerojatnost svaka nenegativna realna funkcija P kojoj je domena neka σ-algebra $\mathcal{F}$ (njene elemente nazivamo događajima) na skupu Ω takva da je $P(\Omega) = 1$ i da je σ-aditivna, tj. da vrijedi

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

za svaku familiju događaja A_i, $i \in \mathbb{N}$ koji se međusobno u parovima isključcuju ($A_i \cap A_j = \emptyset$ za $i \neq j$) [15].

Definicija 1.1.1. *Neka je* $(\Omega, \mathcal{F}, P)$ *dani vjerojatnosni prostor i neka je* $A \in \mathcal{F}$ *događaj.* ***Suprotni događaj*** *događaju* A *je njegov komplement, tj. događaj* $A^c = \Omega \setminus A$.

No, za našu razinu rada nije potreban takav formalni pristup jer će u gotovo svim našim razmatranim primjerima imati **diskretan slučaj**, tj. skup Ω će biti konačan (ili prebrojiv[1]) pa kao σ-algebru $\mathcal{F}$ možemo uzeti cijeli njegov partitivni skup $\mathcal{P}(\Omega)$, a za vjerojatnosti pojedinih događaja ćemo u većini primjera moći uzeti **klasičnu vjerojatnost *a priori***: Vjerojatnost događaja $A \subseteq \Omega$ jednaka je broju za taj događaj povoljnih elementarnih događaja (broj elemenata od A) podijeljen s brojem svih mogućih elementarnih događaja (brojem elemenata od Ω).

Za klasičnu vjerojatnost *a priori* vrijede sljedeće relacije koje se mogu dokazati i u općem slučaju [15]:

Teorem 1.1.1. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor. Tada vrijedi*

(a) $P(\emptyset) = 0$;

(b) Ako su $A_1, \ldots, A_n \in \mathcal{F}$ *međusobno u parovima disjunktni, tada je* $P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$;

(c) Ako su $A, B \in \mathcal{F}$ *i* $A \subseteq B$, *onda je* $P(A) \leq P(B)$;

(d) Ako su $A_n \in \mathcal{F}, n \in \mathbb{N}$, *onda vrijedi* $P(\cup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$;

[1] Skup je prebrojiv ako postoji bijekcija sa skupa $\mathbb{N}$ na taj skup.

(e) Ako su $A, B \in \mathcal{F}$, onda je $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

(f) Za svaki $A \in \mathcal{F}$ vrijedi $P(A^c) = 1 - P(A)$.

Primjer 1.1.2. *U ovom radu Ω će često biti skup svih mogućih brojeva pogodaka koje može dati neka nogometna momčad tijekom jedne utakmice, dakle $\Omega = \mathbb{N}_0$. Naravno, nije realno da će momčad dati više od desetak pogodaka, tako da ćemo u mnogim primjerima moći uzeti npr. $\Omega = \{0, 1, \dots, 10\}$. Vrijediti će, primjerice, da je vjerojatnost da momčad ne postigne zgoditak jednaka 1 minus vjerojatnost da ga postigne, a vjerojatnost da momčad postigne neparnih broj pogodaka bit će jednaka zbrojevima vjerojatnosti da ih postigne 1, 3, 5, ...*

U ovom radu ćemo na mnogim mjestima pretpostavljati nezavisnost događaja:

Definicija 1.1.2. *Za dva događaja A_1 i A_2 kažemo da su **nezavisni događaji** ako vrijedi relacija*

$$P(A_1 \cap A_2) = P(A_1) \cdot P(A_2).$$

1.2 Slučajne varijable

Prilikom izvođenja slučajnih pokusa najčešće izvodimo neka mjerenja, tj. svakom rezultatu slučajnog pokusa pridružujemo neki realan broj. Promatrat ćemo realne funkcije na Ω koje ćemo zvati slučajnim varijablama. Slučajne varijable označavat ćemo velikim tiskanim slovima *X, Y, Z, ...*.

Promotrimo prvo diskretni vjerojatnosti prostor (u ovom radu imat ćemo skoro isključivo tu situaciju) $(\Omega, \mathcal{P}(\Omega), P)$ i neka je on model nekog slučajnog pokusa, primjerice promatranja broja pogodaka koje na pojedinoj utakmici postiže određena momčad.

Definicija 1.2.1. *Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor. **Slučajna varijabla** je proizvoljna realna funkcija definirana na Ω.*

Dakle, X je slučajna varijabla ako je $X : \Omega \to \mathbb{R}$. Slučajna varijabla modelira moguće ishode nekog slučajnog pokusa.

Definicija 1.2.2. *Neka je X diskretna slučajna varijabla. **Razdiobom** od X zovemo niz vrijednosti a_n od X s pozitivnom vjerojatnošću i pripadnih vjerojatnosti p_n. Pišemo*

$$X \sim \begin{pmatrix} a_1 & a_2 & ... & a_n \\ p_1 & p_2 & ... & p_n \end{pmatrix},$$

odnosno

$$X \simeq \begin{pmatrix} a_1 & \dots & a_n & \dots \\ p_1 & \dots & a_n & \dots \end{pmatrix}.$$

Uočimo da mora vrijediti $\sum_n p_n = 1$ [15].

Primjer 1.2.1. *U kontekstu našeg rada, slučajna varijabla X može primjerice biti broj pogodaka koje neka momčad daje u nekoj utakmici. Ta je slučajna varijabla očito diskretna jer poprima vrijednosti u skupu* $\mathbb{N}_0$.

Definicija 1.2.3. *Neka je* $(\Omega, \mathcal{F}, P)$ *vjerojatnosni prostor. Za funkciju* $X : \Omega \to \mathbb{R}$ *kažemo da je* ***kontinuirana slučajna varijabla*** *ako je slika* $\mathcal{R}(X)$ *interval* $I \subseteq \mathbb{R}$, *ako slika ne sadrži izolirane točke i ako je za svaki interval* $I \subseteq \mathbb{R}$ *praslika* $\{\omega \in \Omega : X(\omega) \in I\}$ *familija skupova u* $\mathcal{F}$.

Definicija 1.2.4. *Neka je* $X : \Omega \to \mathbb{R}$ *kontinuirana slučajna varijabla sa slikom* $\mathcal{R}(X)$. *Funkciju* $f : \mathbb{R} \to \mathbb{R}$ *definiranu na sljedeći način:*

$$f(x) = \lim_{\Delta x \to 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x}, \ x \in \mathcal{R}(X)$$

zovemo ***funkcijom gustoće vjerojatnosti*** *kontinuirane slučajne varijable X.*

Definicija 1.2.5. *Neka je* $(\Omega, \mathcal{P}(\Omega), P)$ *vjerojatnosni prostor i* $X_1, X_2, ..., X_k$ *slučajne varijable na* Ω. *Kažemo da su* $X_1, X_2, ..., X_k$ ***nezavisne slučajne varijable*** *ako za proizvoljne* $B_i \subseteq \mathbb{R}, i = 1, ...k$ *vrijedi*

$$P\{X_1 \in B_1, X_2 \in B_2, ..., X_k \in B_k\} = P(\cap_{i=1}^{k} \{X_i \in B_i\}) = \prod_{i=1}^{k} P\{X_i \in B_i\}.$$

Teorem 1.2.2. *Neka su* $X_1, X_2, ..., X_k$ *slučajne varijable na diskretnom vjerojatnosnom prostoru* $(\Omega, \mathcal{P}(\Omega), P)$ *i neka je*

$$X_i \sim \begin{pmatrix} a_1^{(i)} & a_2^{(i)} & ... & a_n^{(i)} & ... \\ p_1^{(i)} & p_2^{(i)} & ... & p_n^{(i)} & ... \end{pmatrix}, i = 1, ..., k.$$

Varijable $X_1, X_2, ..., X_k$ *su nezavisne ako i samo ako vrijedi*

$$P\{X_1 = a_{i_1}^{(1)}, ..., X_k = a_{i_k}^{(k)}\} = \prod_{j=1}^{k} p_{i_j}^{j}.$$

U ovom radu svuda ćemo pretpostavljati nezavisnost slučajnih varijabli koje koristimo, iako je u stvarnom nogometnom kontekstu prava nezavisnost nemoguća. Primjerice, uzimat ćemo da su brojevi pogodaka koje daju dvije momčadi u nekoj međusobnoj utakmici nezavisni, iako naravno da će više ili manje pogodaka koje daje jedna momčad utjecati na igru i time na broj pogodaka druge. Ipak, u elementarnim primjerima, a samo njima se bavimo u ovom radu, pretpostavka nezavisnosti olakšava (pa čak i omogućuje) pristup, a pritom svejedno dobivamo smislene, bar donekle realistične rezultate.

1.3 Očekivanje i varijanca

Neka je $(\Omega, \mathcal{P}(\Omega), P)$ diskretan vjerojatnosni prostor, $\Omega = \{\omega_1, \omega_2, ...\}$ i X slučajna varijabla na Ω, tj. $X : \Omega \to \mathbb{R}$.

Definicija 1.3.1. *Ako red $\sum_{\omega_k \in \Omega} X(\omega_k) P(\{\omega_k\})$ apsolutno konvergira onda njegovu sumu zovemo* ***matematičkim očekivanjem*** *slučajne varijable X i označavamo sa*

$$EX = \sum_{\omega_k \in \Omega} X(\omega_k)\, P(\omega_k).$$

U slučaju konačnog skupa Ω svaka slučajna varijabla X ima očekivanje. Za kontinuirane slučajne varijable X njihovo se očekivanje definira ovako:

Definicija 1.3.2. *Neka je $X : \Omega \to \mathbb{R}$ kontinuirana slučajna varijabla sa slikom $\mathcal{R}(X)$ i funkcijom gustoće vjerojatnosti $f(x)$. Kažemo da kontinuirana slučajna varijabla X ima očekivanje ako integral $\int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x$ konvergira i označavamo*

$$EX = \int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x.$$

Primjer 1.3.1. *Ako primjerice gledamo slučajnu varijablu kao u primjeru 1.2.1, kao njezino očekivanje u ovom ćemo radu uzimati prosječni broj pogodaka koje je promatrana momčad dala po utakmici u nekom relevantnom prethodnom vremenskom intervalu (npr. u prethodnoj sezoni).*

Iz definicije matematičkog očekivanja slijedi da, ako X ima očekivanje i $X \geq 0$, tj. $X(\omega) \geq 0, \omega \in \Omega$, tada je $EX \geq 0$. Odavde i iz gornjeg teorema slijedi da, ako slučajne varijable X i Y imaju očekivanje i ako je $X \geq Y$, tada je $EX \geq EY$. Svojstvo da $X \geq 0$ povlači $EX \geq 0$ zovemo pozitivnost, a svojstvo da $X \leq Y$ povlači $EX \leq EY$ zovemo monotonost matematičkog očekivanja. Uz matematičko očekivanje, u ovom radu trebat će nam i varijanca i standardna devijacija [15].

Definicija 1.3.3. *Neka je X slučajna varijabla i neka EX postoji.* ***Varijanca*** *od X definira se sa*

$$\mathrm{Var}\, X = E[(X - EX)^2],$$

ako to očekivanje postoji.

