UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2019:22

En enkel modell i utslagningsturneringar

Edward Motzi

Examensarbete i matematik, 15 hp
Handledare: Stephan Wagner
Ämnesgranskare: Cecilia Holmgren
Examinator: Veronica Crispin Quinonez
Juni 2019

Department of Mathematics
Uppsala University

Sammanfattning

Denna enkla modell i utslagsturneringar används för att ta reda på hur stora sannolikheterna är för varje spelare att vinna i en turnering. I detta arbete beskrivs modellen och hur man kan tillämpa den. I början tas lite bakgrund upp innan modellen beskrivs och därefter hur man kan räkna problem med den. Senare i arbetet går man igenom hur modellen fungerar när man har ett stort antal spelare i en turnering samt något om seedningsmetoden i diskussionen. Den här modellen är användbar i de fallen där en turnering använder sig utav utslagningar utan efterföljande lottning och resultaten är ganska precisa. Men det finns även vissa kommentarer, som kommer att tas upp i slutet av detta arbete!

This simple model of a knock-out tournament is used to calculate how big the probabilities are for each player to win in a tournament. In this work, the model is described and shows how to use it. In the beginning, a little bit of background is brought up before the model is described more thoroughly and afterwards how to solve problems with it. Later in the work, it is brought up how to use the model when there is a large amount of players in the tournament as well as a little bit on the seeding method in the discussion. This model is useful in those cases where a tournament is using a knock-out system without consecutive draw and the results are quite accurate. But there are some comments as well, which will be taken up in the later part of this work!

Innehåll

1 Introduktion

Att kunna förutsäga vilken spelare eller vilket lag som ska vinna en hel turnering är något som människor har tänkt på länge. Flera tusen år innan vår tidräkning har människor spelat på odds och sannolikhet. Det var på 1650-talet som man började räkna på sannolikhet. Allt började med två frågor som Antoine Gombaud *(f.1607-d.1684)*, Chevalier de Méré, ställde till Blaise Pascal *(f.1623-d.1662)* om sannolikhet. Pascal brevväxlade med Pierre de Fermat *(f.1607-d.1665)* och den klassiska sannolikhetsläran var född [3].

Att slå vad och satsa pengar om olika sporthändelser har engagerat människor i århundraden och det är även aktuellt än idag. Det kan vara allt från en liten Trisslott du köper i kiosken bredvid till att satsa din förmögenhet på en häst i Elitloppet på Solvalla. Vem vinner årets Grand Slam i tennis och vem blir årets skräll? Vilket lag lyfter den ovärderliga trofén över deras huvuden i slutet av säsongen?

Går det verkligen att kunna räkna ut på det? Kan vi använda oss utav sannolikhet för att räkna ut olika händelser och ta reda på med hjälp av formler vilken spelare som vinner över vem? För vi kan se olika versioner av detta, till exempel inom betting, där den spelaren eller laget med lägst odds är den som flest personer satsat sina pengar på, vilket betyder att många tror att det är den spelaren som kommer att vinna, alltså har störst sannolikhet att ta hem segern. Men det finns ingen direkt siffra på hur stor själva sannolikheten är.

2 Bakgrund

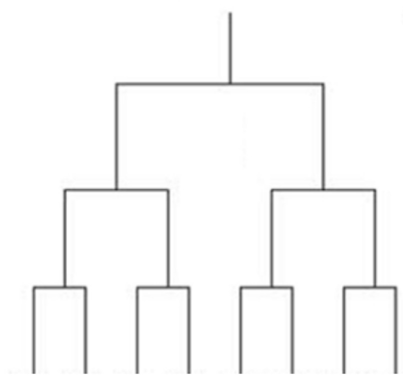
Det här examensarbetet handlar om en enkel modell i utslagningsturneringar. Denna modell kan användas för att räkna ut en given spelare ks väg till finalen och att eventuellt vinna hela turneringen. Det modellen gör är att den räknar ut sannolikheten att spelare k ska gå vidare i turneringen genom att ta hänsyn till om spelare k är den starkare eller svagare spelaren i mötet. Om spelare k är starkare än motståndaren, är sannolikheten stor att gå vidare. Men om spelare k är svagare än motståndaren, är sannolikheten att gå vidare däremot mindre. För att ha störst sannolikhet att komma så långt som möjligt i turneringen, är det bäst att ha alla de bättre spelarna så långt ifrån den givna spelaren k som möjligt.

Hur kan den här studien användas i ett större sammanhang? Denna studie skulle kunna användas för att ta reda på sannolikheter i kommande mästerskap i de sporter och evenemang som använder sig av utslagningar. Det kan tillika vara en grogrund till diskussioner om vadslagningar och idrottshändelser. Ju mer vi tar upp det här ämnet, desto bättre kan det bli.

Den här modellen är påhittad av forskare på Uppsala Universitet och är inte officiellt publicerad någonstans [4]. Detta examensarbete grundar sig mycket på forskarnas kompendier och papper.

3 Beskriva modellen

För ett positivt heltal n, betraktar vi en turnering med 2^n spelare som spelar n rundor av simpel elimination, där vinnaren i mötet avancerar till nästa runda medan förloraren blir utslagen ur turneringen. Spelarna är rankade från 1 till 2^n, där 1 är bäst rankad och 2^n är sämst rankad, och vi gör ett enkelt antagande att den spelare med högre rang slår spelaren med lägre rang med sannolikheten p, där $\frac{1}{2} < p < 1$, och förlorar med sannolikheten $1-p = q$. Likaså vinner den spelare med lägre rang med sannolikheten $1 - p = q$. Den inledande lottningen är bestämd av en likartad slumpmässig permutation av $1, 2, 3, \ldots, 2^n$. Därefter blir det inga fler lottningar under turneringens gång.



Figur 1: Bilden visar en så kallad bracket.

