Primjeri stohastičkih procesa

Dominik, Žarković

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Dominik Žarković

PRIMJERI STOHASTIČKIH PROCESA

ZAVRŠNI RAD

Varaždin, 2018.

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Dominik Žarković

Matični broj: 43279/14–R

Studij: Poslovni sustavi

PRIMJERI STOHASTIČKIH PROCESA

ZAVRŠNI RAD

Mentor:

Dušan Munđar, univ. spec. actuar. math., pred.

Varaždin, rujan 2018.

Dominik Žarković

Izjava o izvornosti

Izjavljujem da je moj završni/diplomski rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristio drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi

Sažetak

Ovaj rad opisuje stohastičke procese, njihovo nastajanje, primjenu te veze među njima. U radu su opisani binomni proces, Poissonov proces, hipergeometrijski proces i proces obnavljanja. Svaki od navedenih procesa ima drugačiju primjenu, kao što je u radu opisano. Binomni proces opisan je na primjeru bacanja novčića i izvršena je simulacija tijeka utakmice pomoću podataka dobivenih analizom prethodnih utakmica pojedinih momčadi. Poissonov proces prikazan je na primjeru predviđanja konačnog rezultata nogometne utakmice. Za predviđanje ishoda utakmice, osim prosječnog broja primljenih i zabijenih golova, u analizu je uključen faktor domaćeg terena. Izračunate su vjerojatnosti za više mogućih ishoda. Također, temeljem dobivenih rezultata, izračunate su vjerojatnosti kojima spomenuta utakmica završi pobjedom domaćina, gosta ili neriješenim rezultatom. Hipergeometrijski proces opisan je primjerom odabira određenih elemenata iz skupa kao i vjerojatnost odabira određenih kombinacija tih elemenata. Kao primjer, izračunata je vjerojatnost da u 10 izvlačenja, nekih od 32 reprezentacije koje su se kvalificirale na svjetsko prvenstvo, izvučemo određenu kombinaciju ekipa iz različitih saveza i kontinenata. Na kraju je opisan proces obnavljanja koji se koristi prilikom izračuna vijeka trajanja strojeva u tvornici, baterija u nekom uređaju i slično te se određuje najprikladnije vrijeme za njihovu zamjenu. Osim što su prikazani primjeri, navedene su i ostale upotrebe ovih stohastičkih procesa koje se koriste u svakodnevnom životu.

Ključne riječi: Stohastički; proces; Binomni proces; Poissonov proces; Hipergeometrijski proces; Proces obnavljanja.

Sadržaj

1. Uvod

U teoriji vjerojatnosti i srodnim područjima, stohastički ili slučajni proces je matematički objekt obično definiran kao zbirka slučajnih varijabli [4]. Stohastički proces također u obzir uzima redoslijed nastalih događaja, ali se češće uzima šira definicija u kojoj se redoslijedu ne pridodaje važnost. U ovom radu razradit ćemo temeljne stohastičke procese: binomni proces, Poissonov proces, hipergeometrijski proces i proces obnavljanja.

Binomnim procesom analizira se nastanak nekog događaja ako je poznata vjerojatnost nastanka tog događaja i broj mogućih nastanaka tj. provedenih pokusa. Navedeno ćemo prikazati na primjeru nogometne utakmice gdje se, temeljem vjerojatnosti postizanja pogodaka, predviđa postizanje pogotka u pojedinoj minuti utakmice te konačan ishod utakmice.

Kod Poissonovog procesa, za razliku od binomnog, postoji konstantna mogućnost za nastankom događaja i promatra se koliko često određeni događaj nastane u nekom vremenskom intervalu. Znajući kolika je učestalost nastanka određenog događaja u nekom vremenskom periodu, moguće je predvidjeti njegovo nastajanje u budućnosti. Takva situacija prikazana je na primjeru nogometne utakmice gdje smo, temeljem analize, prognozirali budući ishod utakmice koji može biti pobjeda, poraz ili neriješeno te najvjerojatniji konačan rezultat.

Hipergeometrijski proces nastaje prilikom izdvajanja pojedinih elemenata iz skupa koji imaju zajedničko svojstvo. Dobro poznavanje ovih procesa pomaže kod rješavanja problema analize rizika, predviđanja rezultata sportskih utakmica, nastanka potresa na određenom području, predviđanje broja prometnih nesreća na nekom raskrižju, analize određenih skupina, predviđanje rezultata na izborima i slično.

Procesi obnavljanja posebna su vrsta Poissonovih procesa i njihova je primjena rasprostranjena u svakodnevnom životu. Također pomoću procesa obnavljanja računa se dotrajalost radnih strojeva u nekoj tvornici te se planira njihova zamjena.

U radu se promatra teorija i pretpostavke kod svakog procesa kao i distribucije koje se koriste pri njihovom modeliranju. Takav pristup pruža nam odličnu priliku da upoznamo brojne važne distribucije i vidimo vezu među njima i njihovu primjenu.

2. Binomni procesi

Binomni proces je nasumični sustav brojenja gdje postoji *n* nezavisnih identičnih pokušaja, od kojih svaki ima istu vjerojatnost uspjeha *p*, koji daje uspjehe *s* iz tih *n* testova (gdje je $0 \leq s \leq n$ i $n > 0$). Postoje, dakle tri veličine $\{(n, p, s)\}$ koje opisuju binomni proces. Sa svakom od tih veličina povezane su tri distribucije koje opisuju njihovu nesigurnost i varijabilnost. Kako bi bilo moguće odrediti jednu od navedenih veličina, potrebno je poznavati ostale dvije.

Najjednostavniji primjer binomnog procesa je bacanje novčića. Ako definiramo „glavu“ kao uspjeh, svako bacanje ima jednaku vjerojatnost uspjeha *p* ($p = 0.5$). Za dani broj pokušaja *n* (bacanja novčića), broj uspjeha će biti *s* (broj dobivenih „glava“). Svaki pokušaj može biti promatran kao nasumična varijabla koja vraća ili 1 („uspjeh“) sa vjerojatnošću *p* ili 0 („neuspjeh“) sa vjerojatnošću (*1-p*). Takav pokus je često poznat kao Bernoullijev pokus, i vjerojatnost (*1-p*) se često označava sa *q*. [1]

Bernoullijev pokus je slučajan pokus čiji ishod može biti uspjeh ili neuspjeh tj. postoje samo dva moguća ishoda. Niz Bernoullijevih pokušaja je niz ponavljanja Bernoullijevih pokusa pri čemu se vjerojatnost uspjeha i neuspjeha ne mijenja.[11] Na primjer, niz Bernoullijevih pokušaja može biti uzastopno bacanje novčića. Neka je *p* vjerojatnost uspjeha u Bernoullijevom pokusu kod beskonačnog niza Bernoullijevih pokušaja. Vjerojatnost za barem jedan uspjeh u prvih *n* pokušaja je $1-(1-p)^n$. Dok je vjerojatnost za točno *k* uspjeha u prvih *n* pokušaja dana foormulom:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

2.1. Broj uspjeha u *n* pokušaja

Istraživanje binomnog procesa počinjemo gledajući vjerojatnost određenog broja uspjeha *s* za dani broj pokušaja *n* s vjerojatnošću uspjeha *p*. Imamo jedno bacanje novčića. Dva ishoda su „glava“ sa vjerojatnošću *p* i „pismo“ sa vjerojatnošću *(1-p)*. Ako imamo dva bacanja novčića postoje četiri moguća ishoda. Ti ishodi imaju vjerojatnosti p^2, $p(1-p)$, $(1-p)p$ i $(1-p)^2$. Ako bacamo novčić sa dva lica svi ovi ishodi imaju jednaku vjerojatnost ostvarivanja koja iznosi 0.25. Binomni proces smatra svaki uspjeh identičnim i prema tome ne razlikuje ishode glava-pismo i pismo-glava: oba ishoda su samo jedan uspjeh u dva bacanja. Vjerojatnost jednog uspjeha u dva bacanja je $2p(1-p)$. U ovoj jednadžbi 2 je broj različitih mogućnosti koji rezultiraju jednim uspjehom u dva pokušaja. Ako bacamo novčić tri puta postoji osam mogućih ishoda. Preciznije, postoji jedna mogućnost dobivanja 3 „glave“, tri mogućnosti dobivanja 2 „glave“, tri mogućnosti dobivanja jedne „glave“ te jedna mogućnost dobivanja nula „glava“. Broj načina kojim možemo dobiti *s* uspjeha iz *n* pokušaja možemo direktno izračunati koristeći binomni koeficijent ${}_nC_s$ koji dobijemo: ${}_nC_s = \binom{n}{s} = \frac{n\,!}{s!(n-s)!}$

Primjerice za tri pokušaja (bacanja novčića):

$$\binom{3}{0} = \frac{3!}{0!\,3!} = 1$$

$$\binom{3}{1} = \frac{3!}{1!\,2!} = 3$$

$$\binom{3}{2} = \frac{3!}{2!\,1!} = 3$$

$$\binom{3}{3} = \frac{3!}{3!\,0!} = 1$$

U navedenom primjeru uzeli smo $n = 3$, te smo vidjeli da odgovara broju kombinacija koji smo prethodno izračunali. Svaki način dobivanja *s* uspjeha u *n* pokušaja ima jednaku vjerojatnost $p^s(1-p)^{n-s}$ pa vjerojatnost promatranja *x* uspjeha u *n* pokušaja dobivamo:

$$P_{Bin}(x) = \binom{n}{x} p^x (1-P)^{n-x}$$

što je funkcija gustoće vjerojatnosti binomne (*n, p*) distribucije.[12] Broj uspjeha *s* promatranih u *n* pokušaja gdje svaki pokušaj ima jednaku vjerojatnost uspjeha dobijemo jednadžbom:

