

SZTOCHASZTIKUS FOLYAMATOK ÉS VÉLETLEN GRÁFOK

Szakdolgozat

Írta:

Faller Beáta

alkalmazott matematikus hallgató

Témavezető:

Dr. Móri Tamás egyetemi docens

Valószínűségelméleti és Statisztika Tanszék



Eötvös Loránd Tudományegyetem Természettudományi Kar

2005

Tartalomjegyzék

1. Bevezetés	1
2. Devroye általános véletlen famodellje	4
3. Véletlen rekurzív fák néhány modellje	7
3.1. A véletlen rekurzív fa definíciója	7
3.2. Szymański modellje	8
3.3. A Pittel-féle famodell	10
4. Fokszámarányos véletlen rekurzív fák	11
4.1. Általános tulajdonságok	12
4.2. Távolságok a fában	14
4.3. A levélszám aszimptotikája	22
4.4. Adott kifokú csúcsok számának aszimptotikus viselkedése	24
4.5. A levelek számának egzakt eloszlása	26
5. A fokszámarányos fák további vizsgálata Pólya-féle urnamodellek felhasználásával	28
5.1. A k -edik ág méretének egzakt és aszimptotikus eloszlása	28
5.2. A k -edik ág levélszámának eloszlása	31
5.3. Az ág levélszámának momentumai	34
5.4. Az ág levélszámának aszimptotikája	38
5.5. A 0, 1 és 2 kifokú csúcsok számának együttes aszimptotikus eloszlása	39
6. A legalsó szint méretének vizsgálata	42
6.1. A program működése	43
6.2. A futtatások eredményei	46
Irodalomjegyzék	49

Ábrák jegyzéke

2.1. Két példa véges véletlen hasadófára	6
3.1. 4 csúcsú fokszámarányos véletlen rekurzív fák és valószínűségeik . . .	9
3.2. 4 csúcsú fokszámarányos véletlen rekurzív fák a Pittel által megadott valószínűségekkel	10
4.1. Harmadrendű fák kiterjesztései	12
5.1. Példa a k -edik ág külső csúcsainak színezésére	31
5.2. Példa a külső csúcsok színezésére	39
6.1. U_{800} értékei a hozzájuk tartozó relatív gyakoriságokkal	47
6.2. U_{1000} értékei a hozzájuk tartozó relatív gyakoriságokkal	47
6.3. U_{1500} értékei a hozzájuk tartozó relatív gyakoriságokkal	47
6.4. U_{2000} értékei a hozzájuk tartozó relatív gyakoriságokkal	48
6.5. $U_{800}, U_{1000}, U_{1500}$ és U_{2000} értékei a hozzájuk tartozó relatív gyakorisá- gokkal	48

1. fejezet

Bevezetés

Ez a dolgozat a véletlen gráfok, azon belül a véletlen rekurzív fák egy fokszám-arányosnak nevezett osztályát vizsgálja, kombinatorikai megfontolásokat és a sztochasztikus folyamatok elméletét felhasználva.

A véletlen gráfok elméletét Erdős Pál és Rényi Alfréd alapozta meg az 1960-as évek elején. A véletlen gráf fogalmát a következőképpen definiálták: Legyen $E_{n,N}$ az a halmaz, mely az összes olyan irányítatlan, egyszerű gráfot tartalmazza, melynek n darab címkézett csúcsa V_1, V_2, \dots, V_n , éleinek száma pedig N . Ha egy gráf eleme az $E_{n,N}$ halmaznak, akkor a gráf élhalmaza az $\binom{n}{2}$ lehetséges él közül választott N darab él, így $E_{n,N}$ elemeinek száma $\binom{\binom{n}{2}}{N}$. Egy véletlen $\Gamma_{n,N}$ gráf egy az $E_{n,N}$ halmazból véletlenszerűen választott gráf, ahol minden $E_{n,N}$ -beli elemet azonos, $1 / \binom{\binom{n}{2}}{N}$ valószínűséggel választhatunk. Erdős és Rényi egy véletlen gráf fejlődését a következő sztochasztikus folyamattal írták le: Legyen adott a csúcsok V_1, V_2, \dots, V_n halmaza. A $t = 1$ időpontban választunk egyet az $\binom{n}{2}$ lehetséges él közül, ahol minden egyes élet azonos valószínűséggel választhatunk. Jelölje a választott élet e_1 . A $t = 2$ időpontban választunk egyet az e_1 -től különböző, $\binom{n}{2} - 1$ lehetséges él közül (minden élet azonos valószínűséggel választhatunk). Folytatva az eljárást, a $t = k + 1$ időpontban választunk egyet az e_1, e_2, \dots, e_k -től különböző, $\binom{n}{2} - k$ lehetséges él közül, ahol minden egyes élet azonos, $\frac{1}{\binom{n}{2} - k}$ valószínűséggel választhatunk. Jelölje $\Gamma_{n,N}$ azt a gráfot, melynek csúcsai V_1, V_2, \dots, V_n , élei pedig e_1, e_2, \dots, e_N . A két definíció ugyanazt a véletlen gráfot adja.

Erdős és Rényi munkássága óta a véletlen gráfok elmélete igen népszerűvé vált, a különböző speciális modellekkel foglalkozó irodalom egyre bővül. Az informatika fejlődésével nőtt az érdeklődés a kifejezetten számítógépes alkalmazású modellek iránt is (ilyen például a bináris keresőfa).

Az internet világméretűvé válása nyomán felmerült a kérdés, hogy milyen a struktúrája. Sokan kezdték vizsgálni a World Wide Web, azaz az interneten elérhető oldalak irányított gráfját, melynek éleit az egymásra való hivatkozások határozzák meg. A WWW mellett számos területen találkozunk olyan gráfokkal, melyek struktúrája nehezen írható le. Ilyen például az a genetikai hálózat, melynek csúcsai fehérjék, élei pedig a köztük létrejövő kémiai reakcióknak felelnek meg. Hasonlóan hatalmas hálózatot alkot az idegrendszer, melynek csúcsai az idegsejtek, élei pedig a köztük kapcsolatot létesítő axonok. Érdekes példa még a tudományos publikációk hivatkozási gráfja, melynek egyik csúcsa éppen ez a szakedolgozat. A hálózatok teljes leírása nagy méretük miatt reménytelen, így merül fel a véletlen gráfokkal való modellezés gondolata.

Barabási Albert-László és Albert Réka bevezetett egy modellt, melynek alapja, hogy a gráfhoz folyamatosan új csúcsokat veszünk, és egy új csúcsot egy már meglévőhöz annak fokszámával arányos valószínűséggel kötünk. E modell alap gondolata korábbi irodalmakban is megjelent, melyek hasonlóan épülő fák vizsgálatával foglalkoznak. Ezen irodalmak szerzői főként kombinatorikai eszközökkel elemzik az ún. fokszámarányos fák különböző tulajdonságait. A dolgozat célja ezen legkorábbi, az 1980-as évek vége és az 1990-es évek eleje között született eredmények összefoglalása. Ha az olvasó érdeklődik a későbbi eredmények iránt, ezekről olvashat például [3]-ban.

Hogy lássunk a dolgozatban vizsgált fákon kívül más véletlen famodellre is példát, a 2. fejezetben ismertetünk egyet, mely elég általános, és több fontos véletlen fa előáll speciális eseteként. E modell részletes vizsgálatára nem térünk ki. A 3. fejezetben pontosan definiáljuk a fokszámarányos véletlen rekurzív fák fogalmát. A 4. és 5. fejezetek e fák különböző tulajdonságait vizsgálják. A 6. fejezet egy problémát vet fel, majd egy ehhez kapcsolódó MATLAB programot mutat be néhány eredménnyel.

Jelölések és elnevezések

A dolgozatban Var jelöli a szórásnégyzetet, $\mathbf{1}_{\{A\}}$ az A eseményhez tartozó indikátor valószínűségi változót, \Rightarrow pedig a sztochasztikus konvergenciát. Mint látni fogjuk, a vizsgált famodell gyökeres fákra épül, így használni fogjuk a szülő, gyerek, testvér, illetve szint elnevezéseket a következő értelemben: Ha az r gyökerű T fában az r -ből x -be vezető út utolsó éle (y, x) , akkor y -t az x szülőjének, x -et az y gyerekének nevezzük. Ha két csúcsnak ugyanaz a szülője, akkor testvéreknek mondjuk őket. Az r gyökérből az x csúcsba vezető út hossza az x szintje (a gyökér a fa 0. szintjén helyezkedik el). Több jelölés és elnevezés a későbbi fejezetekben kerül bevezetésre.

Köszönettel tartozom témavezetőmnek, Móri Tamásnak a dolgozat elkészítésében nyújtott segítségéért, és azért, hogy érdekes témát ajánlott, felkeltve az érdeklődésemet a téma további tanulmányozása iránt. Köszönöm nagymamámnak, Kricskovics Istvánnénak az egyetemi éveim és főként a szakdolgozatírás idején nyújtott támogatását, és barátnőmnek, Aradi Ritának, hogy számítógépét felajánlotta a munka elkészítéséhez.

2. fejezet

Devroye általános véletlen famodellje

Devroye bevezetett egy olyan általános véletlen famodellt, mely a véletlen fák számos fontos családját fogja körül. Ide tartoznak például a véletlen bináris keresőfák, a véletlen m -edrendű keresőfák és a véletlen szimplex fák is. Ezeket a speciális modelleket foglalja össze [4], de számos más irodalomban is olvashatunk róluk. Itt csak az általános modell ismertetésére kerül sor.

Egy b -edrendű T_b *skeleton-fa* egy olyan gyökeres végtelen fa, melyben minden csúcson b gyereke van, 1-től b -ig számozva. Egy b -edrendű *hasadófa* egy olyan b -edrendű skeleton-fa, melyben bizonyos csúcsok „golyókat tartalmaznak”, és melyet ennek megfelelően további két paraméter jellemez: $n \geq 0$ egy a fát jellemző kardinális mennyiség; n golyót osztunk szét a csúcsok között (egy csúcson több golyó is juthat, és nyilván végtelen sok csúcson egy golyó sem jut). $s > 0$ a csúcsok *kapacitása*, s -nél több golyót egy csúcs sem tartalmazhat. u -val jelölve egy csúcsot, $N(u)$ jelöli azon golyók számát, melyeket az u -ban gyökerező részfa tartalmaz. $C(u)$ jelöli az u csúcsban levő golyók számát. Egy u csúcsot *levélnek* nevezünk, ha $C(u) = N(u) > 0$. Ezzel ekvivalens a következő feltétel: $C(u) > 0$ és $N(v) = 0$, u minden v gyerekére. Egy u csúcs *lényegtelen*, ha $N(u) = 0$. Egy olyan csúcsot, mely nem levél és nem is lényegtelen csúcs, *belső csúcsonak* nevezünk. Két, azonos b , s , n paraméterekkel rendelkező hasadófa akkor ekvivalens, ha minden u csúcson az $N(u)$ -k megegyeznek. Egy hasadófaból a lényegtelen csúcsok törlésével előálló véges fát *véges hasadófanak* nevezünk.

Mindezek után definiáljuk a *véletlen hasadófat*, melynek paraméterei b , s , s_0 , s_1 , \mathcal{V} és n . A b , s és n mennyiségek rendre egy csúcs gyerekeinek számát, a csúcskapacitást és a golyók számát jelölik, mint a hasadófák esetében. A golyóeloszlás folyamatának leírásához szükségesek az s_0 és s_1 paraméterek, melyek nemnegatív

egészek, és kielégítik a

$$0 < s, \quad 0 \leq s_0 \leq s, \quad 0 \leq bs_1 \leq s + 1 - s_0$$

egyenlőtlenségeket. Végül, \mathcal{V} valószínűségek egy (V_1, \dots, V_b) véletlen vektora, melyre

$$\sum_i V_i = 1, \quad V_i \geq 0.$$

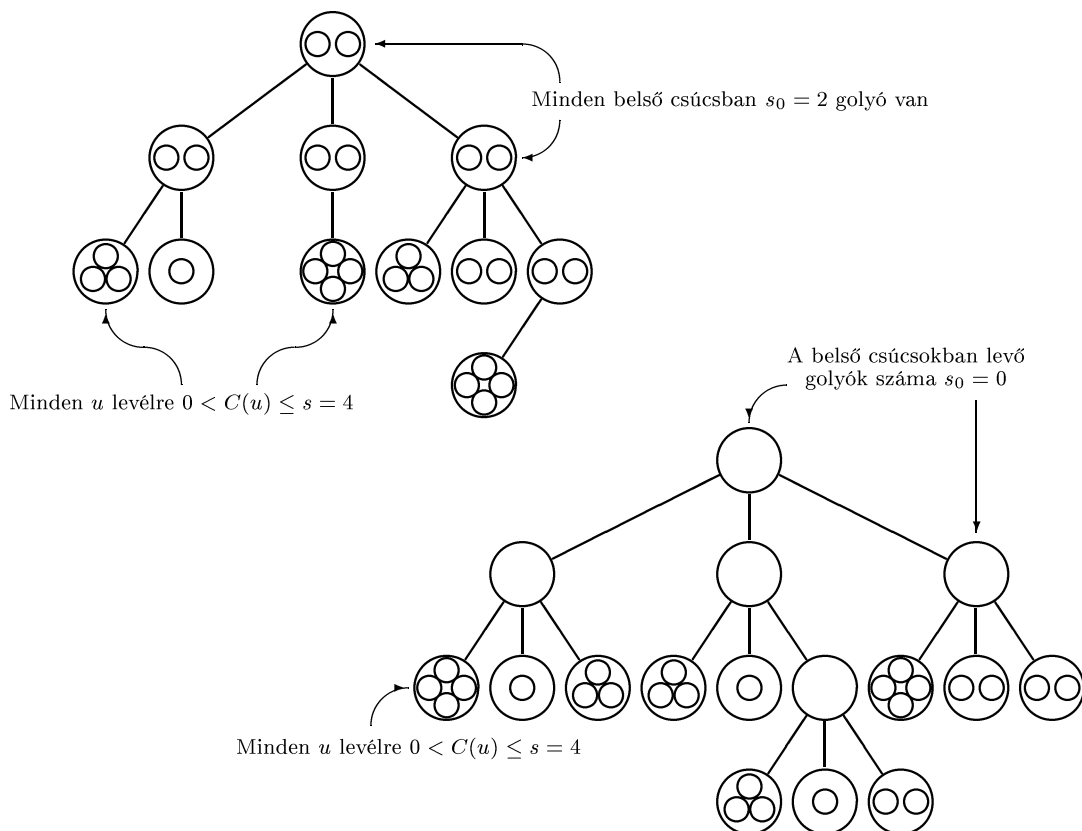
Egy véletlen hasadófa egy olyan T_b szkeleton-fa, melyben az összes u csúcshoz hozzárendeljük \mathcal{V} -nek egy-egy független realizációját, és melyben a csúcsokon n golyó oszlik el az alábbiakban leírt módon, figyelembevételével, hogy minden csúcs kapacitása s . Ez az eloszlás az n golyó egyesével történő folyamatos csatlakoztatásával jön létre. Kezdetben nincs golyó a fában, azaz $C(u) = 0$ minden u csúcsra. Az u -ban gyökerező fához adva egy golyót, a következőképpen járunk el. Legyen (V_1, \dots, V_b) \mathcal{V} -nek az u -hoz tartozó realizációja.