Vrijedi

$$\mathrm{Var}\, X = E(X^2) - (EX)^2.$$

Neka slučajna varijabla X ima varijancu. **Standardna devijacija** σ_x od X nenegativan je kvadratni korijen iz varijance, tj.

$$\sigma_x = \sqrt{VarX}.$$

Primjer 1.3.2. *Ako promatramo utakmice na nekom nacionalnom prvenstvu, temeljem prethodnih prvenstava mogu se procijeniti očekivani broj pogodaka po utakmici i standardna devijacija tog broja.*

Primjerice, u mnogim nacionalnim prvenstvima broj neriješenih utakmica je otprilike četvrtina do trećina svih utakmica (primjerice, u prošloj sezoni turske prve lige Süper lig neodlučeno je završilo 82 *od* 306*, tj. oko* 26,8 % *svih utakmica [2]). Recimo da u nekom prvenstvu broj pogodaka po utakmici modeliramo slučajnom varijablom X i da smo vjerojatnost neriješenog ishoda procijenili s* 0,3*. Budući da je broj pobjeda uvijek jednak broju izgubljenih utakmica (da bi netko pobijedio u nekoj utakmici, netko drugi gubi) uzmimo*[2] *da je vjerojatnost pobjede jednaka vjerojatnosti gubitka, dakle* $\frac{0.7}{2} = 0{,}35$*. Uzmimo da X poprima vrijednosti* 0 *(bodova, toliko momčad osvaja ako izgubi),* 1 *(bod, toliko momčad osvaja ako igra neriješeno) i* 3 *(boda, toliko ih momčad dobije ako pobijedi). Onda je X diskretna slučajna varijabla s distribucijom*

$$X \simeq \begin{pmatrix} 0 & 1 & 3 \\ 0{,}35 & 0{,}3 & 0{,}35 \end{pmatrix}.$$

Slijedi da joj je očekivanje

$$EX = 0 \cdot 0{,}35 + 1 \cdot 0{,}3 + 3 \cdot 0{,}35 = 1{,}35,$$

Stoga je, po definiciji varijance,

$$\mathrm{Var}\, X = (0 - 1{,}35)^2 \cdot 0{,}35 + (1 - 1{,}35)^2 \cdot 0{,}3 + (3 - 1{,}35)^2 \cdot 0{,}35 = 1{,}6275,$$

dakle je standardna devijacija broja pogodaka po utakmici u ovom modelu (približno) 1,2757.

U ovom ćemo radu također često implicitno (kao opravdanje naših računa) koristiti Bernoullijev slabi zakon velikih brojeva, prema kojem relativne frekvencije uspjeha nekog slučajnog pokusa uvijek po vjerojatnosti konvergiraju teorijskoj vjerojatnosti tog pokusa [15].

Definicija 1.3.4. *Kažemo da niz $(X_n,\ n \in \mathbb{N})$ slučajnih varijabli na vjerojatnosnom prostoru $(\Omega, \mathcal{F}, P)$ konvergira po vjerojatnosti prema slučajnoj varijabli X ako za svaki $\varepsilon > 0$ vrijedi:*

$$\lim_{n\to\infty} P(|X_n - X| \geq \varepsilon) = 0.$$

[2]Zanemarujemo prednost domaćeg terena jer svaki klub će jednako puta biti domaćin koliko i gost,

Pišemo

$$(P)\lim_{n\to\infty} X_n = X.$$

Teorem 1.3.3. *Neka je $(X_n,\ n \in \mathbb{N})$ niz nezavisnih slučajnih varijabli čije su varijance konačne i neka je*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \operatorname{Var} X_i = 0.$$

Tada je

$$(P)\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} (X_i - E\,X_i) = 0.$$

Dakle, temeljem ovog zakona uzimat ćemo relativne frekvencije određenih ishoda kao *a posteriori* procjene nepoznatih teorijskih vjerojatnosti.

1.4 Bernoullijeva razdioba, binomna, Poissonova i normalna razdioba

Opišimo najprije Bernoullijevu razdiobu. **Bernoullijeva slučajna varijabla** je slučajna varijabla s dva moguća ishoda, 'uspjeh' 1 i 'neuspjeh' 0 [16]. Njezina razdioba dana je s

$$X \sim \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix},\ p \in \langle 0, 1\rangle,\ q = 1 - p.$$

Kratko pišemo: $X \sim B(p)$. Za $X \sim B(p)$ je $EX = p$ i $\operatorname{Var} X = p\,q$. Prema Bernoullijevom slabom zakonu velih brojeva, ako uzastopno ponavljamo isti Bernoullijev pokus i ako je svaka izvedba nezavisna od prethodnih, onda će relativni broj uspjeha u tim pokusima (dakle, broj uspjeha podijeljen s brojem pokusa) po vjerojatnosti konvergirati prema p. Binomna razdioba vezana je uz nezavisno ponavljanje istog Bernoullijevog pokusa. U takvima pokusima zanima nas je li se neki događaj dogodio (uspjeh) ili se nije dogodio (neuspjeh). Pretpostavimo da pokus ponavljamo n puta i pri tome nas zanima broj uspjeha. Izvođenje pojedinog pokusa može se modelirati istom Bernoullijevom razdiobom $B(p)$.

Definicija 1.4.1. *Neka je $n \in \mathbb{N}$ i $p \in \langle 0, 1\rangle$. Za slučajnu varijablu koja poprima vrijednosti u skupu $\{0, 1, 2, ..., n\}$ s vjerojatnostima*

$$p_i = P\{X = i\} = \binom{n}{i} p^i\, q^{n-i}$$

*kažemo da ima **binomnu razdiobu** s parametrima n i p i pišemo $X \sim B(n, p)$.*

Dakle, kod binomne razdiobe $X \in B(n, p)$ s n je označen broj nezavisnih brojeva ponavljanja pokusa, a s p vjerojatnost uspjeha u pojedinom izvođenju pokusa (i s $q = 1 - p$ vjerojatnost neuspjeha u pojedinom izvođenju pokusa). Za $X \sim B(n, p)$ je $EX = n\,p$ i $\operatorname{Var} X = n\,p\,q$ [15].

Posljednja diskretna razdioba koju ćemo koristiti je Poissonova razdioba. Ona se može primijeniti kao razdioba slučajne varijable koja broji uspjehe, ali ne pri nezavisnom ponavljanju pokusa nego u jediničnom vremenskom intervalu te ako pokus zadovoljava uvjete:

- vjerojatnost da se dogodi uspjeh ne ovisi o tome u kojem će se jediničnom intervalu dogoditi;
- broj uspjeha u jednom intervalu ne ovisi o broju uspjeha u drugom intervalu;
- očekivani broj uspjeha isti je za sve intervale i dan je pozitivnim realnim brojem λ.

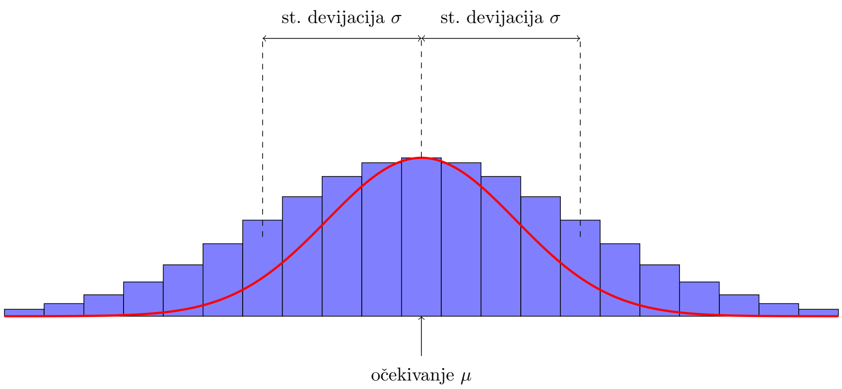
Ako u Bernoullijevoj shemi broj n nezavisnih pokusa teži ka ∞ (jako velik broj pokusa), a vjerojatnost događaja A kojeg promatramo u svakom pokusu teži prema 0 (jako mala vjerojatnost), onda slučajna varijabla X = broj pojavljivanja dogadaja A, ima binomnu razdiobu koju možemo dobro aproksimirati s Poissonovom razdiobom s parametrom $\lambda = np$, tj. $X \simeq P(np)$.

Definicija 1.4.2. *Slučajna varijabla X ima* ***Poissonovu razdiobu*** *s parametrom $\lambda > 0$ i pišemo $X \sim P(\lambda)$ ako poprima vrijednosti u skupu $\mathbb{N}_0$ s vjerojatnostima*

$$p_i = P\{X = i\} = e^{-\lambda}\,\frac{\lambda^i}{i!}.$$

Ako je $X \simeq P(\lambda)$, onda je $EX = \operatorname{Var} X = \lambda$ [15, 16].

Od kontinuiranih razdioba u ovom ćemo radu koristiti samo normalnu. Promatrajući pokus bacanja novčića, povećanjem broja ponavljanja pokusa graf funkcije gustoće binomne slučajne varijable poprima oblik „zvona” kao na slici 1.1. Pojednostavljeno rečeno, ako uzmemo $\mu = n\,p$ i $\sigma = \sqrt{npq}$, gdje su n, p i q parametri binomne slučajne varijable, gdje n teži u beskonačnost, dobijemo normalnu slučajnu varijablu (precizno to formulira de Moivre - Laplaceov granični teorem [15].



Slika 1.1: Binomna razdioba za velike n teži normalnoj razdiobi

Definicija 1.4.3. *Za slučajnu varijablu $X : \Omega \to \mathbb{R}$ kažemo da ima* ***standardnu normalnu razdiobu*** *i označavamo $X \sim \mathcal{N}(0,1)$ ako ima funkciju gustoće vjerojatnosti dana s*

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2},$$

a da ima ***normalnu razdiobu s parametrima*** *μ i σ i pišemo $X \sim \mathcal{N}(\mu, \sigma^2)$ ako joj je funkcija gustoće vjerojatnosti*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}.$$

Za $X \sim \mathcal{N}(\mu, \sigma^2)$ je $EX = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x$ i $\operatorname{Var} X = \int_{-\infty}^{\infty} (x - E(X))^2 f(x)\,\mathrm{d}x$ [15].

Poglavlje 2

Binomna razdioba u nogometu

2.1 Postizanje pogotka

Osnovna primjena binomne razdiobe u nogometu je modeliranje i objašnjavanje vjerojatnosti broja pogodaka na utakmicama. Na temelju prethodnih rezultata utakmica (npr. iz tablice nekog nacionalnog prvenstva) možemo odrediti koliko pogodaka neka momčad prosječno postiže po utakmici u određenom razdoblju. Nogometni stručnjaci, ali i ljubitelji nogometa, vole znati taj podatak. Jednostavnom matematičkom operacijom dijeljenja ukupnog broja postignutih pogodaka s ukupnim brojem odigranih utakmica u nekom razdoblju (primjerice, tijekom prethodnog prvenstva) izračunavamo prosječan broj postignutih pogodaka po utakmici neke nogometne momčadi.