$s = Binomial(n, p)$

2.2. Broj pokušaja potrebnih za ostvarenje *s* uspjeha

Vidjeli smo kako nam binomna distribucija dopušta da modeliramo broj uspjeha koji će nastati u *n* pokušaja gdje znamo vjerojatnost *p*. Ponekad znamo koliko uspjeha želimo imati, znamo vjerojatnost *p* i željeli bi saznati broj pokušaja koji ćemo morati izvesti kako bi ostvarili *s* uspjeha, pretpostavljamo da prestajemo s pokušajima nakon što smo ostvarili željeni broj uspjeha. U tom slučaju *n* je slučajna varijabla. Sada kada imamo binomnu distribuciju možemo odrediti distribuciju za *n* (broj potrebnih pokušaja). Neka je *x* ukupni broj neuspjeha. Ukupni broj pokušaja koji ćemo izvršiti tada je *(s+x)* i do *(s+x-1)*-og pokušaja moramo gledati *(s-1)* uspjeha i *x* neuspjeha (sve do posljednjeg pokušaja koji je uspješan). Vjerojatnost za *(s-1)* uspjeha u *(s+x-1)* pokušaja dobiven je direktno binomnom distribucijom

$$\binom{s+x-1}{s-1} p^{s-1}(1-p)^x$$

Vjerojatnost ovoga praćeno uspjehom dobijemo ako istu jednadžbu pomnožimo sa *p*

$$\binom{s+x-1}{s-1} p^{s}(1-p)^x$$

Koja je masovna funkcija vjerojatnosti negativne binomne distribucije *NegBin (s, p)*. *NegBin (s, p)* distribucija vraća broj neuspjeha prije ostvarivanja *s* uspjeha. Ukupni broj pokušaja *n* je dan:

$$n = s + NegBin(s, p)$$

Ako je $s = 1$ onda distribucija (poznata kao geometrijska distribucija) je vrlo iskrivljena i $p(0) = p$. Vjerojatnost da će biti nula neuspjeha jednaka je *p*, vjerojatnost da će prvi pokušaj biti uspjeh. Također možemo vidjeti da kako se širi negativna binomna distribucija izgleda sve više kao binomna distribucija. Često je slučaj gdje su negativna binomna distribucija i normalna binomna distribucija slične pod određenim uvjetima gdje je *s* velik, kako bi izbjegli računanje velikih faktorijela za *p(x)*. Negativna binomna distribucija mijenja *k* vrijednosti diljem domene ponekad se zove binomna raspodjela vremena čekanja ili Pascal-ova distribucija. [1]
Kako bi izračunali vjerojatnost da primjerice u 15 bacanja novčića ostvarimo 5 uspješnih ako znamo da je vjerojatnost uspjeha $p = 0.5$ koristit ćemo formulu:

$$\binom{s+x-1}{s-1} p^{s}(1-p)^x$$

U navedenoj formuli *s* označava broj uspješnih bacanja, *x* označava broj neuspješnih bacanja dok je *p* oznaka za vjerojatnost uspjeha u pojedinom bacanju. Ako znamo da imamo ukupno 15 bacanja od toga oduzmemo broj uspješnih bacanja $s = 5$ te dolazimo do zaključka da je $x = 10$

Nakon što poznate podatke uvrstimo u formulu, slijedi:

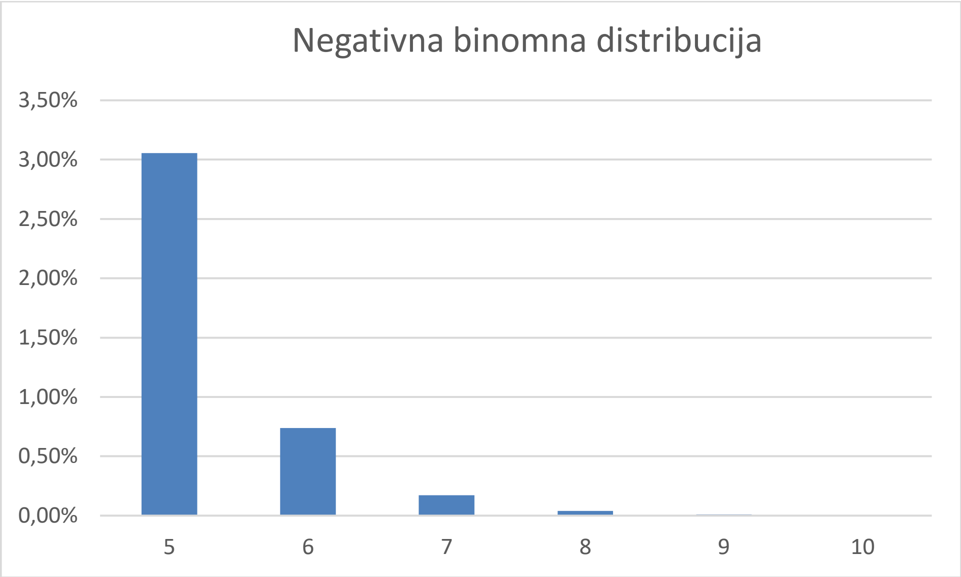
$$\binom{5+10-1}{5-1} 0{,}5^5 (1-0{,}5)^{10}$$

$$= \binom{14}{4} \cdot 0.5^5 \cdot 0.5^{10}$$

$$= 1001 \cdot 0{,}03125 \cdot 0{,}0009765625$$

$$= 0{,}0305 = 3{,}25\%$$

Tako zaključujemo da je vjerojatnost da ćemo u 15 bacanja novčića ostvariti 5 uspjeha i 10 neuspjeha.



Tablica 1 Grafički prikaz negativne binomne distribucije (s=5, p=0.5) (autorski rad)

2.3. Procjena vjerojatnosti uspjeha *p*

Rezultati binomne i negativne binomne distribucije vraćaju distribuciju vjerojatnosti mogućih ishoda. Međutim, gledamo nazad rezultate binomnog procesa i želimo odrediti jedan od parametara. Na primjer, možemo promatrati *n* pokušaja od kojih je *s* uspješnih i iz te informacije želimo procijeniti vrijednost *p*. Ta binomna vjerojatnost je temeljno svojstvo stohastičkog sustava i nikada se ne može točno odrediti, ali prikupljanjem podataka postupno možemo postati sigurniji o njezinoj stvarnoj vrijednosti. Kao što će biti vidljivo, nesigurnost o vjerojatnosti *p* može se lako kvantificirati pomoću beta distribucije. Ukratko, ako nemamo prethodnih informacija o *p*, ili ne želimo prihvatiti bilo koje prijašnje informacije o *p*, onda je poprilično prirodno koristiti konstantnu funkciju za *p*. Putem Bayes-ovog teorema imamo jednadžbu:

$$F(x) = \frac{x^s(1-x)^{n-s}}{\int_0^1 t^s(1-t)^{n-s}\,\mathrm{d}t}$$

koja je Beta (s+1,n-s+1) distribucija, pa

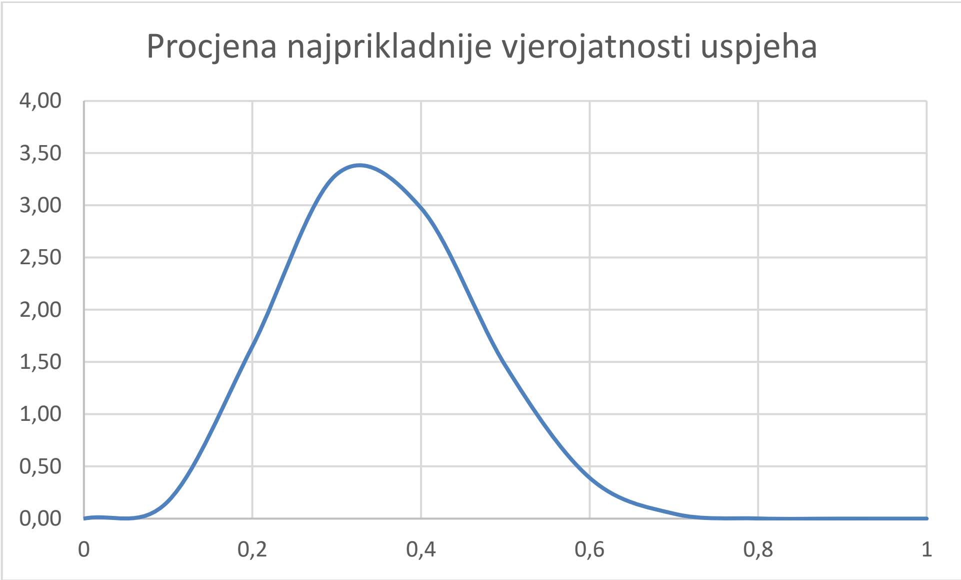
$$p=Beta(s+1,n-s+1)$$

Beta distribucija također se može koristiti u događajima za koje imamo mišljenje o vrijednosti *p* temeljeno na prikupljanju podataka. U takvim slučajevima, možemo razumno pretpostaviti vrijednost *p* koristeći beta distribuciju u formi Beta (*a*, *b*), zadnja strana postaje beta distribucije *(a+s, b+n-s)* budući da je beta distribucija konjugirana s binomnom distribucijom. [1]

Koristeći tablični kalkulator Excel procijenit ćemo prikladnost vjerojatnosti uspjeha *p* vrijednošću beta distribucije za slučaj 5 uspjeha (s=5) u 13 pokušaja (n=15). [10]

p	alpha (s+1)	beta (n-s+1)	Distribucija
0	6	11	0,00
0,1	6	11	0,17
0,2	6	11	1,65
0,3	6	11	3,30
0,4	6	11	2,98
0,5	6	11	1,47
0,6	6	11	0,39
0,7	6	11	0,05
0,8	6	11	0,00
0,9	6	11	0,00
1	6	11	0,00

Tablica 2 Parametri i izračun beta distribucije (autorski rad)



Tablica 3 Grafički prikaz procjene prikladnosti vjerojatnosti uspjeha (s=5, n=15) (autorski rad)

Na grafu je prikazana promjena veličine beta distribucije koja služi za procjenu prikladnosti parametra *p* za slučaj nastanka događaja sa 5 uspjeha u 15 pokušaja. Grafički i tablični rezultat ukazuje da bi najveći rezultat trebao biti između 0.3 i 0.4. Za konkretni slučaj, izvjesno je da je najprikladnija vjerojatnost 5/15=1/3. Beta distribucija u ovom slučaju upozorava da i procesi s drugačijom *p* vrijednošću mogu imati isti ishod.