Vi är speciellt intresserade i sannolikheten att spelare k vinner turneringen. Det vi kan se är att den högst rankade spelaren, spelare 1, vinner turneringen med sannolikheten p^n, på grund av att varje match är oberoende av varandra och är även oberoende av den inledande lottningen. På samma sätt vinner spelare 2^n turneringen med sannolikheten q^n [4].

Låt $P(k, n)$ (där $k \in (1, 2, ..., 2^n)$) beteckna sannolikheten att spelare k vinner en utslagningsturnering med n-omgångar, enligt detta antagande. Den exakta sannolikheten kan räknas ut rekursivt enligt följande: $P(1, 0) = 1$ och

$$P(k,n) = \frac{1}{\binom{2^n-1}{2^{n-1}-1}} \sum_{j=1}^{\min(k,2^{n-1})} \binom{k-1}{j-1} \binom{2^n-k}{2^{n-1}-j} P(j,n-1)$$
$$\left(q \sum_{r=1}^{k-j} P(r,n-1) + p \sum_{r=k-j+1}^{2^{n-1}} P(r,n-1) \right)$$

eller ekvivalent

$$P(k,n) = \frac{1}{\binom{2^n-1}{2^{n-1}-1}} \sum_{j=1}^{\min(k,2^{n-1})} \binom{k-1}{j-1} \binom{2^n-k}{2^{n-1}-j} P(j,n-1)$$
$$\left(p + (1-2p) \sum_{r=1}^{k-j} P(r,n-1) \right).$$

Nämnaren i formeln förklarar på hur många olika sätt det går att dela upp resterande spelare. Där $2^n - 1$ är totala antalet spelare i hela turneringen och $2^{n-1} - 1$ är antalet andra spelare i samma halva som spelare k.

Därefter beskriver formeln att man tar summan, från när $j = 1$, när spelare k är rankad som bästa spelare i sin halva, till $\min(k, 2^{n-1})$, som beror på hur många bättre spelare det finns i samma halva som spelare k. Här kollar vi vilket tal som är minst av de två siffrorna inom parentesen. Det här förklarar vilken rang spelare k kan ha som lägst, i sin halva.

Om vi tar ett exempel där $k = 5$ och $n = 4$, så har spelare k rangen 5 av 16 spelare i en turnering med 4 omgångar, där 7 spelare finns i samma halva. Spelare k kan då vara som sämst 5:a, för vi vet att 5 är mindre än 8, så $\min(5, 2^{4-1}) = \min(5, 8) = 5$. Om $k = 11$ och $n = 4$, då är $\min(11, 2^{4-1}) = \min(11, 8) = 8$, vilket betyder att spelare k i detta fall inte kan vara sämre än 8:a. Det här menas att spelare k inte kan vara sämre än k-värdet, så j-värdet är antingen lika med eller mindre än k-värdet ($j \leq k$). Det j-värdet representerar är rankningen av spelare k i sin halva. Om det finns två starkare spelare i samma halva som spelare k, är j lika med 3.

Det går även att modifiera den nedre gränsen i summan från $j = 1$ till $\max(1, k - 2^{n-1})$. I detta fall kollar vi vilket tal som är störst av de två siffrorna inom parentesen. Det här förklarar vilken rang spelare k kan ha som högst, i sin halva. Om vi tar samma exempel med $k = 11$ och $n = 4$, så har vi att spelare k inte kan vara högre rankat än 3:a, då $\max(1, 11 - 2^{4-1}) = \max(1, 11 - 8) = \max(1, 3) = 3$. Då har vi att spelare $k = 11$ inte kan vara högre rankad än 3:a och inte sämre rankad än 8:a. Denna ersättning fungerar bäst om spelare k är mycket lågt rankad ($k > 2^{n-1}$).

Formeln summerar kombinationen av $\binom{k-1}{j-1}$, där $k-1$ är antalet starkare spelare i hela turneringen och $j-1$ är antalet starkare spelare i samma halva som den givna spelaren k. Multiplicerat med den andra kombinationen $\binom{2^n-k}{2^{n-1}-j}$, där $2^n - k$ är antalet svagare spelare i hela turneringen och där $2^{n-1} - j$ är antalet svagare spelare i samma halva som den givna spelaren k. Dessa två kombinationer multiplicerat med varandra, förklarar hur många olika sätt det går att fördela de kvarvarande spelarna, så att spelare k är den j:te i dess halva.

Därefter multipliceras allt med $P(j, n-1)$ som beskriver sannolikheten att den givna spelaren k tar sig till final.

Den sista delen i formeln beskriver sannolikheten att finalmotståndaren är starkare eller svagare än spelare k. Sannolikheten att finalmotståndaren är starkare än spelare k, är $\sum_{r=1}^{k-j} P(r, n-1)$. I detta fall vinner spelare k finalen med sannolikheten q. Sannolikheten att motståndaren är svagare, är $\sum_{r=k-j+1}^{2^{n-1}} P(r, n-1)$. I detta fall vinner spelare k finalmatchen med sannolikheten p.

Den här sista delen i formeln kan se ut på två olika sätt. Vi kan räkna ut hur vi får den andra varianten genom att uttrycka den på detta sätt, då alla sannolikheter måste adderas upp till 1:

$$\sum_{r=1}^{k-j} P(r, n-1) + \sum_{r=k-j+1}^{2^{n-1}} P(r, n-1) = \sum_{r=1}^{2^{n-1}} P(r, n-1) = 1.$$

Därför får vi:

$$q \cdot \sum_{r=1}^{k-j} P(r, n-1) + p \cdot \sum_{r=k-j+1}^{2^{n-1}} P(r, n-1) = q \cdot \sum_{r=1}^{k-j} P(r, n-1) + p \cdot (1 - \sum_{r=1}^{k-j} P(r, n-1))$$

$$= p + (q - p) \sum_{r=1}^{k-j} P(r, n-1) = p + (1 - 2p) \sum_{r=1}^{k-j} P(r, n-1).$$

3.1 Att räkna problem med modellen

Om vi vill räkna vägen till att vinna turneringen för spelare k, så måste vi veta vad k-värdet är. Vi vet att k-värdet ligger mellan 1 och 2^n, $(1 \leq k \leq 2^n)$. Om vi vill ta reda på spelare 1, har vi att $k = 1$, spelare 2, $k = 2$, osv. Sedan måste vi också ta reda på hur många omgångar det krävs att vinna turneringen. Om det är få deltagare är n-värdet lågt, medan n-värdet är högre ju fler deltagare det finns i turneringen. När vi vet våra k- och n-värden, lägger vi in värdena i formeln vi gick igenom i föregående avsnitt och räknar ut.

