МАТЕМАТИЧКИ ОМНИБУС, 7 (2020), 135 – 152
http://im-pmf.weebly.com/matematicki-omnibus-kniga-7.html

БРИЕРОВИОТ СКОР И ПРЕДВИДУВАЧКИТЕ ТУРНИРИ

Марко Димовски [1]

Ирена Стојковска [1]

Неизвесноста на идните настани и желбата да се знае што ќе се случи, кај секого буди интерес за предвидување. Интересот е помасовен, ако исходите од тие настани имаат глобално значење. „Дали Велика Британија ќе ја напушти Европската унија до 1 февруари 2020 година?“, „Дали Ирак ќе организира парламентарни избори пред 1 декември 2020 година?“ и „Дали годишната стапка на инфлација во Еврозоната, за секој месец меѓу октомври 2019 година и декември 2020 година, ќе биде еднаква или поголема од 2%?“ се примери за геополитички прашања кои побудуваа интерес во декември 2019 година. „Колку случаи на заболени од COVID-19 ќе бидат регистрирани до 1 април 2021 година?“, „Кога ќе биде одобрен медицински лек за лечење од COVID-19?“ и „Кој ќе победи на претседателските избори во САД во 2020 година?“ беа прашања од интерес во август 2020 година. Вакви и слични актуелни прашања за настани во блиска иднина се поставуваат на предвидувачките турнири на кои индивидуалци или тимови се обидуваат да дадат веројатносна проценка за нивна реализација.

Како е организиран еден предвидувачки турнир? Учесниците на предвидувачкиот турнир одговараат на n прашања од глобален геополитички карактер. Потпирајќи се на свои или јавно достапни податоци, со помош на познати или нови предвидувачки методи и техники, тие ги изразуваат своите предвидувања во вид на веројатност за случување на одредениот настан. Затоа, предвидувачките турнири, од математички аспект, претставуваат одлична можност за тестирање на иновативни предвидувачки техники. Од политички аспект, тие може да послужат и како алатки за деполаризација на политичките дебати, [6].

Победник на предвидувачки турнир е тимот кој има најголема предвидувачка способност, која пак се мери со помош на Бриеровиот скор. Главна карактеристика на Бриеровиот скор е тоа што оние кои прават помали грешки при предвидувањата имаат подобар скор и истовремено ги принудува предвидувачите да ги прикажуваат своите вистински предвидувања, а не да лажираат веројатности за да ја оптимизираат очекуваната вредност на скорот. Но, Бриеровиот скор има и свои недостоци.

Во продолжението на овој труд ќе го изложиме концептот на предвидувачките геополитички турнири и ќе разгледаме дел од победничките методологии на турнирите. Централна проблематика на овој труд е анализата на Бриеровиот скор како мерка за мерење на предвидувачката способност. Ќе биде анализирана правилноста на Бриеровиот скор и ќе биде изложен парадоксот на предвидувачките турнири.

1. ПРЕДВИДУВАЧКИ ТУРНИРИ И ПОБЕДНИЧКИ ПРЕДВИДУВАЧКИ МЕТОДОЛОГИИ

Во теоријата на одлучување, донесената одлука е добра колку и предвидувањата на кои таа е основана. Затоа, приватните и јавните сектори значително вложуваат во предвидувањата. На пример, финансиските сервиси се интересираат да го предвидат трендот на корисниците, промените во геополитичките кругови и измените во регулативите. Енергетските компании вложуваат во предвидувањето на цената на електричната енергија, а владите се интересираат за политички настани со глобални последици. Дури и невладините организации сакаат да ги предвидат исходите на критичните прашања со кои се засегнати нивните клиенти, за да им помогнат да ја остварат нивната мисија – на пример: „Дали оваа година ќе има масовни убиства во Јужен Судан?“ или „Колку бегалци ќе пристигнат на брегот на Европа во следните три месеци?“, [12].

Турнирите за предвидување се натпревари во кои учествуваат индивидуалци или тимови и даваат веројатносни предвидувања за исходите на одреден број конкретни идни настани. Прашањата кои ги одговараат се со геополитичко значење и најчесто се со дихотомни бинарни одговори (да или не), но има и прашања со повеќе понудени одговори. Натпреварувачите, покрај тоа што ќе треба да го предвидат исходот, треба да дадат и проценка за веројатноста за реализација на предвидениот исход. Правилно дизајнираните предвидувачки турнири, не се натпревари на кои победува најсреќниот, туку претставуваат моќна алатка за наоѓање на одговори на прашања кои во услови на нормално расудување невозможно е да се одговорат, [9], [10].