2.4. Procjena broja pokušaja *n* koji su izvršeni sa *s* uspjeha

Uzmimo u obzir situaciju gdje smo promatrali *s* uspjeha i znali vjerojatnost uspjeha *p*, ali željeli bi znati koliko bacanja je bilo potrebno za ostvarivanje tog broja uspjeha. Želimo procijeniti fiksnu vrijednost, pa nam je potrebna distribucija koja predstavlja našu nesigurnost o tome koja je istinita vrijednost. Postoje dvije moguće situacije: ili znamo da je bacanje zaustavljeno na pokušaju *s* ili ne znamo. Ako znamo da je bacanje zaustavljeno na pokušaju *s* možemo modelirati našu nesigurnost o pravoj vrijednosti *n* na način *n=NegBin(s, p)+s*.

Ako u drugom slučaju ne znamo da je zadnji pokušaj bio uspjeh (iako je mogao biti) onda našu nesigurnost o *n* modeliramo kao *n=NegBin(s+1,p)+s*.

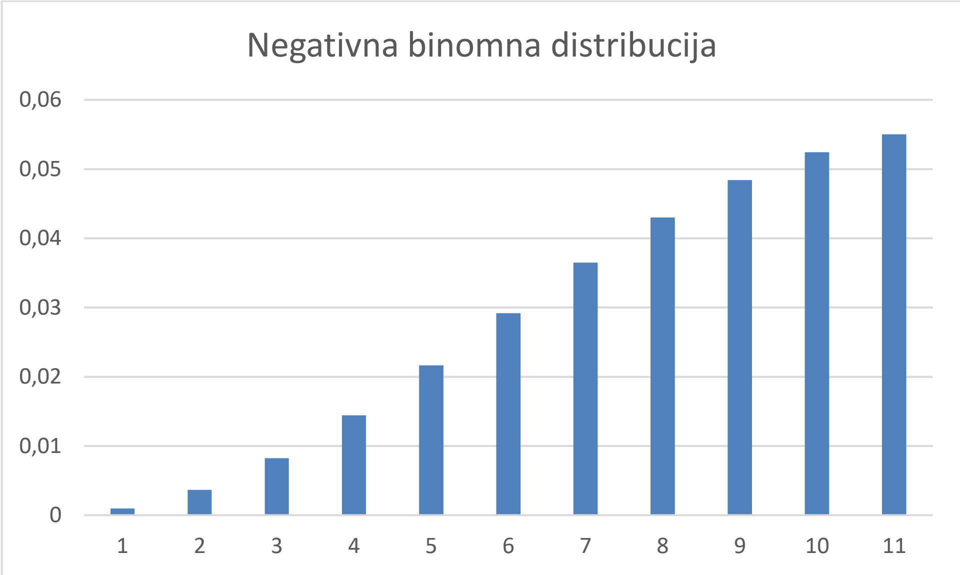
Sada ćemo izvesti ta dva rezultata koristeći standardan Bayesov zaključak. Neka je *x* broj neuspjeha izvršenih prije uspjeha *s*, uzet ćemo pretpostavku da je *x* primjerice *p(x)=c* i iz binomne distribucije vjerojatnost funkcije je vjerojatnost da u *(s+x-1)*-om pokušaju nastaju *(s-1)* uspjeha i onda je *(s+x)*-ti pokušaj bio uspješan što je negativna binomna funkcija vjerojatnosti:

$$p(x) = \binom{s+x-1}{s-1} p^s (1-p)^x.$$

Uzmemo li za primjer slučaj gdje je vjerojatnost uspjeha *p = 0.25*, broj uspješnih pokusa *s = 5* i broj neuspješnih x = 0,1,2…,10. Nakon uvrštavanja veličina u navedenu formulu u stupcu p(x), prikazali smo izračunate vjerojatnosti o ostvarenju *s* uspjeha i *x* neuspjeha.

p	s	x	p(x)
0,25	5	0	0,000977
0,25	5	1	0,003662
0,25	5	2	0,00824
0,25	5	3	0,01442
0,25	5	4	0,021629
0,25	5	5	0,0292
0,25	5	6	0,0365
0,25	5	7	0,043017
0,25	5	8	0,048394
0,25	5	9	0,052427
0,25	5	10	0,055049

Tablica 4 Primjer izračuna negativne binomne distribucije (s=5, p=0.25) (autorski rad)



Tablica 5 Grafički prikaz negativne binomne distribucije(autorski rad)

Na prikazanom grafu može se primijetiti rast vjerojatnosti ostvarivanja događaja s obzirom na promjenu parametara opisanih u primjeru.

2.5. Primjer

U istraživanju je analizirano 256 nogometnih utakmica koje su odigrane za vrijeme kvalifikacija za svjetsko prvenstvo u Rusiji 2018. godine te utakmice odigrane za vrijeme samog prvenstva. Podaci su preuzeti sa stranice www.rezultati.com

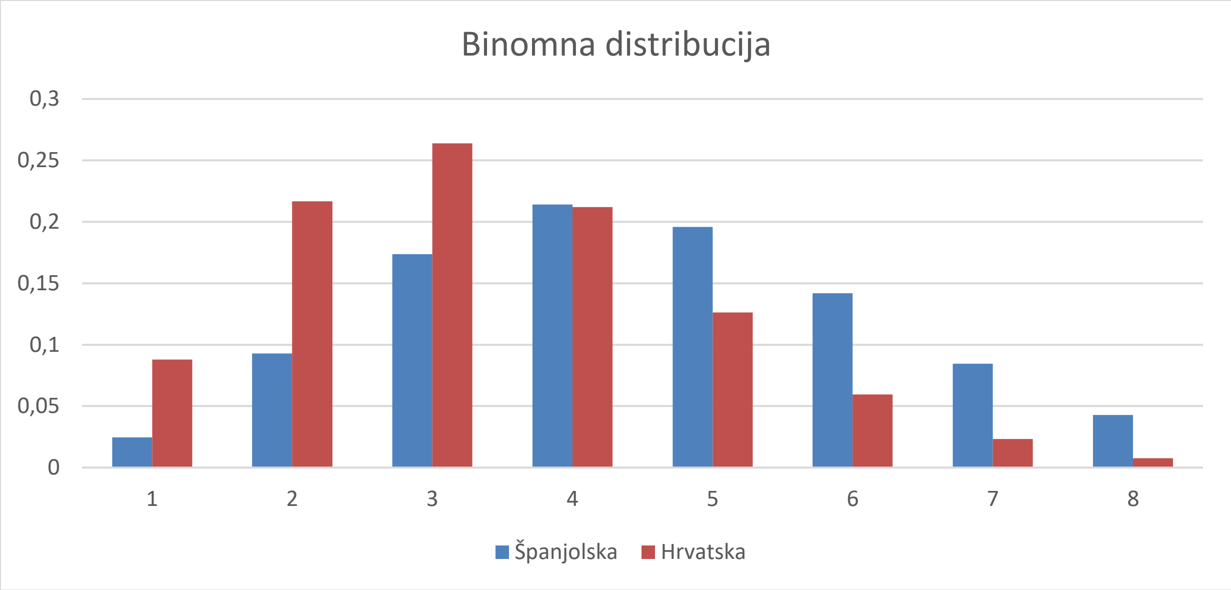
Međunarodne oznake	Momčadi	Broj utakmica	Zabijeni golovi	Primljeni golovi	Prosjek zabijenih golova	Prosjek primljenih golova
ARG	Argentina	12	11	13	0,92	1,08
AUS	Australija	12	14	16	1,17	1,33
BEL	Belgija	9	22	8	2,44	0,89
BRA	Brazil	12	24	7	2,00	0,58
DNK	Danska	12	20	6	1,67	0,50
EGY	Egipat	7	6	9	0,86	1,29
ENG	Engleska	11	21	10	1,91	0,91
FRA	Francuska	10	15	7	1,50	0,70
HRV	Hrvatska	12	18	10	1,50	0,83
IRN	Iran	7	8	4	1,14	0,57
ISL	Island	9	13	6	1,44	0,67
JPN	Japan	11	17	10	1,55	0,91
KOR	Južna Koreja	9	6	7	0,67	0,78
COL	Kolumbija	12	14	10	1,17	0,83
CRI	Kostarika	13	12	18	0,92	1,38
MAR	Maroko	7	13	4	1,86	0,57
MEX	Meksiko	15	21	13	1,40	0,87
NGA	Nigerija	7	10	6	1,43	0,86
DEU	Njemačka	9	29	8	3,22	0,89
PAN	Panama	14	12	23	0,86	1,64
PER	Peru	12	12	9	1,00	0,75
POL	Poljska	10	20	18	2,00	1,80
PRT	Portugal	11	24	7	2,18	0,64
RUS	Rusija	4	10	7	2,50	1,75
SAU	Saudijska Arabija	10	13	15	1,30	1,50
SEN	Senegal	7	12	7	1,71	1,00
SRB	Srbija	8	12	9	1,50	1,13
ESP	Španjolska	10	28	9	2,80	0,90
SWE	Švedska	12	28	11	2,33	0,92
CHE	Švicarska	11	20	10	1,82	0,91
TUN	Tunis	7	13	12	1,86	1,71
URY	Urugvaj	11	13	13	1,18	1,18

Tablica 6 Podaci o odigranim utakmicama (izvor:www.rezultati.com, autorski rad)

Distribucija se odnosi na zabijene golove tijekom utakmice (uključujući sučevu nadoknadu). Procijenjena vjerojatnost da će ekipa Španjolske zabiti u jednoj minuti iznosi 3.82% dok vjerojatnost da će Hrvatska u svakoj minuti zabiti gol iznosi 2.53%. Vjerojatnost da će ekipa zabiti određen broj golova na utakmici ovisi o golovima koje ekipa daje te o broju golova koje protivnička ekipa prima. Binomna distribucija prikazana grafom pokazuje vjerojatnost da će ekipa postići određen broj pogodaka sa izračunatom vjerojatnošću uzimajući po jedan pokušaj u svakoj minuti utakmice uz sučevu nadoknadu vremena, ukupno 95 minuta. U ovom slučaju, vjerojatnost da će Španjolska, kao domaćin, svake minute zabiti pogodak iznosi 3.82%, dok za Hrvatsku ta vjerojatnost iznosi 2.53%. Navedene vjerojatnosti izračunali smo tako što smo uzeli prosječan broj zabijenih golova koje po utakmici odabrana ekipa prima te prosječan broj golova po utakmici koje protivnička ekipa prima, te dvije vrijednosti smo zbrojili i nakon toga podijelili sa 95 (prosječan broj minuta po utakmici) kako bi dobili vjerojatnost zabijenog gola po minuti. Primjerice za Španjolsku $p = \frac{2.8+0.83}{95} \cdot 100\%$. Binomna distribucija izračunata je formulom $\binom{n}{k} p^k (1-p)^{n-k}$ pomoću tabličnog kalkulatora Excel gdje je kao parametre potrebno navesti broj uspjeha, broj pokušaja te vjerojatnost za uspjeh. Želimo li izračunati šanse da će Španjolska zabiti 2 gola uvrstimo u formulu, dobivamo $\binom{95}{2} 0.0382^2 (1-0.0382)^{95-2} = 0.09$.