3.2 Vandermondes identitet

Identiteten är döpt efter den franske matematikern Alexandre-Théophile Vandermonde *(f.1735-d.1796)*. Vandermondes identitet beskriver följande identitet av binomialkoefficienter:

$$\sum_j \binom{a}{j}\binom{b}{c-j} = \binom{a+b}{c}.$$

Man kan räkna ut beviset av Vandermondes identitet på tre olika sätt, genom algebraisk, kombinatorisk eller geometrisk bevis. Här använder vi oss av det algebraiska beviset.

Bevis. Genom att använda binomialsatsen för exponenterna a och b, samt formel för polynomialprodukten får vi följande [1, 2]:

$$\begin{aligned}\sum_{c=0}^{a+b} \binom{a+b}{c} x^c &= (x+1)^{a+b} = (x+1)^a (x+1)^b \\ &= \sum_{j=0}^{a} \binom{a}{j} x^j \cdot \sum_{r=0}^{b} \binom{b}{r} x^r = \sum_{c=0}^{a+b} \sum_{j=0}^{c} \binom{a}{j}\binom{b}{c-j} x^c.\end{aligned}$$

$\square$

Vi kan även bevisa genom att utföra det kombinatoriska beviset.

Bevis. Tänk dig en kvalificeringsomgång som består av a duktiga och b mindre duktiga spelare. På hur många olika sätt kan man bilda en turnering som innehåller c antal spelare? Svaret är $\binom{a+b}{c}$. Svaret kan man få på liknade sätt genom att ta summan över alla möjliga värden på j, av antalet spelare som består av j duktiga och $c-j$ mindre duktiga, alltså $\sum_j \binom{a}{j}\binom{b}{c-j}$. □

Vandermondes identitet tillämpas likaså i den formel som beskrivits tidigare i denna text:

$$\sum_j \binom{k-1}{j-1}\binom{2^n-k}{2^{n-1}-j} = \binom{2^n-1}{2^{n-1}-1}.$$

Nämnaren i denna formel, alltså binomialen $\binom{2^n-1}{2^{n-1}-1}$, är det totala antalet möjligheter att distribuera spelarna i turneringen, förutom spelare k, mellan de två halvorna. Det som måste göras är att beräkna över alla kombinationer, vilket är varför vi dividerar med det hela. Det vi kan då se är att summan av produkterna är exakt lika med nämnaren. Detta är ett specialfall som kallas Vandermondes identitet.

4 Kan vi hitta en formel för spelare 2?

Vi vet att spelare 1, rankad som den bästa spelare i hela turneringen, vinner turneringen med sannolikheten $P(1,n) = p^n$, samt att spelare 2^n, rankad som den sämsta spelaren i hela turneringen, kan vinna turneringen med sannolikheten $P(2^n, n) = q^n$. Kan vi ta reda på med hjälp av en formel hur spelare 2, eller någon annan spelare mellan 1 och 2^n, vinner turneringen? Vi kan göra det med hjälp av två olika bevis.

4.1 Bevis 1

Av induktion på n, genom användandet av rekursion. Matematisk induktion är en metod för att bevisa påståenden med talföljder som innehåller naturliga tal som är större än eller lika med ett startvärde ($n \geq 0$). Induktionsbeviset delas upp i tre steg.

1. Man ska visa att påståendet gäller för startvärdet.
2. Man ska anta att påståendet gäller för något heltal $n-1$.
3. Man ska visa att om det gäller för $n-1$, då måste påståendet också gälla för heltalet n.

Om alla tre steg uppfylls, då har påståendet bevisats. Därefter används rekursion i induktionssteget. Vi får följande formler för olika k-värden:

$$P(1,n) = p^n.$$

$$P(2,n) = \frac{1}{2^n - 1}(p^{n-1} + (2^n - 1)p^n - 2^n p^{2n-1}).$$

Genom att ersätta k-värdet från 2 till 3 och utföra beviset på samma sätt får man följande formel för $P(3,n)$:

$$P(3,n) = \frac{1}{p^2(1+2p)(2^n-1)(2^n-2)}(3 \cdot 2^{2n}p^{3n} - 2^{2n+2}p^{2n+2} + 2^{n+3}p^{2n+2} - 2^{2n+1}p^{2n+1} - 2^{n-1}p^{2n+1} - 3 \cdot 2^n p^{2n} + 2^{2n+1}p^{n+3} + 2^{2n}p^{n+2} - 3 \cdot 2^{n+1}p^{n+3} + 2^n p^{n+2} + 2^{n+1}p^{n+1} + 4p^{n+3} - 6p^{n+2} + 2p^{n+1}).$$

Bevis. Först kontrollerar vi att formeln är korrekt för $n = 1$.

$$P(2,1) = \frac{1}{1}(1 + p - 2p) = 1 - p = q.$$

Detta är korrekt, då spelare 2 vinner med sannolikheten q i rond 1 i en turnering med endast två spelare.