Еден таков предвидувачки турнир е IARPA (Intelligence Advanced Research Projects Activity), организиран од американската агенција за разузнавање, кој започнал во 2011 година и траел 4 години. Целта на овој турнир била откривање на најдобрите предвидувачки тимови и предвидувачки методологии. Како мерка за предвидувачката способност на натпреварувачите користен е Бриеровиот скор, односно сумата на квадратите од отстапувањата на предвидената веројатност од реалноста, при што за реалноста се става 1, за настаните коишто се случиле и 0 за настаните коишто не се случиле. На пример, нека за еден настан, натпреварувачот предвидел дека има 80% шанси да се случи, односно предвидел дека веројатноста за реализација на настанот е 0,8. Настанот се случил, па затоа Бриеровиот скор кој одговара на тоа предвидување е $(1-0,8)^2=0,04$. Ова значи, колку е поточно предвидувањето, толку е помал Бриеровиот скор. Најдобра вредност на Бриеовиот скор за едно прашање е 0, а најлоша 1. Победник на еден предвидувачки турнир се прогласува тимот со најмал вкупен Бриеров скор, [10].

При тие услови на мерење на предвидувачката способност, победник на IARPA турнирот од 2011 годнина е тимот на The Good Judgement Project (GJP), [12]. Тајната за победата на овој тим е во специфичноста на нивниот пристап. Составен од статистичари и психолози, GJP тимот организирал турнир во самиот турнир, во кој учествувале 2400 американци со најразлична демографска и професионална позадина. Потоа, своите предвидувања GJP тимот ги градел врз основа на предвидувањата на масата. Така, тие имаат дојдено до следните заклучоци, [10]:

1) Предвидувањата основани на агрегација на резултатите од предвидувачките анкети и користење на најразлични методи, од проста аритметичка средина, до помодерни методи од типот на логаритамско-веројатносно оптимизирање на тежинската средина, се подобри од останатите методи за предвидувања.

2) Победничкиот алгоритам во секоја од годините на предвидувачкиот турнир е логаритамската трансформација на тежинската средина

од веројатносните проценки на испитаниците, [2]. Имено, при предвидување на *вистинската веројатност* f за реализација на некој настан, со p ја означуваме средната вредност од веројатносните проценки на испитаниците која претставува *предвидена веројатност* за реализација на настанот. Тогаш, со помош на трансформацијата

$$t(p)=\frac{p^a}{p^a+(1-p)^a},$$

агрериранатa веројатност p станува поекстремна, поблиска до 0 или 1 и со тоа поблиска до вистинската веројатност за реализација на настанот. За подобар квалитет на агрерираните веројатности, за параметарот a се зема вредност поголема од 1, на пример $a=2,5$. Со следната логаритамска трансформација

$$\log\frac{t(p)}{1-t(p)}=\log\frac{p^a}{(1-p)^a}=a[\log p-\log(1-p)],$$

се овозможува уште пореална оценка на вистинската веројатност.

3) Некои од предвидувачите се изненадувачки константно подобри од другите. Нивните лични профили откриваат надпросечен „отворен ум“ т.е. волја да ги третираат туѓите убедувања како хипотези за тестирање, а не како убедувања кои веднаш треба да се отфрлат. Тоа се луѓе кои веруваат дека предвидувањето е способност која може да се развива и подобрува и во него треба да се вложува. Најуспешните предвидувачи имаат и високо општо познавање на политиката, што и не е од толку суштинска полза, затоа што прашањата на предвидувачкиот турнир се од најразлични области.

4) Учењето, а со тоа и напредокот, е возможен и покрај тоа што светот на меѓународната политика и економија не е пријателски наклонет кон учењето. Постојат показатели дека предвидувачите се сè подобри во предвидувањата на глобалните настани, кои најчесто се единствени од својот вид и не постојат податоци на слични настани од минатото врз основа на кои би се направиле идните предвидувања за овие настани.

Со помош на овие заклучоци, GJP тимот успеал да генерира предвидувања кои биле околу 30% подобри од предвидувањата на врвните аналитичари од агенцијата за разузнавање кои имаат пристап до тајните класифицирани податоци. Затоа, IARPA предвидувачкиот турнир може да се смета за полезна алатка за генерирање на знаења. Со самото тоа што во процесот на предвидување им се посветува подеднакво внимание на сите ставови и убедувања, предвидувачките турнири може да се сметаат и за алатка за деполаризирање на непотребно поларизираните политички дебати, [6].

2. МЕРЕЊЕ НА ПРЕДВИДУВАЧКАТА СПОСОБНОСТ. БРИЕРОВ СКОР

При правење на предвидувања, важно е и обезбедување соодветни мерки за евалуација на направените предвидувања. Во претходниот дел го спомнавме Бриеровиот скор како мерка за мерење на предвидувачката способност на натпреварувачите на еден предвидувачки турнир. Бриеровиот скор е само една мерка од множеството мерки за евалуација на веројатносните предвидувања. Во продолжение ќе изложиме теориска рамка на мерките за евалуација на веројатносните предвидувања, [3].