Broj pogodaka	Španjolska	Hrvatska
p	3,82%	2,53%
0	0,02	0,09
1	0,09	0,22
2	0,17	0,26
3	0,21	0,21
4	0,20	0,13
5	0,14	0,06
6	0,08	0,02
7	0,04	0,01

Tablica 7 Tablični prikaz binomne distribucije (autorski rad)



Tablica 8 Grafički prikaz vjerojatnosti zabijanja broja golova u utakmici prema binomnoj distribuciji za Španjolsku i Hrvatsku (autorski rad)

Temeljem dostupnih podataka, nakon obavljene analize utakmica u tabličnom kalkulatoru Excel, napravljena je simulacija koja generira slučajne brojeve za svaku minutu utakmice te ukoliko oni ispunjavaju uvjete vjerojatnosti, bilježi pogodak.

UTAKMICA	**Španjolska**	**Hrvatska**			
p	3,82%	2,53%			
minuta	**E1**	**E2**	**minuta**	**E1**	**E2**
1	0	0	**48**	1	2
2	0	0	**49**	1	2
3	0	0	**50**	1	2
4	0	0	**51**	1	2
5	0	0	**52**	1	2
6	0	0	**53**	1	2
7	0	0	**54**	1	2
8	0	1	**55**	1	2
9	0	1	**56**	1	2
10	0	1	**57**	1	2
11	0	1	**58**	1	2
12	0	2	**59**	1	2
13	0	2	**60**	1	2
14	0	2	**61**	1	2
15	0	2	**62**	1	2
16	0	2	**63**	1	2
17	0	2	**64**	1	2
18	0	2	**65**	1	2
19	0	2	**66**	1	2
20	0	2	**67**	1	2

21	0	2	**68**	1	2
22	0	2	**69**	1	2
23	0	2	**70**	1	2
24	0	2	**71**	1	2
25	0	2	**72**	1	2
26	0	2	**73**	1	2
27	0	2	**74**	1	2
28	0	2	**75**	1	2
29	0	2	**76**	1	2
30	0	2	**77**	1	2
31	0	2	**78**	1	2
32	0	2	**79**	1	2
33	0	2	**80**	1	2
34	0	2	**81**	1	2
35	0	2	**82**	1	2
36	0	2	**83**	1	2
37	0	2	**84**	1	2
38	0	2	**85**	1	2
39	0	2	**86**	1	2
40	0	2	**87**	1	2
41	0	2	**88**	1	2
42	0	2	**89**	1	2
43	0	2	**90**	1	2
44	0	2	**91**	1	2
45	0	2	**92**	1	2
45+1	1	2	**93**	1	3
46	1	2	**94**	1	3
47	1	2	**KONAČNI REZULTAT**	**1**	**3**

Tablica 9 Simulacija tijeka utakmice (autorski rad)

U tablici je prikazana simulacija nogometne utakmice Španjolska – Hrvatska u trajanju od ukupno 95 minuta. Prikazana je svaka minuta utakmice te je za nju simulirano postizanje pogotka pomoću slučajne varijable i vjerojatnosti postizanja pogotka. U ovoj simulaciji možemo vidjeti da je prvi pogodak zabila Hrvatska u 8. minuti te je ubrzo nakon toga u 12. minuti povećala svoje vodstvo na 0-2. Španjolska vodstvo Hrvatske smanjuje u sučevoj nadoknadi prvog poluvremena na 1-2 te se rezultat ne mijenja do sučeve nadoknade drugog poluvrijemena gdje u 93. minuti Hrvatska svoje vodstvo povećava na 1-3, što na kraju ostaje konačan rezultat.

3. Poissonov proces

U binomnom procesu postoji *n* diskretnih prilika za nastanak događaja. U Poissonovom procesu postoji kontinuirana i stalna mogućnost nastanka događaja. Udar groma za vrijeme oluje možemo smatrati primjerom Poissonovog procesa. To znači da u bilo kojem trenutku za vrijeme oluje postoji šansa da nastane udar groma.

Poissonov proces je jedan od najčešće korištenih procesa brojenja. Obično se koristi u slučajevima gdje se broje pojave određenih događaja za koje se čini da nastaju u određenim intervalima, ali je u potpunosti nasumično i bez određene strukture. Uzimajući u obzir povijesne podatke, saznajemo da se na nekom području potresi javljaju dva puta mjesečno, dok se ostale informacije, kao vrijeme nastanka čine potpuno nasumične. Tako dolazimo do zaključka da Poissonov proces može biti dobar model za previđanje nastanka potresa, a koristi se i za predviđanje broja automobilskih nesreća na određenom mjestu ili na nekom području. Za lokacije korisnika bežične mreže, zahtjeva za određeni dokument na serveru, nastanak sukoba i slično. [1]

Za razliku od binomnog procesa, kod Poissonovog procesa postoji stalna mogućnost nastanka događaja. Teoretski je moguće da, za određeni broj prilika, nastane bilo koji broj događaja uz zadanu vjerojatnost nastanka događaja bez obzira koliko je malo vrijeme promatranja.

Distribucije koje opisuju Poissonov i binomni proces međusobno su čvrsto povezane. U binomnoj distribuciji opisni ključni parametar je *p*, vjerojatnost nastanka događaja koji uvijek jednak i čini događaje međusobno nezavisnima. Ključni parametar kod Poissonove distribucije je λ, prosječan broj događaja koji će nastati po jedinici izloženosti. Jedinica izloženosti također se smatra konstantom u odnosu na ukupnu količinu izloženosti *t*. Bez obzira na to jeli se neki događaj upravo dogodio ili je već prošlo neočekivano puno vremena od njegova nastanka, postoji konstantna vjerojatnost da će događaj nastati. Takve karakteristike pripadaju i binomnim i Poissonovim procesima. Parametar *p* iz binomnog procesa i parametar λ iz Poissonovog procesa dobivaju se proučavanjem stvarnih događaja. U stohastičkim procesima *p* i λ nisu varijable, ali poznavanje njihovih vrijednosti prikazujemo distribucijama. U Poissonovom procesu promatramo broj događaja koji mogu nastati u periodu *t* vremenom koje mora proteći za promatranje α događaja, i prosječnog broja događaja λ poznat kao Poissonov intenzitet. [1]

3.1. Izvod Poissonove distribucije iz binomne distribucije

Kod binomnog procesa broj pokušaja teži ka beskonačnosti, a broj istovremenih uspjeha teži nuli sa ograničenjem da binomna distribucija ostane konačno velika. [1] Funkcija binomne distribucije može se mijenjati radi što točnijeg predviđanja broja uspjeha koji će se dogoditi pod takvim uvjetima glasi:

$$p\ (X = x) = {}_nC_x p^x (1-p)^{n-x}$$

Uzmemo li da je $\lambda t = np$,

$$\text{p (X=x)} = \frac{n!}{x!(n-x)!}\left(\frac{\lambda t}{n}\right)^x\left(1-\frac{\lambda t}{n}\right)^{n-x} = \frac{n(n-1)\ldots(n-x+1)}{n^x}\ \frac{(\lambda t)^x}{x!}\frac{(1-\lambda t/n)^n}{(1-\lambda t/n)^x}$$

Ukoliko je n velik a p malen

$$\left(1-\frac{\lambda t}{n}\right)^n \approx e^{-\lambda t}$$

$$\frac{n(n-1)\ldots(n-x+1)}{n^x} \approx 1$$

$$\left(1-\frac{\lambda t}{n}\right)^x \approx 1$$

Čime se jednadžba pojednostavljuje na:

$$p(X = x) \approx \frac{e^{-\lambda t}(\lambda t)^x}{x!}$$

Time smo dobili smo funkciju Poissonove (λt)distribucije.

Broj nastalih događaja α u vremenskom periodu t =*Poisson* $(\lambda\, t)$ u kojoj je prosječan broj događaja nastalih u vremenskom periodu označen kao λ. Uzmimo za primjer tvornicu koja proizvodi baterije. Prilikom proizvodnje postoji 0,02 šanse da će baterija imati grešku. Ukoliko pretpostavimo da promatramo proizvodnju 100 komada baterija, možemo predvidjeti broj neispravnih baterija koristeći i binomnu i Poissonovu distribuciju. Time te se možemo uvjeriti da su podaci iz obje distribucije gotovo identični.