Nu antar vi att formeln är korrekt för $P(2, n-1)$ och sedan bekräftar vi att det är likaledes korrekt för $P(2, n)$. Till det andra steget använder vi oss av rekursion från den föregående sektionen i induktionssteget. I steget mellan det andra och det tredje uttrycket använder vi oss utav induktionshypotesen till $P(2, n-1)$.

$$\begin{aligned}
P(2,n) &= \frac{1}{\binom{2^n-1}{2^{n-1}-1}} \sum_{j=1}^{2} \binom{1}{j-1}\binom{2^n-2}{2^{n-1}-j} P(j,n-1) \\
&\qquad \left(p + (1-2p)\sum_{r=1}^{2-j} P(r,n-1)\right) \\
&= \frac{1}{\binom{2^n-1}{2^{n-1}-1}} \binom{2^n-2}{2^{n-1}-1} P(1,n-1)\left(p+(1-2p)P(1,n-1)\right) \\
&\qquad + \frac{1}{\binom{2^n-1}{2^{n-1}-1}} \binom{2^n-2}{2^{n-1}-2} P(2,n-1)\cdot p \\
&= \frac{2^{n-1}}{2^n-1} p^{n-1}\left(p+(1-2p)p^{n-1}\right) \\
&\qquad + \frac{2^{n-1}-1}{2^n-1}\frac{1}{2^{n-1}-1}(p^{n-2} + (2^{n-1}-1)p^{n-1} - 2^{n-1}p^{2n-3})\cdot p \\
&= \frac{1}{2^n-1} 2^{n-1}p^{n-2}(p^n + p^2 - 2p^{n+1}) \\
&\qquad + \frac{1}{2^n-1}(p^{n-1} + 2^{n-1}p^n - p^n - 2^{n-1}p^{2n-2}) \\
&= \frac{1}{2^n-1}(p^{n-1} + (2^n-1)p^n - 2^n p^{2n-1}).
\end{aligned}$$

$\square$

Alla dessa formler härstammar från samma ursprungliga formel ifrån avsnitt 2 och kan härledas på samma sätt genom att ändra på k-värdet. När man väl har fått fram formeln och vet sina n-värden, lägger man in sina värden och löser ut sannolikheten för varje spelare i turneringen.

4.1.1 Tabeller och grafer

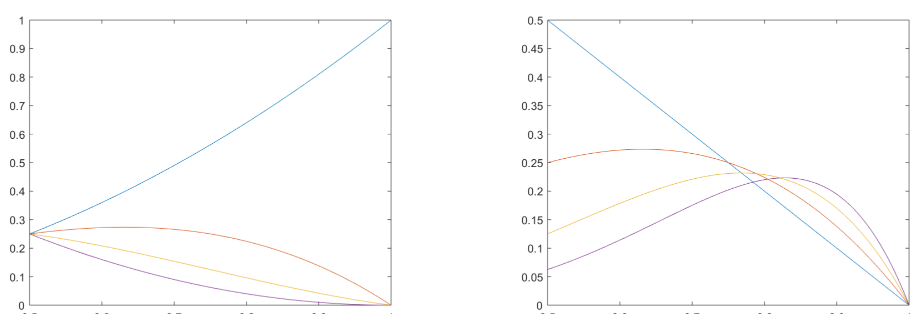
I delavsnittet visas två tabeller på värden man får genom att kombinera olika k- och n-värden. Notera att dessa tabeller är avskalade.

	$n=1$	$n=2$	$n=3$
$k=1$	p	p^2	p^3
$k=2$	$q=1-p$	$\frac{1}{3}(p+3p^2-4p^3)$	$\frac{1}{7}(p^2+7p^3-8p^5)$
$k=3$	-	$\frac{1}{3}(5p-9p^2+4p^3)$	$\frac{1}{7}(3p^2+5p^3-24p^5+16p^6)$
$k=4$	-	$q^2=(1-p)^2$	$\frac{1}{35}(p+26p^2+29p^3-120p^4$ $-16p^5+144p^6-64p^7)$

	$n=4$
$k=1$	p^4
$k=2$	$\frac{1}{15}(p^3+15p^4-16p^7)$
$k=3$	$\frac{1}{105}p^3(17+99p+12p^2-48p^3-176p^4-96p^5+192p^6$
$k=4$	$\frac{1}{455}(p^2+126p^3+405p^4+60p^5-528p^6$ $-864p^7-1440p^8+3520p^9-768p^{10}-512p^{11})$

Vi kan även notera att $P(8,3)=q^3=(1-p)^3$ och $P(16,4)=q^4=(1-p)^4$.

Det vi kan se är att sannolikheten för att en svagare spelare ska vinna turneringen minskar avsevärt mycket ju fler spelare och ronder det finns. Om man adderar ihop alla värden i vardera kolumn kommer man få summan 1.

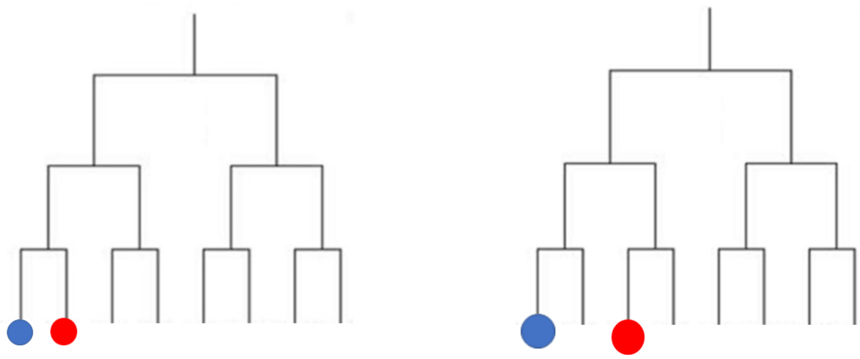


(a) Fixt $n=2$ (b) Fixt $k=2$

Vi kan också se i graferna att ju högre n-värdet är, desto mer åt höger rör sig topparna och konvergerar mot en gräns.

4.2 Bevis 2

Spelare k ($2 \leq k \leq 2^n - 1$) vinner med sannolikheten p, såvida inte spelaren möter en starkare motståndare. Sannolikheten beror på var den starkare spelaren befinner sig i utslagningsplanet.