1. Ha u belső csúcs (láttni fogjuk, hogy ekkor $C(u) = s_0$), válasszunk u gyerekei közül egyet. A j -edik gyereket V_j valószínűséggel választhatjuk. Ha az i -edik gyereket választottuk, növeljük $N(u)$ -t 1-gyel, és rekurzív módon adjuk a golyót az i -ben gyökerező részfához.
2. Ha u egy levél és $C(u) = N(u) < s$, adjuk a golyót u -hoz, növeljük $N(u)$ és $C(u)$ mindegyikét 1-gyel, és kész vagyunk.
3. Ha u egy levél és $C(u) = N(u) = s$, a csúcs kapacitását kihasználtuk, így az új golyó nem maradhat u -ban. Ebben az esetben legyen $N(u) = s + 1$ és $C(u) = s_0$. Helyezzünk el $s_0 \leq s$ véletlenszerűen választott golyót u -ban, és továbbítsuk a maradék $s + 1 - s_0$ golyót u gyerekeinek a következőképpen: először adjunk minden gyereknek s_1 darab véletlenszerűen választott golyót, és ennek megfelelően állítsuk be a gyerekek golyószámláló függvényeinek értékeit. A további $s + 1 - s_0 - bs_1$ golyó mindegyike számára válasszunk egymástól függetlenül egy-egy gyereket az u -hoz tartozó (V_1, \dots, V_b) vektornak megfelelően, majd alkalmazzuk az „adjuk a golyót a választott gyerekekben gyökerező fához” eljárást.

Vegyük észre, hogy e 3. pontban leírt lépés egy golyó csatlakoztatásánál akár többször is megismétlődhet, ha $s_0 = 0$, amennyiben mind az $s+1$ golyó egyetlen gyerek felé továbbítódik. Ha $s_0 > 0$, egy gyereknek sem juthat az s kapacitásnál több golyó, így csak egyszer hajtható végre a 3. eset művelete.

4. Ha u egy lényegtelen csúcs, adjuk a golyót u -hoz, növeljük $N(u)$ és $C(u)$ mindegyikét 1-gyel, és kész vagyunk.

Vegyük észre, hogy belső u csúcsok esetében $C(u) = s_0$ áll fenn, míg minden v levélre $0 < C(v) \leq s$. Egy $b, s, s_0, s_1, \mathcal{V}$ és n paraméterű *véges véletlen hasadófa* egy ugyanilyen paraméterű véletlen hasadófából áll elő a lényegtelen csúcsok elhagyásával. A fent leírt golyóeloszlásnak számtalan előnye van. A legfontosabb az, hogy e modell sok olyan speciális fával hozható kapcsolatba, melyek gyakran előfordulnak természetes adatszerkezetként. Az s, s_0 és s_1 paramétereknek különböző értékeket adva speciális esetként sokféle modell előáll. A 2.1. ábra két véges véletlen hasadófát ábrázol. Az első fa paraméterei $(b, s, s_0, s_1, n) = (3, 4, 2, 0, 27)$, a második fa pedig $(b, s, s_0, s_1, n) = (3, 4, 0, 1, 26)$ paraméterű.



2.1. ábra. Két példa véges véletlen hasadófára

3. fejezet

Véletlen rekurzív fák néhány modellje

3.1. A véletlen rekurzív fa definíciója

Legyen R egy olyan n csúcsú fa, melyben a csúcsok 1-től n -ig vannak címkézve. Az R fát *rekurzív fának* nevezzük, ha

1. a fa *gyökeres*, azaz az 1-es címkével ellátott csúcsnak kitüntetett szerepe van a többihez képest; ezt a csúcsot a fa *gyökerének* vagy *gyökércsúcsának* nevezzük,
2. továbbá minden $k = 2, \dots, n$ -re, a gyökérből a k -as címkéjű csúcsba vezető egyértelmű út által bejárt csúcsok címkéi $\{1, 2, \dots, n\}$ -nek egy növekvő részsorozatát alkotják.

Két n csúcsú rekurzív fát akkor tekintünk azonosnak, ha minden $k = 2, \dots, n$ -re a definíció 2. pontjában említett részsorozatok a két fában megegyeznek. Jegyezzük meg, hogy a gráfelméletben két gráfot akkor tekintünk azonosnak, ha azok *izomorfak*. Mint látjuk, rekurzív fák esetében viszont két izomorf fát gyakran különbözőnek tekintünk.

Egy ilyen fa előáll csúcsok egymást követő csatlakoztatásával, ahol az i -edik csúcs a már meglévő $i - 1$ csúcs valamelyikéhez csatlakozik ($i = 2, 3, \dots, n$). Egy n csúcsú *véletlen rekurzív fa* egy, az összes n csúcsú rekurzív fák családjából véletlenszerűen választott fa, ahol a különböző fák választásának valószínűségei különbözőek is lehetnek. Egy olyan $[\mathcal{R}_n, \mathcal{A}, P]$ valószínűségi mezőt vizsgálunk, ahol az \mathcal{R}_n eseménytér az összes n pontú rekurzív fából áll. Mivel ezek száma $(n-1)!$, \mathcal{R}_n -nek $(n-1)!$ eleme van. Ezek után a véletlen rekurzív fák egyes modelljei a P valószínűségi mérték pontos

meghatározásával adódnak. $R \in \mathcal{R}_n$ esetén $P(R)$ egyértelműen meghatározott, ha a fa épülése során az egyes csúcsok választásának valószínűségeit ismerjük. Példaként következzen két, lényegében azonos modell. Az elsőt Jerzy Szymański ismertette először egy publikálatlan, majd egy 1987-ben megjelent cikkében [19]. Később e modellt többen tanulmányozták. Második példaként Boris Pittel megközelítését írjuk le röviden [18].

3.2. Szymański modellje

Szymański olyan véletlen rekurzív fákkal foglalkozott, melyekben egy új csúcs csatlakozásakor a már meglévő csúcsok választásának valószínűségei kizárólag ezen csúcsok fokaitól függenek. Ha tehát az $n + 1$ -es címkéjű csúcs csatlakozik egy n csúcsú fához, akkor rögzített d esetén minden egyes d fokú csúcsot p_{nd} valószínűséggel választhatunk ($d = 0, \dots, n - 1$). Itt fel kell tennünk, hogy

$$\sum_{i=1}^n p_{n, D_{ni}(R)} = 1 \quad (3.1)$$

teljesül minden $R \in \mathcal{R}_n$ -re, ahol $D_{ni}(R)$ jelöli az n csúcsú R fa i címkével ellátott csúcsának fokát.

Ez a modell p_{nd} konkretizálásával tovább specializálható. Szymański megfogalmazott több olyan eredményt, melyek p_{nd} tetszőleges, a feltételeknek eleget tevő megválasztása mellett érvényesek, majd két speciális esetet vizsgált részletesebben, melyek közül a korábbi irodalmak csak az első esettel foglalkoztak. E két modell rövid bemutatása következik.

Egyenletes véletlen rekurzív fák

Abban az esetben, ha $p_{nd} = \frac{1}{n}$ a d fokszámtól függetlenül, adódik az *egyenletes véletlen rekurzív fák* modellje. Ekkor minden $R \in \mathcal{R}_n$ -re $P(R) = \frac{1}{(n-1)!}$ teljesül. Ezzel a modellel foglalkozott korábban többek között Na és Rapaport [15], Moon [14], Gastwirth [5], Najock és Heyde [16], valamint Meir és Moon [13].

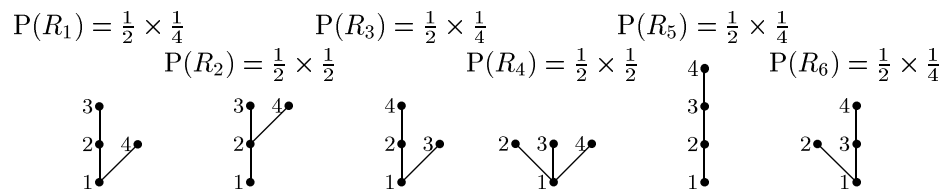
Fokszámarányos véletlen rekurzív fák

Ha $p_{nd} = \frac{d}{2n-2}$, akkor annak valószínűsége, hogy az új csúcs az i címkével ellátott csúcshoz csatlakozik, arányos az i -edik csúcs fokával. Vegyük észre, hogy (3.1) ebben

az esetben is teljesül. Ekkor [19] szerint minden $R \in \mathcal{R}_n$ -re $n \geq 2$ esetén

$$P(R) = \frac{\prod_{i=1}^n (D_{ni}(R) - 1)!}{2^{n-2}(n-2)!}$$

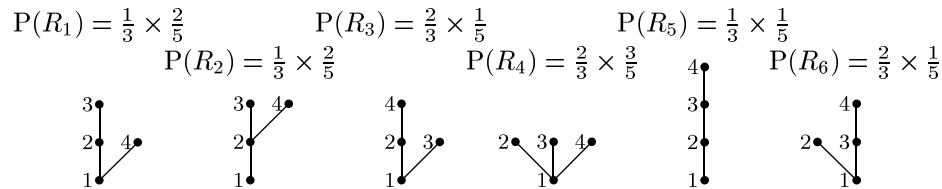
teljesül. Ez lényegében megegyezik azzal a modellel, mellyel a 4. és 5. fejezetek részletesebben foglalkoznak. A két modell szinte teljesen azonos, és mint azt látni fogjuk, mindkét esetben találó a *fokszámarányos véletlen rekurzív fák* modellje elnevezés. A 3.1. ábra példaként szemlélteti $\mathcal{R}_4 = \{R_1, \dots, R_6\}$ elemeit a hozzájuk tartozó valószínűségekkel.



3.1. ábra. 4 csúcsú fokszámarányos véletlen rekurzív fák és valószínűségeik

3.3. A Pittel-féle famodell

E modell csupán annyiban tér el az előzőtől, hogy az új csúcs számára szülőnek választható csúcsok valószínűségeit nem a foksámozatoktól, hanem a gyerekek számától (azaz a kifoktól) teszi függővé. $D(v)$ -vel jelölve a v csúcs gyerekeinek számát, az új csúcs számára tetszőleges v csúcsot $aD(v) + 1$ valószínűségi aránnyal választhatunk szülőnek, ahol a egy nemnegatív konstans. $a = 0$ esetén kapjuk az egyenletes véletlen rekurzív fák modelljét. Az $a = 1$ választás lényegében az előző szakaszban bemutatott második modellhez vezet, ezért ebben az esetben is a foksámarányos véletlen rekurzív fák elnevezést használjuk (mindkét modellben teljesül, hogy tetszőleges nemgyökér v csúcsot $D(v) + 1$ valószínűségi aránnyal választhatunk; az r gyökér viszont Szymański modelljében $D(r)$, Pittel modelljében pedig $D(r) + 1$ valószínűségi aránnyal választható). Kis méretű fák esetén a két foksámarányosnak nevezett famodell közötti eltérés megmutatkozik. A 3.2. ábra $\mathcal{R}_4 = \{R_1, \dots, R_6\}$ elemeit ábrázolja, a hozzájuk tartozó Pittel-féle valószínűségekkal. Hasonlítsuk össze ezt a 3.1. ábrával. A 4. és 5. fejezetek vizsgálatai során Pittel modelljét vesszük alapul.



3.2. ábra. 4 csúcsú foksámarányos véletlen rekurzív fák a Pittel által megadott valószínűségekkal

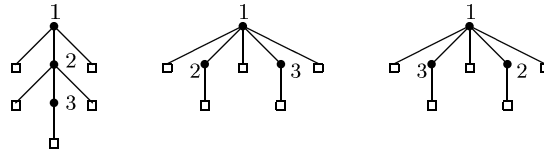
4. fejezet

Fokszámarányos véletlen rekurzív fák

A 3.3. szakaszban definiáltuk a *fokszámarányos véletlen rekurzív fák* fogalmát. Ebben a fejezetben ezek részletesebb vizsgálatára kerül sor. Bevezetünk néhány jelölést és elnevezést, melyekre szükségünk lesz. A fa csúcsainak számát a fa *rendjének* vagy *méretének* nevezzük. Egy n -edrendű fában a k -as címkéjű csúcsból, mint gyökérből induló részfa a *k -adik ág*, jele $T_n^{(k)}$. Az utód nélküli csúcsok a *levelek*, a többi csúcs neve *belső csúcs*. A levelek számát n méretű fában L_n jelöli. Ha a fa éleit úgy irányítjuk, hogy minden él a szülőből az utódba mutat, akkor egy csúcs *kifoka* az ily módon kapott irányított fában a belőle kimutató irányított élek száma. A levelek tehát a 0 kifokú csúcsok. Egy n méretű fában a v csúcs kifokát $D_n(v)$ -vel jelöljük. Az n -es címkével ellátott csúcs gyökértől való távolságát az n -edik csúcs *mélységének* is nevezzük (korábban erre a *szint* elnevezést vezettük be), és D_n -nel jelöljük. Az összes csúcs mélységei közül a maximális a fa *magassága*, melynek jelölésére H_n -t használjuk.