3.2. Vrijeme potrebno za *s* događaja

Kod Poissonovog procesa uzimamo pretpostavku prema kojoj postoji konstantna vjerojatnost da će određeni događaj nastati u nekom vremenskom razdoblju. Uzmemo li u obzir manji dio vremena Δt, onda je vjerojatnost da će događaj nastati u baš tom dijelu vremena $k\,\Delta t$ gdje je k konstanta. Uzmimo da je *P(t)* vjerojatnost da događaj neće nastati u vremenu *t*. Vjerojatnost da će događaj nastati prvom prilikom tijekom manjeg intervala $k\,\Delta t$ nakon vremena t tada je $k\,\Delta t P(t)$. Što je jednako *P(t) – P(t +* Δt*),* iz toga dobijemo

$$\left[\frac{P(t+\Delta t)-P(t)}{P(t)}\right] = -k\Delta t$$

Uzmemo li da je Δt iznimno mali to postaje diferencijalna jednadžba:

$$\frac{dP(T)}{P(t)} = -k\,dt$$

Integracijom dobivamo:

$$ln[P(t)] = -kt$$

$$P(t) = e^{-kt}$$

Definiramo li *F(t)* kao vjerojatnost da će događaj nastati u određenom vremenu dobijemo:

$$F(t) = 1 - e^{-\mathrm{kt}}$$

Što je kumulativna distribucijska funkcija eksponencijalne distribucije Exp *(1/k)* sa srednjom vrijednošću *1/k*. Slijedi da je *1/k* srednje vrijeme između nastanka događaja, *k* je srednji broj događaja po jedinici vremena čime se dobiva Poissonov parametar λ. Parametar $1/\lambda$ je prosječno vrijeme između nastanka događaja označen je s β. Vrijeme proteklo prije nastanka prvog događaja za Poissonovu distribuciju određuje se

$$t_1 = Exp(\beta)$$

$$\beta = \frac{1}{\lambda}$$

Također je moguće pokazati da je vrijeme do nastanka *s* događaja dano Gamma distribucijom:

$$t_\alpha = Gamma\ (s, \beta)$$

Eksponencijalna (β) distribucija je pojednostavljeni slučaj Gamma distribucije:

$$Gamma(1, \beta) = Exp\ (\beta) \qquad [1]$$

3.3. Primjer

U primjeru su analizirane utakmice svjetskog prvenstva u Rusiji 2018 i kvalifikacijske utakmice za prvenstvo odigrane u 2017. godini. U analizu su uključene 32 reprezentacije koje su se kvalificirale na svjetsko prvenstvo. U tom razdoblju odigrano je 256 utakmica. Pomoću Poissonove distribucije kreiran je model prema kojem se predviđa broj golova određene ekipe po utakmici. Cilj je razviti model koji omogućuje predviđanje ishoda utakmice prilikom budućeg susreta odabranih ekipa. Podaci o ekipama, kao što su odigrane utakmice i njihovi ishodi, dostupni su na stranici *www.rezultati.com*.

Broj zabijenih golova računali smo pomoću Poissonove slučajne varijable, formulom:

$$p(X = x) = \frac{e^{-\lambda_t}(\lambda_t)^x}{x!}$$

Prema Munđar i Šimić [3] Parametar λ_t ovisi o broju primljenih i zabijenih golova u prijašnjim utakmicama. Pokazalo se da tim koji je domaćin zabija više golova od tima koji igra u gostima pa se uračunava prednost domaćeg terena dodavanjem faktora *h*. Ukoliko imamo 2 tima, tim A kao domaćin i tim B kao gost λ_t računamo kao:

$$\lambda_A = \lambda(1 + h)(1 + t_A)$$

$$\lambda_B = \lambda(1 - h)(1 + t_B)$$

U navedenim formulama λ je prosječan broj zabijenih golova po timu u svim utakmicama dok su t_A i t_B faktori specifični za tim. Golovi domaćina i gosta dani na utakmicama su modelirani zasebno. Ukoliko je *X* broj golova koji je zabio tim A i *Y* broj golova zabijen od strane tima B vjerojatnost rezultata *n:m* dana je:

$$P(X = n, Y = m|\lambda_A, \lambda_B) = P(X = n|\lambda_A)P(Y = m|\lambda_B)$$

Gdje su X i Y Poissonove slučajne varijable sa parametrima $\lambda_A\ i\ \lambda_B$.

Momčadi	Utakmica doma	Gol doma	Utakmica u gostima	Gol u gostima	Primili doma	Primili u gostima
ARG	6	3	6	8	5	8
AUS	5	10	7	4	7	9
BEL	4	13	5	9	3	5
BRA	6	12	6	12	3	4
DNK	6	10	6	10	2	4
EGY	3	3	4	3	2	7
ENG	5	11	6	10	3	7
FRA	5	9	5	6	2	5
HRV	5	9	7	9	2	8
IRN	5	6	2	2	4	0
ISL	5	7	4	6	2	4
JPN	5	10	6	7	3	7
KOR	5	4	4	2	2	5
COL	6	6	6	8	8	2
CRI	5	5	8	7	5	13
MAR	3	9	4	4	1	3
MEX	8	13	7	8	5	8
NGA	4	8	3	2	2	4
DEU	5	20	4	9	3	5
PAN	6	11	8	1	7	16
PER	6	8	6	4	6	3
POL	5	11	5	9	8	10
PRT	6	16	5	8	4	3
RUS	3	10	1	0	4	3
SAU	4	5	6	8	1	14
SEN	2	2	5	10	2	5
SRB	5	6	3	6	5	4
ESP	5	13	5	15	5	4
SWE	7	18	5	10	4	7
CHE	4	11	7	9	4	6
TUN	3	3	4	10	3	9
URY	8	10	3	3	12	1

Tablica 10 Golovi u domaćim i gostujućim utakmicama (izvor: www.rezultati.com , autorski rad)

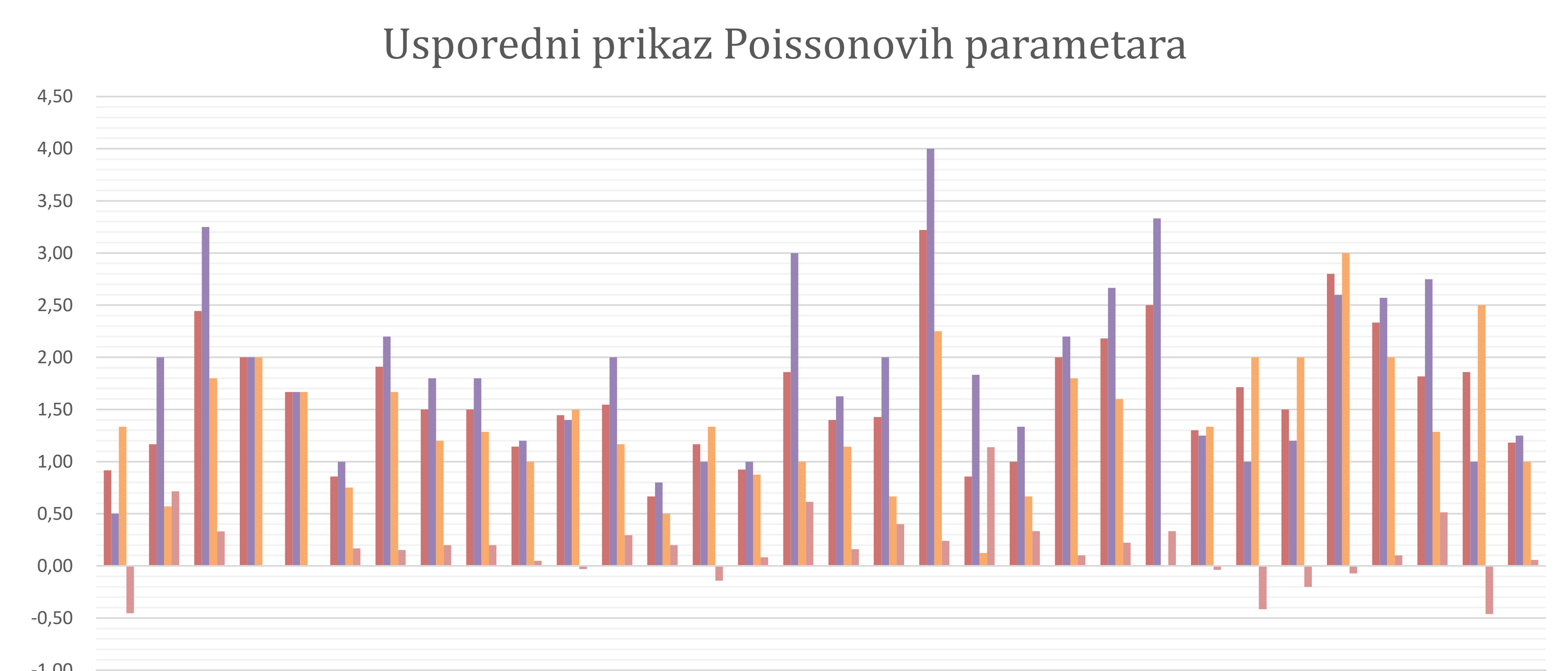
Bonus domaćina h se računa kao $h = \frac{\lambda_h}{\lambda} - 1$. U ovoj formuli λ_h označava prosječan broj golova domaćina zabijen u svim utakmicama. Faktor t specifičan za svaki tim koji je igrao n utakmica kod kuće i m utakmica u gostima određuje se na način $t = \frac{g}{e} - 1$. U danoj formuli g je prosječan broj golova koji tim zabija dok je e očekivani broj golova koji prosječni tim zabije u n domaćih i m gostujućih utakmica dobiven formulom: $e = \frac{\lambda_h n + \lambda_a m}{n+m}$.

Na temelju dostupnih podataka određeni su Poissonovi parametri λ_h i λ_a za domaće i gostujuće timove, bonus domaćina h i timski specifični faktori t.