(a) Möts i första ronden

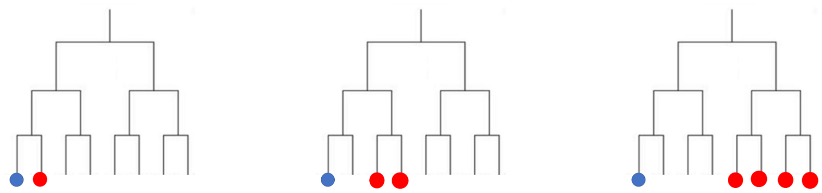
(b) Möts i andra ronden

I detta exempel är spelare *blå* rankad tvåa och spelare *röd* rankad etta, alltså bäst i turneringen.

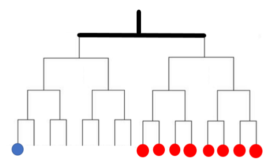
I det första spelet har spelare *blå* sannolikheten q att vinna. Om spelare *röd* blir utslagen tidigt, till exempel i rond 1, har spelare *blå* sannolikheten p att vinna nästa spel. Det är för att spelare *blå* är den bäst rankade spelaren i hela turneringen, på grund av att den spelare som tidigare var rankad bäst är nu utslagen. Då har spelare *blå* sannolikheten p^n att vinna resten av turneringen med n-antal ronder.

Om spelarna möts i rond m. Sannolikheten att spelare *röd* bli utslagen innan rond m är, $1-p^{m-1}$, och att spelare *röd* avancerar till rond m är, p^{m-1}. I det första fallet, där spelare *röd* blir utslagen innan rond m, har spelare *blå* sannolikheten p^n att vinna turneringen. Medan i det andra fallet, där spelare *röd* avancerar till rond m, har spelare *blå* sannolikheten $p^{n-1}q$ att vinna turneringen. Vi kommer få en geometrisk summa som kan räknas ut. Totalt har vi:

$$(1 - p^{m-1})p^n + p^{m-1}p^{n-1}q.$$



(a) Rond 1 - 1 position (b) Rond 2 - 2 positioner (c) Rond 3 - 4 positioner



Figur 5: Rond 4 - 8 positioner

Antalet möjligheter att möta den bättre spelaren under turneringen gång, specifikt i rond m, är 2^{m-1}. Då kan alltså den bättre spelaren ha 2^{m-1} möjliga positioner i turneringen, om spelarna ska mötas i rond m. Det totala antalet positioner spelare 1 kan ha i turneringen jämfört med spelare 2 är $2^n - 1$.

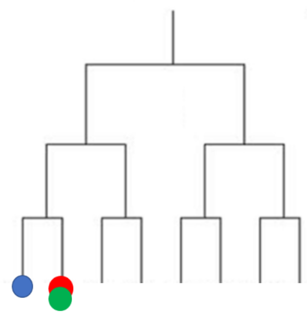
Bevis.

$$\begin{aligned}
&\sum_{m=1}^{n} \frac{2^{m-1}}{2^n - 1}((1 - p^{m-1})p^n + p^{m-1}p^{n-1}q) \\
&\quad = \sum_{m=1}^{n} \frac{p^n}{2^n - 1} 2^{m-1} - \sum_{m=1}^{n} \frac{2^{m-1}p^{m-1}p^n}{2^n - 1} + \sum_{m=1}^{n} \frac{2^{m-1}p^{m-1}p^{n-1}q}{2^n - 1} \\
&\quad = \frac{p^n}{2^n - 1} \sum_{m=1}^{n} 2^{m-1} - \frac{p^n}{2^n - 1} \sum_{m=1}^{n} 2^{m-1}p^{m-1} + \frac{p^{n-1}q}{2^n - 1} \sum_{m=1}^{n} 2^{m-1}p^{m-1} \\
&\quad = \frac{p^n}{2^n - 1}(2^n - 1) - \frac{p^n}{2^n - 1} \frac{2^n p^n - 1}{2p - 1} + \frac{p^{n-1}(1-p)}{2^n - 1} \frac{2^n p^n - 1}{2p - 1} \\
&\quad = \frac{1}{2^n - 1}(p^{n-1}(1 + (2^n - 1)p - 2^n p^n)) \\
&\quad = \frac{1}{2^n - 1}(p^{n-1} + (2^n - 1)p^n - 2^n p^{2n-1}).
\end{aligned}$$

□

4.2.1 Fler än 1 bättre spelare

Om vi ska reda på spelare 3, $P(3, n)$, måste vi ta hänsyn till både spelare 1 och spelare 2 i turneringen. Det blir fler och fler spelare som måste tas till hänsyn ju lägre rankat en spelare man är intresserad av är.



Figur 6: Då det finns fler än en bättre spelare.

Här kan antingen spelare *röd* eller spelare *grön* (som är andra bästa spelare i detta fall) möta spelare *blå* som är rankat 3:a, i rond 1. Det som kan konstateras, är att sannolikheten att möta en bättre spelare i rond 1 är dubbelt så stort om det är två bättre spelare än en. Om man däremot är rankad som sämst (2^n), möter man garanterat alltid en starkare motståndare i varje rond oavsett position.

5 Mycket stort antal spelare i turneringen

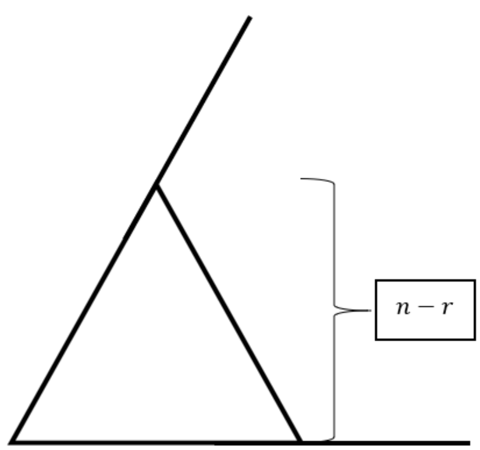
Detta avsnitt handlar om en turnering där det finns ett mycket stort antal spelare ($n \to \infty$). Det vi kunde se från graferna i avsnitt 4.1.1. på sidan 12, är en trend där topparna går mot höger och tenderar att konvergera mot 1. Denna sats bygger på just detta.