A vizsgált faosztályban egy n méretű T_n fa egy $n - 1$ méretű T_{n-1} fából áll elő úgy, hogy T_{n-1} -ben választunk egy csúcsot (szülőt), és ehhez csatlakoztatjuk az új, n -es címkével ellátott csúcsot. Az egyes csúcsok választásának valószínűsége kizárólag a csúcsok kifokától függ (lásd 3.3. szakasz). A fa növekedésének folyamata másképpen is felfogható: Adott egy $n - 1$ méretű rekurzív fa, ehhez csatlakozik az új csúcs. Ebben egy csúcs gyerekeinek sorrendjét tetszőlegesen változtatva a kapott rekurzív fa az eredetivel azonos marad (lásd 3.1. szakasz). Ha ezeket az eddigiekben azonosnak tekintett fákat különbözőknek tekintjük, akkor korábbi modellünktől eltérően az n -edik csúcs nem $n - 1$ csúcs közül választ különböző valószínűségekkal, hanem, mint később látni fogjuk, $2n - 3$ „rés” közül választ azonos valószínűséggel (tehát az sem mindegy, hogy az n -edik csúcsot a választott szülő korábbi gyerekeinek adott sorrendjébe hova szúrjuk be – tehetjük bármely két gyerek közé, a sor

elejére, vagy végére). Az új csúcs tehát a választott szülő közvetlen utódai közötti bármely beszúrási pozíciót elfoglalhatja, a gyerekek közötti minden egyes rész választása ebben az értelemben más-más fát eredményez. Nyilván ehhez a modellhez más valószínűségi mező tartozik. Ebben az esetben \mathcal{R}_n klasszikus valószínűség algebra. Elemeinek, azaz az n méretű fáknek a száma, mint később látni fogjuk, nem $(n-1)!$, hanem $(2n-3)!!$, melyek mindegyike azonos valószínűséggel adódhat. Célszerű lesz a későbbiekben ezt az értelmezést alapul venni, és két fa különbözőségét ebben az értelemben használni. Ezt a felfogásmódot követve a későbbiekben sokszor megkönnyíti munkánkat, ha a fa ún. *kiterjesztésével* dolgozunk. Ezt úgy kaphatjuk meg, ha az eredeti fa beszúrási pozícióinak jelölésére bevezetünk egy új csúcsosztályt, melynek elemeit *külső csúcsoknak* nevezzük. A 4.1. ábra bemutatja az összes 3 csúcsú fa kiterjesztését. A fa csúcsai a címkézett körök, a külső csúcsokat pedig négyzet jelöli. Ha ezen fák bármelyikét tovább építenénk, a 4. csúcs az 5 külső csúcs egyikét választhatná, mindegyiket $\frac{1}{5}$ valószínűséggel.



4.1. ábra. Harmadrendű fák kiterjesztései

4.1. Általános tulajdonságok

A következő egyszerű lemmák állításaira már utaltunk. Ezeket a későbbiekben gyakran felhasználjuk.

4.1.1. Lemma. *Egy n méretű fokszámárányos véletlen rekurzív fa kiterjesztésében a külső csúcsok száma $2n - 1$.*

Bizonyítás: Ha v egy adott n méretű T_n fa egy csúcsa, akkor T_n kiterjesztésében v -nek $D_n(v) + 1$ közvetlen leszármazottja lesz külső csúcs. Így a kiterjesztett fának

$$\sum_{v \in T_n} (D_n(v) + 1) = 2n - 1$$

külső csúcsa van. ■

4.1.2. Lemma. *Az n méretű fokszámárányos véletlen rekurzív fák száma a csúcsok gyerekeinek különböző sorrendjével előálló fákat különbözőknek tekintve $(2n - 3)!!$.*

Bizonyítás: Az állítást $n \geq 2$ szerinti indukcióval bizonyítjuk. Tegyük fel, hogy az n méretű fák száma $p_n = (2n - 3)!!$. Az előző lemma szerint, ha T_n egy n méretű fa, akkor neki $2n - 1$ beszúrási pozíciója (külső csúcsa) van. Így T_n -ből $2n - 1$ különböző $n + 1$ méretű fa adódhat. Tehát $p_{n+1} = (2n - 1)p_n = (2n - 1)!!$

■

Jelölje $D_n(r)$ egy n méretű véletlen fa gyökerének (ki)fokát. A $\binom{2n}{n}$ binomiális együtthatók meg fognak jelenni néhány későbbi formulában. Aszimptotikus viselkedésüket ismerjük. Az egyszerűbb jelölés érdekében legyen

$$u_n = \frac{4^n}{\binom{2n}{n}} = \prod_{i=1}^n \frac{2i}{2i-1} = \frac{(2n)!!}{(2n-1)!!}.$$

Megmutatható [19], hogy $n \rightarrow \infty$ esetén

$$u_n = \sqrt{\pi n} \left(1 + \frac{1}{8n} + \frac{1}{128n^2} + O\left(\frac{1}{n^3}\right) \right). \quad (4.1)$$

4.1.1. Tétel. *Ha $n \geq 2$, akkor*

$$E[D_n(r)] = u_{n-1} - 1.$$

Bizonyítás: Nézzük először annak valószínűségét, hogy a gyökér kifoka 1:

$$\begin{aligned} P(D_n(r) = 1) &= \frac{2n-5}{2n-3} P(D_{n-1}(r) = 1) \\ &= \frac{2n-5}{2n-3} \times \frac{2n-7}{2n-5} P(D_{n-2}(r) = 1) \\ &\vdots \\ &= \frac{1}{2n-3}; \end{aligned} \quad (4.2)$$

felhasználva, hogy $P(D_2(r) = 1) = 1$. A gyökér nagyobb fokszámának valószínűségére k , $n \geq 2$ esetén

$$\begin{aligned} P(D_{n+1}(r) = k) &= P(D_{n+1}(r) = k | D_n(r) = k) P(D_n(r) = k) \\ &\quad + P(D_{n+1}(r) = k | D_n(r) = k-1) P(D_n(r) = k-1) \\ &= \left(1 - \frac{k+1}{2n-1} \right) P(D_n(r) = k) + \frac{k}{2n-1} P(D_n(r) = k-1) \end{aligned}$$

adódik. Ebből a várható érték definíciója alapján

$$\begin{aligned} E[D_{n+1}(r)] &= \sum_{k=1}^n k \left(\frac{2n-k-2}{2n-1} P(D_n(r) = k) + \frac{k}{2n-1} P(D_n(r) = k-1) \right) \\ &= \frac{1}{2n-1} \left(\sum_{k=1}^n k(2n-k-2) P(D_n(r) = k) + \sum_{k=1}^n (k+1)^2 P(D_n(r) = k) \right) \\ &= \frac{1}{2n-1} \sum_{k=1}^n (2nk+1) P(D_n(r) = k). \end{aligned}$$

Kapjuk tehát az

$$E[D_{n+1}(r)] = \frac{2n}{2n-1} E[D_n(r)] + \frac{1}{2n-1}$$

rekurzív egyenlőséget az $E[D_2(r)] = 1$ kezdeti feltétellel, melynek megoldása

$$E[D_n(r)] = \sum_{k=1}^{n-1} \frac{1}{2k-1} \prod_{i=k+1}^{n-1} \frac{2i}{2i-1};$$

ahol a szorzat értéke 1, ha i lehetséges értékeinek halmaza üres. A jobb oldal felírható az u_j -k segítségével a következőképpen:

$$E[D_n(r)] = u_{n-1} \sum_{j=1}^{n-1} \frac{1}{u_j(2j-1)}.$$

Indukcióval ellenőrizhető a

$$\sum_{j=1}^{n-1} \frac{1}{u_j(2j-1)} = 1 - \frac{2n}{u_n(2n-1)}$$

egyenlőség, melyet helyettesítve adódik a tétel állítása. ■

A 4.1.1. Tétel és (4.1) felhasználásával adódik:

4.1.1. Következmény. *Ha $n \rightarrow \infty$, akkor*

$$E[D_n(r)] \sim \sqrt{\pi n}.$$

Megjegyzés: Egyenletes rekurzív fák esetében a gyökér fokszámának várható értéke aszimptotikusan logaritmikus nagyságrendű [12].

4.2. Távolságok a fában

A fejezet elején bevezettük a D_n jelölést az n -edik csúcs gyökértől való távolságára, azaz n mélységére. Ebben a szakaszban meghatározzuk a gyökértől k távolságra levő külső csúcsok átlagos számát, majd ennek segítségével megadjuk D_n eloszlását, várható értékét és szórásnégyzetét. Megmutatjuk, hogy D_n aszimptotikusan

normális, és *martingálok* alkalmazásával vizsgáljuk a külső csúcsok mélységeinek összegét. Eredményeinkből arra következtethetünk, hogy az n -edik csúcs mélysége aszimptotikusan az egyenletes fákra kapott érték felével egyenlő.

Ahogy a fokszámarányos véletlen rekurzív fa csúcsok folyamatos beszúrásával növekszik, két kumulatív valószínűségi változó jellemzi a fa egyes állapotainak szerkezetét. Ezek egyike a csúcsok mélységeinek összege, melyet a továbbiakban *mélységösszegnek* nevezünk. Ez n méretű fa esetén a következő:

$$N_n = \sum_{j=1}^n D_j.$$

N_n számos tulajdonsága a csúcsok mélységeinek tulajdonságaiból meghatározható. A másik kumulatív változó a külső csúcsok mélységeinek összege, melyet *külső mélységösszegnek* nevezünk, és X_n -nel jelölünk. Ennek felírásához tegyük fel, hogy a külső csúcsok 1-től $2n - 1$ -ig vannak számozva, és a megfelelő mélységek x_1, \dots, x_{2n-1} . Ekkor

$$X_n = \sum_{j=1}^{2n-1} x_j.$$

A távolságok elemzését egy a fa magasságának aszimptotikáját leíró tétellel zárjuk.

Egy szint külső csúcsainak átlagos száma

Jelölje Y_{nk} egy n -edrendű fokszámarányos véletlen rekurzív fában a gyökértől k távolságra levő külső csúcsok számát. Ebben a szakaszban először e véletlen változó várható értékét határozzuk meg. Formulánk tartalmazni fogja az *előjel nélküli elsőfajú Stirling-számokat*. Jelölje $s(n, k)$ azon lehetőségek számát, ahogyan n különböző elemet k ciklusba lehet elrendezni. Az $s(n, k)$ számokat előjel nélküli elsőfajú Stirling-számoknak nevezzük. Bővebben olvashatunk róluk [6]-ban, ahol $\left[\begin{smallmatrix} n \\ k \end{smallmatrix} \right]$ jelöli őket (általánosan elfogadott jelölésük nincs). Nekünk $s(n, k)$ -nak csupán azon tulajdonságára lesz szükségünk, hogy $s(n, k)$ egyenlő az $\langle x \rangle_n = x(x+1) \cdots (x+n-1)$ szorzat x^k tagjának együtthatójával. Használni fogjuk továbbá a $\langle z \rangle_n$ függvény $z = \frac{1}{2}$ helyen vett első és második deriváltját, melyekben szerepelnek az

$$\alpha_n^{(j)} = 1 + \frac{1}{3^j} + \frac{1}{5^j} + \cdots + \frac{1}{(2n-1)^j}, \quad j = 1, 2$$

mennyiségek. Ezeknek a harmonikus számokkal való kapcsolatát fejezi ki az

$$\alpha_n^{(j)} = H_{2n-1}^{(j)} - \frac{1}{2^j} H_{n-1}^{(j)}$$

összefüggés, ahol $H_r^{(j)} = 1 + \frac{1}{2^j} + \dots + \frac{1}{r^j}$. ($j = 1$ esetén szokásos a felső index elhagyása. Az egyszerűbb jelölés érdekében ezt követjük.) Első tételünk bizonyításából közvetlenül adódik majd D_n eloszlása.

4.2.1. Tétel. *Egy n -edrendű fokszámarányos véletlen rekurzív fában a gyökértől k távolságra levő külső csúcsok számának várható értéke*

$$E[Y_{nk}] = \frac{2^{n-k}}{(2n-3)!!} s(n, k), \quad n \geq 2.$$

Bizonyítás: Egy n méretű T_n fa úgy áll elő egy T_{n-1} fából, hogy az n -edik csúcs a D_n -edik szint egy külső csúcsát elfoglalva kapcsolódik a fához. A j -edik szinten való elhelyezkedés tehát a szint egy külső csúcsát eredeti csúcscsá alakítja, miközben ennek jobbján és balján két új külső csúcs megjelenését eredményezi. A $j+1$ -edik szinten az új csúcs egyetlen gyerekeként megjelenik egy további külső csúcs. Ezek alapján Y_{nj} értéke 1-gyel nő. Ha az n -edik csúcs a $j-1$ -es szintet választja, Y_{nj} értéke ugyancsak 1-gyel nő. Más szint választása nem változtatja meg a j -edik szinten levő külső csúcsok számát. Mindebből következik, hogy

$$E[Y_{nj}|D_n = k] = \begin{cases} E[Y_{n-1,j}] + 1, & \text{ha } k = j, \\ E[Y_{n-1,j}] + 1, & \text{ha } k = j - 1, \\ E[Y_{n-1,j}], & \text{különben.} \end{cases}$$

Várható értéket véve adódik

$$E[Y_{nj}] = E[Y_{n-1,j}] + P(D_n = j) + P(D_n = j - 1). \quad (4.3)$$

T_{n-1} minden külső csúcsa egyenlő valószínűséggel választható, és mivel ezek száma $2n-3$, ezért

$$P(D_n = r|Y_{n-1,r}) = \frac{Y_{n-1,r}}{2n-3},$$

amiből pedig

$$P(D_n = r) = \frac{E[Y_{n-1,r}]}{2n-3}. \quad (4.4)$$

Így (4.3) a következőképpen alakul:

$$E[Y_{nj}] = \frac{2n-2}{2n-3} E[Y_{n-1,j}] + \frac{1}{2n-3} E[Y_{n-1,j-1}]. \quad (4.5)$$

E rekurzió megoldásához használjuk fel az $E[Y_{nj}]$ sorozat generátorfüggvényét:

$$L_n(z) = \sum_{j=0}^{\infty} E[Y_{nj}] z^j.$$

Szorozzuk be (4.5) mindkét oldalát z^j -nel, majd összegezzünk j szerint. Ezzel egy rekurziót kapunk $L_n(z)$ -re:

$$L_n(z) = \frac{2n-2+z}{2n-3} L_{n-1}(z).$$

$L_1(z) \equiv z$ figyelembevételével ennek megoldása

$$L_n(z) = \frac{2^n}{(2n-3)!!} \frac{1}{2} z \left(\frac{1}{2}z + 1\right) \cdots \left(\frac{1}{2}z + n - 1\right).$$

Ebben z^k együtthatója megegyezik $E[Y_{nk}]$ értékével. Felhasználva, hogy a jobb oldalon megjelent $\langle \frac{1}{2}z \rangle_n$ -ben $(\frac{1}{2}z)^k$ együtthatója $s(n, k)$, adódik a tétel állítása. ■

Egy csúcs mélységének eloszlása

A 4.2.1. Tétel formuláját (4.4)-be helyettesítve kapjuk a következő állítást.