Tim	λ	λ_h	λ_a	g	e	t=g/(e-1)	h
ARG	0,92	0,50	1,33	0,64	0,92	-0,30	-0,45
AUS	1,17	2,00	0,57	1,89	1,17	0,62	0,71
BEL	2,44	3,25	1,80	1,38	2,44	-0,43	0,33
BRA	2,00	2,00	2,00	1,00	2,00	-0,50	0,00
DNK	1,67	1,67	1,67	1,00	1,67	-0,40	0,00
EGY	0,86	1,00	0,75	1,17	0,86	0,36	0,17
ENG	1,91	2,20	1,67	1,13	1,91	-0,41	0,15
FRA	1,50	1,80	1,20	1,27	1,50	-0,15	0,20
HRV	1,50	1,80	1,29	1,14	1,50	-0,24	0,20
IRN	1,14	1,20	1,00	1,14	1,14	0,00	0,05
ISL	1,44	1,40	1,50	1,00	1,44	-0,31	-0,03
JPN	1,55	2,00	1,17	1,33	1,55	-0,14	0,29
KOR	0,67	0,80	0,50	1,14	0,67	0,71	0,20
COL	1,17	1,00	1,33	0,86	1,17	-0,27	-0,14
CRI	0,92	1,00	0,88	1,08	0,92	0,17	0,08
MAR	1,86	3,00	1,00	1,86	1,86	0,00	0,62
MEX	1,40	1,63	1,14	1,25	1,40	-0,11	0,16
NGA	1,43	2,00	0,67	1,83	1,43	0,28	0,40
DEU	3,22	4,00	2,25	1,71	3,22	-0,47	0,24
PAN	0,86	1,83	0,13	2,71	0,86	2,17	1,14
PER	1,00	1,33	0,67	1,40	1,00	0,40	0,33
POL	2,00	2,20	1,80	1,14	2,00	-0,43	0,10
PRT	2,18	2,67	1,60	1,50	2,18	-0,31	0,22
RUS	2,50	3,33	0,00	3,67	2,50	0,47	0,33
SAU	1,30	1,25	1,33	0,92	1,30	-0,29	-0,04
SEN	1,71	1,00	2,00	0,58	1,71	-0,66	-0,42
SRB	1,50	1,20	2,00	0,82	1,50	-0,45	-0,20
ESP	2,80	2,60	3,00	0,90	2,80	-0,68	-0,07
SWE	2,33	2,57	2,00	1,35	2,33	-0,42	0,10
CHE	1,82	2,75	1,29	1,38	1,82	-0,24	0,51
TUN	1,86	1,00	2,50	0,54	1,86	-0,71	-0,46
URY	1,18	1,25	1,00	1,18	1,18	0,00	0,06
Prosjek:	**1,62**	**1,85**	**1,34**	**1,31**	**1,62**	**-0,09**	**0,15**

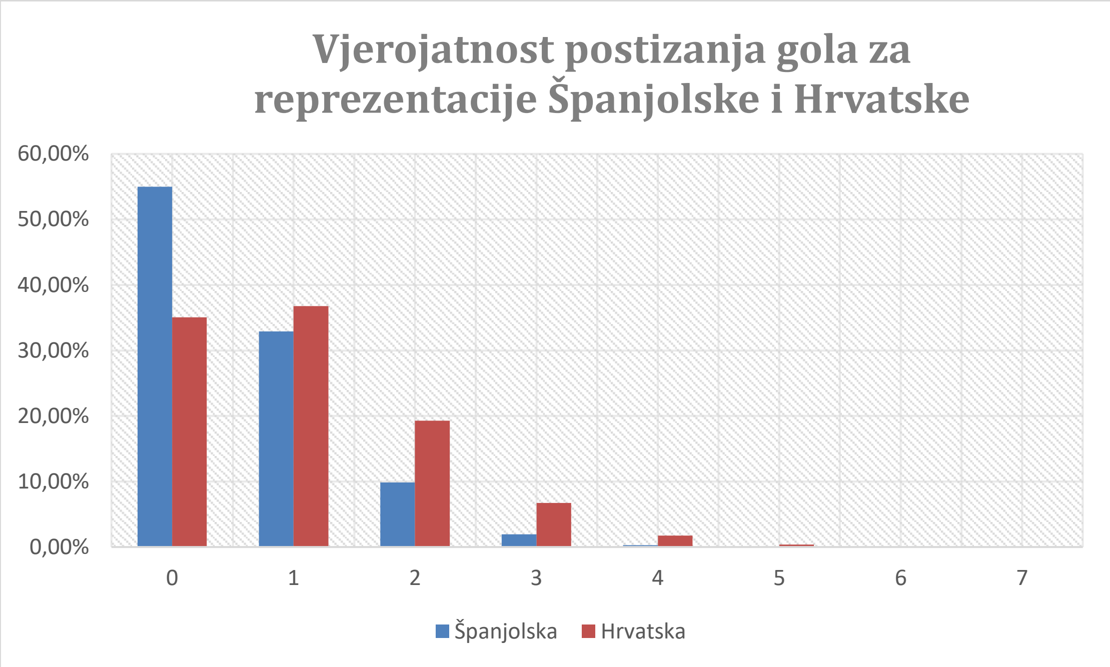
Tablica 11 Izračun poissonovih parametara (autorski rad)

Tijekom kvalifikacijskog ciklusa za svjetsko prvenstvo 2017. godine i svjetskog prvenstva 2018. godine, kvalificirane momčadi odigrale su 160 utakmica na domaćem terenu i 163 utakmice na gostujućem terenu. Prosječan broj zabijenih golova po timu λ bio je 1.62. Timovi koji su igrali na gostujućem terenu u prosjeku λ_a su zabijali 1.34 gola, dok su ekipe na domaćim terenima u prosjeku λ_h zabijale 1.85 golova. Domaćinski bonus iznosio je 0.15 u prosjeku, što znači da su ekipe domaćini u prosjeku zabijale 15% više golova nego ekipe koja igra na neutralnom terenu. U tablicama 11 i 12 prikazan je utjecaj domaćeg i gostujućeg terena na rezultat pojedinih momčadi. Iz prikupljenih podataka može se vidjeti utjecaj domaćeg i gostujućeg terena na rezultat utakmice. U pravilu, domaći teren donosi prednost, ali vidljivo je da za pojedine momčadi to nije slučaj.



	ARG	AUS	BEL	BRA	DNK	EGY	ENG	FRA	HRV	IRN	ISL	JPN	KOR	COL	CRI	MAR	MEX	NGA	DEU	PAN	PER	POL	PRT	RUS	SAU	SEN	SRB	ESP	SWE	CHE	TUN	URY
λ	0,92	1,17	2,44	2,00	1,67	0,86	1,91	1,50	1,50	1,14	1,44	1,55	0,67	1,17	0,92	1,86	1,40	1,43	3,22	0,86	1,00	2,00	2,18	2,50	1,30	1,71	1,50	2,80	2,33	1,82	1,86	1,18
λ_h	0,50	2,00	3,25	2,00	1,67	1,00	2,20	1,80	1,80	1,20	1,40	2,00	0,80	1,00	1,00	3,00	1,63	2,00	4,00	1,83	1,33	2,20	2,67	3,33	1,25	1,00	1,20	2,60	2,57	2,75	1,00	1,25
λ_a	1,33	0,57	1,80	2,00	1,67	0,75	1,67	1,20	1,29	1,00	1,50	1,17	0,50	1,33	0,88	1,00	1,14	0,67	2,25	0,13	0,67	1,80	1,60	0,00	1,33	2,00	2,00	3,00	2,00	1,29	2,50	1,00
h	-0,4	0,71	0,33	0,00	0,00	0,17	0,15	0,20	0,20	0,05	-0,0	0,29	0,20	-0,1	0,08	0,62	0,16	0,40	0,24	1,14	0,33	0,10	0,22	0,33	-0,0	-0,4	-0,2	-0,0	0,10	0,51	-0,4	0,06

Tablica 12 Poissonovi parametri i domaćinski bonus (autorski rad)



Tablica 13 Grafički prikaz vjerojatnosti zabijanja golova prema Poissonovoj distribuciji za Španjolsku i Hrvatsku (autorski rad)

Na prethodnom grafu prikazana je Poissonova distribucija za nadolazeću utakmicu novo osnovanog natjecanja Europske nogometne organizacije pod imenom Liga nacija. Hrvatska prvu utakmicu na ovom natjecanju igra kao gost protiv Španjolske. Analizom kvalifikacijskih utakmica u 2017. godini i natjecateljskih u 2018. godini, dobiveni su podaci prikazani na grafu koji pokazuju vjerojatnost koju ekipe imaju za postizanje pogodaka. U vjerojatnost su uračunati parametri za prosječan broj pogodaka te bonus domaćeg terena.

Izračunate veličine prikazane u tablici Poissonovi parametri uvrstili smo u formule:

$$\lambda_A = \lambda(1+h)(1+t_A)$$
$$\lambda_B = \lambda(1-h)(1+t_B)$$

U navedenim formulama λ_A označava prosječan broj golova domaćina u ovom slučaju reprezentacija Španjolske za koju on iznosi

$$\lambda_A = 1.62(1+0.15)(1+(-0.68)) = 0.598.$$

Dok se prosječan broj gostujuće ekipe označava parametrom λ_B. U ovom primjeru za gostujuću ekipu uzeli smo reprezentaciju Hrvatske te za nju taj parametar iznosi

$$\lambda_B = 1.62\,(1-0.15)(1+(-0.24)) = 1.049$$

Pomoću tako izračunatih parametara, odredili smo Poissonovu distribuciju koja je prethodno grafički prikazana, a na njoj možemo vidjeti vjerojatnosti postizanja pogodaka svake od dviju ekipa. Dobivene vjerojatnosti iskoristili smo kako bi predvidjeli ishod utakmice i konačan rezultat. Umnoškom vjerojatnosti za postizanjem određenog broja golova svake ekipe u donjoj tablici označeno je 5 najvjerojatnijih ishoda utakmice. Najvjerojatniji ishod utakmice Španjolska – Hrvatska prema prikupljenim podacima iznosi 0:1 i iznosi 20.2%. Do tog postotka došli smo umnoškom vjerojatnosti da će Španjolska kao domaćin zabiti 0 golova i Hrvatska kao gost zabiti 1 gol.

		Hrvatska							
		0	1	2	3	4	5	6	7
Španjolska	0	19,3%	20,2%	10,6%	3,7%	1,0%	0,2%	0,0%	0,0%
	1	11,5%	12,1%	6,3%	2,2%	0,6%	0,1%	0,0%	0,0%
	2	3,4%	3,6%	1,9%	0,7%	0,2%	0,0%	0,0%	0,0%
	3	0,7%	0,7%	0,4%	0,1%	0,0%	0,0%	0,0%	0,0%
	4	0,1%	0,1%	0,1%	0,0%	0,0%	0,0%	0,0%	0,0%
	5	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	6	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
	7	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Tablica 14 Prikaz vjerojatnosti konačnog ishoda utakmice Španjolska - Hrvatska (autorski rad)

Nakon što smo izračunali vjerojatnosti za ostvarivanje konačnog rezultata utakmice, koji su prikazani u gornjoj tablici, možemo izračunati kolika je vjerojatnost da će utakmica završiti pobjedom domaćina, gosta ili neriješenim rezultatom. Vjerojatnost da će utakmica Španjolske i Hrvatske završiti pobjedom domaćina računa se tako da se zbroje vrijednosti ispod dijagonale tablice i ona iznosi 20.7%. Za neriješeni rezultat vjerojatnost se računa tako što se zbroje vrijednosti svih ishoda u kojem obje ekipe zabijaju jednak broj golova, a one su u tablici prikazane na dijagonali te ukupno iznose 33.4 %. Ukoliko zbrojimo vrijednosti iznad dijagonale, koje prikazuju vjerojatnost da će reprezentacija Hrvatske zabiti više golova, dobijemo vrijednost koja iznosi 45.9%.