Sats 1. *Anta att sannolikheten* $p = p(n)$ *beror på* n *på det viset att* $\lim_{n\to\infty} p^n = a \in [0, 1]$. *Då har vi, för varje fixt* k*:*

$$\lim_{n\to\infty} P(k, n) = a(1 - a)^{k-1}.$$

Med andra ord, rankningen av den vinnande spelaren följer en geometrisk distribution i gränsvärdet, given att $a \in (0, 1)$ [5].

Bevis. Om $\lim_{n\to\infty} p^n = 1$, då har vi $\lim_{n\to\infty} P(1, n) = \lim_{n\to\infty} p^n = 1$ och således automatiskt $\lim_{n\to\infty} P(k, n) = 0$, för alla $k > 1$. Om $\lim_{n\to\infty} p^n = 0$, då har vi $\lim_{n\to\infty} P(1, n) = \lim_{n\to\infty} p^n = 0$ och således $\lim_{n\to\infty} P(k, n) = 0$ för alla $k > 1$, eftersom $P(k, n) \leq P(1, n)$. Härigenom är fallen $a = 0$ och $a = 1$ avklarade och vi kan anta i det efterföljande att $a \in (0, 1)$.



Figur 7: Figuren visar en del av en simplifierad bracket.

Notera nu att $\lim_{n\to\infty} p^n = a > 0$ implicerar $\lim_{n\to\infty} p = 1$ således också $\lim_{n\to\infty} q = 0$. Under turneringen gång kan tre olika scenarier inträffa. Här förklaras dessa tre scenarier och hur de räknas ut.

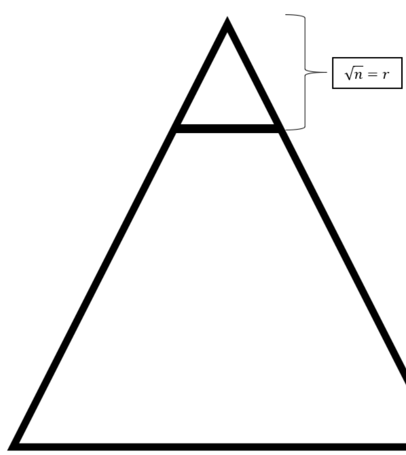
Vi betraktar tre olika scenarier, S_1, S_2 och S_3:

S_1: En av de $k-1$ bästa spelarna tar sig till de sista r ronderna, där $r = \lfloor\sqrt{n}\rfloor$ och där n tenderar mot oändlighet. Sannolikheten att den bästa spelaren, bland alla de spelare som tar sig till rond $n-r+1$, även vinner turneringen är p^r. Efter $n-r$ ronder är inte spelare k den bästa kvarvarande spelaren och därför är sannolikheten att spelare k ska vinna turneringen som mest $1-p^r$. Vi vet att $p^n \to a$ och vi antar exempelvis att $p = a^{\frac{1}{n}}$. Då har vi $p^r = (p^n)^{\frac{r}{n}}$. Då p^n konvergerar mot a, konvergerar exponenten $\frac{r}{n}$ mot 0. Således kommer p^r gå mot 1, vilket leder till att $1-p^r \to 0$.

S_2: Åtminstone två av de topp k spelarna möts någonstans under de första $n-r$ ronderna. Att två topp k spelare skulle möta varandra innan rond r är sannolikheten $\frac{2^{n-r}-1}{2^n-1}$. Där $2^{n-r}-1$ är antal spelare som kan mötas innan rond $n-r$, exklusive spelare k och 2^n-1 är det totala antalet spelare i hela turneringen, förutom spelare k. Vi kan likaså utveckla formeln:

$$\frac{2^{n-r}-1}{2^n-1} = 2^{-r}\frac{2^n-2^r}{2^n-1} \leq 2^{-r}.$$

Därför är sannolikheten att S_2 ska hända inte större än $\binom{k}{2}2^{-r}$. Då $n \to \infty$ härleder $r \to \infty$, vilket slutligen leder till att $\binom{k}{2}2^{-r} \to 0$.



Figur 8: Figuren visar en simplifierad version av en bracket med n-omgångar.

S_3: Att ingen av scenarierna S_1 eller S_2 händer, alltså att inga av de topp $k-1$ spelarna tar sig till de sista r ronderna, genom att bli utslagna tidigare än rond r, eller att inga topp k spelare möter varandra under de första $n-r$ ronderna. Sannolikheten att inga av de topp $k-1$ spelarna tar sig till rond $n-r$ är $(1-p^{n-r})^{k-1}$, och detta går mot $(1-a)^{k-1}$. Vi vet även att $p^{n-r} = \frac{p^n}{p^r}$. Från S_1 vet vi att p^r går mot 1, då går $\frac{p^n}{p^r} \to a$. Därför får vi att $(1-p^{n-r})^{k-1} \to (1-a)^{k-1}$. Sannolikheten att inga topp k spelare möter varandra innan rond r är $1 - \binom{k}{2}2^{-r}$, som går mot 1, då $n \to \infty$, vilket har bevisats i S_2.