4.2.2. Tétel.

$$P(D_n = k) = \frac{2^{n-k-1}}{(2n-3)!!} s(n-1, k).$$

Ebből D_n várható értéke közvetlenül adódik:

$$E[D_n] = \frac{2^{n-1}}{(2n-3)!!} \sum_{k=0}^{n-1} \frac{k}{2^k} s(n-1, k).$$

Használjuk most fel, hogy az előjel nélküli elsőfajú Stirling-számok említett tulajdonsága alapján $\langle z \rangle_{n-1} = \sum_{k=0}^{n-1} s(n-1, k) z^k$, és így a $\langle z \rangle_{n-1}$ függvény $z = \frac{1}{2}$ helyen vett első deriváltját kétféleképpen meghatározva kapjuk, hogy

$$\sum_{k=0}^{n-1} s(n-1, k) \frac{k}{2^{k-1}} = \left[1 + \frac{1}{3} + \frac{1}{5} + \cdots + \frac{1}{2n-3} \right] \frac{(2n-3)!!}{2^{n-2}}.$$

Vegyük észre, hogy a jobb oldalon megjelent $\alpha_{n-1}^{(1)}$. Mindezek alapján a várható értékre kapott kifejezés tovább egyszerűsödik:

$$E[D_n] = \alpha_{n-1}^{(1)} = H_{2n-3} - \frac{1}{2} H_{n-2}.$$

Mivel a harmonikus számok aszimptotikus viselkedése ismert (megtalálható [6]-ban), a várható érték aszimptotikája is nagy pontossággal meghatározható:

$$E[D_n] = \frac{1}{2} \ln n + \ln 2 + \frac{1}{2} \gamma + O\left(\frac{1}{n}\right),$$

ahol γ Euler-konstans néven ismeretes, közelítő értéke $\gamma = 0.5772156649 \dots$. Hasonlóan, ezúttal a $\langle z \rangle_{n-1}$ függvény $z = \frac{1}{2}$ helyen vett második deriváltját kétféleképpen meghatározva adódik, hogy

$$\text{Var}[D_n] = \alpha_{n-1}^{(1)} - \alpha_{n-1}^{(2)} = \frac{1}{2} \ln n + \ln 2 + \frac{1}{2}\gamma - \frac{1}{8}\pi^2 + O\left(\frac{1}{n}\right).$$

A Csebisev-egyenlőtlenség felhasználásával igazolható, hogy

$$\frac{D_n}{\ln n} \Rightarrow \frac{1}{2}.$$

A mélységösszeg várható értéke a definíció alapján

$$E[N_n] = \sum_{k=1}^n E[D_k],$$

melyből a jobb oldal ismert várható értékeinek összegzésével kapjuk, hogy

$$E[N_n] = (H_{2n-3} - \frac{1}{2}H_{n-2})(n - \frac{1}{2}) - \frac{1}{2}(n - 1).$$

Ebből $n \rightarrow \infty$ esetén $E[N_n] \sim \frac{1}{2}n \ln n$ adódik.

Egy csúcs mélységének határeloszlása

4.2.3. Tétel. *Definiáljuk a D_n^* normalizált véletlen változót a következőképpen:*

$$D_n^* = \frac{D_n - \frac{1}{2} \ln n}{\sqrt{\frac{1}{2} \ln n}}.$$

Legyen továbbá D egy 0 várható értékű, 1 szórásnégyzetű, normális eloszlású valószínűségi változó. Ekkor D_n^ eloszlásban tart D -hez.*

Bizonyítás: Jelöljük $M_n(t)$ -vel D_n^* momentumgeneráló függvényét. Ekkor

$$M_n(t) = E[e^{D_n^* t}]$$

minden rögzített valós t -re. Az egyszerűség kedvéért vezessük be az $a_n = \frac{1}{2} \ln n$ jelölést. Mivel D_n eloszlását a 4.2.2. Tételben meghatároztuk, kapjuk, hogy

$$\begin{aligned} M_n(t) &= \sum_{k=0}^{\infty} \exp\left\{\frac{k - a_n}{\sqrt{a_n}} t\right\} P(D_n = k) \\ &= \frac{2^{n-1} \exp\{-t\sqrt{a_n}\}}{(2n-3)!!} \sum_{k=0}^{n-1} \left(\frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\}\right)^k s(n-1, k) \\ &= \frac{2^{n-1} \exp\{-t\sqrt{a_n}\}}{(2n-3)!!} \left\langle \frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\} \right\rangle_{n-1} \\ &= \frac{2^{n-1} \exp\{-t\sqrt{a_n}\}}{(2n-3)!!} \times \frac{\Gamma\left(n + \frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\} - 1\right)}{\Gamma\left(\frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\}\right)}. \end{aligned}$$

Itt felhasználtuk a $\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt$, $\alpha > 0$ által definiált *Euler-féle gamma-függvény* azon tulajdonságát, hogy $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$. Ha most azt is kihasználjuk, hogy $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ és $(2n - 3)!! \sim \frac{(2n)!}{2^{n+1}n!n}$, valamint alkalmazzuk a Stirling formulát, azt kapjuk, hogy $n \rightarrow \infty$ esetén

$$M_n(t) \sim \frac{n!4^n n \exp\{-t\sqrt{a_n}\}}{(2n)!\sqrt{\pi}} \Gamma(n) n^{\frac{1}{2}} \exp\{\frac{t}{\sqrt{a_n}}\}^{-1}.$$

Mivel

$$\frac{n!n\Gamma(n)}{(2n)!} = \frac{1}{\binom{2n}{n}} \sim \frac{\sqrt{\pi n}}{4^n},$$

ezért

$$M_n(t) \sim \sqrt{n} \exp\left\{-t\sqrt{a_n} + \left(\frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\} - 1\right) \ln n\right\}.$$

Az $\frac{1}{2} \exp\left\{\frac{t}{\sqrt{a_n}}\right\}$ függvény Taylor-sorba fejtése után adódik, hogy

$$M_n(t) \sim \sqrt{n} \exp\left\{-t\sqrt{a_n} + \left[\left(\frac{1}{2} + \frac{t}{2\sqrt{a_n}} + \frac{t^2}{4a_n} + O\left(\frac{1}{\sqrt{a_n^3}}\right)\right) - 1\right] \ln n\right\} \sim e^{\frac{t^2}{2}}.$$

Mivel egy valószínűségi változó a momentumgeneráló függvénye által egyértelműen meg van határozva, és a jobb oldalon megkaptuk a standard normális eloszlás momentumgeneráló függvényét, a tétel állítása igaz. ■

A külső mélységösszeg

Ebben a szakaszban martingálok segítségével vizsgáljuk a külső csúcsok mélységeinek összegét. Egy n -edrendű T_n fa egy $n - 1$ méretű T_{n-1} fából áll elő az n -edik csúcs D_n -edik szinten való megjelenésével. Mint tudjuk, minden külső csúcs azonos, $\frac{1}{2^{n-3}}$ valószínűséggel választható. Minden egyes beszúrásnál 2 új külső csúcs megjelenésével és 1 elvesztésével 1-gyel nő a választott szint külső csúcsainak száma, és ugyancsak 1-gyel nő a D_n -edik szint alatti szinten levő külső csúcsok száma, az új csúcs gyerekeinek megjelenésével. Így a külső mélységösszeg növekedésének mértéke:

$$2D_n - D_n + (D_n + 1) = 2D_n + 1.$$

Legyen (\mathcal{F}_n) , $n \in \mathbb{N}$ σ -algebrák monoton növekvő családja. Az \mathcal{F}_n σ -algebra reprezentálja azt az információt, amelyet az n -edik beszúrásig összegyűjtöttünk. Az \mathcal{F}_n -be tartozó eseményekről a T_n fa alakja alapján el tudjuk dönteni, hogy bekövetkeztek-e vagy sem. \mathcal{F}_n tehát a T_n fa által generált σ -algebra. Ha a T_{n-1} fát ismerjük, a külső csúcsok $x_1, \dots, x_{2^{n-3}}$ mélységei is adottak. Ezek D_n lehetséges értékei, D_n ezek

mindegyikét $\frac{1}{2n-3}$ valószínűséggel veszi fel. Ezen megfontolások alapján felírhatjuk az

$$E[X_n | \mathcal{F}_{n-1}] = \frac{1}{2n-3} \sum_{j=1}^{2n-3} (X_{n-1} + 2x_j + 1) = X_{n-1} + 1 + \frac{2}{2n-3} \sum_{j=1}^{2n-3} x_j$$

feltételes várható értéket, melyből

$$E[X_n | \mathcal{F}_{n-1}] = \frac{2n-1}{2n-3} X_{n-1} + 1 \quad (4.6)$$

adódik. Várható értéket véve kapjuk az

$$E[X_n] = \frac{2n-1}{2n-3} E[X_{n-1}] + 1 \quad (4.7)$$

rekurziót, melynek megoldása az $E[X_1] = 1$ feltétellel

$$E[X_n] = (2n-1)\alpha_n^{(1)}.$$

Azt kaptuk, hogy $E[X_n] \sim n \ln n$, azaz aszimptotikus viselkedésüket tekintve az átlagos külső mélységösszeg az átlagos mélységösszeg kétszerese.

A következő tétel bizonyítása során használni fogjuk a jól ismert *martingálkonvergencia-tételt*.

4.2.4. Tétel. *Létezik egy olyan X valószínűségi változó, hogy*

$$\frac{X_n - n \ln n}{2n} \rightarrow X$$

1 valószínűséggel és L^2 -ben.

Bizonyítás: Megmutatjuk, hogy

$$Z_n = \frac{X_n - E[X_n]}{2n-1}$$

olyan (\mathcal{F}_n) szerinti martingál, melynek második momentumai egyenletesen korlátozottak. $E[|Z_n|] < \infty$ teljesül minden n -re, minthogy $E[X_n]$ létezik minden n -re. (4.6) és (4.7) alapján adódik

$$E[Z_n | \mathcal{F}_{n-1}] = \frac{X_{n-1}}{2n-3} - \frac{E[X_n] - 1}{2n-1} = Z_{n-1}.$$

Tehát (Z_n) valóban martingál. Utóbbi egyenlőségből $E[Z_1] = 0$ figyelembevételével az is következik, hogy $E[Z_n] = 0$ minden n -re. Z_n második momentumának meghatározásához először egy rekurziót adunk a következőképpen. Helyettesítsük a Z_n -t definiáló képletben X_n -t $X_{n-1} + 2D_n + 1$ -gyel. Ezzel

$$\begin{aligned} Z_n &= \frac{X_{n-1} + 2D_n + 1 - E[X_{n-1} + 2D_n + 1]}{2n-1} \\ &= \frac{2n-3}{2n-1} Z_{n-1} + \frac{2}{2n-1} (D_n - E[D_n]) \end{aligned}$$

adódik. Mindkét oldalt négyzetre emelve, majd várható értéket véve kapjuk, hogy

$$\begin{aligned} \mathbb{E}[Z_n^2] &= \left(\frac{2n-3}{2n-1}\right)^2 \mathbb{E}[Z_{n-1}^2] + \frac{4}{(2n-1)^2} \text{Var}[D_n] \\ &\quad + \frac{4(2n-3)}{(2n-1)^2} \mathbb{E}[Z_{n-1}(D_n - \mathbb{E}[D_n])]. \end{aligned} \quad (4.8)$$

Mivel a jobb oldali $\mathbb{E}[Z_{n-1} \mathbb{E}[D_n]]$ kifejezés értéke nulla, csupán $\mathbb{E}[Z_{n-1} D_n]$ -t kell meghatároznunk. Ennek értéke

$$\mathbb{E}[Z_{n-1} D_n] = \mathbb{E}[\mathbb{E}[Z_{n-1} D_n | \mathcal{F}_{n-1}]] = \mathbb{E}[Z_{n-1} \mathbb{E}[D_n | \mathcal{F}_{n-1}]].$$

A T_{n-1} fa ismeretében D_n várható értéke

$$\mathbb{E}[D_n | \mathcal{F}_{n-1}] = \sum_{j=1}^{2n-3} \frac{x_j}{2n-3} = \frac{X_{n-1}}{2n-3},$$

így

$$\mathbb{E}[Z_{n-1} D_n] = \mathbb{E}[Z_{n-1}^2].$$

Visszahelyettesítve a kapott kifejezést a (4.8) egyenlőségbe kapjuk az

$$\mathbb{E}[Z_n^2] = \frac{(2n-3)(2n+1)}{(2n-1)^2} \mathbb{E}[Z_{n-1}^2] + \frac{4}{(2n-1)^2} \text{Var}[D_n]$$

rekurziót. Vezessük be a

$$Q_n = \frac{(2n-1) \mathbb{E}[Z_n^2]}{2n+1}$$

jelölést. Q_n felhasználásával rekurzióink linearizálható:

$$Q_n = Q_{n-1} + \frac{4}{(2n-1)(2n+2)} \text{Var}[D_n].$$

D_n szórásnégyzetét ismerjük, így e rekurzió megoldásából $\mathbb{E}[Z_n^2]$ értékére kapjuk, hogy

$$\mathbb{E}[Z_n^2] = \frac{4(2n+1)}{2n-1} \sum_{j=2}^n \frac{\alpha_{j-1}^{(1)} - \alpha_{j-1}^{(2)}}{(2j-1)(2j+1)}.$$

A jobb oldal $n \rightarrow \infty$ esetén konvergens, így a második momentumok egyenletesen korlátosak. Z_n -re teljesülnek a martingálkonvergencia-tétel feltételei, ezért Z_n konvergál valamilyen valószínűségi változóhoz 1 valószínűséggel és L^2 -ben is. A tétel állítása ebből következik. ■

A fa magasságának aszimptotikája

4.2.5. Tétel. *A fokszámarányos véletlen rekurzív fák H_n magasságára teljesül, hogy $n \rightarrow \infty$ esetén*

$$\frac{H_n}{\ln n} \rightarrow \frac{1}{2\gamma}$$

1 valószínűséggel, ahol γ a $\gamma e^{\gamma+1} = 1$ egyenlet megoldása.