Zamislimo međusobni susret dviju momčadi s prosjecima pogodaka po utakmici r_1 odnosno r_2. Pretpostavimo da prednost domaćeg terena neznatno utječe na učinkovitost momčadi, da su brojevi pogodaka koje pojedine momčadi daju po utakmici međusobno bar približno nezavisni, da su prosjeci relevantni[1] te da je $r_1 > r_2$. Računanje će nam olakšati posebno označavanje omjera prosjeka pogodaka dviju momčadi:

$$R = \frac{r_1}{r_2}.$$

Prosječan broj uspjeha (u našem slučaju broja pogodaka po utakmici) u nekom nizu koristit ćemo za procjenu vjerojatnosti ponovnog uspjeha (davanja sljedećeg pogotka na novoj utakmici). To znači da pretpostavljamo da je u susretu naših momčadi najvjerojatniji rezultat $r_1 : r_2$, što je ekvivalentno pretpostavci da je

$$p : q = r_1 : r_2,$$

ako s p označimo vjerojatnost da sljedeći po redu pogodak daje prva momčad, a s $q = 1 - p$ da ga daje druga [9]. Izražavanjem p iz prethodne jednakosti dobivamo:

[1]To može biti više ili manje točno, no u ovom radu dajemo samo osnovne modele.

$$p = \frac{r_1}{r_1 + r_2} = \frac{R}{R+1},$$

$$q = 1 - p = \frac{1}{R+1}.$$

Time smo događaj „daje sljedeći po redu pogodak” sveli na Bernoullijev pokus. Ako pretpostavimo da taj sljedeći po redu pogodak ne utječe na prosjeke[2] te da svaki pogodak na utakmici pada neovisno o prethodnima,[3] a uz te su pretpostavke dodajemo ključnu pretpostavku da su rezultati utakmica posljedice slučajnih procesa (dakle, ignoriramo faktore izvan čiste vjerojatnosti), onda su vjerojatnosti brojeva pogodaka koje pojedine momčadi daju (uz pretpostavku poznavanja ukupnog broja pogodaka n na utakmici) opisane binomnom razdiobom. Ako je zbroj pogodaka na utakmici n, tada je vjerojatnost da je domaćin postigao sve pogotke jednaka p^n, ili da su gosti postigli sve pogotke $q^n = (1-p)^n$. Ako je utakmica završila s ukupno n pogodaka, a nas zanima koja je vjerojatnost da je domaćin postigao točno k od tih n pogodaka ($k < n$), tj. da je rezultat $k : (n-k)$, ta će vjerojatnost biti jednaka:

$$\binom{n}{k} p^n \, q^{n-k}.$$

Pogledajmo to na konkretnom primjeru iz Prve slovenske nogometne lige u sezoni 2018./2019 [6]. Zamislimo da Triglav i Gorica nakon odigranoga cijelog prvenstva moraju igrati razigravanje za ostanak u prvoj ligi. Obje su momčadi odigrale po 36 utakmica u prvenstvu. Triglav je postigao 51 pogodak i time je bio malo učinkovitiji od Gorice, koja je u sezoni postigla 44 pogotka. Dakle, prosjeci broha pogodaka po utakmici su $r_1 = \frac{51}{36} \approx 1,42$ (za Triglav) te $r_2 = \frac{44}{36} \approx 1,22$ (za Goricu). Dalje je

$$R = \frac{51}{44} \approx 1{,}16$$

pa je vjerojatnost da na međusobnoj utakmici Triglava i Gorice sljedeći pogodak postigne Triglav jednaka $p = \frac{51}{95} \approx 53{,}68\%$, a za Goricu ta vjerojatnost iznosi $q = 1 - p = \frac{44}{95} \approx 46{,}32\%$.

Pogledajmo prvo situaciju utakmice u kojoj je svih n pogodaka postigla jedna odnosno druga momčad, odnosno utakmice s rezultatom $n : 0$ ili $0 : n$. Gledat ćemo samo n do 6 jer su rijetke utakmice s više pogodaka, ali lako bi bilo proširiti račune i na veće n. Znamo da je vjerojatnost rezultata $n : 0$ jednaka p^n, a za rezultat $0 : n$ iznosi q^n. U tablici 2.1 se nalaze vjerojatnosti da su Triglav i Gorica od $n \in \{0, 1, 2, 3, 4, 5, 6\}$ pogodaka na utakmici postigle sve pogotke.

---

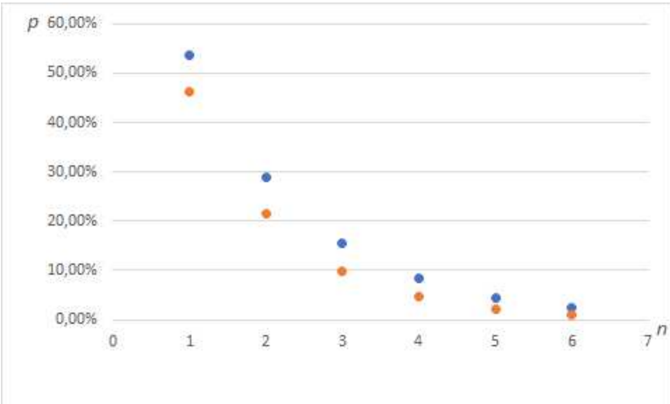
[2] To naravno nije istina.

[3] Ni to nije baš sasvim točno.

n	1	2	3	4	5	6
Triglav	53,68%	28,82%	15,47%	8,31%	4,46%	2,39%
Gorica	46,32%	21,45%	9,94%	4,60%	2,13%	0,99%

Tablica 2.1: Vjerojatnost postizanja svih n pogodaka

Grafički isti podaci izgledaju kao na slici 2.1 (plavi kružići prikazuju vjerojatnosti za Triglav, a narančasti za Goricu). Vidimo, a to znamo i iz navedenih formula, da vjerojatnosti da je jedna momčad postigla sve pogotke eksponencijalno padaju s porastom broja pogodaka u utakmici.



Slika 2.1: Grafički prikaz vjerojatnosti postizanja svih n pogadaka

Pogledajmo sad i druge moguće rezultate osim $n : 0$ i $0 : n$. Princip ćemo opisati na primjeru $n = 6$, tj. ako je na utakmici ukupno palo šest pogodaka. Vjerojatnosti da u takvoj utakmici prva momčad, tj. Triglav, postigne $k = 0$ do 6 pogodaka, a ostale postigne Gorica iznose:

$$\binom{6}{k} p^k q^{6-k} = \binom{6}{k} \frac{51^k \cdot 44^{6-k}}{95^6}.$$

Kad uvrstimo različite k od 0 do 6 dobivamo vjerojatnosti u tablici 2.2.

Rezultat utakmice	0:6	1:5	2:4	3:3	4:2	5:1	6:0
Vjerojatnost	0,99%	6,87%	19,89%	30,74%	26,73%	12,39%	2,39%

Tablica 2.2: Vjerojatnosti rezultata na hipotetskoj utakmici Triglava i Gorice na kojoj je ukupno palo 6 pogodaka

Vidimo da je, ako znamo da je na utakmici bilo ukupno šestt pogodaka, najvjerojatniji rezultat 3:3. Također vidimo da je više nego dvostruko vjerojatnije da Triglav pobijedi s 6:0, nego obratno, što je, naravno, u skladu s intuicijom jer Triglav ima veći prosjek broja pogodaka po utakmici [9].

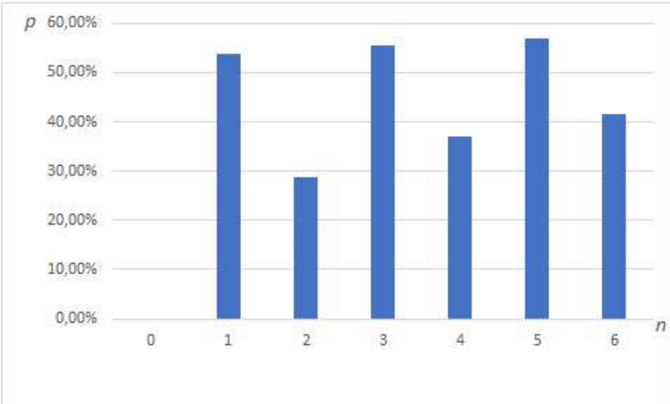
Ukupan broj pogodaka	Pogotci Triglava: 0	1	2	3	4	5	6
0	100,00%						
1	46,32%	53,68%					
2	21,45%	49,73%	28,82%				
3	9,94%	34,55%	40,04%	15,47%			
4	4,60%	21,34%	37,09%	28,66%	8,31%		
5	2,13%	12,35%	28,63%	33,19%	19,23%	4,46%	
6	0,99%	6,87%	19,89%	30,74%	26,73%	12,39%	2,39%

Tablica 2.3: Vjerojatnost postizanja pogotka Triglava u utakmici s n pogodaka

Da bismo izveli općenitiji zaključak, pogledajmo kako se ponašaju vjerojatnosti postizanja k od 0 do n pogodaka za Triglav, za utakmice s n od 0 do 6 pogodaka. Iz tablice 2.3 redom lako izračunamo vjerojatnosti da će Triglav pobijediti Goricu. Naime, različiti rezultati se međusobno isključuju pa se odgovarajuće vjerojatnosti mogu zbrajati. Stoga je vjerojatnost da Triglav pobijedi jednaka

- 0 ako je na utakmici palo 0 pogodaka,
- vjerojatnosti rezultata 1 : 0, tj. 53,68 % ako je na utakmici pao 1 pogodak,
- vjerojatnosti rezultata 2 : 0, tj. 28,82 % ako su na utakmici pala 2 pogotka,
- zbroju vjerojatnosti rezultata 3 : 0 i 2 : 1, tj. 55,52 % ako su na utakmici pala 3 pogotka,
- zbroju vjerojatnosti rezultata 4 : 0 i 3 : 1, tj. 36,97 % ako su na utakmici pala 4 pogotka,
- zbroju vjerojatnosti rezultata 5 : 0, 4 : 1 i 3 : 2, tj. 56,88 % ako je na utakmici palo 5 pogodaka,
- zbroju vjerojatnosti rezultata 6 : 0, 5 : 1 i 4 : 2 tj. 41,51 % ako je na utakmici palo 6 pogodaka.

Prikažimo te vjerojatnosti grafički (slika 2.2).



Slika 2.2: Vjerojatnosti pobjede Triglava nad Goricom za utakmice s $n = 0$ do 6 pogodaka

Posljednje opažanje možemo generalizirati na sve n pa izreći ovako: Što je više pogodaka, to je manja šansa da objektivno slabija momčad ne izgubi. To potvrđuje poznata statistička opažanja da su u nogometu, u kojemu brojevi pogodaka po utakmici nisu veliki, češći neočekivani rezultati (da jači nije pobijedio) nego u sportovima u kojima na utakmicama pada više zgoditaka [10].