Нека $\{E_1, E_2, ..., E_m\}$ е множество од попарно дисјунктни настани кои формираат разбивање на просторот од елементарни настани, т.е.

$$\sum_{i=1}^{m} P(E_i) = 1,\ E_i \cap E_j = 0,\ i \neq j,$$

каде $P(E_i)$ е предвидување на веројатноста за реализација на настанот E_i. Со $\mathbf{p}$ го означуваме m-димензионалниот вектор од веројатносни предвидувања за реализација на даденото множество настани т.е.

$$\mathbf{p} = (P(E_1), ..., P(E_m))^T .$$

Низата од n прогнози за веројатностите на настаните $E_1, E_2, ..., E_m$ ја претставуваме со помош на $m \times n$ матрица $\mathbf{P}$ чија k-та колона е векторот $\mathbf{p}_k$ на k-тата прогноза. За означување на низата настани кои се реализирале во n различни ситуации, го користиме n-димензионалниот вектор $\boldsymbol{\varepsilon}$ со целобројни компоненти, така што $\varepsilon_k = i$, ако i-тиот настан се реализирал во k-тата ситуација. На пример, при предвидување на времето за n различни денови, за секој ден разгледуваме m различни временски услови и доделуваме веројатносни предвидувања за секој од временските услови во секој од деновите. Трите основни особини кои бараме да ги поседува една скор функција $S_n(\mathbf{P})$ со која се евалуираат веројатносните предвидувања се:

1) адитивност т.е. при додавање на нова прогноза новата вредност на скор функцијата е:
$$S_{n+1}(\mathbf{P},\mathbf{p}_{n+1}) = S_n(\mathbf{P}) + S_1(\mathbf{p}_{n+1}),$$
2) зависност само од физичките набљудувања т.е. само од веројатностите на настаните кои се реализирале:
$$S_n(\mathbf{P}) = S_n(\mathbf{p}_1(\varepsilon_1),\mathbf{p}_2(\varepsilon_2),...,\mathbf{p}_n(\varepsilon_n)),$$
3) строго правилно однесување, што подразбира ако сите прогнози за веројатностите на настаните се еднакви т.е. $\mathbf{p}_k = \mathbf{p}$ за сите k, тогаш скор функцијата достигнува екстремна вредност кога сите предвидени веројатности се еднакви на набљудуваните вистински веројатности на настаните и во никој друг случај.

Која е оправданоста за овие особини? Барањето за адитивност повлекува дека секое поединечно предвидување влијае на вкупниот скор на ист начин како и секое друго предвидување. Второто барање за зависност само од веројатностите на настаните кои се реализирале доаѓа од таму што моделирањето на природните појави не може да зависи од тоа што би можело да биде, но не е набљудувано. Строгата правилност на скор функцијата обезбедува искрени предвидувања, а не наместени само заради подобра вредност на скорот.

Се покажува дека, во случај на повеќе од два настани т.е. $m > 2$, функција која ги поседува особините 1) - 3) има логаритамски облик:
$$S(x) = a\ln x + b, \tag{1}$$
каде што a и b се константи и секој друг поинаков избор на фунцкција повлекува нарушување на некоја од особините 1) – 3), [3].

Кога имаме два настани т.е. $m = 2$, тогаш станува збор за спротивни настани E и $\overline{E}$, на пример „ќе врне“ и „нема да врне“, односно станува збор за предвидување на еден настан, затоа што предвидувачот треба да одреди само една веројатност $P(E) = p$, од каде што ќе следува дека предвидената веројатност за спротивниот настан е $P(\overline{E}) = 1 - p$. И во овој случај се покажува дека скор функцијата којашто ги поседува особините 1) – 3), исто така има логаритамски облик:

$$S(x) = a \ln x,$$

каде што a е константа, [3].

Ако го прилагодиме обликот на скор функцијата (1) за дадена низа од n прогнози ја добиваме следната скор функција:

$$S_n(\mathbf{P}) = nb + a \sum_{k=1}^{n} \ln \mathbf{p}_k(\varepsilon_k),$$

Но, природно е скорот да претставува средна вредност од направените предвидувања, па затоа делиме со n и ја добиваме средната вредност на скорот:

$$\langle S_n(\mathbf{P}) \rangle = b + \frac{a}{n} \sum_{k=1}^{n} \ln \mathbf{p}_k(\varepsilon_k). \tag{2}$$

Со помош на теоријата на информации наоѓаме најсоодветен избор за константите a и b во (2), а така добиениот скор се нарекува *фер скор*, [3]:

$$FS = \ln m + \frac{1}{n} \sum_{k=1}^{n} \ln \mathbf{p}_k(\varepsilon_k).$$

Во случај на перфектни предвидувања, кога предвидувачот доделува веројатност 1 на настани кои потоа се случуваат т.е. $\mathbf{p}_k(\varepsilon_k) = 1$, за сите k, тој добива фер скор $FS = \ln m$, кој претставува максимален скор. На овој начин, не само искреноста, прецизноста и точноста, се промовира и решавачкиот капацитет, во смисла, колку е поголемо множеството од настани кои се предвидуваат т.е. m, толку е поголем и фер скорот.