Megjegyzés: A 4.2.5. Tételre Pittel adott bizonyítást [18]-ban. Korábban Mahmoud egy publikálatlan cikkében belátta, hogy

$$\limsup \frac{H_n}{\ln n} \leq \frac{1}{2\gamma}$$

1 valószínűséggel.

4.3. A levélszám aszimptotikája

L_n -nel jelöltük a fa leveleinek számát. Vizsgáljuk meg e mennyiség várható értékét és szórásnégyzetét.

4.3.1. Tétel. *Egy $n \geq 2$ méretű fokszámarányos véletlen rekurzív fa levelei számának várható értéke és szórásnégyzete*

$$E[L_n] = \frac{2n-1}{3}$$

és

$$\text{Var}[L_n] = \frac{2n(n-2)}{9(2n-3)} = \frac{n}{9} - \frac{1}{18} - \frac{1}{6(2n-1)}.$$

Bizonyítás: Keressünk először rekurziót $E[L_n]$ -re. Ha egy új csúcsot adunk az n -edrendű T_n fához, a levelek száma vagy változatlan marad:

$$P(L_{n+1} = L_n | T_n) = \frac{L_n}{2n-1},$$

vagy 1-gyel nő:

$$P(L_{n+1} = L_n + 1 | T_n) = 1 - \frac{L_n}{2n-1}.$$

Ebből adódik a feltételes várható értékre

$$\begin{aligned} E[L_{n+1} | T_n] &= L_n \times \frac{L_n}{2n-1} + (L_n + 1) \times \left(1 - \frac{L_n}{2n-1}\right) \\ &= \frac{2n-2}{2n-1} L_n + 1, \end{aligned}$$

melyből ismét várható értéket véve kapjuk az

$$E[L_{n+1}] = \frac{2n-2}{2n-1} E[L_n] + 1$$

rekurziót. Ennek megoldása az $E[L_2] = 1$ feltétel felhasználásával adja a tétel első állítását. Hasonlóan,

$$\begin{aligned} E[L_{n+1}^2] &= \frac{2n-3}{2n-1} E[L_n^2] + \frac{4n-3}{2n-1} E[L_n] + 1 \\ &= \frac{2n-3}{2n-1} E[L_n^2] + \frac{4}{3}n, \end{aligned} \quad (4.9)$$

melyből $E[L_3^2] = 3$ felhasználásával

$$E[L_n^2] = \frac{8n^3 - 18n^2 + 10n - 3}{9(2n-3)}.$$

Kapjuk, hogy

$$\text{Var}[L_n] = \frac{2n(n-2)}{9(2n-3)} = \frac{n}{9} - \frac{1}{18} - \frac{1}{6(2n-1)},$$

ahogyan állítottuk. ■

4.3.1. Következmény. *Fokszámárányos véletlen rekurzív fák 1 fokú csúcsainak átlagos száma*

$$\frac{2n-1}{3} + \frac{1}{2n-3}.$$

Bizonyítás: Jelölje az 1 fokú csúcsok számát V_n és legyen

$$I_n = \begin{cases} 1, & \text{ha a gyökér foka 1,} \\ 0, & \text{különben.} \end{cases}$$

Ekkor

$$V_n = L_n + I_n,$$

ezért

$$E[V_n] = E[L_n] + P(D_n(r) = 1).$$

A jobb oldali kifejezések korábban kapott értékeit helyettesítve adódik az állítás. ■

A Csebisev-egyenlőtlenség alkalmazásával igazolható a következő eredmény.

4.3.2. Következmény.

$$\frac{L_n}{n} \Rightarrow \frac{2}{3}.$$

Megjegyzés: Ha összevetjük eredményünket az egyenletes rekurzív fákra vonatkozó eredménnyel, azt látjuk, hogy utóbbiaknak $n \rightarrow \infty$ esetén aszimptotikusan $\frac{n}{2}$ levelük van, ami tehát kevesebb, mint a most kapott érték. A következménybeli konvergencia 1 valószínűséggel is teljesül [2].

4.4. Adott kifokú csúcsok számának aszimptotikus viselkedése

A vizsgált csúcsok kifoka az előbbieken 0 volt. Az általános eset vizsgálatához vezessük be az X_{nj} jelölést a j kifokú csúcsok száma számára. Ekkor X_{n0} és L_n ugyanazt jelölik. Előbb rekurziót adunk $E[X_{nj}]$ -re, majd megoldjuk azt a $j = 1, 2$ értékekre.

4.4.1. Lemma. *Ha $n \geq 2$ és $k < n$, akkor*

$$E[X_{n+1,k}] = \frac{2n-k-2}{2n-1} E[X_{nk}] + \frac{k}{2n-1} E[X_{n,k-1}].$$

Bizonyítás: Ha az n -edrendű T_n fát az $n+1$ -es címkéjű csúcs csatlakoztatásával növeljük, a k kifokú csúcsok száma vagy 1-gyel csökken, vagy változatlan marad, vagy pedig 1-gyel nő. A csökkenés valószínűsége $\frac{(k+1)X_{nk}}{2n-1}$, a növekedése $\frac{kX_{n,k-1}}{2n-1}$. Ezt felhasználva kapjuk az

$$\begin{aligned} E[X_{n+1,k}|T_n] &= (X_{nk} - 1) \times \frac{(k+1)X_{nk}}{2n-1} + X_{nk} \times \left(1 - \frac{(k+1)X_{nk}}{2n-1} - \frac{kX_{n,k-1}}{2n-1}\right) \\ &\quad + (X_{nk} + 1) \times \frac{kX_{n,k-1}}{2n-1} \end{aligned}$$

feltételes várható értéket. Egyszerűsítés után várható értéket véve adódik a kívánt rekurzió. ■

4.4.1. Következmény. *Ha $n \geq 3$, akkor*

$$E[X_{n1}] = \frac{n}{6} - \frac{1}{12} + \frac{3}{4(2n-3)}.$$

Bizonyítás: A 4.4.1. Lemma állítását $k = 1$ -ra alkalmazva, $E[X_{n0}]$ helyébe $E[L_n]$ ismert értékét helyettesítve kapjuk az

$$E[X_{n+1,1}] = \frac{2n-3}{2n-1} E[X_{n1}] + \frac{1}{3}$$

rekurziót, melyet az $E[X_{3,1}] = \frac{2}{3}$ kezdeti feltétellel megoldva adódik a kívánt egyenlőség. ■

4.4.2. Következmény. *Ha $n \geq 3$, akkor*

$$E[X_{n2}] = \frac{n}{15} - \frac{1}{30} + \frac{3}{4n-6}.$$

Bizonyítás: Alkalmazzuk ismét a 4.4.1. Lemmát, most a $k = 2$ értékre. Használjuk a 4.4.1. Következmény eredményét, így az

$$E[X_{n+1,2}] = \frac{2n-4}{2n-1} E[X_{n2}] + \frac{2n^2-4n+6}{3(2n-1)(2n-3)}$$

rekurzióhoz jutunk. $E[X_{4,2}] = \frac{8}{15}$ ismerete mellett az állítás indukcióval bizonyítható. ■

A 4.3.1. Tétel, valamint a 4.4.1. és 4.4.2. Következmény egzakt eredményeket ad $E[X_{nk}]$ értékére $k = 0, 1$ és 2 esetén. Következzen egy állítás, mely $E[X_{nk}]$ aszimptotikus viselkedését minden rögzített $k \geq 0$ esetre megadja.

4.4.1. Tétel. *Rögzített k és $n \rightarrow \infty$ esetén*

$$E[X_{nk}] = \frac{4n-2}{(k+1)(k+2)(k+3)} + O\left(\frac{1}{n}\right).$$

Bizonyítás: Megmutatjuk, hogy az

$$\alpha_{nk} \stackrel{\text{def}}{=} E[X_{nk}] - \frac{4n-2}{(k+1)(k+2)(k+3)}$$

egyenlőséggel definiált kifejezés aszimptotikusan elhanyagolható, pontosabban azt, hogy minden rögzített k -ra $n \rightarrow \infty$ esetén $\alpha_{nk} = O\left(\frac{1}{n}\right)$. Azt kell tehát megmutatnunk, hogy minden k -hoz létezik olyan $c > 0$, mellyel $|\alpha_{nk}| \leq \frac{c}{n}$ teljesül minden elég nagy n -re. A bizonyítást $k \geq 2$ szerinti indukcióba ágyazott n szerinti indukcióval végezzük. $k = 2$ -re igaz az állítás, ahogyan ezt a 4.4.2. Következmény kimondja. Tegyük fel, hogy valamely $k - 1$ -hez ($k \geq 3$) találtunk olyan pozitív c -t, mellyel $|\alpha_{n,k-1}| \leq \frac{c}{n}$ teljesül minden $n \geq k$ -ra. Megmutatjuk, hogy k -hoz is található megfelelő pozitív konstans. Ezt $n \geq k$ szerinti indukcióval végezzük. A 4.4.1. Lemma és α_{nk} értelmezése adja az

$$\alpha_{n+1,k} = \frac{2n-k-2}{2n-1} \alpha_{n,k} + \frac{k}{2n-1} \alpha_{n,k-1} \quad (4.10)$$

rekurziót, melyből a feltevés szerint

$$|\alpha_{n+1,k}| \leq \frac{2n-k-2}{2n-1} |\alpha_{n,k}| + \frac{ck}{n(2n-1)}. \quad (4.11)$$

Legyen A olyan, hogy $A > \max\{2, ck\}$. Ekkor

$$\frac{A(2n-k-2)}{n(2n-1)} + \frac{ck}{n(2n-1)} < \frac{A}{n+1} \quad (4.12)$$

érvényes minden n -re. Definíció alapján

$$\alpha_{kk} = -\frac{4k-2}{(k+1)(k+2)(k+3)},$$

melyből $|\alpha_{kk}| < \frac{2}{k} < \frac{A}{k}$. Felhasználva ezt és (4.12)-t, a (4.11) egyenlőtlenség $n = k$ esetén adja az állítást $n = k + 1$ -re:

$$|\alpha_{k+1,k}| \leq \frac{(k-2)A}{(2k-1)k} + \frac{ck}{k(2k-1)} < \frac{A}{k+1}.$$

Ha feltesszük most, hogy $|\alpha_{nk}| < \frac{A}{n}$ teljesül valamely n -re, ismét (4.11) adja

$$|\alpha_{n+1,k}| < \frac{A}{n+1}$$

érvényességét, így állításunk igaz k esetén is minden $n \geq k$ -ra, tehát minden rögzített k esetén minden elég nagy n -re. ■

Megjegyzés: A 4.1.1. Tételben a gyökér adott fokszámának valószínűségére kapott rekurzió és (4.10) azonos alakúak, így a 4.4.1. Tétel bizonyításának gondolatmenete használható annak igazolására, hogy rögzített k -ra $n \rightarrow \infty$ esetén

$$P(D_n(r) = k) = O\left(\frac{1}{n}\right).$$

4.5. A levelek számának egzakt eloszlása

Megadjuk egy n méretű fa levelei számának eloszlását. Formulánk tartalmazni fogja a *másodfajú Euler-számokat*. Vegyük az $\{1,1,2,2,\dots,n,n\}$ multihalmaz azon permutációit, amelyekben minden $m = 1, 2, \dots, n$ -re az m két előfordulása közötti helyeken csupa m -nél nagyobb szám áll. Az ilyen permutációk közül azok számát, amelyekben az emelkedések száma k (azaz k darab olyan j index van, amelyre $\pi_j < \pi_{j+1}$), jelöljük $\langle\langle n \rangle\rangle_k$ -val. Az $\langle\langle n \rangle\rangle_k$ számokat másodfajú Euler-számoknak nevezzük. Fennáll a következő rekurzív összefüggés:

$$\langle\langle n \rangle\rangle_k = (k+1) \langle\langle n-1 \rangle\rangle_k + (2n-1-k) \langle\langle n-1 \rangle\rangle_{k-1}.$$

A kezdeti feltétel $\langle\langle n \rangle\rangle_0 = 1$, teljesül továbbá az

$$\langle\langle n \rangle\rangle_k = \begin{cases} 1, & \text{ha } n = 0, \\ 0, & \text{különben} \end{cases}$$

egyenlőség. A következő tétel eredményét a későbbiekben felhasználjuk, amikor egy adott ág levelei számának egzakt eloszlását megadjuk. Az így előálló formula igen bonyolult lesz, ezért belőle a részfa levelei számának várható értéke és szórása nehezen meghatározható. Egyszerűbb megoldást a Pólya-féle urnamodellek szolgáltatnak majd.