2.2 Vjerojatnost pobjede

U nastavku ćemo iskoristiti binomnu razdiobu kao jedan od modela za određivanje vjerojatnosti pobjede za odabranu od dvije nacionalne momčadi koje se susreću na nekoj utakmici [13]. Uzmimo da na vjerojatnost pobjede utječu sljedeća tri faktora: prijašnji međusobni rezultati susreta tih momčadi, ukupna gol-razlika u tim prijašnjim utakmicama te pozicija na FIFA-inoj ljestvici.[4] Ako momčadi nikada nisu igrale jedna protiv druge koriste se samo FIFA-ini bodovi, a inače možemo prema vlastitoj intuiciji ili drugim argumentima odabrati s kojom težinom ćemo uzeti svaki od triju faktora. Primjerice, možemo odabrati da sva ta tri faktora imaju istu težinu; u tom slučaju ako smo izračunali da su vjerojatnosti pobjede temeljem pojedinog od tih faktora jednake redom p_1, p_2 i p_3, naša konačna procjena vjerojatnosti pobjede bit će $\frac{1}{3}\,p_1 + \frac{1}{3}\,p_2 + \frac{1}{3}\,p_3$. Općenito kao težine možemo uzeti bilo koja tri nenegativna broja w_1, w_2 i w_3 čiji zbroj iznosi 1, a konačna procjena vjerojatnosti bit će onda $w_1\,p_1 + w_2\,p_2 + w_3\,p_3$.

Najpoznatija i najpopularnija, iako ne i baš točna, metoda predviđanja ishoda nogometne utakmice je korištenje povijesnih rezultata. Poznata nam je situacija kad komentator

[4] FIFA World Ranking, `https://www.fifa.com/fifa-world-ranking/`.

kaže: „Hrvatska već 20 godina nije izgubila od Slovenije" — tada navijači očekuju da će Hrvatska i ovaj puta postići „pozitivan" rezultat, tj. nastaviti niz.

Ostanimo na primjeru susreta Hrvatska - Slovenija i za njega iskoristimo binomnu raspodjelu i navedene faktore da dobijemo procjenu vjerojatnosti pobjede Hrvatske. Do sada su ove reprezentacije odigrale 9 utakmica, od čega je Hrvatska pobijedila u njih 6, a 3 su utakmice završile neriješeno. Iz ovih podataka lako izračunamo prvu procjenu vjerojatnosti pobjede Hrvatske

$$P(1) = \frac{6}{9} \approx 66{,}7\%.$$[5]

Analogno je prema prethodnim rezultatima vjerojatnost da Hrvatska ne izgubi $P(1\mathrm{X}) = 100\%$, tj. ako bismo gledali samo prethodne rezultate, bilo bi „sigurno" da Hrvatska neće izgubiti. Vjerojatnost neriješenog rezultata je očito $P(\mathrm{X}) = \frac{3}{9} \approx 33,3\%$, a vjerojatnost pobjede Slovenije je u ovom slučaju $P(2) = 0\%$. Očigledna je mana ovog modela: Sigurno nije nemoguće da Hrvatska izgubi. Neovisno o tome, čak i da u povijesnim utakmicama ima slučajeva da je Hrvatska izgubila, jasno je da prethodni rezultati, pogotovo oni stari više godina, imaju više psihološki nego stvarni značaj i da nisu dovoljni kao parametri za procjenu vjerojatnosti pobjede.

Kao što smo spomenuli na početku, sljedeći faktor za koji uzimamo da utječe na vjerojatnost pobjede je gol-razlika u prijašnjim susretima. Gol-razlika u prijašnjim susretima za naš primjer je 16:8 u korist Hrvatske, prosječni broj pogodaka u utakmici tih dviju reprezentacija je $g = \frac{24}{9} = \frac{8}{3} \approx 2{,}67$ (zaokruženo na najbliži cijeli broj: na utakmicama Hrvatske i Slovenije obično padaju oko 3 pogotka).

Ako pretpostavimo da je vjerojatnost postizanja jednog pogotka p, dobit ćemo vjerojatnost postizanja k pogodaka iz n pokušaja (udaraca):

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

Temeljem prethodnih rezultata, tj. prethodne gol-razlike $16 : 8$, pretpostavljamo da je vjerojatnost za postizanje pogotka za Hrvatsku dvostruko veća nego za Sloveniju. Također, uzmimo kao prosječnu procjenu vjerojatnosti uspjeha u pogotku onu navedenu u [13], tj. da oko 16 % svih udaraca završava pogotkom. Stoga možemo uzeti da je vjerojatnost postizanja pogotka Hrvatske iznosi $P_{\mathrm{H}} = \frac{2}{10} = \frac{1}{5}$, a Slovenije $P_{\mathrm{S}} = \frac{1}{10}$. Uvrštavanjem $n = 10$ i $p = P_{\mathrm{H}}$ odnosno $p = P_{\mathrm{S}}$ u formulu binomne razdiobe dobivamo tablicu 2.4. U toj tablici su razmatrani neodlučeni rezultati $k : k$ samo za $k < 10$ jer su za veće k vjerojatnosti zanemarivo male. Također, vjerojatnost rezultata $k : k$ smo računali kao umnožak vjerojatnosti da pojedina momčad dade k pogodaka, tj. opet smo pretpostavili da su brojevi pogodaka koje momčadi postižu nezavisne slučajne varijable.

[5] S $P(1)$ označavamo vjerojatnost pobjede domaćina, a s $P(1\mathrm{X})$ vjerojatnost da domaćin nije izgubio, s $P(2)$ vjerojatnost pobjede gostujuće momčadi i t.d.

$n = 10$	Hrvatska $P(10, k, \frac{1}{5})$	Slovenija $P(10, k, \frac{1}{10})$	Rezultat	Vjerojatnost $P_H \cdot P_S$
$k = 0$	0,107374	0,348678	0:0	3,74%
$k = 1$	0,268435	0,387420	1:1	10,40%
$k = 2$	0,301990	0,193710	2:2	5,85%
$k = 3$	0,201366	0,057396	3:3	1,16%
$k = 4$	0,088080	0,011160	4:4	0,10%
$k = 5$	0,026424	0,001488	5:5	0,00%
$k = 6$	0,005505	0,000138	6:6	0,00%
$k = 7$	0,000786	$8,75 \cdot 10^{-6}$	7:7	0,00%
$k = 8$	0,000074	$3,65 \cdot 10^{-7}$	8:8	0,00%
$k = 9$	0,000004	$9 \cdot 10^{-9}$	9:9	0,00%
Zbroj	100%	100%		21,25%

Tablica 2.4: Vjerojatnost izjednačenog rezultata $k : k$ temeljem gol-razlike u prethodnim susretima

Budući da se rezultati $k : k$ za različite k međusobno isključuju, iz tablice 2.4 vidimo da je vjerojatnost neriješenog rezultata jednaka 21,25%. Međutim, i dalje ne znamo kolika je vjerojatnost pobjede Hrvatske ili Slovenije. Vjerojatnost da će uz navedene pretpostavke Slovenija postići pogodak jednaka je $1 - P(10, 0, \frac{1}{10}) \approx 65{,}1\%$. Vjerojatnost da Hrvatska neće postići pogodak je približno 10,74%. Slijedi da je vjerojatnost pobjede Slovenije bez primljenog pogotka jednaka

$$P(2) = P_H(X = 0) \cdot P_S(X > 0) = 0{,}1074 \cdot 0{,}651 \approx 7\%.$$

Pogledajmo sada tablicu 2.5 sa svim mogućim ishodima (rezultatima) susreta, opet računato temeljem pretpostavke o nezavisnosti broja pogodaka koje daju Hrvatska i Slovenija. U tablici imamo rezultate s najviše 12 postignutih pogodaka (to je i više nego dovoljno jer gotovo nikad na profesionalnim utakmicama ne pada više od 12 pogodaka). Lako se provjeri da je za sve ishode s više od 12 pogodaka vjerojatnost jako mala, gotovo jednaka nuli.

	Hrvatska	0	1	2	3	4	5	6
Slovenija		0,1074	0,2684	0,3020	0,2014	0,0881	0,0264	0,0055
0	0,3487	3,74%	9,36%	10,53%	7,02%	3,07%	0,92%	0,19%
1	0,3874	4,16%	10,40%	11,70%	7,80%	3,41%	1,02%	0,21%
2	0,1937	2,08%	5,20%	5,85%	3,90%	1,71%	0,51%	0,11%
3	0,0574	0,62%	1,54%	1,73%	1,16%	0,51%	0,15%	0,03%
4	0,0112	0,12%	0,30%	0,34%	0,22%	0,10%	0,03%	0,01%
5	0,0015	0,02%	0,04%	0,04%	0,03%	0,01%	0,00%	0,00%
6	0,0001	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%	0,00%

Tablica 2.5: Vjerojatnosti svih mogućih ishoda susreta Hrvatske i Slovenije temeljem gol-razlike prethodnih susreta

Dakle, iz tablice 2.5 vidimo da je, na primjer, vjerojatnost pobjede Hrvatske rezultatom 2:1 jednaka 11,70%. Ako zbrojimo sve vjerojatnosti u kojima Hrvatska pobjeđuje onda ćemo dobiti vjerojatnost pobjede, opet jer se različiti rezultati međusobno isključuju. Vjerojatnost pobjede Hrvatske iznosi otprilike $P(1) \approx 62{,}19\%$. Već smo izračunali da je $P(\mathrm{X}) \approx 21{,}25\%$ pa je $P(2) = 1 - P(1) - P(X) \approx 16{,}56\%$.

Naposlijetku, uzet ćemo u obzir i bodove dviju reprezentacija na FIFA-inoj rang ljestvici: Hrvatska ih u trenutku pisanja ovog rada ima $F_{\mathrm{H}} = 1617$, a Slovenija $F_{\mathrm{S}} = 1379$. Radi lakšeg računanja uzmimo da je očekivani broj postignutih pogodaka na utakmicama konstatan i iznosi 3. Dakle, uzimamo da vrijedi $g_{\mathrm{H}} + g_{\mathrm{S}} = 3$, gdje je g_{H} očekivani broj postignutih pogodaka Hrvatske, a g_{S} Slovenije. Uvodimo dodatnu pretpostavku da FIFA-bodovi dobro odražavaju relativne snage momčadi i da je stoga u susretu neke dvije momčadi očekivani omjer postignutih brojeva pogodaka uvijek jednka omjeru njihovih FIFA-bodova. Za Hrvatsku i Sloveniju dakle pretpostavljamo:

$$g_{\mathrm{H}} + g_{\mathrm{S}} = 3$$

$$\frac{g_{\mathrm{H}}}{g_{\mathrm{S}}} = \frac{1617}{1379}.$$

Iz gore navedenih jednadžbi lako izračunamo

$$g_{\mathrm{H}} = \frac{693}{428} \approx 1{,}62, \quad g_{\mathrm{S}} = \frac{591}{428} \approx 1{,}38.$$

Gornji model tada se može koristiti za određivanje vjerojatnosti pobjede na temelju FIFA-inih bodova. Vjerojatnost postizanja pogotka Hrvatske iznosi $P_{\mathrm{H}} = \frac{\frac{693}{428}}{10} = \frac{693}{4280} \approx 0{,}162$, a Slovenije $P_{\mathrm{S}} = \frac{\frac{591}{428}}{10} = \frac{591}{4280} \approx 0{,}138$. Uvrštavanjem $n = 10$ i $p = P_{\mathrm{H}}$ odnosno $p = P_{\mathrm{S}}$ u formulu binomne razdiobe, dakle isto kao u prethodnom slučaju, ali s drugim iznosima za p, dobivamo tablicu 2.6. Vidimo da je vjerojatnost neriješenog rezultata izračunata temeljem FIFA-bodova jednaka 25,16%.