Од 1950 година па наваму, Бриеровиот скор е една од најчесто користените алатки за верификација на веројатносните предвидувања, [4]. Согласно ознаките воведени тука, Бриеровиот скор се пресметува според формулата:

$$BS = \frac{1}{n} \sum_{k=1}^{n} \| \mathbf{p}_k - \hat{\mathbf{e}}_k \|^2, \tag{3}$$

каде што $\|\cdot\|$ е Евклидската норма, а $\hat{\mathbf{e}}_k$ е соодветниот вектор на реализација на настаните чии m компоненти се 0, освен i-тата која е 1 и покажува дека i-тиот настан E_i се реализирал во k-тата ситуација. Се покажува дека Бриеровиот скор е строго правилен скор, но не ја задоволува особината 2), затоа што зависи од веројатностите кои се доделуваат на настани кои никогаш не се случиле. Но, и покрај овој недостаток, Бриеровиот скор е широко прифатен од предвидувачката заедница и се користи со децении. Едно можно рационално објаснување за широката примена на Бриеровиот скор, покрај едноставната лесно пресметлива формула, може да биде и тоа што Бриеровиот скор за предвидувања на еден настан ($m=2$), претставува апроксимација од втор ред на логаритамскиот скор во точката $p=1/2$, [3]. Но, и покрај својата успешност за време на овие децении, Бриеровиот скор манифестира ограничувања кога се применува за предвидувања на многу ретки (или многу чести) настани, како и за предвидувања со слаб квалитет.

3. ЗА ПРАВИЛНОСТА НА БРИЕРОВИОТ СКОР

Правилноста на една скор функција подразбира подобро вреднување на веројатносните предвидувања кои се најблиску до вистинските веројатности на настаните, пред да се случат самите настани. Па, скор функцијата достигнува екстрем во случај кога веројатносните предвидувања се еднакви на вистинските веројатности на настаните, а секој поинаков избор на веројатности доведува до полош скор. На тој начин, предвидувачите се охрабруваат, доколку сакаат да постигнат најдобар скор, да бидат искрени, да ја прикажат вистинската предвидена веројатност, [5], [7].

Бриеровиот скор (3) е строго правилен скор. Да го разгледаме Бриеровиот скор за едно бинарно предвидување ($m=1, n=1$):

$$BS=\begin{cases}(1-p)^2, & \text{ако настанот се случил}\\ p^2, & \text{ако настанот не се случил}\end{cases}, \tag{4}$$

каде што p е веројатносното предвидување за реализација на настанот. Нека, f е вистинската веројатност за реализација на настанот, тогаш очекуваната вредност на Бриеровиот скор е:

$$E(BS) = f \cdot (1-p)^2 + (1-f) \cdot p^2 = f - 2fp + p^2 . \qquad (5)$$

За да ја минимизира очекуваната вредност на Бриеровиот скор, затоа што во тој случај ќе значи дека направил подобро предвидување на веројатностите, предвидувачот треба да го одбере веројатносното предвидување p, така што:

$$\frac{\partial E(BS)}{\partial p} = 2(p-f) = 0 .$$

Со ова се покажува правилноста на Бриеровиот скор, имено предвидувачот треба да одбере веројатносно предвидување p еднакво на вистинската веројатност f за реализација на настанот. Строгата правилност е затоа што $p = f$ е единствената вредност за p која го минимизира очекуваниот Бриеров скор. Понатаму, секоја мерка која е линеарна функција од Бриеровиот скор, е исто така строго правилна.

Заклучокот дека строгата правилност на скорот ги принудува предвидувачите да ги прикажуваат вистинските предвидени веројатности, лежи на претпоставката дека предвидувачите се ризик неутрални во однос на скорот, односно дека функцијата на корисност е линеарна функција од скорот. Да видиме од каде произлегува таа претпоставка. Имено, корисноста е нумеричка мерка за допаѓање (преферирање) која е во основата на теоријата на корисност на фон Нојман и Моргенштерн, [8], [11]. Ако $U(A)$ и $U(B)$ се соодветно корисности на настаните A и B, и настанот A се преферира повеќе од настанот B, тогаш $U(A) > U(B)$. Со помош на досегашните ознаки, очекуваната корисност од предвидувањето на еден настан E е

$$EU = U(E) \cdot f + U(\overline{E}) \cdot (1-f) .$$

Носењето одлуки според теоријата на корисност, подразбира максимизирање на очекуваната корисност, додека правилноста на Бриеровиот скор е основана на претпоставката дека предвидувачот се обидува да ја оптимизира очекуваната вредност на Бриеровиот скор (5). Затоа, имлицитно се претпоставува дека корисноста на предвидувачот е линеарна функција од скорот. Ако корисноста на предвидувачот не е линеарна функција од скорот, тогаш тој нема да ги носи одлуките само на основа на очекуваната вредност на скорот, туку ќе ја земе предвид и потенцијалната варијабилност на скорот. Така, предвидувачите кои не сакаат да преземаат ризици ќе се обидуваат да ја намалат варијабилноста, а тие кои немаат проблеми со преземање ризици, ќе се стремат кон зголемување на варијабилноста. За дисперзијата (варијансата) на Бриеровиот скор за едно бинарно предвидување (4), користејќи ја неговата очекувана вредност (5), добиваме дека е:

$$V(BS) = f \cdot (1-f) \cdot (1-2p)^2 . \tag{6}$$

Од формулата (6) согледуваме дека варијансата е најмала за предвидена веројатност $p = 1/2$, а најголема варијанса се добива за веројатносни предвидувања $p = 0$ и $p = 1$. Ризик неутралноста на предвидувачот е претпоставка за правилноста на Бриеровиот скор, но кога се работи за човечки предвидувања, тешко може да се каже дека ризик неутралноста е одржлива. За поимите корисност и ризик неутралност може да се прочита повеќе во [8], [11].

Пример за функција на корисност која не е линеарна функција од скорот е скалестата функција на корисност дефинирана со

$$U = \begin{cases} 1, & BS \le K \\ 0, & BS > K \end{cases}, \tag{7}$$

каде K е некоја целна вредност на Бриеровиот скор BS. При вака дефинирана корисност U, очекуваната корисност EU е веројатноста на настанот $BS \le K$. Ако целта е да се достигне одредена целна вредност на скорот, предвидувачот треба да се обиде да ја максимизира веројатниста на настанот $BS \le K$. Се покажува дека оптималната вредност на веројатносното предвидување е ([7])

$$p^* = \begin{cases} 1/2 + \sqrt{1/4 - K}, & \text{ако } f > 0,5 \\ 1/2 - \sqrt{1/4 - K}, & \text{ако } f < 0,5 \end{cases} .$$

Ова значи дека правилноста на Бриеровиот скор ќе биде нарушена, ако целта на предвидувачот е да ја максимизира веројатноста за постигнување на целна вредност K на Бриеровиот скор.

4. ПАРАДОКС НА ПРЕДВИДУВАЧКИТЕ ТУРНИРИ

Да се навратиме повторно на разгледување на предвидувачките турнири, затоа што Бриеровиот скор е најчесто користена мерка за евалуација на предвидувачката способност на учесниците на овие турнири. Од математички аспект, во предвидувачките турнири се тру-диме да дадеме веројатносен одговор на n прашања, од типот „Предивиди ја веројатноста да одреден светски настан се реализира пред одреден датум“. Во реалноста, на учесниците на овие предвидувачки турнири, им е дозволено повремено да ја обновуваат и менуваат првично дадената предвидена веројатност. За поедноставување, во продолжение ќе претпоставиме дека менување на првично одредената веројатност не е дозволено и прашањата се дихотомни (бинарни), односно се одговараат само со да или не. Бодувањето, односно мерењето на квалитетот на направените предвидувања се врши со помош на Бриеровиот скор (4). Вкупниот скор на крајот од турнирот претставува збир од бодовите освоени на секое од n-те прашања за предвидување. Слично како во голфот, натпреварувачот се обидува да оствари што е можно помал скор, односно победник е оној чиј Бриеров скор е најмал.

Нека S_i е Бриеровиот скор кој натпреварувачот го остварил на i-тото прашање. Тогаш, S_i е случајна променлива која прима вредности $(1-p_i)^2$ или p_i^2, во зависност од тоа дали настанот се реализирал или не, каде p_i е предвидената веројатност за реализирање на настанот од i-тото прашање. Нека f_i е вистинската веројатност за реализација на i-тиот настан. Тогаш, за математичкото очекување на S_i имаме:

$$E(S_i) = (1-p_i)^2 f_i + p_i^2 (1-f_i) = f_i(1-f_i) + (p_i - f_i)^2 .$$

Ако со S го означиме скорот од целиот предвидувачки турнир, тогаш неговата очекувана вредност ќе ја добиеме со сумирање на очекуваните вредности $E(S_i)$, $i = 1,...,n$.

$$E(S) = \sum_{i=1}^{n} E(S_i) = \sum_{i=1}^{n} [f_i(1-f_i) + (p_i - f_i)^2] = \sum_{i=1}^{n} f_i(1-f_i) + n\sigma^2 , \quad (8)$$

каде што

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(p_i - f_i)^2$$

е средно-квадратната грешка при предвидувањето на веројатностите. Да издвоиме неколку работи од оваа едноставна формула за математичкото очекување на скорот, [1]:

- Првиот собирок во изразот за очекуваниот скор (8) е еднаков за сите учесници на турнирот;
- Оваа формула ни покажува дека соодветен начин за мерење на точноста на предвидувањата е со помош на коренот на средно-квадратната грешка, односно со σ;
- Вистинскиот скор е случаен, односно

  $$S = \sum_{i=1}^{n} f_i(1 - f_i) + n\sigma^2 + \xi\,,$$

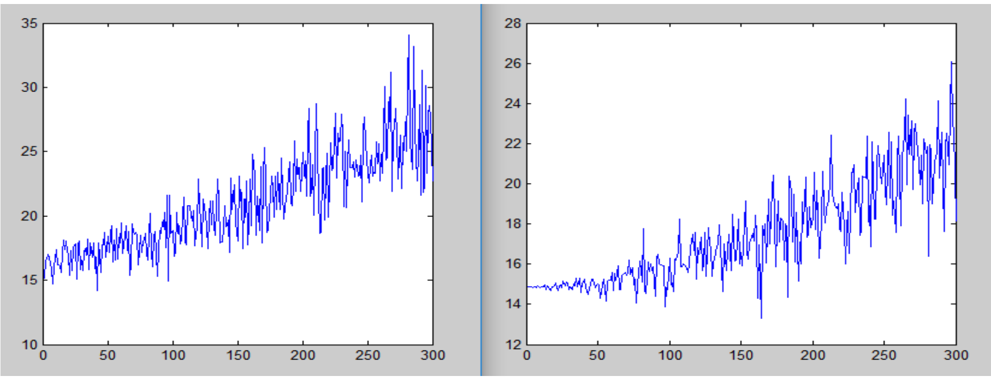
  каде што шумот ξ има очекување 0 т.е. $E(\xi) = 0$.

Не е за очекување дека натпреварувачите ќе ги погодат точните веројатности, затоа, за да можеме да измоделираме еден вид на реален турнир, ќе треба да ја измоделираме неточноста при предвидувањата. Како што кажавме претходно, средно-квадратната грешка σ^2 е онаа која влијае врз скорот, така што ќе ја параметризираме „неточноста“ во предвидувањето, преку коренот на средно-квадратната грешка σ.

Да разгледаме турнир со 300 натпреварувачи, кои одговараат на 100 прашања чии што вистински веројатности се 0,05, 0,10, 0,15,…, 0,95, каде што секоја од овие веројатности се повторува 10 пати по дадениот редослед. Вистинските турнири имаат проближно толку прашања. Од сите возможни модели, ќе земеме еден модел кој можеби е наједноставен. Во случај кога вистинската веројатност за реализација на i-тиот настан е f_i, претпоставуваме дека еден натпреварувач предвидува дека настанот ќе се реализира со веројатност $p_i = f_i \pm \sigma_i$, при што овие веројатности не треба да излегуваат надвор од интервалот [0,1], а во случај да излезат ги проектираме на интервалот $[0,1]$, односно ако $p_i < 0$, ставаме $p_i = 0$, а ако $p_i > 1$, ставаме $p_i = 1$.

Ако секој натпреварувач има подеднаква способност за предвидување, тогаш еднакво веројатно е било кој од нив да биде победник на турнирот. Моделирањето на варијабилноста при точноста на предвидувањето, исто така претставува тешка задача, па од таа причина одново ќе разгледаме еден едноставен модел. Ако способноста ја мериме преку коренот од средно-квадратната грешка, тогаш истата ја земаме да биде рамномерно распределена на интервал со произволно земена должина од $0,3$.

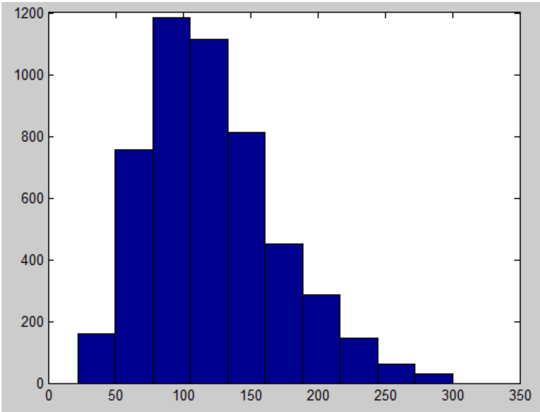
Накратко, за j-тиот натпреварувач имаме дека $\sigma_j = \sigma_0 + 0,3\frac{j}{m}$, $j = 1,\ldots,m$, каде што m е бројот на натреварувачи, што во нашиот случај е 300. Гледано во однос на коренот од средно-квадратната грешка, првиот натпреварувач има најдобра предвидувачка способност, а последниот има најлоша. Со бројка од 300 натпреварувачи, според начинот на кој е моделирано σ_j, способноста на првиот натпреварувач не се разликува премногу од онаа на вториот или пак на третиот. Вообичаената логика ни вели дека победникот треба да го бараме помеѓу најдобро рангираните натпреварувачи, односно оние кои имаат најдобра предвидувачка способност според нашиот модел. Слично е и во спортот, најчесто шампионот е некоја од екипите или индивидуалците кои што се фаворити на почетокот од натпреварувањето.