4.5.1. Tétel. *Az n -edrendű fokszámarányos véletlen rekurzív fák levelei számának egzakt eloszlása*

$$P(L_n = k) = \frac{\left\langle \left\langle \begin{matrix} n-1 \\ k-1 \end{matrix} \right\rangle \right\rangle}{(2n-3)!!}, \quad k = 1, 2, \dots, n.$$

Bizonyítás: Jelölje $t(n, k)$ azon n -edrendű fák számát, melyeknek pontosan k levelük van. Rekurziót adunk $t(n, k)$ -ra, követve Najock és Heyde módszerét [16], akik az egyenletes rekurzív fák vizsgálatánál jártak el hasonlóan. Az $n+1$ -edrendű, k levelet tartalmazó T_{n+1} fa egy n méretű T_n fából a következő két lehetőség egyike szerint áll elő:

- (a) T_n -nek $k-1$ levele van és az $n+1$ -es címkéjű csúcs egy belső csúcsot választ szülőnek, így $n+1$ egy új levél, a T_{n+1} fában a k -adik. Az n méretű fák közül $t(n, k-1)$ azok száma, melyeknek $k-1$ levelük van. Minden ilyen fában $k-1$ külső csúcs kapcsolódik levélhez, így esetünk $((2n-1) - (k-1))t(n, k-1)$ féleképpen adódhat.
- (b) T_n -nek k levele van és az $n+1$ -es címkéjű csúcs ezek egyikét választja szülőnek. T_{n+1} leveleinek száma ugyancsak k . Ez $kt(n, k)$ féleképp fordulhat elő.

Mindebből kapjuk a

$$t(n+1, k) = kt(n, k) + (2n-k)t(n, k-1)$$

rekurziót, mely megegyezik a másodfajú Euler-számok által kielégített rekurzióval. Könnyen látható, hogy a kezdeti feltételek is azonosak, és

$$t(n, k) = \left\langle \left\langle \begin{matrix} n-1 \\ k-1 \end{matrix} \right\rangle \right\rangle.$$

Az összes n -edrendű fa száma a 4.1.2. Lemma szerint valóban $(2n-3)!!$. Ezzel állításunkat igazoltuk. ■

5. fejezet

A fokszámarányos fák további vizsgálata Pólya-féle urnamodellek felhasználásával

Ebben a fejezetben a fokszámarányos véletlen rekurzív fák néhány tulajdonsága és a különböző Pólya-féle urnamodellek összetétele közötti kapcsolatokat vizsgáljuk. Ezek segítségével meghatározzuk a k -adik ág méretének, valamint az ág levélszámának egzakt és aszimptotikus eloszlását, végül a 0, 1 és 2 kifokú csúcsok számának együttes aszimptotikus eloszlását.

5.1. A k -adik ág méretének egzakt és aszimptotikus eloszlása

Az általánosított Pólya-urna k -féle golyót tartalmaz C_1, \dots, C_k színekkel. Kezdetben az urnában levő egyes típusok száma ismert. Húzzunk egy golyót véletlenszerűen (minden egyes urnában levő golyót egyenlő valószínűséggel választhatunk ki), megfigyeljük a színét, és ha ez a szín C_i , akkor a golyót további α_{ij} darab C_j színű golyóval tesszük vissza az urnába ($j = 1, \dots, k$). Ismételjük meg ezt az eljárást n -szer. A legáltalánosabb esetben az α_{ij} -k véletlen változók, melyek pozitívak és negatívak is lehetnek. Mi célunk érdekében a következő modellekben mégis determinisztikus számoknak választjuk őket. Vizsgálandó az urna összetétele n húzás után.

A fokszámarányos rekurzív fák esetében a kiterjesztett fák külső csúcsai felelnek meg az azonos valószínűséggel húzható golyóknak. A k -adik ág (melyre korábban a $T_n^{(k)}$ jelölést vezettük be) méretét jelölje $S_n^{(k)}$. Ennek vizsgálatához a kiterjesztett fa

növekedését a klasszikus Pólya-urnával modellezzük, mely kétféle golyótípust tartalmaz, így a külső csúcsokat két osztályba sorolja a következőképpen: a k -adik ághoz tartozó külső csúcsok felelnek meg a fekete, az ágon kívüliek pedig a fehér golyóknak. A továbbiakban, ha külső csúcs színéről beszélünk, ez a neki megfelelő golyóosztály színét jelenti. A kiterjesztett T_n fa külső csúcsai közül választunk egyet véletlenszerűen, majd átalakítjuk őt egy $n + 1$ -es címkéjű levéllé. Ezáltal két új külső csúcs jelenik meg az új levél jobb és bal oldalán, egy pedig az ő utódjaként. Így összességében kettővel nő a külső csúcsok száma. Nyilvánvaló, hogy a két új külső csúcs színe megegyezik a választott csúcs színével, tehát a visszatevés mátrixa a következő:

$$\begin{array}{c} \text{új golyók} \\ \text{húzott golyó} \end{array} \begin{array}{c} W \\ B \end{array} \begin{array}{c} W \\ B \end{array} \begin{pmatrix} W & B \\ 2 & 0 \\ 0 & 2 \end{pmatrix}. \quad (5.1)$$

Jelölje W_n és B_n az urnában n húzás után levő fehér, illetve fekete golyók számát. Mikor a fa mérete k , külső csúcsainak száma $2k - 1$. Ezek közül k utódja az egyedüli fekete csúcs, a többi mind fehér. Ezt az állapotot tekintjük a Pólya-urna kezdeti összetételének, ahol tehát $W_0 = 2k - 2$ és $B_0 = 1$. Ettől a ponttól kezdve számoljuk a húzásokat, így n húzás után a kiterjesztett fának $n+k$ csúcsa és $2n+2k-1 = W_n+B_n$ külső csúcsa lesz. Utóbbiak közül B_n tartozik az ághoz és W_n van rajta kívül. Jelölje Y_n azt a számot, ahányszor n húzás során fekete golyót kapunk. Y_n 2 paraméterű Pólya-eloszlású valószínűségi változó. Mindahányszor fekete golyót húzunk, 2-vel nő az urnában levő fekete golyók száma, ezért

$$B_n = 2Y_n + 1,$$

mivel kezdetben 1 fekete golyónk volt. Másfelől B_n a $T_{n+k}^{(k)}$ fa külső csúcsainak száma, és mivel e fa mérete $S_{n+k}^{(k)}$, ezért a 4.1.1. Lemma alapján

$$B_n = 2S_{n+k}^{(k)} - 1, \quad (5.2)$$

így

$$S_{n+k}^{(k)} = 1 + Y_n;$$

azaz a k -adik ág méretének eloszlása ugyanolyan, mint egy 1-gyel növelt 2 paraméterű Pólya-eloszlású véletlen változóé. Ebből, vagy közvetlen megfontolással $n \geq k > 1$ -re adódik

$$P(S_n^{(k)} = j) = \frac{(2k-3)!! \binom{n-k}{j-1} (2j-3)!! (2k-2)(2k) \cdots (2n-2j-2)}{(2n-3)!!},$$

azaz

$$P(S_n^{(k)} = j) = \frac{(2j-3)!!(2k-2)(2k)\cdots(2n-2j-2)}{(2k-1)\cdots(2n-3)} \binom{n-k}{j-1},$$

$$j = 1, \dots, n-k+1;$$

ahol a $(2k-2)(2k)\cdots(2n-2j-2)$ szorzat értékét $n=k$, vagy $n>k$ és $j=n-k+1$ esetén 1-nek tekintjük, valamint a $(2j-3)!!$ szemi-faktoriális értéke $j=1$ esetén ugyancsak 1-nek tekintendő. [8, 179. old.] szerint

$$E[S_n^{(k)}] = \frac{n+k-1}{2k-1}$$

$$\sim \frac{n}{2k-1} \quad (5.3)$$

és

$$\text{Var}[S_n^{(k)}] = \text{Var}[Y_{n-k}]$$

$$= \frac{(n-k)(4n-2)(k-1)}{(2k-1)^2(2k+1)}$$

$$\sim \frac{4(k-1)}{(2k-1)^2(2k+1)} n^2. \quad (5.4)$$

Megjegyzés: A fokszámarányos véletlen rekurzív fák k -adik ágának mérete nagy k esetén átlagosan $\frac{n}{2k}$. Ha összehasonlítjuk ezt az eredményt az egyenletes fák esetével, azt látjuk, hogy utóbbiban a k -adik ág átlagos mérete a mi eredményünk kétszerese.

[1, 220. old.] egy tétele kimondja, hogy $k > 1$ esetén

$$\frac{B_n}{B_n + W_n} \rightarrow \beta \left(\frac{B_0}{2}, \frac{W_0}{2} \right) = \beta \left(\frac{1}{2}, k-1 \right)$$

1 valószínűséggel, ahol a jobb oldal egy $(\frac{1}{2}, k-1)$ paraméterű béta-eloszlású valószínűségi változót jelöl ($k=1$ esetén pedig nyilván $\frac{B_n}{B_n+W_n} = 1$). Másfelől, $B_n + W_n$ egy $n+k$ méretű fa külső csúcsainak száma, így $B_n + W_n = 2n + 2k - 1$ a 4.1.1. Lemma alapján. A számláló az (5.2) összefüggés alapján felírható $S_{n+k}^{(k)}$ függvényében, melyből adódik a

$$\frac{2S_{n+k}^{(k)} - 1}{2n + 2k - 1} \rightarrow \beta \left(\frac{1}{2}, k-1 \right)$$

1 valószínűségű konvergencia. Az eddigiek alapján igaz a következő állítás.

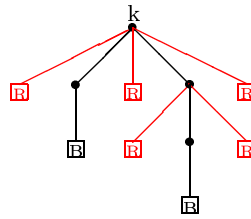
5.1.1. Tétel. *Rögzített k és $n \rightarrow \infty$ esetén az $\frac{S_n^{(k)}}{n}$ véletlen változó 1 valószínűséggel konvergál egy $(\frac{1}{2}, k-1)$ paraméterű béta-eloszlású valószínűségi változóhoz.*

5.2. A k -adik ág levélszámának eloszlása

Vizsgáljuk most a k -adik ág levélszámának eloszlását egy másik urnamodell segítségével, melyben az adott ág külső csúcsait két színnel színezzük (az előző modellben ezek mind feketék voltak). Jelölje $L_n^{(k)}$ a k -adik ág leveleinek számát. Ezúttal az urnában levő golyók színe háromféle lehet, azaz a kiterjesztett fa külső csúcsainak megkülönböztetésére három színt használunk:

1. **Fekete** (B) színnel jelöljük a $T_n^{(k)}$ leveleihez kapcsolódó külső csúcsokat;
2. **Piros** (R) színnel jelöljük a $T_n^{(k)}$ belső csúcsaihoz kapcsolódó külső csúcsokat;
3. **Fehér** (W) színnel pedig a k -adik ágon kívüli külső csúcsokat.

Az 5.1. ábra egy lehetséges részfa ily módon történő színezését ábrázolja (a k -adik ágon kívüli összes külső csúcs, mint mondtuk, fehér).



5.1. ábra. Példa a k -adik ág külső csúcsainak színezésére

Mikor a fa mérete k , külső csúcsainak száma $2k - 1$, melyek mindegyike fehér, kivéve k utódját, ami fekete. Ez legyen a Pólya-urnánk kezdeti összetétele: $2k - 2$ fehér, 1 fekete és 0 piros golyó. Ettől az állapottól számolva a húzásokat, jelölje W_n , B_n és R_n az n húzás után az urnában levő fehér, fekete, illetve piros golyók számát. Ezzel a kezdeti állapot a $W_0 = 2k - 2$, $B_0 = 1$ és $R_0 = 0$ alakban írható. n húzás során a fa csúcsainak száma $n + k$ -ra, külső csúcsainak száma $2n + 2k - 1$ -re nő, melyekből $B_n + R_n$ külső csúcs van a k -adik ágon és W_n darab van rajta kívül ($W_n + B_n + R_n = 2n + 2k - 1$). A kiterjesztett fa növekedésének módját adja meg a következő mátrix:

$$\begin{array}{c} \text{új golyók} \\ W \ B \ R \\ \text{húzott golyó} \end{array} \begin{array}{c} W \\ B \\ R \end{array} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}. \quad (5.5)$$

Vizsgáljuk B_n -t, az n húzás után előálló $n + k$ méretű fa k -adik ágában levő levelek számát.