Så om S_3 skulle inträffa, kommer spelare k aldrig att spela mot en starkare motståndare genom hela turneringen och kommer därför att vinna med sannolikheten p^n, $\lim_{n\to\infty} p^n = a$.

$$\mathbb{P}(S_3) : (1 - \mathbb{P}(S_2))(1 - p^{n-r})^{k-1}.$$

Låt X vara händelsen att spelare k vinner turneringen. Med hjälp av satsen om total sannolikhet har vi [6]:

$$P(k,n) = \mathbb{P}(X) = \mathbb{P}(X \mid S_1)\mathbb{P}(S_1) + \mathbb{P}(X \mid S_2 \setminus S_1)\mathbb{P}(S_2 \setminus S_1) + \mathbb{P}(X \mid S_3)\mathbb{P}(S_3).$$

Genom argumenten ovan: $\mathbb{P}(X \mid S_1) \to 0$ där $n \to \infty$, är den första termen försumbar:

$$\mathbb{P}(X \mid S_1) \leq 1 - p^r \to 0, n \to \infty.$$

$$\mathbb{P}(S_1) \leq 1.$$

På liknande sätt:

$$\mathbb{P}(X \mid S_2 \setminus S_1) \leq 1.$$

$$\mathbb{P}(S_2 \setminus S_1) \leq \binom{k}{2} 2^{-r} \to 0, n \to \infty.$$

Slutligen:

$$\lim_{n\to\infty} \mathbb{P}(S_3) = (1-a)^{k-1}.$$

$$\lim_{n\to\infty} \mathbb{P}(X \mid S_3) = \lim_{n\to\infty} p^n = a.$$

Vi drar slutsatsen att:

$$\lim_{n\to\infty} P(k,n) = a(1-a)^{k-1}.$$

□

Vi vet att a måste vara en konstant. I så fall måste p^n konvergera mot en konstant. Om p^n går mot a, är p approximativt $a^{\frac{1}{n}}$, vilket går närmare och närmare mot 1 ju mer n växer, då exponenten går mot 0. Det kan vi se i graferna i avsnitt 4.1.1. på sidan 12, är att ju högre n-värdet är, desto mer åt 1 konvergerar topparna på kurvorna. Därför kan vi alltså dra slutsatsen att värdet på p närmar sig 1 ju högre n-värdet är.

Det här resultatet kan användas för att bevisa en annan sats om den maximala sannolikheten att spelare k vinner.

Följdsats 1.1. *Vi har* $\lim_{n\to\infty} \max_p P(k,n) = \frac{(k-1)^{k-1}}{k^k}$ *för varje bestämd* k*.*

Bevis. Notera att maximum av $a(1-a)^{k-1}$ är $\frac{(k-1)^{k-1}}{k^k}$, som är uppnådd för $a = \frac{1}{k}$. Så om vi har $p = p(n) = k^{-\frac{1}{n}}$, då $p^n \to \frac{1}{k}$ och konsekvent $\lim_{n\to\infty} P(k,n) = \frac{(k-1)^{k-1}}{k^k}$.

Vi kan då dra slutsatsen att:

$$\liminf_{n\to\infty} \max_p P(k,n) \geq \frac{(k-1)^{k-1}}{k^k}.$$

Å andra sidan, låt n_r vara någon ökande sekvens vilket $\lim_{r\to\infty} \max_p P(k,n_r) = \limsup_{n\to\infty} \max_p P(k,n)$. Låt p_r vara den korresponderande sekvensen av värden för p där maxima $\max_p P(k,n_r)$ är uppnådd. Vi kan anta att $p_r^{n_r}$ tenderar till en gräns (om inte, kan vi då välja en undersekvens för vilket det här är fallet) $\lim_{r\to\infty} p_r^{n_r} = a$. Med hjälp av satsen ovan kan vi finna att:

$$\limsup_{n\to\infty} \max_p P(k,n) = a(1-a)^{k-1} \leq \frac{(k-1)^{k-1}}{k^k}.$$

Genom att kombinera de två olikheterna följer det önskade resultatet [4]! $\square$

6 Diskussion och slutsats

Det vi kan rakt av säga om sannolikheter och turneringar är att den högre rankade spelaren alltid har en större chans att vinna över den lägre rankade spelaren. Vi kan även med hjälp av en formel kunna ta reda på vad sannolikheten att vinna för respektive spelare i turneringen är. Då måste vi ta hänsyn till hur många bättre spelare det finns i turneringen och framförallt i sin egna halva. Ju lägre rankat spelare k är, desto fler högre rankade spelare måste man ta hänsyn till. Det som vore det mest ideala är att alla de bättre spelarna finns på den andra halvan. Det gör att sannolikheten att möta en av de bättre spelarna är mindre, då de kan slå ut varandra på vägen mot final. Istället för att möta bättre spelare tre-fyra gånger på vägen till final, så kanske du bara möter en bättre spelare i finalen.

Det som också kan påpekas är att det egentligen inte spelar någon roll om det finns ett mycket stort antal spelare i turneringen. Toppspelarna kommer i vilket fall som helst kunna slå ut många av de lågt rankade spelarna.

Det finns även något som borde uppmärksammas, och det är att sannolikheten p inte är konstant under hela turneringen. I denna modell som tagits upp i detta arbete är sannolikheten p en konstant då vi räknar med att den starkare spelaren vinner med just p över vilken sämre motståndare som helst. Men om det vore i verkligheten så är inte sannolikheten en konstant. I modellen så vinner spelare 1 över spelare 2 med sannolikheten p. Men spelare 1 vinner tillika över spelare 100 med sannolikheten p, vilket i verligheten inte skulle vara helt sant. Då spelare 2 är avsevärt mycket bättre än spelare 100, borde p vara högre när spelare 1 möter spelare 100 än mot spelare 2. Den här modellen tar bara hänsyn till om spelaren är bättre eller sämre, inte hur mycket bättre eller hur mycket sämre.

Vad har vi då för likheter och skillnader mellan formlerna som tagits upp i detta arbete? Vi har nämnt två formler som handlar om sannolikheter i utslagningsturneringar. Den ena formeln är den generella och den andra är för turneringar med ett mycket stort antal spelare. De likheter jag ser med formlerna är att man kan använda de på ett enkelt sätt för att ta reda på sannolikheter för olika spelares chanser att vinna turneringar. Det man behöver göra är att lägga in de kända k- och n-värdena i formlerna och då får man det förväntade svaret. Nackdelarna med formlerna är ganska många. Jag har tidigare i detta avsnitt nämnt att p inte är en konstant vilket gör att sannolikheten p kan vara allt mellan 0 och 1, som gör det ganska omöjligt att räkna den riktiga sannolikheten. I avsnittet nedan, påpekar jag andra aspekter som skulle kunna ha möjliga nackdelar på modellen.