4. Hipergeometrijski procesi

Hipergeometrijski proces nastaje prilikom uzorkovanja nekog skupa te izdvajanja elemenata koji imaju zajedničko svojstvo. Kao najčešći primjer izdvajaju se analize na određenoj populaciji. U većini slučajeva ukupna populacija je puno veća u odnosu na promatrani uzorak sa zajedničkim karakteristikama. Zbog toga možemo pretpostaviti da, ukoliko se određeni uzorak odabere te vrati natrag u ukupnu populaciju, šanse za ponovnim odabirom istog uzorka su vrlo male. U tom slučaju svaki odabir iz ukupnog skupa ima jednaku vjerojatnost za odabirom uzorka sa specifičnim karakteristikama.

Zbog jednake vjerojatnosti uspjeha prilikom svakog odabira hipergeometrijski proces postaje binomni proces. Zbog toga postoji pravilo koje govori da bi ukupna populacija trebala biti manje od 10 puta veća od promatranog uzorka. Kada slučaj zadovoljava navedeno pravilo nije moguće napraviti binomnu aproksimaciju.[1]

Hipergeometrijskim procesom se smatra skup od M elemenata od kojih D ima određene karakteristike pri čemu zadovoljavaju uvjet da je $D \leq M$. Nasumični odabir n elemenata iz skupa M bez dodavanja zamjena prilikom čega svaki od elemenata iz skupa M ima jednaku vjerojatnost da bude odabran. [2]

Kao primjer navodi se vreća koja u sebi sadrži 7 loptica, od kojih su tri crvene dok su ostale četiri plave. Kolika je vjerojatnost da će prilikom odabira 3 loptice iz vreće 2 biti crvene boje. Vjerojatnost da će druga odabrana loptica biti crvena ovisi o boji prvoizabrane loptice. Ukoliko je prva odabrana loptica crvene boje sa vjerojatnošću $\frac{3}{7}$, u vreći su ostale samo 2 crvene od ukupno 6 loptica. Iz toga slijedi da je vjerojatnost da je i druga odabrana loptica crvene boje $\frac{1}{3}$. Svaka loptica koja je ostala u vreći ima jednake šanse da bude odabrana. Što znači da svaki događaj koji rezultira da bude odabrano ukupno x crvenih loptica ima jednaku vjerojatnost. Stoga se u obzir uzimaju samo moguće različite kombinacije.

Ukoliko je u vreći 7 loptica, postoji $\binom{7}{3} = 35$ različitih kombinacija za odabir 3 loptice. Za odabir 2 crvene loptice od ukupno 3 odabranih loptica je $\binom{3}{2} = 3$ te $\binom{4}{1} = 4$ kombinacije za odabir 1 od 4 plavih loptica u vreći. Od mogućih 35 kombinacija odabira loptica samo je $\binom{3}{2}\binom{4}{1} = 3 \cdot 4 = 12$ kombinacija koje bi rezultirale izvlačenjem 2 crvene od ukupno 3 loptice.

Vjerojatnost za izvlačenjem 2 crvene loptice iznosi $\frac{12}{35} = 34.29\%$. Općenito za skup veličine *M* od kojih *D* elemenata ima određenu karakteristiku prilikom nasumičnog odabira *n* elemenata bez dodavanja zamjena u skup vjerojatnost za odabir *x* elementa sa traženim karakteristikama računa se prema:

$$p(x) = \frac{\binom{D}{x}\binom{M-D}{n-x}}{\binom{M}{n}}$$

Uz uvjet: $0 \leq x \leq n, x \leq D, n \leq M$

Hipergeometrijska distribucija također se može koristiti ukoliko postoje više od dvije vrste specifičnih elemenata u skupu. Vjerojatnost za izdvajanje s_1 iz skupa D_1, s_2 iz skupa D_2 prilikom uzimanja uzorka *n* računa se formulom:

$$p(s_1, s_2, \dots, s_k) = \frac{\binom{D_1}{s_1}\binom{D_2}{s_2}\cdots\binom{D_k}{s_k}}{\binom{M}{n}}$$

Uz uvjet: $\sum_{i=1}^{k} s_i = n, \sum_{i=1}^{k} D_i = M, D_i \geq s_i \geq 0, M > D_i > 0.$

Za primjer ćemo uzeti svjetsko nogometno prvenstvo u Rusiji održano 2018. godine koje je organizirala krovna svjetska nogometna organizacija FIFA. Na prvenstvu su sudjelovale 32 reprezentacije od ukupno 211 mogućnih sudionika. [5] Svoje mjesto na svjetskom prvenstvu izborilo je 5 reprezentacija od 47 iz azijskog AFC nogometnog saveza, 5 reprezentacija iz afričkog CAF nogometnog saveza od ukupno 56 reprezentacija članica. Nogometni savez CONCAF koji okuplja reprezentacije iz Sjeverne i Srednje Amerike te Kariba imao je 3 predstavnika od 41 reprezentacije. CONMEBOL, nogometni savez Južne Amerike koji ima 10 zemalja članica imao je 5 predstavnika na svjetskom prvenstvu. Kao sudionik s najviše predstavnika na prvenstvu je sudjelovao europska nogometna organizacija UEFA koja je imala 14 predstavnika od 55 zemalja članica. [6] Vjerojatnost da ćemo od 32 reprezentacije koje su sudjelovale na svjetskom prvenstvu u 10 izvlačenja izvući po 2 reprezentacije saveza AFC, CAF i CONMEBOL, 1 reprezentaciju saveza CONCAF te 3 reprezentacije iz UEFA-e dobivamo na način:

$$p = \frac{\binom{5}{2}\binom{5}{2}\binom{3}{1}\binom{5}{2}\binom{14}{3}}{\binom{32}{10}} = \frac{10 \cdot 10 \cdot 3 \cdot 10 \cdot 364}{64512240} = 0{,}0169$$

Vjerojatnost da ćemo izvući baš ovu kombinaciju reprezentacija iznosi 1.69%.

4.1. Binomna aproksimacija hipergeometrijske distribucije

Ukoliko, nakon odabira, svaki element bude nadomješten u skupu, vjerojatnost da pojedini element ima određene karakteristike računa se kao $\frac{D}{M}$. Tada se broj uzoraka uzet iz skupa *D* računa binomnom (*n, D/M*) distribucijom. Ukoliko je *M* jako velik u odnosu na *n*, šanse za odabir istog elementa više od jednog puta uz dodavanje elemenata u skup nakon svakog odabira je vrlo mala. Za velike skupove *M*, velikim skupovima obično se smatraju skupovi koji su $n < 0.1\ M$ vrijedi da će biti mala razlika prilikom uzorkovanja bez obzira nadomještaju li se uzeti uzorci iz skupa ili ne. U tom slučaju je moguće, zbog puno lakšeg računanja, umjesto Hypergeo (*n, D, M*) distribucije koristiti Binomnu (*n, D/M*). [1]

4.2. Broj uzoraka potrebnih za odabir specifičnih elementa

Odabiru se elemenata bez dodavanja zamjenskih elemenata iz skupa *M*, koji sadrži *D* elemenata sa određenim karakteristikama dok se ne zadovolji uvjet od *s* odabranih elemenata sa željenim karakteristikama. Distribucija broja neuspjeha nastalih prije ostvarivanja *s* uspjeha može se izračunati jednostavnom primjenom negativne binomne distribucije. Vjerojatnost promatranja (s-1) uspjeha u (x+s-1) pokušaja dobije se direktnom primjenom hipergeometrijske distribucije uvrštavanjem u formulu:

$$p(x, s-1) = \frac{\binom{D}{s-1}\binom{M-D}{x}}{\binom{M}{x+s-1}}$$

Vjerojatnost *p* promatranog uspjeha u sljedećem (s+x)-tom pokušaju računa se brojem ostalih D elemenata podijeljenih s ostatkom ukupne populacije *M*:

$$p = \frac{D-s+1}{M-x-s+1}$$

Vjerojatnost nastanka točno *x* neuspjeha, prije ostvarenja *s* uspjeha, gdje se prestaje s uzorkovanjem nakon ostvarenja tog *s*-tog uspjeha je produkt prethodne dvije vjerojatnosti:

$$p(x) = \frac{\binom{D}{s-1}\binom{M-D}{x}(D-s+1)}{\binom{M}{x+s-1}(M-x-s+1)}$$

Neka postoji skup veličine M koji se sastoji od 79 nogometnih reprezentacija koje su sudjelovale na svjetskom nogometnom prvenstvu. [5] Od tih svih reprezentacija, postoji 8 njih koje su to svjetsko prvenstvo osvojile te one spadaju u podskup D. Ukoliko želimo saznati kolika je vjerojatnost da odaberemo 3 reprezentacije koje su osvojile prvenstvo, uz mogućnost odabira jedne koja ga nije osvojila, uvrstimo podatke u gore navedenu formulu i dobijemo da ona iznosi:

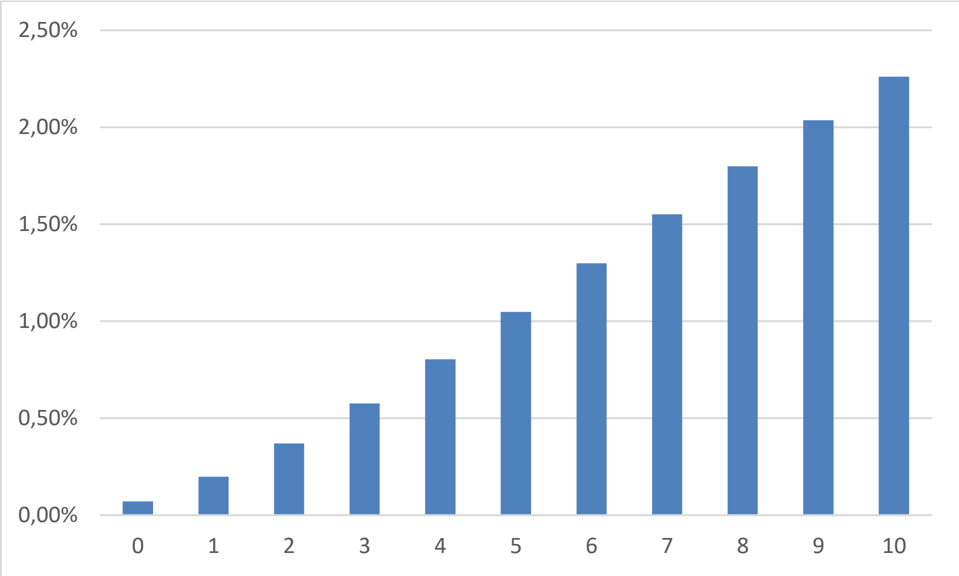
$$p(x) = \frac{\binom{D}{s-1}\binom{M-D}{x}(D-s+1)}{\binom{M}{x+s-1}(M-x-s+1)}$$

$$= \frac{\binom{8}{2}\binom{71}{1}(8-3+1)}{\binom{79}{3}(79-1-3+1)} = 0{,}001985 = 0{,}20\%$$

M	***D***	***s***	***x***	***p(x)***
79	8	3	0	0,07%
79	8	3	1	0,20%
79	8	3	2	0,37%
79	8	3	3	0,58%
79	8	3	4	0,80%
79	8	3	5	1,05%
79	8	3	6	1,30%
79	8	3	7	1,55%
79	8	3	8	1,80%
79	8	3	9	2,04%
79	8	3	10	2,26%

Tablica 15 Izračun vjerojatnosti odabira s elemenata uz x neuspjeha (autorski rad)

U tablici je prikazan izračun vjerojatnosti ostvarivanja događaja u pojedinim slučajevima. Svi izračuni dobiveni pomoću prethodno prikazane formule gdje je prikazan izračun za s = 3 uspjeha i x = 1 neuspjeha.



Tablica 16 Grafički prikaz vjerojatnosti odabira s elemenata uz x neuspjeha (autorski rad)

Na priloženom grafu nalazi se prikaz rasta vjerojatnosti ostvarivanja *s* uspjeha uz rast broja neuspješnih izvlačenja.

5. Procesi obnavljanja

U Poissonovim procesima, vremena između uspješnih događaja, opisana su nezavisnim, identičnim, eksponencijalnim distribucijama. U procesu obnavljana, vremena između uspješnih događaja, također su nezavisna i identična kao i u Poissonovom procesu, ali se mogu opisati bilo kojom distribucijom. Zbog toga je Poissonov proces poseban slučaj obnavljajućeg procesa. [1] Teorija obnavljanja je vrsta teorije vjerojatnosti koja generalizira Possonov proces za potrebe određivanja proizvoljnog vremena zadržavanja. Procesi obnavljanja koriste se kod proračuna najbolje strategije za zamjenu dotrajalih strojeva u nekoj tvornici i kod procjena trajanja baterije ili vijeka trajanja električne žarulje. [7] Proces obnavljanja je proces u kojem su intervali između nastalih događaja slučajne varijable koje su međusobno nezavisne, pozitivne i jednako raspoređene.[8]

Neka je $\{X_n,\ n=1,2,3,4..\}$ niz pozitivnih, nezavisnih slučajnih varijabli s uobičajenom distribucijom F. Kako bi izbjegli trivijalnost, pretpostavimo da vrijedi $F(0) = P(X_n = 0) < 1$. X_n označava vrijeme između (n-1) i n-tog događaja.

$$\mu = E[X_n] = \int_0^\infty x\, dF(x)$$

Neka je μ iz gore prikazane formule oznaka za prosječno vrijeme između uspješnih događaja. Iz toga slijedi da pod pretpostavkom $X_n \geq 0$ i $F(0) < 1$, vrijedi $0 < \mu \leq \infty$.

$$S_0 = 0,\ S_n = \sum_{t=1}^{n} X_t,\ n \geq 1$$

Oznakom S_n prikazano je vrijeme trajanja n-tog događaja. Broj događaja tijekom vremena t biti će jednak najvećoj vrijednosti n. Ako je n-ti događaj nastao prije ili u trenutku t vrijedi da $N(t)$, broj događaja u vremenu t određen:

$$N(t) = sup\{n: S_n \leq t\} \qquad [9]$$

Recimo da je n-ti događaj nastao u vrijeme S_n . Kako je vrijeme između dva događaja nezavisno i jednako distribuirano slijedi da svakim novim događajem proces počinje ispočetka.

$$\frac{S_n}{n} \to \mu\,, n \to \infty$$

Gornja pretpostavka nam pokazuje da beskonačan broj događaja ne može nastati u konačnom vremenu. S obzirom da je $\mu > 0$ znači da S_n teži u beskonačnost kako n teži u beskonačnost. Stoga S_n može biti manji ili jednak t za konačan broj vrijednosti n te zaključujemo da $N(t)$ mora biti konačan.

$$N(t) = max\{n: s_n \leq t\}$$

6. Zaključak

U radu su opisani slučajevi stohastičkih procesa, a to su binomni proces, Poissonov proces, hipergeometrijski proces i proces obnavljanja. Svaki od tih procesa je detaljno opisan zajedno sa svojim parametrima. Navedeni su načini određivanja pojedinih parametara te su navedeni primjeri u kojima se isti koriste.

Tako smo mogli vidjeti opis i primjenu binomnog procesa. Binomni proces prikazan je na primjeru određivanja vjerojatnosti da pojedina ekipa zabije gol u svakoj minuti nogometne utakmice. Na temelju podataka o zabijenim golovima po minuti utakmice, izračunata je vjerojatnost o ukupnom broju pogodaka određene ekipe po utakmici.

Pomoću Poissonovog procesa predviđali smo rezultat utakmice uz uračunatu prednost domaćeg ili gostujućeg terena prilikom odigravanja natjecateljskih nogometnih utakmica. Moguće je primijetiti kako pojedine ekipe nisu nužno efikasnije na domaćem terenu te smo pomoću izračunatih parametara pokušali previdjeti konačan ishod utakmice. Na konkretnom primjeru prikazani su najvjerojatniji ishodi te je izračunata vjerojatnost za pobjedu, poraz ili neriješeni rezultat.

Hipergeometrijski proces opisan je pomoću pokusa izvlačenja broja loptica iz vreće. Osim pokazanih primjera navedeni su dodatni slučajevi primjene ovih procesa čime se ukazuje na njihovu važnost. Hipergeometrijski proces iskoristili smo kako bi izračunali vjerojatnost za odabir određene kombinacije, jedne ili više iz skupine od 32 reprezentacije, koje su se kvalificirale na svjetsko prvenstvo, iz 5 različitih svjetskih nogometnih saveza za koju smo imali 10 pokušaja izvlačenja.

Objašnjene su poveznice između tih procesa te su dani specifični slučajevi kada je moguće, radi jednostavnosti, umjesto jednog koristiti drugi proces. Stohastički procesi su često korišteni i imaju široku primjenu u svakodnevnom životu te omogućuju niz analiza i zanimljivih podataka koji mogu biti od velike važnosti.

Popis literature

1. Vose, D. Risk analysis: a quantitative quide, John Wiley & Sons, 2000.
2. Glumac, Z. Vjerojatnost i statistika kratak uvod [Na internetu]. Dostupno: http://gama.fizika.unios.hr/~zglumac/uvs.pdf [pristupano 20.07.2018.].
3. Munđar, D., Šimić D. „Croatian First Football League: Team's performance in the championship“, Croatian Review of Economic, Business and Social Statistics (1849-8531) rujan 2016, [Na internetu].Dostupno: Hrcak, https://hrcak.srce.hr/index.php?show=clanak&id_clanak_jezik=245359 [pristupano 20.07.2017].
4. Stohastic process, Wikipedia, free Encyclopedia. Dostupno: https://en.wikipedia.org/wiki/Stochastic_process [pristupio: 29.08.2018]
5. FIFA World Cup,Wikipedia, free Encyclopedia. Dostupno: https://en.wikipedia.org/wiki/FIFA_World_Cup [pristupio 30.08.2018]
6. 2018 FIFA WORL CUP RUSSIA Dostupno: https://www.fifa.com/worldcup/teams/ [pristupio: 29.08.2018]
7. Renewal theory, Wikipedia, free Encyclopedia. Dostupno: https://en.wikipedia.org/wiki/Renewal_theory [pristupio 03.09.2018]
8. Discrete Stohastic Processes, Chapter 4: Renewal processes Dostupno: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-262-discrete-stochastic-processes-spring-2011/course-notes/MIT6_262S11_chap04.pdf [pristupio 03.09.2018]
9. STOCHASTIC PROCESSES Second Edition Sheldon M. Ross University of California, Berkeley Dostupno: https://mcdtu.files.wordpress.com/2017/03/stochastic-processes-ross_2.pdf [pristupio 03.09.2018]
10. Beta Distribution: Definition, Calculation, Statistic How To Dostupno: http://www.statisticshowto.com/beta-distribution/ [pristupio 07.09.2018]
11. The Bernoulli and Poisson process, Dostupno: https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-436j-fundamentals-of-probability-fall-2008/lecture-notes/MIT6_436JF08_lec20.pdf [pristupio: 07.09.2018]
12. The Binomial Distribution: A Probability Model for a Discrete Outcome, Dostupno: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Probability/BS704_Probability7.html [pristupio 07.09.2017]

Popis tablica

Prilog

Radu se prilaže Excel datoteka u kojoj su provedeni izračuni.