Слика 1. Бриеровиот скор на учесниците на два симулирани предвидувачки турнири, за $\sigma_0 = 0$.

На Слика 1 се прикажани два симулирани предвидувачки турнири, за $\sigma_0 = 0$. Симулациите се направени во софтверот Python. Иако, со зголемувањето на индексот (т.е. редниот број на натпреварувачот), се намалува неговата предвидувачка способност, сепак забележуваме дека во првиот симулиран турнир, најмал Бриеров скор има натпреварувачот кој е рангиран околу 50-тото место, а во вториот симулиран турнир, победникот е натпревувачот рангиран околу 170-тото место.

На Слика 2 е прикажан хистограм од честоти на рангот во кој се наоѓа победникот на турнирот изработен за 5000 симулирани предвидувачки турнири, при веќе наведените услови и $\sigma_0 = 0$.



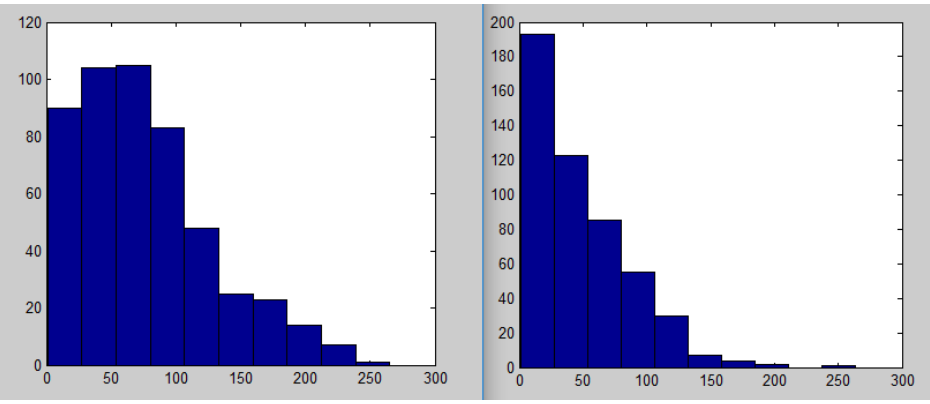
Слика 2. Хистограм од честоти за рангот на победниците на 5000-те симулирани турнири, за $\sigma_0 = 0$.

Како што се забележува од Слика 2, победникот најчесто бил меѓу натпреварувачите кои според предвидувачките способности биле рангирани од 75-то до 100-то место. Ниту еднаш не се случило да победи некој од натпреварувачите кои биле најдобро рангирани на почетокот на турнирот, што е целосен контраст со нашата почетна интуиција и претставува *парадокс на предвидувачките турнири*.

Дали со промена на вредноста на σ_0, би имале голема разлика во резултатите? На Слика 3 се дадени соодветните хистограми, при $\sigma_0 = 0,05$ и $\sigma_0 = 0,1$ и 500 извршени симулации.

Како што можеме да забележиме, хистограмите на Слика 3 имаат поинаков облик, но сè уште не е исполнето она што се очекува, односно дека најдобро рангираните би биле најголеми фаворити за победа на турнирот.

Парадоксот што го добивме е лесен за објаснување. Група од подобро рангираните натпреварувачи според моделираната способност за предвидување врши скоро еднакви предвидувања, како да има само еден натпреварувач со иста способност. Но, ако гледаме кај натпреварувачите со корен на средно-квдратната грешка од 0,1, тие прават за нијанса поразлични предвидувања, во просек добивајќи помалку бодови, но кај некои прогнозери постои шанса дека нивните предвидувања одат во насока на реалниот исход на настанот, па овие натпреварувачи би добиле подобар скор само поради имањето среќа.



Слика 3. Хистограм од честоти за рангот на победниците на 500-те симулирани турнири, при $\sigma_0 = 0,05$ односно $\sigma_0 = 0,1$.

Ако направиме една физичка аналогија, да замислиме дека имаме натпреварувачи кои стрелаат последователно во 100 различни црвени кругови на средината од една вообичаена мета. Да замислиме дека постои син круг кој се наоѓа во близина на црвениот круг и дека официјалните лица ги бодуваат натпреварувачите според просечното растојание од погодокот во метата и синиот круг. Оние натпреварувачи кои се поспособни и повешти, би имале стрелања кои би биле во или околу црвената мета на кругот, така што грубо кажано, тие ќе имаат сличен скор. Натпреварувачите со помала техника и вештини, вообичаено би имале полош скор, меѓутоа постои шанса нивните грешки да бидат повеќе во насока на синиот круг, така што како резултат на случајноста би имале подобар скор.