$n = 10$	Hrvatska $P(10, k, \frac{591}{4280})$	Slovenija $P(10, k, \frac{591}{4280})$	Rezultat	Vjerojatnost $P_H \cdot P_S$
$k = 0$	0,170953	0,226281	0:0	3,87%
$k = 1$	0,330277	0,362516	1:1	11,97%
$k = 2$	0,287139	0,261348	2:2	7,50%
$k = 3$	0,147932	0,111652	3:3	1,65%
$k = 4$	0,050015	0,031302	4:4	0,16%
$k = 5$	0,011595	0,006018	5:5	0,01%
$k = 6$	0,001867	0,000803	6:6	0,00%
$k = 7$	0,000206	$7,35 \cdot 10^{-5}$	7:7	0,00%
$k = 8$	$1,49 \cdot 10^{-5}$	$4,42 \cdot 10^{-6}$	8:8	0,00%
$k = 9$	$6,41 \cdot 10^{-7}$	$1,57 \cdot 10^{-7}$	9:9	0,00%
Zbroj	100%	100%		25,16%

Tablica 2.6: Vjerojatnost izjednačenog rezultata $k : k$ prema FIFA-inim bodovima

U tablici 2.7 vidimo sve moguće ishode susreta Hrvatske i Slovenije.

	Hrvatska	0	1	2	3	4	5	6
Slovenija		0,1710	0,3303	0,2871	0,1479	0,0500	0,0116	0,0019
0	0,2263	3,87%	7,47%	6,50%	3,35%	1,13%	0,26%	0,04%
1	0,3625	6,20%	11,97%	10,41%	5,36%	1,81%	0,42%	0,07%
2	0,2613	4,47%	8,63%	7,50%	3,87%	1,31%	0,30%	0,05%
3	0,1117	1,91%	3,69%	3,21%	1,65%	0,56%	0,13%	0,02%
4	0,0313	0,54%	1,03%	0,90%	0,46%	0,16%	0,04%	0,01%
5	0,0060	0,10%	0,20%	0,17%	0,09%	0,03%	0,01%	0,00%
6	0,0008	0,01%	0,03%	0,02%	0,01%	0,00%	0,00%	0,00%

Tablica 2.7: Vjerojatnosti svih mogućih ishoda susreta Hrvatske i Slovenije prema FIFA-inim bodovima

Zbrojimo sve vjerojatnosti u kojima Hrvatska pobjeđuje i dobit ćemo vjerojatnost pobjede. Vjerojatnost pobjede Hrvatske iznosi otprilike 43,11%. Već smo izračunali da je $P(\mathrm{X}) \approx 25{,}16\%$ pa preostaje da je $P(2) \approx 31{,}73\%$.

Prema danom modelu sada se može odrediti koliko će pojedinačni čimbenici biti bitni. Najveću težinu 0,50 dodijelit ćemo FIFA-inim bodovima (jer oni od naša tri faktora imaju najveću aktualnost) dok će prethodni rezultati te gol-razlika imati težinu 0,25. U tablici 2.8 smo izračunali vjerojatnost pojedinog ishoda susreta Hrvatske i Slovenije.

	Težina	Prethodni rezultati	Težina	Gol-razlika	Težina	FIFA-bodovi	Ukupno
$P(1)$	0,25	67%	0,25	62,19%	0,50	43,11	53,85%
$P(\mathrm{X})$	0,25	33,3%	0,25	21,25%	0,50	25,16%	26,22%
$P(2)$	0,25	0,00%	0,25	16,56%	0,50	31,73%	20,01%

Tablica 2.8: Vjerojatnost pobjede Hrvatske

Prema ovom modelu Hrvatska će (uz očekivana tri pogotka ukupno na utakmici i sve ostale navedene pretpostavke) pobijediti uz vjerojatnost od otprilike 53,85%. U 2021. godini Hrvatska i Slovenija će odigrati dvije utakmice u sklopu kvalifikacija za nadolazeće Svjetsko prvenstvo 2022. godine pa ćemo vidjeti koliko su naše pretpostavke i zaključci bili realistični.

Poglavlje 3

Poissonova razdioba u nogometu

Posebno često se za predviđanje rezultata nogometnih utakmica koristi Poissonova razdioba. Ona u stvarnosti pokazuje određena ograničenja, ali se može koristiti kao polazni i ne prezahtjevan model, koji se onda naprednijim vjerojatnosno-statističkim tehnikama može dorađivati.

U ovom ćemo poglavlju opisati neke primjere korištenja Poissonove razdiobe u analizi i predviđanju nogometnih rezultata. Parametar Poissonove razdiobe bit će nam prosječni broj pogodaka r koje po utakmici daje neka momčad. Stoga nam je jedinica vremena „jedna utakmica", iako naravno ne traju sve utakmice sasvim jednako. Pritom uzimamo, ne sasvim realistično, da su zadovoljeni svi teorijski uvjeti za korištenje Poissonove razdiobe [12]. Prvo, jasno je da je vjerojatnije da će momčad dati pogodak u duljem nego u kraćem razdoblju, ali a priori ne postoji razlog da pretpostavimo da je vjerojatnost pojavljivanja jednog događaja „momčad daje pogodak" (dakle, vjerojatnost da momčad dade jedan pogodak) točno proporcionalna broju utakmica (vremenu). S druge strane, realistično je da je nemoguće da dva pogotka padnu istovremeno. Naposlijetku, samo uvjetno je realistično pretpostaviti da su brojevi pogodaka koje postiže promatrana momčad nezaivsni za različite utakmice. Usprkos tome, Poissonova razdioba daje raznolike i korisne rezultate za našu temu.

3.1 Vjerojatnost od n pogodaka u vremenu t

U ovom odjeljku bavimo se izračunavanjem vjerojatnosti da određena nogometna momčad postigne određeni broj n pogotaka tijekom jedne utakmice, dakle za $t \in [0, 1]$. Neka je r parametar Poissonove razdiobe i neka je on jednak prosječnom broju pogodaka koje promatrana momčad prosječno daje po utakmici (prosjek računat za neki relevantni broj prethodno odigranih utakmica) [9]. Prema Poissonovoj razdiobi, vjerojatnost da u vremenu

t momčad dade $n \in \mathbb{N}_0$ pogodatak jednaka je

$$P_n(t) = \frac{(rt)^n}{n!} \cdot e^{-rt}, 0 \leq t \leq 1.$$

Za svaki odabrani n ta je vjerojatnost najveća ako je $t = \frac{n}{r}$ i tada ima vrijednost

$$P_{\max} = \frac{n^n}{n!} \cdot e^{-n}.$$

Dokažimo to. Derivirajmo P_n:

$$P_n'(t) = \frac{r^n}{n!}(n - r\,t)t^{n-1}\, e^{-rt}.$$

Kritične točke su stoga 0 i 1 te stacionarna točka $T = \frac{n}{r}$. Za $0 < t < T$ je $P_n'(t) > 0$, a za $T < t < 1$ je $P_n'(t) < 0$ pa je T točka (globalnog) maksimuma funkcije P_n.

U tablici 3.1 nalaze se maksimalne vjerojatnosti za postizanje od jednoga do pet pogodaka tijekom jedne utakmice. Vidimo da iznos maksimalne vjerojatnosti ne ovisi eksplicitno o prosjeku r, ali zato o tom prosjeku ovisi vrijeme $\frac{n}{r}$ koje treba proteći do te maksimalne vjerojatnosti postizanja n pogodaka.

n	1	2	3	4	5
P_{max}	36,8%	27,1%	22,4%	19,5%	17,5%

Tablica 3.1: Maksimalne vjerojatnosti postizanja n pogodaka

Primjerice, ako je $r = 2$ (momčad prosječno daje dva pogotka po utakmici), a zanima nas do koje minute sljedeće utakmice je najvjerojatnije da će dati jedan pogodak ($n = 1$), računamo da će se to dogoditi nakon $t = \frac{1}{2}$ utakmice (dakle, do kraja prvog poluvremena) i da je vjerojatnost da se to dogodi $P = e^{-1} \approx 36{,}8\%$. Manje je vjerojatno da će se to dogoditi prije, a vjerojatnost da će postići samo jedan pogodak nakon poluvremena pada. Vjerojatnost da će ta momčad postići jedan pogodak do polovice prvog poluvremena ($t = \frac{1}{4}$) iznosi približno 30,33%, a da će tijekom cijele utakmice postići samo jedan pogodak ($t = 1$) je „samo” 27,07% [9].

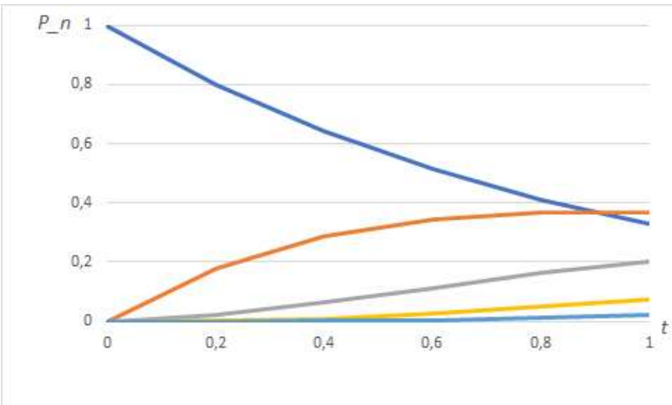
		Pobjeda	Izjednačeno	Poraz	Pogotci	Bodovi
1	Dinamo	29	5	2	74:20	92
2	Rijeka	19	10	7	70:36	67
3	Osijek	18	8	10	61:36	62
4	Hajduk	17	11	8	59:39	62
5	Gorica	17	8	11	57:46	59
6	Lokomotiva	10	9	13	51:43	49
7	Slaven	7	16	13	41:53	37
8	Inter	9	4	23	40:84	31
9	Istra	6	7	23	31:73	25
10	Rudeš	3	5	28	26:80	14

Tablica 3.2: Prva HNL, sezona 2018./19. [5]

Zanimljivo je usporediti i vjerojatnosti da neka momčad tijekom jedne utakmice postigne različite brojeve n pogodaka, ako je poznat njezin prosječni broj r pogodaka po utakmici. Ako to želimo provjeriti na primjeru Intera, s prosjekom od $r = \frac{40}{36} = \frac{10}{9} \approx 1{,}11$ pogodaka po utakmici (tablica 3.2), trebamo usporediti funkcije

$$P_n(t) = \frac{(\frac{10}{9}t)^n}{n!} \cdot e^{-\frac{10}{9}t}$$

za različite n. Na slici 3.1 prikazani su grafovi funkcija P_n za $n = 0, 1, 2, 3$ i 4. Uočavamo da se samo za 0 pogodaka s vremenom smanjuje vjerojatnost, a za sve ostale brojeve prema kraju utakmice raste vjerojatnost da će ih Inter postići. No, što više pogodaka očekujemo od Intera, manja je ukupna vjerojatnost da će se to i dogoditi.