Vilken av formlerna är då lämpligast att använda och till vilket tillfälle? Denna fråga kan ställas då vi vet hur många deltagare det finns i turneringen. Om det inte är ett stort antal deltagare så fungerar den generella modellen allra bäst, men den blir svår att använda sig av om det är ett stort antal spelare, vilket vi använder oss av den formeln i avsnitt 5. Slutligen, finns det andra modeller som inte tagits upp som går att använda i liknande sammanhang? Då detta område är påhittat ganska nyligen finns fortfarande många saker som inte har upptäckts eller hittats på ännu, vilket gör att denna fråga är svår att besvara. Då jag endast haft ca två månader att skriva detta arbete har jag inte fått all information om detta ämne. Det jag har skrivit om i detta arbete är just nu det viktigaste och det hittills upptäckta. I framtiden kanske det dyker upp mer information. Forskningen pågår i hösta grad fortfarande vid detta tillfälle.

6.1 Relevanta aspekter

Det finns en aspekt som borde noteras, vilket är att alla spelare som nämns i detta arbete är fullt friska, har inga skador och har inga åkommor. Då vi inte tar hänsyn till detta, betraktas alla spelare som fullt speldugliga. Om man skulle ta hänsyn till skador och besvär skulle möjligheterna att räkna på sannolikheter vara helt av omöjliga. Det vi endast är intresserade av är sannolikheter i turneringar där allt är perfekt.

En aspekt man skulle kunna ta upp, om vi tar ett exempel inom tennis. Wimbledon är en av de största händelserna inom tennis. Där spelar 256 spelare om att vinna turneringen. De senaste tio åren har fyra olika spelare vunnit turneringen. Dessa fyra spelare klassas som de allra bästa inom sporten tennis. Om vi skulle utöka Wimbledon till att innehålla över 1000000 spelare. Då skulle dessa fyra spelare vara likaså de allra bästa och troligtvis vinna turneringen. Varför då? Dessa fyra spelare är bland de 256 bästa i hela världen. Om det skulle komma till 1000000 fler spelare, då skulle ändå dessa 256 spelare vara just de 256 bästa i hela världen och därför skulle världsrankningen inte alls ändras så dramatiskt. Visserligen kan det dyka upp specialfall där nya spelare slår ut några toppspelare, men det skulle inte påverka topp fyra spelarna avsevärt mycket. Detta arbete tar likaledes ingen hänsyn till spelarnas tidigare prestationer.

En tredje aspekt som kan nämnas är att den här studien utgår från tidigare matematisk forskning vid Uppsala univeristet där deras arbeten inte är officiellt utgivna. Då detta examensarbete pågick, under våren 2019, är arbetet med utslagningsturneringar fortfarande i full gång och studeras fortfarande på universtetet mellan flera forskare. Om ett par år eller senare kan arbetet som pågår i detta ämne officiellt publiceras och då kan mycket av det som tas upp i detta examensarbete vara annorlunda eller ändrat.

6.2 Seedningsmetoden

Detta avsnitt har jag bara gått igenom ytligt under dessa två månader och inte haft tillräckligt med tid att skriva om. Det skulle ha blivit ett helt avsnitt, men för tillfället får detta vara en del av diskussionen.

Seedning innebär inom sport/turneringssammanhang att placera i rangordning efter skicklighet/styrka i syfte att undvika att de bästa deltagarna slår ut varandra i ett tidigt stadium. Alltså får de högre seedade spelarna ett privilegium i att komma så långt som möjligt utan större motstånd.

Spelarna i turneringen kan även delas in i olika seedningsgrupper beroende på hur starka dem är. Jämnstarka spelare förväntas hamna i samma grupp eller i två olika grupper som inte är långt ifrån varandra styrkemässigt. Seedningsgrupperna skulle kunna se ut på olika sätt, till exempel (1),(2),(3,4),(5,6,7,8) eller (1,2),(3,4,5,6),(7,8,...) osv. Seedning brukar ta hänsyn till spelarens tidigare prestationer. Ju fler poäng man tagit eller ju bättre man tidigare presterat, desto högre upp i seedningen hamnar man.

I dessa fall skulle spelarna i respektive seedningsgrupp ha ungefär lika stora chanser att vinna över en annan spelare i samma seedningsgrupp. Likaså skulle alla spelare ha lika stora chanser att slå en annan spelare från en annan seedningsgrupp, då spelarna som är grupperade i samma seedningsgrupp är jämnstarka och borde således ha lika stor vinstmöjlighet som de andra i gruppen.

Ett exempel. En spelare i seedningsgrupp a slår en spelare från seedningsgrupp b med sannolikheten $\frac{1}{2}x^{a-b}$, där $x \geq 1$. Det finns fler möjliga modeller om seedningar som inte tas upp i detta arbete.

Referenser

[1] Vandermonde identity proof. http://www.cut-the-knot.org/arithmetic/algebra/VandermondeConvolution.shtml, 1996. [Online].

[2] How to prove vandermonde identity. https://math.stackexchange.com/questions/337923/how-to-prove-vandermondes-identity-sum-k-0n-binomrk-binommn-k?noredirect=1&lq=1, 2013. [Online].

[3] Victor J. Katz. *A History of Mathematics: An Introduction*. Addison-Wesley, 3rd edition, 1993, 2009.

[4] Svante Janson; Stephan Wagner och Cecilia Holmgren. Knock-out tournaments. [Uppsala university, Department of Mathematics, Unpublished].

[5] Gunnar Blom; Jan Enger; Gunnar Englund; Jan Grandell och Lars Holst. *Sannolikhetsteori och statistikteori med tillämpningar*. Lund: Studentlitteratur, 7th edition, 2017.

[6] Sven Erick Alm och Tom Britton. *Stokastik*. Stockholm: Liber, 1st edition, 2008, 2011.