Jegyezzük meg, hogy $k = 1$ esetén minden csúcs az ághoz tartozik, így modellünk csak két szintet használ annak megkülönböztetésére, hogy egy külső csúcs levélhez kapcsolódik-e vagy sem. Ezt az esetet korábban már vizsgáltuk, most csak a $k \geq 2$ esettel foglalkozunk. Használni fogjuk az előző részben a k -adik ág méretének eloszlására kapott eredményünket. Vezessünk be egy új színosztályt a k -adik ág külső csúcsainak megkülönböztetés nélküli jelölésére. Legyen az új színünk neve kombinált szín (C). Egy külső csúcs tehát akkor kombinált színű, ha vagy piros, vagy fekete. Legyen n húzás után $C_n = B_n + R_n$. Az (5.5) mátrixból kiolvasható, hogy a fekete golyók száma piros golyó húzása esetén 1-gyel nő, míg más szín húzása esetén nem változik. Ezért [11] gondolatmenetét követve kapjuk, hogy:

$$\begin{aligned} P(B_{n+1} = j, C_n = m) &= P(B_{n+1} = j, C_n = m, B_n = j) \\ &\quad + P(B_{n+1} = j, C_n = m, B_n = j - 1) \\ &= P(B_{n+1} = j | C_n = m, B_n = j) P(B_n = j | C_n = m) P(C_n = m) \\ &\quad + P(B_{n+1} = j | C_n = m, B_n = j - 1) P(B_n = j - 1 | C_n = m) P(C_n = m). \end{aligned} \tag{5.6}$$

A $C_n = m$ és $B_n = j$ feltétel mellett $T_{n+k}^{(k)}$ kiterjesztésében új csúcs csatlakoztatása esetén a fekete színű külső csúcsok száma akkor lesz ismét j , ha $T_{n+k}^{(k)}$ -nak vagy fekete, vagy fehér külső csúcsát választjuk. Mivel összesen $2n + 2k - 1$ külső csúcs van, az első feltételes valószínűség (5.6)-ban $\frac{2n+2k-1-(m-j)}{2n+2k-1}$. Ha feltételünk $C_n = m$ és $B_n = j - 1$, akkor annak valószínűsége, hogy a fekete színű külső csúcsok száma j -re nő, megegyezik annak valószínűségével, hogy az $n + k + 1$ -es címkéjű csúcs számára piros színű külső csúcsot választunk. Így ez utóbbi feltételes valószínűség értéke $\frac{m-(j-1)}{2n+2k-1}$. (5.6)-ból ezért azt kapjuk, hogy:

$$\begin{aligned} P(B_{n+1} = j, C_n = m) &= \frac{P(C_n = m)}{2n + 2k - 1} \{ (2n + 2k - 1 - m + j) P(B_n = j | C_n = m) \\ &\quad + (m - j + 1) P(B_n = j - 1 | C_n = m) \}. \end{aligned} \tag{5.7}$$

Az ág minden levelének megfelel az urnában egy fekete golyó, így minden i egészre

$$P(B_n = i | C_n = m) = P(L_{n+k}^{(k)} = i | C_n = m).$$

Ha a fa mérete $n + k$, akkor a k -adik ág mérete $S_{n+k}^{(k)}$, így a 4.1.1. Lemma alapján $C_n = 2S_{n+k}^{(k)} - 1$. Tehát $C_n = m$ esetén m alakja $2p - 1$, valamely $1 \leq p \leq n + 1$

egésszel. Más alakú m -re $P(C_n = m) = 0$ teljesül. Ha még figyelembe vesszük, hogy a részfák az eredeti fa viselkedését „utánozzák”, észrevesszük, hogy $P(B_n = j|C_n = m)$ és $P(B_n = j - 1|C_n = m)$ csupán annak valószínűségei, hogy egy p -edrendű fának j ill. $j - 1$ levele van. Így

$$P(B_n = i|C_n = 2p - 1) = P(L_p^{(1)} = i).$$

A jobb oldal a 4.5.1. Tétel szerint

$$\frac{\langle\langle p-1 \rangle\rangle}{(2p-3)!!}.$$

(5.7) utolsó meghatározandó komponense

$$P(C_n = m) = P(C_n = 2p - 1) = P(S_{n+k}^{(k)} = p).$$

Felhasználva a k -adik ág méretének korábban meghatározott eloszlását (5.7)-ből adódik, hogy

$$\begin{aligned} P(B_{n+1} = j) &= \frac{1}{2n+2k-1} \sum_{p=1}^{n+1} \binom{n}{p-1} \frac{(2k-2) \cdots (2k+2n-2p-2)}{(2k-1) \cdots (2k+2n-3)} \\ &\times \left\{ (2n+2k+j-2p) \langle\langle p-1 \rangle\rangle + (2p-j) \langle\langle p-1 \rangle\rangle \right\}. \end{aligned} \quad (5.8)$$

Eredeti célunk az volt, hogy megkapjuk egy n -edrendű fa k -adik ágában levő levelek számának eloszlását. Foglaljuk össze eddigi eredményeinket:

$$\begin{aligned} L_n^{(k)} &= a \text{ } k\text{-adik ág leveleinek száma, ha a fa rendje } n, \\ &= \text{az urnában levő fekete golyók száma, ha az összes golyó száma } 2n - 1, \\ &= \text{az urnában } n - k \text{ húzás után levő fekete golyók száma.} \end{aligned}$$

Tehát

$$L_n^{(k)} = B_{n-k}, \quad (5.9)$$

így (5.8)-ből adódik a keresett eloszlás:

$$\begin{aligned} P(L_n^{(k)} = j) &= \frac{(2k-3)!!}{(2n-3)!!} \sum_{p=1}^{n-k} \binom{n-k-1}{p-1} (2k-2) \cdots (2n-2p-4) \\ &\times \left\{ (2n+j-2p-2) \langle\langle p-1 \rangle\rangle + (2p-j) \langle\langle p-1 \rangle\rangle \right\}, \\ &j = 1, \dots, n-k; \end{aligned} \quad (5.10)$$

ahol $n \geq k \geq 2$. $j = 1$ esetén $\langle\langle p-1 \rangle\rangle = 1$, valamint $n = k$ esetén $\binom{n-k-1}{p-1} = 1$.

5.3. Az ág levélszámának momentumai

Mint már említettük, az előző szakasz végén kapott eloszlás bonyolult képlete helyett a második Pólya-féle urnamodellünk további vizsgálata vezet a momentumok zárt formuláihoz.

Definiáljuk az ω_n , β_n és ρ_n segéd indikátorokat a következőképpen:

$$\omega_n = \begin{cases} 1, & \text{ha az } n\text{-edik húzásnál fehér golyó adódik,} \\ 0, & \text{különben,} \end{cases}$$
$$\beta_n = \begin{cases} 1, & \text{ha az } n\text{-edik húzásnál fekete golyó adódik,} \\ 0, & \text{különben,} \end{cases}$$
$$\rho_n = \begin{cases} 1, & \text{ha az } n\text{-edik húzásnál piros golyó adódik,} \\ 0, & \text{különben.} \end{cases}$$

Az (5.5) mátrixnak megfelelően

$$\begin{aligned} W_{n+1} &= W_n + 2\omega_{n+1}, \\ B_{n+1} &= B_n + \rho_{n+1}, \\ R_{n+1} &= R_n + 2\beta_{n+1} + \rho_{n+1}. \end{aligned}$$

Várható értéket véve kapjuk, hogy

$$\begin{aligned} E[W_{n+1}] &= E[W_n] + 2 E[\omega_{n+1}], \\ E[B_{n+1}] &= E[B_n] + E[\rho_{n+1}], \\ E[R_{n+1}] &= E[R_n] + 2 E[\beta_{n+1}] + E[\rho_{n+1}]. \end{aligned}$$

Az ω_{n+1} indikátor várható értéke az

$$E[\omega_{n+1}|W_n, B_n, R_n] = \frac{W_n}{W_n + B_n + R_n} = \frac{W_n}{2n + 2k - 1}$$

feltételes várható értékből

$$E[\omega_{n+1}] = \frac{E[W_n]}{2n + 2k - 1}.$$

Hasonlóan,

$$E[\beta_{n+1}] = \frac{E[B_n]}{2n + 2k - 1}$$

és

$$E[\rho_{n+1}] = \frac{E[R_n]}{2n + 2k - 1}.$$

Visszahelyettesítve ezeket fenti egyenleteinkbe adódnak az

$$\begin{aligned} E[W_{n+1}] &= \frac{2n+2k+1}{2n+2k-1} E[W_n], \\ E[B_{n+1}] &= E[B_n] + \frac{1}{2n+2k-1} E[R_n], \\ E[R_{n+1}] &= \frac{2n+2k}{2n+2k-1} E[R_n] + \frac{2}{2n+2k-1} E[B_n] \end{aligned}$$

rekurziók, melyek közül az első megoldása $E[W_0] = 2k - 2$ -vel

$$E[W_n] = \frac{2(k-1)(2n+2k-1)}{2k-1} \sim \frac{4(k-1)}{2k-1}n.$$

Ahhoz, hogy a másik két egyenletből $E[B_n]$ -t kifejezzük, használjuk fel, hogy a kombinált színű golyók száma n húzás után $C_n = B_n + R_n = 2S_{n+k}^{(k)} - 1$, és így (5.3)-ból

$$E[B_n] + E[R_n] = 2E[S_{n+k}^{(k)}] - 1 = \frac{2n+2k-1}{2k-1}.$$

Ezzel az egyenlőséggel eliminálhatjuk $E[R_n]$ -t a második rekurzióból, és kapjuk a csupán B_n -eket tartalmazó

$$E[B_{n+1}] = \frac{2n+2k-2}{2n+2k-1} E[B_n] + \frac{1}{2k-1}$$

rekurziót. Ennek megoldása $n \geq 2$ -re figyelembe véve, hogy $E[B_0] = 1$:

$$\begin{aligned} E[B_n] &= \frac{1}{2k-1} + \frac{4^{n-2}}{(2n+2k-3) \binom{2n+2k-4}{n+k-2}} \\ &\quad \times \left\{ 2k \binom{2k}{k} + \frac{4^k}{2k-1} \sum_{i=k}^{n+k-3} \frac{2i+1}{4^i} \binom{2i}{i} \right\}. \end{aligned}$$

Indukcióval igazolható, hogy

$$\sum_{i=0}^{s-1} \frac{2i+1}{4^i} \binom{2i}{i} = \frac{s(2s+1)}{6 \times 4^{s-1}} \binom{2s}{s},$$

így egyszerűsítés után $n \geq 2$ esetére kapjuk, hogy

$$E[B_n] = \frac{1}{2k-1} + \frac{4(n+k-2)}{6(2k-1)} + \frac{4^n \binom{2k}{k} k(k-1)}{6(2k-1)(2n+2k-3) \binom{2n+2k-4}{n+k-2}}.$$

Mivel rögzített k és $n \rightarrow \infty$ mellett $\frac{4^n}{(2n+2k-3) \binom{2n+2k-4}{n+k-2}} = O\left(\frac{1}{\sqrt{n}}\right)$, ezért

$$E[B_n] \sim \frac{2}{3(2k-1)}n.$$

Az $E[W_n] + E[B_n] + E[R_n] = 2n + 2k - 1$ egyenlőségből $E[R_n]$ pontos értéke kifejezhető, mely alapján a piros golyók átlagos számának aszimptotikája

$$E[R_n] \sim \frac{4}{3(2k-1)}n.$$

Az előző szakasz (5.9) összefüggéséből adódik, hogy a k -adik ág leveleinek átlagos száma

$$\begin{aligned} E[L_n^{(k)}] &= \frac{1}{2k-1} + \frac{4(n-2)}{6(2k-1)} + \frac{4^{n-k} \binom{2k}{k} k(k-1)}{6(2k-1)(2n-3) \binom{2n-4}{n-2}}, \quad n \geq 2 \\ &\sim \frac{2n}{3(2k-1)}. \end{aligned} \quad (5.11)$$

Mivel a k -adik ág méretének várható értéke aszimptotikusan $\frac{n}{2k-1}$, nagy méretű fákban várhatóan a részfa csúcsainak $\frac{2}{3}$ -a levél. Éppen ezt az eredményt vártuk, hiszen ez a teljes fára is igaz, és a részfák levélszámának eloszlása megegyezik az eredeti fa levélszámeloszlásával.

Egzakt formulát a részfa levélszámának szórásnégyzetére nem adunk, mert az igen bonyolult lenne (tartalmazná (5.11) négyzetét is). Viszont felhasználva a teljes fa levélszámának a 4.3.1. Tételben meghatározott várható értékét és szórásnégyzetét, valamint azt a tulajdonságot, hogy a részfák és az eredeti fa szerkezete megegyezik, megadhatjuk a k -adik ág levélszámának aszimptotikus varianciáját. Először írjuk fel B_n -re a teljes szórásnégyzet tételét:

$$\text{Var}[B_n] = \text{Var}[E[B_n | S_{n+k}^{(k)}]] + E[\text{Var}[B_n | S_{n+k}^{(k)}]]. \quad (5.12)$$

Alakítsuk (5.12) jobb oldalának első kifejezését a 4.3.1. Tétel állításának felhasználásával a következőképpen:

$$\begin{aligned} E[B_n | S_{n+k}^{(k)}] &= E[B_n \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} | S_{n+k}^{(k)}] + E[B_n (1 - \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}) | S_{n+k}^{(k)}] \\ &= E[B_n | S_{n+k}^{(k)}] \times \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} + (1 - \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}) E[B_n | S_{n+k}^{(k)}] \\ &= E[L_{n+k}^{(k)} | S_{n+k}^{(k)}] \times \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} + (1 - \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}) E[B_n | S_{n+k}^{(k)}] \\ &= \frac{2S_{n+k}^{(k)} - 1}{3} \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} + 1 - \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}. \end{aligned}$$

Az 5.1.1. Tétel alapján az $\mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}$ indikátor valószínűségi változó 1 valószínűséggel 1-hez tart, így (5.4) alapján

$$\text{Var}[E[B_n | S_{n+k}^{(k)}]] \sim \frac{4}{9} \text{Var}[S_{n+k}^{(k)}] \sim \frac{16(k-1)}{9(2k-1)^2(2k+1)}n^2.$$

Most alakítsuk (5.12) jobb oldalának második kifejezését a 4.3.1. Tételt felhasználva:

$$\begin{aligned}
\text{Var}[B_n|S_{n+k}^{(k)}] &= \text{Var}[B_n|S_{n+k}^{(k)}] \times \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} \\
&= \text{Var}[L_{n+k}^{(k)}|S_{n+k}^{(k)}] \times \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} \\
&= \frac{2(S_{n+k}^{(k)})^2 - 4S_{n+k}^{(k)}}{9(2S_{n+k}^{(k)} - 3)} \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}} \\
&= \left(\frac{S_{n+k}^{(k)}}{9} - \frac{1}{18} - \frac{1}{6(2S_{n+k}^{(k)} - 1)} \right) \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}.
\end{aligned}$$

Az 5.1.1. Tétel szerint $\frac{S_{n+k}^{(k)}}{n} \rightarrow \beta\left(\frac{1}{2}, k-1\right)$ 1 valószínűséggel, amiből következik a

$$\frac{\text{Var}[B_n|S_{n+k}^{(k)}]}{n} \rightarrow \frac{1}{9}\beta\left(\frac{1}{2}, k-1\right)$$

1 valószínűségű konvergencia. Ha az $\{S_{n+k}^{(k)} \geq 2\}$ esemény bekövetkezik, akkor a $2S_{n+k}^{(k)} - 3$ véletlen változót $cS_{n+k}^{(k)}$ tetszőlegesen választott $c \in (0, \frac{1}{2})$ -del alulról korlátozza. Azt kapjuk, hogy

$$\frac{\text{Var}[B_n|S_{n+k}^{(k)}]}{n} \leq \frac{2S_{n+k}^{(k)} \mathbf{1}_{\{S_{n+k}^{(k)} \geq 2\}}}{9cn} \leq \frac{2}{9c}, \quad \text{ha } S_{n+k}^{(k)} \leq n+1.$$

A majorálhatóságot felhasználva

$$\begin{aligned}
\mathbb{E}[\text{Var}[B_n|S_{n+k}^{(k)}]] &\sim \frac{1}{9}n \mathbb{E}\left[\beta\left(\frac{1}{2}, k-1\right)\right] \\
&\sim \frac{n}{9(2k-1)}
\end{aligned} \tag{5.13}$$

adódik. Mindezek alapján azt látjuk, hogy $k > 1$ esetén $\text{Var}[B_n]$ nagyságrendjét az (5.12) egyenlőség jobb oldalának első kifejezése határozza meg, azaz

$$\text{Var}[L_n^{(k)}] = \text{Var}[B_{n-k}] \sim \frac{16(k-1)}{9(2k-1)^2(2k+1)}n^2.$$

$k = 1$ -re (5.12) első kifejezése azonosan nulla, a második aszimptotikusan $\frac{n}{9}$ -cel egyenlő. Ezzel visszakaptuk az eredeti fa levélszámának szórásnégyzetére a 4.3.1. Tételben kapott eredményünket. Vegyük észre, hogy a k -adik ág levélszámának szórásnégyzete egy nagyságrenddel nagyobb, mint az eredeti fa levélszámának szórásnégyzete. Ez annak köszönhető, hogy magának az ágméretnek is nagy a varianciája.