Slika 3.1: Vjerojatnost postizanja 0, 1, 2, 3 i 4 pogotka momčadi Intera

Vjerojatnost da neka nogometna momčad postigne n pogodaka u vremenu t najzornije se prikazuje na primjeru jedne sezone neke nogometne lige. Za primjer nam je i ovdje poslužila Prva hrvatska nogometna liga, sezona 2018./2019. (tablica 3.2). Iz službene statistike udruge profesionalnih klubova prve Hrvatske nogometne lige [5] uzeti su sljedeći podatci: broj odigranih utakmica (O) i broj postignutih pogodaka (G+) u sezoni te je na temelju toga izračunat prosjek postignutih pogodaka po utakmici (r) za sve momčadi koje su sudjelovale. Očekivani broj pogodaka R, za koje su izračunate vjerojatnosti P_R i $P_{\max}$, dobiven je zaokruživanjem prosječnog broja pogodaka po utakmici naviše. U stupcu P_0 su vjerojatnosti da momčad s prosjekom r u sljedećoj utakmici ($t = 1$) uopće ne postigne pogodak ($n = 0$); ona za prosjek r iznosi e^{-r}. U predzadnjem su stupcu vjerojatnosti P_R da pojedina momčad u sljedećoj utakmici postigne svoj zaokruženi prosječan broj pogodaka R. Dakle, riječ je o vjerojatnosti P_R izračunatoj za $t = 1$:

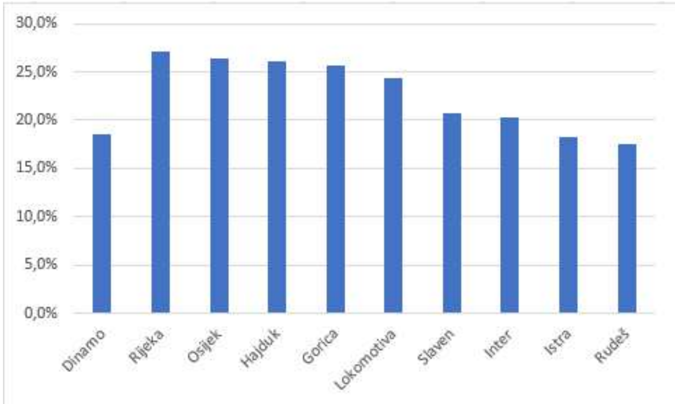
$$P_R(1) = \frac{r^R}{R!}e^{-r}.$$

Dobiveni rezultati prikazani su tablicom 3.3.

	Klub	O	G+	r	R	P_0	P_R	$P_{\max}$
1	Dinamo	36	74	2,06	3	12,8%	18,5%	22,4%
2	Rijeka	36	70	1,96	2	14,3%	27,0%	27,1%
3	Osijek	36	61	1,69	2	18,4%	26,4%	27,1%
4	Hajduk	36	59	1,64	2	19,4%	26,1%	27,1%
5	Gorica	36	57	1,58	2	20,5%	25,7%	27,1%
6	Lokomotiva	36	51	1,42	2	24,3%	24,3%	27,1%
7	Slaven	36	41	1,14	2	32,0%	20,8%	27,1%
8	Inter	36	40	1,11	2	32,9%	20,3%	27,1%
9	Istra	36	31	0,86	1	42,3%	18,2%	36,8%
10	Rudeš	36	26	0,72	1	48,6%	17,5%	36,8%

Tablica 3.3: Vjerojatnosti postizanja očekivanog broja pogodaka u sljedećoj utakmici

Iz dobivenih podataka možemo vidjeti kako se te vjerojatnosti ponašaju. Momčadi s većim prosjekom pogodaka imaju manju maksimalnu vjerojatnost da ostvare taj prosjek i manju vjerojatnost da utakmicu završe bez postignutog pogotka [9]. Dobiveni rezultati za P_R, tj. vjerojatnost da svaka momčad u sljedećoj utakmici postigne svoj (zaokruženi) prosječni broj pogodaka grafički su prikazani sljedećim dijagramom 3.2.



Slika 3.2: Vjerojatnosti postizanja prosječnog broja pogodaka

Ako pogledamo malo bolje, vidjet ćemo da dobiveni rezultati prate stvarno stanje na terenu. Svatko tko je odigrao ili gledao barem jednu nogometnu utakmicu u životu zna da nije lako postići jedan, a posebno pet pogodaka. Iz dobivenih rezultata vidljivo je da je vjerojatnost P_R za postizanje čak i najvjerojatnijeg (očekivanog) broja pogodaka u jednoj utakmici manja od 30%. Intuicija također sugerira našim računom potvrđenu činjenicu da momčadi s većim prosjekom pogodaka po utakmici s mnogo manjom vjerojatnošću uopće ne postižu pogodak na utakmici. No matematika daje i podatak više: ta vjerojatnost eksponencijalno pada, tj. povećanjem prosjeka pogodaka nesrazmjerno će se povećati vjerojatnost da momčad ne završi sljedeću utakmicu bez postignutog pogotka [9].

3.2 Vjerojatnosti pojedinačnih rezultata unutar nacionalne lige

U ovom odjeljku iskoristit ćemo Poissonovu razdiobu kao jednostavan model za izračunavanje vjerojatnosti pojedinačnih rezultata utakmica. Taj model je osnova stvarnih modela koje koriste kladionice za odabir koeficijenata za klađenje. Naime, u teoriji bi pošteni kladionički koeficijent za događaj koji ima vjerojatnost p trebao iznositi točno $\frac{1}{p}$. Pošteni koeficijent ovdje znači koeficijent koji bi osigurao da dugoročno, nakon mnoštva ponavljanja slučajnog pokusa, osigurava da u prosjeku niti kladionica niti onaj tko se kladi niti gubi niti dobiva.

Primjer 3.2.1. *Ako primjerice bacamo pravednu kockicu, vjerojatnost da padne šestica je* $\frac{1}{6}$. *Možemo to reći i ovako – ili će pasti šestica (jedna mogućnost) ili neki od drugih brojeva (pet mogućnosti). Dakle, šansa da padne šestica je* $1 : 5$. *Zamislimo li da se kladimo s prijateljem da će pasti šestica (ja tvrdim da hoće, on da neće). To znači da on ima pet* *puta veću šansu biti u pravu. Ako bismo htjeli biti pravedni, tada bi naš ulog trebao biti pet puta manji od njegovog. Drugim riječima, ako mi uložimo* 1 *kunu, on bi trebao uložiti njih* 5*. Ako padne šestica, mi dobijemo našu* 1 *kn natrag i još* 5 *kn od njega: koeficijent za naše klađenje je* $1+5=6$*. Ako bismo istu igru ponavljali „u beskonačnost", naš očekivani dobitak je* $-1\cdot\frac{5}{6}+5\cdot\frac{1}{6}=0$ *jer s vjerojatnošću* $\frac{1}{6}$ *dobivamo* 5 *kuna, a s vjerojatnošću* $\frac{5}{6}$ *gubimo jednu.*

Ako bi vam netko ponudio veći koeficijent od 6*, značilo bi da je podcijenio vjerojatnost vašeg uspjeha, odnosno bilo bi vjerojatnije da ćete profitirati na njegovu štetu. S druge strane, ako bi vam netko ponudio manji koeficijent od* 6*, značilo bi da je precijenio vjerojatnost vašeg uspjeha, odnosno vaš eventualni dobitak je manje vjerojatan i manji nego što je realno.*

Upravo tako funkcionira i klađenje na sportske rezultate. Pritom će kladionica pokušati precijeniti vjerojatnost uspjeha igrača (ponuditi manji koeficijent od realističnog) kako bi sebi povećala očekivanu dobit, ali da bi to mogla, mora prvo imati pouzdanu metodu za procjenu same vjerojatnosti. Osnovni princip te procjene opisat ćemo na konkretnom primjeru, prateći izvor [3].

Analizirajmo utakmicu Prve HNL između Osijeka i Slavena; uzet ćemo da je Osijek domaćin, a Slaven gostujuća momčad. Osijek i Slaven odigrali su mnoštvo utakmica kroz povijesta HNL-a, a prvi međusobni susret čeka ih u veljači 2021. Najprije ćemo izračunati očekivani broj postignutih pogodaka svake momčadi. U obzir ćemo uzeti broj postignutih pogodaka, ali i broj primljenih pogodaka. Pri računanju moramo pripaziti da odaberemo odgovarajuće podatke, tj. upotrijebimo relevatni datumski raspon. Uzmimo za primjer sezonu 2018./2019. Prve HNL (tablica 3.2).

Prva korak je izračunati prosječan broj postignutih pogodaka po utakmici u Prvoj HNL, dakle za čitavu ligu, i to zasebno za domaćine i gostujuće momčadi. U navedenoj sezoni momčadi su u domaćim utakmicama postigle 271 pogodak u 180 utakmica dok momčadi kao gosti postigle 239 pogodaka u istom broju utakmica: prosječan broj postignutih pogodaka domaćina po utakmici jednak

$$D=\frac{271}{180}\approx 1{,}51,$$

a prosječan broj postignutih pogodaka gostiju je

$$G=\frac{239}{180}\approx 1{,}33.$$

Sljedeći je korak utvrđivanje prosječnog broja primljenih pogodaka po utakmici — i za domaće i za gostujuće momčadi — što je inverzno prosjeku postignutih pogodaka po utakmici. Zamijenimo gore dobivene podatke i dobijemo prosječan broj primljenih pogodaka domaćina po utakmici $d = G$, a prosječan broj primljenih pogodaka gostiju je $g = D$.

U sljedećem koraku računamo te prosjeke samo za momčadi koje nas zanimaju, i to u ulozi koju će imati u utakmici za koju računamo vjerojatnost. Dakle, trebaju nam prosječni broj O pogodaka koje Osijek postiže kao domaćin. Osijek je u 18 domaćih utakmica postigao 34 pogotka:

$$O = \frac{34}{18} \approx 1{,}89,$$

odnosno uočavamo da Osijek, u prosjeku, postiže više pogodaka kao domaćin nego što je prosjek lige. Sada definiramo jačinu napada J^+ Osijeka kao kvocijent prosječnog broja pogodaka domaće momčadi po domaćoj utakmici s prosječnim brojem pogodaka lige po domaćoj utakmici, tj. kao O/D:

$$J^+ = \frac{34}{18} : \frac{271}{180} \approx 1{,}255.$$

Ovaj račun nam pokazuje da je Osijek postigao 25,5% više pogodaka kod kuće u odnosu na prosjek lige u toj sezoni.

Na analogan način izračunat ćemo jačinu J^- gostujuće obrane. Najprije ćemo izračunati prosjek s primljenih pogodaka Slavena u gostujućim utakmica. Slaven je primio 28 pogodaka u 18 gostujućih utakmica. Slijedi

$$s = \frac{28}{18} \approx 1{,}56.$$

Jačinu gostujuće obrane definiramo kao količnik prosječnog broja primljenih pogodaka gostujuće momčadi po utakmici s prosječnim brojem primljenih pogodaka gostiju u ligi. Dobijemo

$$J^- = \frac{28}{18} : \frac{271}{180} \approx 1{,}033.$$

To pokazuje da je Slaven primio 3,4% više pogodaka u gostima u odnosu na prosjek lige u toj sezoni.

Sada možemo izračunati očekivani broj pogodaka domaćina (Osijek) u sljedećoj utakmici. Pomnožimo jačinu napada Osijeka, jačinu obrane Slavena i prosječan broj pogodaka domaćina u ligi.

$$r_O = J^+ \cdot J^- \cdot D \approx 1{,}255 \cdot 1{,}033 \cdot 1{,}51 \approx 1{,}952.$$

Ovime smo dobili očekivanje da će Osijek postići 1,952 pogodaka protiv Slavena u domaćoj utakmici.