Да разгледаме турнир со 100 прашања во кој вистинската веројатност за реализација на секој настан е 0,5. Во овој случај, ако имаме натпреварувач кој направил перфектни прогнози, тогаш неговиот скор ќе биде $100 \cdot 0,5^2 = 25$. Да претпоставиме дека еден натпреварувач одговарал со веројатности 0,4 или 0,6 на секое од прашањата. Тогаш, неговиот скор е случаен со математичко очекување 26, според формулата (8), и стандардна девијација 0,98, па може да се пресмета дека има околу 15% шанса да го победи перфектниот предвидувач, [1]. Во типичен спортски натпревар, очекувано е дека победникот ќе биде некој од најдобрите екипи/натпреварувачи. Затоа, треба да се сфати дека нашиот турнир е концептуално различен од спортските натпревари. Во спортот грешките најчесто нѐ чинат скапо и многу ретко имаме среќа да профитираме од нив (на пример, некој лош удар може да се одбие од некој играч во голот), за разлика од предвидувачките турнири каде што предвидување од 0,6 или 0,4 кога вистинската веројатност е 0,5 е скоро подеднакво профитирачко или штетно. 100 грешки на спортски натпревари ќе не чинат многу повеќе од 100 грешки во предвидувачките турнири, кои некогаш може да бидат чист бенефит. За предвидените веројатности, очекуваната штета од малите грешки се скалира како квадратот од грешката, додека стандардната девијација се скалира како грешката.

Се поставува прашањето дали нашиот модел е реален? Две од нашите претпоставки во едноставниот модел се нереални. Првата е дека натпреварувачите немаат систематска пристрасност кон доволно високите или доволно ниските предвидувања. Второто е дека грешките се независни и во однос на прашањата и во однос на натпреварувачите. Во реалноста, ако натпреварувачите ги носат одлуките на слично множество од податоци, тогаш нивната одлука ќе биде пристрасна во иста насока на дадено прашање. Во нашиот модел очигледно е дека оваа претпоставка за независност на грешките ни значи дека различни натпреварувачи би имале избор кој е рамномерно распределен на некоја низа од веројатности блиски до вистинската, што во реалноста не е така. Сепак, парадоксот на предвидувачките турнири изложен со овој модел, претставува доволен поттик за понатамошни истражувања на полето на примената на Бриеровиот скор за мерење на предвидувачката способност на натпреварувачите на предвидувачките турнири.

5. ЗАКЛУЧОК

Математиката може да понуди модел за секој реален проблем. Гледано од аспект на математичкиот јазик, некои од моделите се поедноставни, некои посложени. Сложеноста на моделот не е секогаш предуслов за негова примена. Бриеровиот скор е пример за едноставна математичка алатка за мерење на предвидувачката способност која успешно се користи со децении. Во овој труд ги анализиравме својствата на Бриеровиот скор, неговата оправданост, правилност и веродостојност. Случајната природа на проблемот е една од причините да заклучиме дека и покрај оправданоста, постои простор за подобрување на Бриеровиот скор.

ЛИТЕРАТУРА

[1] D. Aldous, *A prediction Tournament Paradox*, March 5 2019, arXiv:1903.02131.

[2] J. Baron, B. A. Mellers, P. E. Tetlock, E. Stone, L. H. Ungar, *Two Reasons to Make Aggregated Probability Forecasts More Extreme*, Decision Analysis 11 (2) (2014), 133−145.

[3] R. Benedetti, *Scoring Rules for Forecast Verification*, Monthly Weather Review 138(1) (2010), 203−211.

[4] G. W. Brier, *Verification of forecasts expressed in terms of probability*, Mon. Wea. Rev. 78 (1950), 1−3.

[5] T. Gneiting, A. E. Raftery, *Strictly Proper Scoring Rules, Prediction, and Estimation,* Journal of the American Statistical Association 102 (447) (2007), 359–378.

[6] B. Mellers, P.Tetlock, H. R. Arkes, *Forecasting tournaments, epistemic humility and attitude depolarization*, Cognition 188 (2019), 19−26.

[7] M. S. Roulston, *Performance targets and the Brier score*, Meteorological Applications 14 (2007), 185−194.

[8] И. Стојковска, *Одлучување во услови на ризик и неизвесност*, Математички омнибус 5 (2019), 33−55.

[9] P. E. Tetlock, B. A. Mellers, N. Rohrbaugh, E. Chen, *Forecasting Tournaments: Tools for Increasing Transparency and Improving the Quality of Debate*, Current Directions in Psychological Science 23(4) (2014), 290−295.

[10] P. E. Tetlock, B. A. Mellers, J. P. Scoblic, *Bringing probability judgments into policy debates via forecasting tournaments*, Science 355 (2017), 481–483.

[11] J. Von Neuman, O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, 1944.

[12] *Good Judgment*, https://goodjudgment.com/.

[1] Универзитет „Св. Кирил и Методиј", Скопје
Природно-математички факултет
Архимедова 3, 1000 Скопје, Р. Северна Македонија
e-mail: mdimovski16@gmail.com,
e-mail: irena.stojkovska@gmail.com

Примен: 26.8.2020
Поправен: 23.11.2020
Одобрен: 1.12.2020
Објавен на интернет: 6.12.2020