Da bismo izračunali očekivanje koliko će Slaven postići pogodaka na utakmici moramo najprije izračunati obrambenu snagu j^- Osijeka i napadačku snagu j^+ Slavena. Postupak je analogan prethodnom. Slaven je u 18 utakmica u gostima postigao 20 pogodaka pa je

$$S = \frac{20}{18} \approx 1{,}11,$$

te je jačina napada Slavena

$$j^+ = \frac{20}{18} : \frac{239}{180} \approx 0{,}837$$

To pokazuje da je Slaven postigao 16,3% manje pogodaka u gostima u odnosu na prosjek lige.

Sada ćemo još izračunati jačinu obrane j^- domaće momčadi. Opet kao gore računamo količnik broja primljenih pogodaka kod kuće s ukupnim brojem domaćih utakmica. Osijek je u 18 domaćih utakmica primio 18 pogodaka pa slijedi

$$o = \frac{18}{18} = 1,$$

te je jačina obrane Osijeka

$$j^- = \frac{18}{18} : \frac{239}{180} \approx 0{,}753$$

To pokazuje da je Osijek primio 24,7% manje pogodaka kod kuće u odnosu na prosjek lige.

Sada možemo izračunati očekivani broj pogodaka Slavena u sljedećoj utakmici protiv Osijeka. Pomnožimo jačinu napada Slavena, jačinu obrane Osijeka i prosječan broj pogodaka gostiju u ligi:

$$r_S = j^+ \cdot j^- \cdot G \approx 0{,}837 \cdot 0{,}753 \cdot 1{,}33 \approx 0{,}837$$

Ovime smo dobili očekivanje da će Slaven postići 0,837 pogodaka protiv Osijeka u gostujućoj utakmici. Nasamo, očito je da nijedna nogometna utakmica ne završava 1,952 : 0,837, to je samo matematičko očekivanje. Pogledajmo tablicu 3.5 koja prikazuje vjerojatnost postizanja n pogodaka za Osijek i Slaven.

Broj pogodaka	0	1	2	3	4	5	6
Osijek	14,20%	27,72%	27,05%	17,60%	8,59%	3,35%	1,09%
Slaven	43,14%	36,10%	15,10%	4,21%	0,88%	0,15%	0,02%

Tablica 3.4: Vjerojatnost postizanja n pogodaka

Sad treba te očekivane brojeve pogodaka r_O i r_S iskoristiti kao parametre Poissonove razdiobe. Ako pretpostavimo da su brojevi pogodaka koje daju momčadi međusobno nezavisni, onda je vjerojatnost rezultata $n : m$ jednaka umnošku vjerojatnosti da domaćin postigne n pogodaka i vjerojatnosti da gost postigne m pogodaka. Ako te vjerojatnosti procijenimo s Poissonovom razdiobom s parametrima r_O odnosno r_S, znači da vjerojatnost rezultata $n : m$ procjenjujemo formulom

$$P_{n:m} = \frac{r_O^n \, r_S^m}{n! \, m!} \, e^{-r_O - r_S}.$$

Uvrštavanjem dobivamo vjerojatnosti u tablici 3.5:

	Osijek						
Slaven	0	1	2	3	4	5	6
0	6,15%	12,01%	11,72%	7,62%	3,72%	1,45%	0,47%
1	5,15%	10,05%	9,80%	6,38%	3,11%	1,21%	0,40%
2	2,15%	4,20%	4,10%	2,67%	1,30%	0,51%	0,17%
3	0,60%	1,17%	1,14%	0,74%	0,36%	0,14%	0,05%
4	0,13%	0,25%	0,24%	0,16%	0,08%	0,03%	0,01%
5	0,02%	0,04%	0,04%	0,03%	0,01%	0,00%	0,00%
6	0,00%	0,01%	0,01%	0,00%	0,00%	0,00%	0,00%

Tablica 3.5: Vjerojatnosti svih mogućih ishoda susreta Osijeka i Slavena

Zbrajanjem povoljnih ishoda (rezultata) za Osijek u tablici 3.5 izračunamo vjerojatnost pobjede Osijeka; ona iznosi $\sum_{\substack{n,m=0 \\ n>m}}^{\infty} P_{n,m} = 63{,}12\%$. Analogno izračunamo vjerojatnost neriješenog ishoda te pobjede Slavena (tablica 3.6).

Ishod	Vjerojatnost
Pobjeda Osijeka	63,12%
Izjednačeno	21,13%
Pobjeda Slavena	15,75%

Tablica 3.6: Vjerojatnost ishoda susreta Osijeka i Slavena

Poglavlje 4

Normalna razdioba u nogometu

Nogometni klubovi žude za trofejima kao simbolom svog uspjeha. Većini klubova je najveći uspjeh osvojiti nacionalno prvenstvo. Ako momčad osvoji prvenstvo, onda je najbolja u državi, zar ne? Ipak, ako je prvenstvo osvojeno samo jednim bodom prednosti pred drugoplasiranim, onda je moguće razmišljati o „nesretnim" događanjima tijekom sezone (vratareva ili sudačka pogreška, kazneni udarac,...) te da je drugoplasirana momčad također mogla osvojiti prvenstvo. Primjerice, u njemačkoj lizi *Bundesliga*, Bayern je sezone 2018./19. osvojio prvenstvo sa samo dva boda prednosti pred Dortmundom [1, 9, 16]. Provjerit ćemo s kojom vjerojatnošću je Bayern zaista bio najbolja momčad u Njemačkoj te sezone te je li zasluženo postao prvak.

Počnimo sa zamišljenom ligom u kojoj su sve momčadi jednako dobre, recimo da ih je 18 kao u mnogim nacionalnim ligama poput *Bundeslige*. Na prvi, amaterski, pogled, mogli bismo očekivati da će sve momčadi na kraju prvenstva imati otprilike jednak broj bodova te da neće biti prevelike bodovne razlike. Kako smo već naveli u uvodu, u primjeru 1.3.2, za takvu ligu ishod utakmice možemo modelirati slučajnom varijablom X s razdiobom

$$X \simeq \begin{pmatrix} 0 & 1 & 3 \\ 0{,}35 & 0{,}3 & 0{,}35 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 3 \\ \frac{7}{20} & \frac{6}{20} & \frac{7}{20} \end{pmatrix}.$$

Budući da u prvenstvu s 18 klubova svaki odigra po 34, prema primjeru 1.3.2 očekivani konačni broj bodova za svaku momčad ovog zamišljenog prvenstva iznosi $34 \cdot 1{,}35 = 45{,}9 \approx 46$. Ako pak pretpostavimo nekoreliranost[1] rezultata utakmica, onda je standardna devijacija za takvo prvenstvo $\sqrt{34} \cdot 1{,}6275 \approx 9{,}5$ bodova, tj. prema normalnoj razdiobi očekujemo da će oko 68 % (dakle, oko 12) momčadi ostvariti bodove u rasponu 36–55 bodova [9]. Sada možemo odigrati zamišljeno prvenstvo sa zamišljenim momčadima. Za svaku utakmicu, kao ishod, dodijelimo jedan broj od 1 do 20, najlakše pomoću nekog

[1] Za slučajne varijable X i Y kažemo da su nekorelirane ako je $\mathrm{cov}(X, Y) = 0$. Pritom $\mathrm{cov}(X, Y)$ označava kovarijancu od X i Y, a to je veličina koja govori o zavisnosti varijabli; $\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$.

računalnog programa. Ako je dobiven broj od 1 do 7, pobijedio je domaćin; ako je dobiven broj od 8 do 13, susret je završio neriješeno, a ako je dobiven broj veći od 13, pobijedio je gost [8]. U našoj konkretnoj simulaciji dobili smo konačnu tablicu našeg zamišljenog ujednačenog prvensta prikazane u tablici 4.1.

Pozicija	Pobjeda	Izjednačeno	Poraz	Bodovi
1.	19	6	9	63
2.	17	11	6	62
3.	14	15	5	57
4.	16	7	11	55
5.	13	12	9	51
6.	13	12	9	51
7.	13	11	10	50
8.	11	10	13	43
9.	10	13	11	43
10.	11	9	14	42
11.	9	13	12	40
12.	9	13	12	40
13.	9	13	12	40
14.	9	12	13	39
15.	9	11	14	38
16.	8	13	13	37
17.	9	9	16	36
18.	7	8	19	29

Tablica 4.1: Konačna tablica zamišljenog ujednačenog prvenstva

Vidimo da je prva momčad osvojila prvenstvo sa 63 boda te da je raspon između prve i zadnje momčadi 34 boda, odnosno da je usprkos tome da smo pretpostavili da je liga ujednačena dobili velik raspon rezultata. Čak 13 od 18 klubova je unutar standradne devijacije od očekivanog rezultata, što je malo više od očekivanih 68%. Sve to jednostavno je objašnjivo normalnom razdiobom, koja je primjenjiva u opisanim uvjetima. Naravno, ona nije primjenjiva za stvarne lige (odnosno, rezultate stvarnih prvenstava mogli bismo statistički testirati na normalnost), no u ovom našem radu u kojem opisujemo samo vrlo pojednostavljene modele uzet ćemo karakteristike normalne razdiobe i temeljem njih za konkretnu ligu argumentirati koliko je prvak stvarno bio najbolji. Napominjemo da ovdje zanemarujemo mnoge precizne detalje jer nam je cilj opisati samo osnovne ideje.

Prije nego pogledamo je li Bayern sezone 2018./19. stvarno zasluženo osvojio prvenstvo, izvest ćemo još jednu simulaciju zamišljenog prvenstva, ali ćemo sad uzeti da je jedna momčad jača od ostalih. Pretpostavimo da ta bolja momčad ima 1,5 puta veću vjerojatnost da dobije utakmicu nego da ju izgubi, a vjerojatnost izjednačenog rezultata neka je i za nju jednaka 30%. Dakle, rezultat utakmice (osvojeni bodovi) te zamišljene bolje momčadi modeliran je slučajnom varijablom Y s razdiobom

$$Y \sim \begin{pmatrix} 0 & 1 & 3 \\ \frac{14}{50} & \frac{15}{50} & \frac{21}{50} \end{pmatrix}.$$

Budući da će ta najbolja momčad, kao i sve ostale, odigrati 34 utakmice, njezin očekivani broj bodova na kraju prvenstva je

$$34 \cdot \left(3 \cdot \frac{21}{50} + 1 \cdot \frac{15}{50} + 0 \cdot \frac{14}{50}\right) = 53{,}04 \approx 53$$

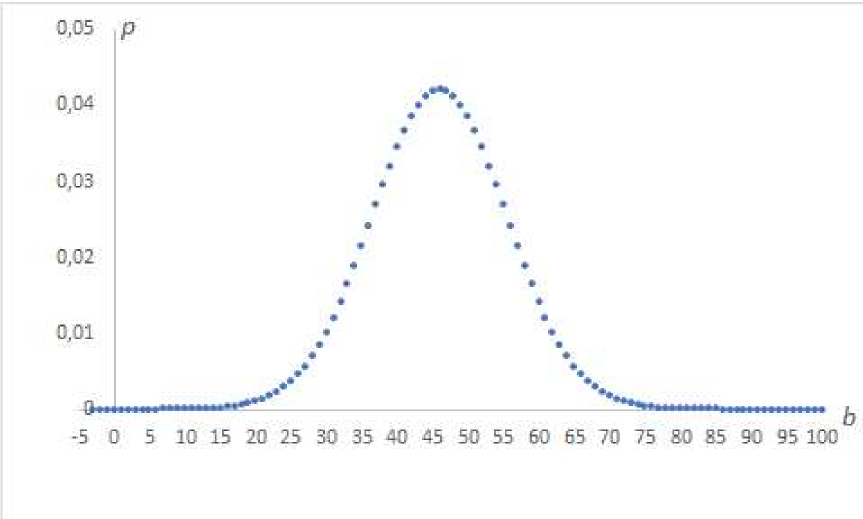
boda. Simulacijom takvog prvenstva sa 17 ujednačenih i jednom boljom momčadi dobili smo tablicu 4.2.

Pozicija	Pobjeda	Izjednačeno	Poraz	Bodovi	
1.	14	11	9	53	
2.	15	7	12	52	
3.	14	9	11	51	
4.	12	14	8	50	„najbolja momčad”
5.	12	13	9	49	
6.	12	12	10	48	
7.	12	11	11	47	
8.	11	13	10	46	
9.	11	11	12	44	
10.	9	17	8	44	
11.	11	9	14	42	
12.	11	9	14	42	
13.	10	12	12	42	
14.	10	11	13	41	
15.	9	11	14	38	
16.	8	14	12	38	
17.	7	13	14	34	
18.	6	15	13	33	

Tablica 4.2: Konačna tablica zamišljenog prvenstva s najboljom momčadi

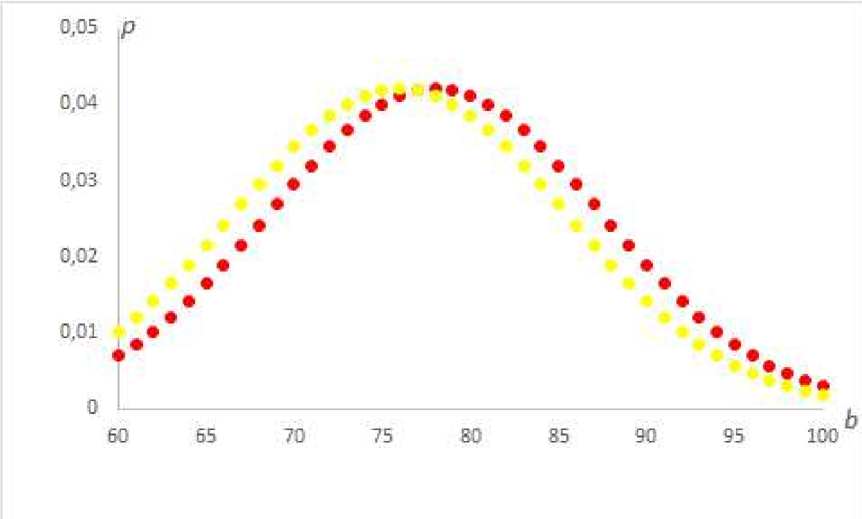
Vidimo da najbolja momčad nije osvojila prvenstvo iako je bila bolja od svih ostalih momčadi lige, štoviše u našoj simulaciji ispala je četvrta. Naravno, trebalo bi izvesti puno takvih simulacija za dobivanje statistički korektnog zaključka, ali već ovako vidimo da dok je glavni faktor uspjeha slučajnost, nije nevjerojatno da realno najbolji ne pobijedi.

Uzmimo da jedino što utječe na rezultat je slučajnost i da za svaku momčad rezultati svih njezinih utakmica budu nezavisni i modelirani istom slučajnom varijablom poput varijabli X i Y u prethodnom dijelu ovog poglavlja. Za fiksiranu momčad u prvenstvu s 18 momčadi uz te pretpostavke vjerojatnost osvajanja nekog broja b bodova između 0 i $34 \cdot 3 = 102$ bodova može se aproksimativno opisati normalnom raspodjelom pri čemu za sve momčadi uzimamo istu standardnu devijaciju, a kao očekivanje uzimamo stvarno ostvarene bodove. Primjerice, za momčad za koju rezultat utakmice možemo modelirati našom slučajnom varijablom X (pa joj očekivani broj bodova aproksimativno modeliramo normalnom razdiobom $N(46, 9.5)$) vjerojatnosti p pojedinih konačnih bodova b biti će raspoređene kao na slici 4.1.



Slika 4.1: Normalna razdioba vjerojatnosti bodova za $\mu = 46$, $\delta = 9{,}5$

Pogledajmo sad normalnu razdiobu $N(78, 9.5)$ vjerojatnosti konačnih bodova Bayerna (crveni kružići) i $N(76, 9.5)$ Dortmunda (žuti kružići) na slici 4.2.



Slika 4.2: Normalna razdioba bodova Bayerna i Dortmunda

Da bismo procijenili vjerojatnost da je Bayern stvarno bio bolji od Dortmunda uzet ćemo slučajne varijable $B \sim N(78, 9{,}5)$ i $D \sim N(76, 9{,}5)$ koje opisuju, uz pretpostavku da na rezultat utječe samo slučaj, osvojeni broj bodova Bayerna odnosno Dortmunda. Za oba kluba smo očekivani broj bodova uzeli stvarno osvojeni iznos (78 odnosno 76). S F_{B} i F_{D} označavamo pripadne funkcije razdiobe.[2]

Vjerojatnost da je Bayern bio bolji od Dortmunda računamo kao zbroj svih vjerojatnosti da je Bayern ostvario b bodova ($b \in 0, 1, \ldots, 102$) i pritom bio bolji od Dortmunda [16].

Ta vjerojatnost P_{BD}, aproksimiramo $P(B = b)$ s $F_{\mathrm{D}}(b)$, temeljem de Moivre - Laplace-ovog graničnog teorema, iznosi

$$P_{BD} = \sum_{b=0}^{102} p(B = b)\, p(D \leq b-1) \approx \sum_{b=0}^{102} p(B = b)\, F_{\mathrm{D}}(b-1) = \sum_{b=0}^{102} \frac{1}{9{,}5\,\sqrt{2\pi}} \exp^{-\frac{(b-78)^2}{1805}} \cdot F_{\mathrm{D}}(b-1).$$

Dakle, s vjerojatnošću 53,95% ispada da je Bayern stvarno bio bolji od Dortmunda.

[2] Ako je f funkcija gustoće vjerojatnosti neke kontinuirane slučajne varijable, pripadna funkcija razdiobe je dana s $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$.

Bibliografija

[1] *Bundesliga*, https://www.bundesliga.com/en/bundesliga/news/title-champions-league-europa-league-relegation-2018-19-final-day-permutations-3577, svibanj 2020.

[2] *Fussballaten.de*, https://www.fussballdaten.de/, svibanj 2020.

[3] *How to calculate Poisson distribution for football betting*, https://help.smarkets.com/hc/en-gb/articles/115001457989-How-to-calculate-Poisson-distribution-for-football-betting, ožujak 2020.

[4] *Poisson distribution*, https://www.sbo.net/strategy/football-prediction-model-poisson-distribution/, ožujak 2020.

[5] *Prva HNL*, http://prvahnl.hr/povijest/rezultati-i-poretci/?sid=28, svibanj 2020.

[6] *Prva liga*, https://www.prvaliga.si/tekmovanja/default.asp?action=lestvica&id_menu=102&id_sezone=2019, svibanj 2020.

[7] *Soccer goal probabilities*, https://blog.annabet.com/soccer-goal-probabilities-poisson-vs-actual-distribution/, siječanj 2020.

[8] Franka Miriam Brückler, *Matematičke pripreme za Svjetsko nogometno prvenstvo u Brazilu–nastavak drugi: O simuliranju*, Priroda **3** (2014), 39–44.

[9] Franka Miriam Brückler i Dinko Cicvarić, *Nogometna matematika i fizika*, Školska knjiga (2008), 58–91.

[10] Franka Miriam Brückler, *Matematičke pripreme za SNP u Brazilu–nastavak treći: Binomna razdioba u nogometu*, Priroda **5** (2014), 37–41.

[11] Singfat Chu, *Using soccer goals to motivate the Poisson process*, INFORMS Transactions on Education **3** (2003), br. 2, 64–70.

[12] Miljenko Huzak, *Vjerojatnost i matematička statistika*, (2006).

[13] Matthias Ludwig, *Die Mathematik hinter Siegwahrscheinlichkeiten*, (2014), https://docplayer.org/22946909-Die-mathematik-hinter-siegwahrscheinlichkeiten.html, svibanj 2020.

[14] Drew Olsen, *Predicting goals scored using the binomial distribution*, https://www.americansocceranalysis.com/home/2014/03/24/predicting-goals-scored-using-the-binomial-distribution, siječanj 2020.

[15] Nikola Sarapa, *Teorija vjerojatnosti*, Školska knjiga, 1987.

[16] John Wesson, *The science of soccer*, IOP Publishing LTD, 2002.

Sažetak

U ovom su radu opisane neke osnovne primjene vjerojatnosnih razdioba u jednom od najpopularnijih sportova, nogometu. Rad je podijeljen u četiri poglavlja u kojima su opisane primjene binomne, Poissonove i normalne razdiobe. U prvom poglavlju su opisani i definirani osnovni pojmovi iz teorije vjerojatnosti. U drugom poglavlju se opisuje primjena binomne razdiobe na konkretne nogometne situacije, točnije na izračunavanje vjerojatnosti određenih rezultata utakmica uz pretpostavku poznavanja ukupnog broja pogodaka i prosječnih brojeva pogodaka koje su momčadi dale u nekom prethodnom razdoblju. U trećem poglavlju koristi se Poissonova razdioba za prognoziranje vjerojatnosti ishoda nogometne utakmice. U posljednjem poglavlju koristimo normalnu razdiobu za procjenu vjerojatnosti da je prvak neke lige stvarno bio najbolji.

Summary

In this thesis we describe some basic applications of probability distributions in one of the most popular sports, football (soccer). The thesis is divided in four chapters, giving applications of binomial, Poisson and normal distribution. In the first chapter fundamental probability concepts are described. In the second chapter we describe the application of the binomial distribution in concrete football situations, more precisely for calculating probabilities of specific results given a total number of goals scored and average numbers of goals of the teams in a chosen previous period. In the third chapter we use the Poisson distribution to estimate probabilities of outcomes of a football match. In the last chapter we use the normal distribution to estimate the probability that the league champion was truly the best team.

Životopis

Rođen sam 13. rujna 1994. godine u Zagrebu. Pohađao sam Osnovnu školu Pavla Štoosa u Kraljevcu na Sutli, srednjoškolsko obrazovanje steko sam u srednjoj školi ”Ban Josip Jelačić”, smjer opća gimnazija u Zaprešiću. Maturirao sam 2013. godine i iste godine upisao sam preddiplomski studij na Prirodnoslovno-matematičkom fakultetu u Zagrebu. U srpnju 2017. godine završio sam preddiplomski studij Matematike, smjer: nastavnički i upisao diplomski studij Matematike, smjer: nastavnički, na istom fakultetu.