5.4. Az ág levélszámának aszimptotikája

Elsőként a teljes fa levélszámának határeloszlását vizsgáljuk. Legyen $k = 1$, és tekintsük az (5.5) mátrixhoz tartozó urnát. Ekkor az első ág levelei és a teljes fa levelei ugyanazok, ezért nincs szükségünk fehér golyókra. Az urna kezdeti összetételét az az állapot adja, mikor a fa mérete $k = 1$. Ekkor $B_0 = 1$ és $R_0 = 0$. Az így kapott urnát Bagchi és Pal vizsgálták [2]. Eredményeikre támaszkodva [12] kimondja, hogy a fa levélszáma kielégíti a centrális határeloszlás-tételt, azaz

$$\frac{3(L_n - \frac{2n}{3})}{\sqrt{n}} \rightarrow \mathcal{N}(0, 1)$$

eloszlásban.

Következő célunk B_n aszimptotikájának meghatározása az általános $k \geq 1$ esetben. Ehhez szükségünk lesz a $B_n + R_n$ összeg és a $2B_n - R_n$ különbség vizsgálatára. Ha a fekete és piros golyókat ismét a C kombinált színnel jelöljük, és használjuk az [1, 220. old.] eredményt, kapjuk, hogy

$$\frac{C_n}{C_n + W_n} \rightarrow \beta\left(\frac{C_0}{2}, \frac{W_0}{2}\right)$$

majdnem mindenütt, azaz

$$\frac{B_n + R_n}{2n + 2k - 1} \rightarrow \beta\left(\frac{1}{2}, k - 1\right) \quad (5.14)$$

1 valószínűséggel. (Nyilván $k = 1$ esetén $\frac{B_n + R_n}{2n + 2k - 1} = 1$.)

A következő tétel $X_n \equiv 2B_n - R_n$ aszimptotikus viselkedését adja rögzített k -ra.

5.4.1. Tétel. *Rögzített $k > 1$ és $n \rightarrow \infty$ esetén*

$$\frac{X_n}{\sqrt{\frac{n}{2}}} \rightarrow Z$$

eloszlásban, ahol a Z valószínűségi változó karakterisztikus függvénye

$$E\left[\exp\left\{-\left(2\beta\left(\frac{1}{2}, k - 1\right)\right) \times \frac{t^2}{2}\right\}\right] = \frac{\Gamma(k - \frac{1}{2})}{(k - 2)!\sqrt{\pi}} \int_0^1 e^{-yt^2} y^{-\frac{1}{2}} (1 - y)^{k-2} dy.$$

5.4.1. Következmény. *Rögzített k és $n \rightarrow \infty$ esetén*

$$\frac{L_n^{(k)}}{n} \Rightarrow \frac{2}{3}\beta\left(\frac{1}{2}, k - 1\right).$$

5.4.2. Következmény. *Rögzített k esetén a k -adik ág belső csúcsainak számára adódik, hogy*

$$\frac{S_n^{(k)} - L_n^{(k)}}{n} \Rightarrow \frac{1}{3}\beta\left(\frac{1}{2}, k - 1\right),$$

ha $n \rightarrow \infty$.

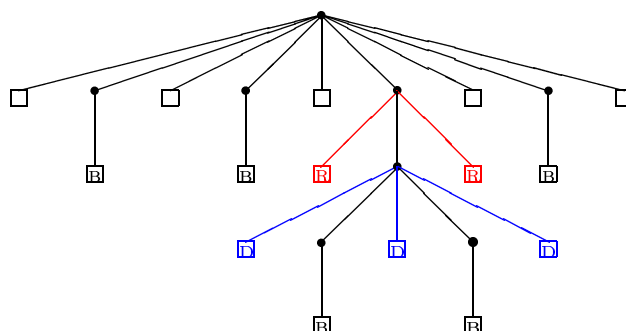
Ezen eredmények bizonyításai megtalálhatók [12]-ben.

5.5. A 0, 1 és 2 kifokú csúcsok számának együttes aszimptotikus eloszlása

Egy újabb urnamoddellel való kapcsolat adja e szakasz eredményeit. Ismét a különböző típusú külső csúcsok felelnek meg a különböző színű golyóknak, de ebben a modellben egy-egy húzás után bizonyos színű golyók száma csökkenhet is. Négy színt használunk:

1. **Fekete** (B) színnel jelöljük azon külső csúcsokat, melyek levélhez kapcsolódnak;
2. **Piros** (R) színnel jelöljük azon külső csúcsokat, melyek 1 kifokú csúcshoz kapcsolódnak (minden 1 kifokú csúcshoz két piros színű külső csúcs kapcsolódik);
3. **Sötétkék** (D) színnel jelöljük azon külső csúcsokat, melyek 2 kifokú csúcshoz kapcsolódnak (minden 2 kifokú csúcshoz három sötétkék színű külső csúcs kapcsolódik);
4. **Fehér** (W) színnel jelöljük az összes többi külső csúcsot (a fehér színű külső csúcsok pontosan azokhoz a csúcsokhoz kapcsolódnak, melyek kifoka ≥ 3).

Az 5.2. ábra egy lehetséges 8 csúcsú fa ilyen módon való színezését mutatja (a fehér színű külső csúcsokat üres négyzet jelöli).



5.2. ábra. Példa a külső csúcsok színezésére

Mint korábbi modelljeinkben, most is egy új csúcs csatlakoztatását, pontosabban egy külső csúcs választását feleltetjük meg egy golyó urnából való húzásának (ezzel eleget téve az urnamoddell azon követelményének, hogy minden golyó azonos valószínűséggel húzható). Jelölje B_n , R_n , D_n és W_n az urnában n húzás után levő fekete,

piros, kék, illetve fehér színű golyók számát. Amikor a fa mérete 1, azaz az egyetlen csúcs a gyökér, akkor egy külső csúcsunk van, aminek a színe fekete. Ez legyen az urna kezdeti összetétele: $B_0 = 1, R_0 = D_0 = W_0 = 0$. $n - 1$ húzás során a csúcsok száma n -re, a külső csúcsok száma pedig $2n - 1$ -re nő. Könnyen ellenőrizhető, hogy a fa épülésének megfelelően a visszatevés mátrixa a következő:

$$\begin{array}{c} B \quad R \quad D \quad W \\ \begin{array}{l} B \\ R \\ D \\ W \end{array} \begin{pmatrix} 0 & 2 & 0 & 0 \\ 1 & -2 & 3 & 0 \\ 1 & 0 & -3 & 4 \\ 1 & 0 & 0 & 1 \end{pmatrix}. \end{array} \quad (5.15)$$

Következzenek azok az eredmények, melyek a most bevezetett urnamodell felhasználásával igazolhatók. Legyen

$$\begin{aligned} X_i &= B_i - \frac{2i+1}{3}; \\ Y_i &= R_i - \frac{2i+1}{6}; \\ Z_i &= D_i - \frac{2i+1}{10}; \end{aligned} \quad (5.16)$$

$i = 0, 1, \dots$. Az alábbi tétel ezeket a jelöléseket használja.

5.5.1. Tétel. *Tekintsük a következő 3-dimenziós valószínűségi vektorváltozót:*

$$\left(\frac{X_i}{\sqrt{i}}, \frac{Y_i}{\sqrt{i}}, \frac{Z_i}{\sqrt{i}} \right).$$

$i \rightarrow \infty$ esetén e vektorváltozó eloszlása egy olyan 3-dimenziós normális eloszláshoz konvergál, melynek várható értéke 0 , kovarianciamátrixa pedig

$$\begin{pmatrix} \frac{1}{9} & -\frac{8}{45} & -\frac{1}{15} \\ -\frac{8}{45} & \frac{23}{45} & -\frac{11}{105} \\ -\frac{1}{15} & -\frac{11}{105} & \frac{179}{350} \end{pmatrix}.$$

5.5.1. Következmény. *Jelölje X_{n0}, X_{n1} és X_{n2} egy n -edrendű foksámarányos véletlen rekurzív fában a 0, 1, illetve 2 kifokú csúcsok számát. Legyen*

$$\begin{aligned} \tilde{X}_{n0} &= X_{n0}; \\ \tilde{X}_{n1} &= X_{n1} - \frac{2n+1}{12}; \\ \tilde{X}_{n2} &= X_{n2} - \frac{2n+1}{30}. \end{aligned}$$

Ekkor $n \rightarrow \infty$ esetén az

$$\left(\frac{\tilde{X}_{n0}}{\sqrt{n}}, \frac{\tilde{X}_{n1}}{\sqrt{n}}, \frac{\tilde{X}_{n2}}{\sqrt{n}} \right)$$

valószínűségi vektorváltozó eloszlása egy olyan 3-dimenziós normális eloszláshoz konvergál, melynek várható értéke $\underline{0}$, kovarianciamátrixa pedig

$$\begin{pmatrix} \frac{1}{9} & -\frac{4}{45} & -\frac{1}{45} \\ -\frac{4}{45} & \frac{23}{180} & -\frac{11}{630} \\ -\frac{1}{45} & -\frac{11}{630} & \frac{179}{3150} \end{pmatrix}.$$

E két utolsó eredmény bizonyítása szintén [12]-ben olvasható.

6. fejezet

A legalsó szint méretének vizsgálata

A 4. fejezet 4.2.5. Tétéle szerint a fokszámárányos véletlen fák H_n magasságára teljesül, hogy $n \rightarrow \infty$ esetén

$$\frac{H_n}{\ln n} \rightarrow \frac{1}{2\gamma} = 1.79556$$

1 valószínűséggel, ahol γ a $\gamma e^{\gamma+1} = 1$ egyenlet megoldása. Móri Tamás felvetett egy ezzel kapcsolatos problémát, melyet itt megfogalmazunk: Jelölje U_n egy n -edrendű fa legalsó szintjének méretét (korábbi jelölésünkkel $U_n = Y_{n, H_n+1}$). Mivel a fa magassága 1-gyel nő, ha a legalsó szinthez kapcsolódó külső csúcsok közül választunk neki pozíciót, és nem változik, ha a többi külső csúcs valamelyikét választjuk, ezért adódik az

$$E[H_{n+1} - H_n | \mathcal{F}_n] = \frac{U_n}{2n - 1}$$

feltételes várható érték. H_n tehát olyan növekvő adaptált sorozat, melynek differenciái egyenletesen korlátosak, ezért [17, 152. old.] szerint

$$H_n \sim \sum_{i=1}^{n-1} \frac{U_i}{2i - 1},$$

amiből kapjuk, hogy

$$\lim_{n \rightarrow \infty} \frac{1}{\ln n} \sum_{i=1}^n \frac{U_i}{i} = \frac{1}{\gamma} = 3.591$$

1 valószínűséggel. Felvetődik a kérdés, vajon van-e az U_n sorozatnak határeloszlása, azaz létezik-e minden k pozitív egészre $\lim_{n \rightarrow \infty} P(U_n = k)$?

Ha az aszimptotikus eloszlás logaritmikus periodicitást mutatna (ami feltehető, mivel n lépés alatt körülbelül $\ln n$ -szer nő a magasság, ezért a szomszédos növekedési idők közötti távolság exponenciálisan nő, és a növekedési időpontokban U_n mindig

visszaáll 1-re, közben pedig monoton növekszik), akkor az a kérdés, hogy van-e U_n -nek ún. 1 valószínűségű határeloszlása, vagyis létezik-e

$$\lim_{n \rightarrow \infty} \frac{1}{\ln n} \sum_{i=1}^n \frac{\mathbf{1}_{\{U_i=k\}}}{i}$$

1 valószínűséggel?

Ez a problémafelvetés ihlette a dolgozathoz megírt program elkészítését. A MATLAB program az U_n sorozatot szimulálja, nem célja a fa egyéb tulajdonságainak vizsgálata, de alkalmazható lenne például a szintek méretének és ezáltal a fa szélességének elemzésére is.

6.1. A program működése

A program „generál” egy f méretű fát, legalsó szintjének csúcsait megszámlolja, majd ezt végrehajtja k -szor. A kapott adatokat egy vektorban tárolja.

1. Input adatok:

- (a) k , az iterációk száma;
- (b) f , a fa mérete.

2. Lehetséges output adatok:

- (a) \mathbf{a} , egy k dimenziós vektor, mely az egyes fák legalsó szintjein levő csúcsok számait, azaz az U_f értékeket tartalmazza;
- (b) $\max \mathbf{a}$, az \mathbf{a} vektor maximális eleme;
- (c) \mathbf{b} , egy $\max \mathbf{a}$ dimenziós vektor, mely \mathbf{a} elemeinek relatív gyakoriságait tartalmazza.

A program egy olyan mátrixot használ az egyes szinteken levő csúcsok számának tárolására, melynek minden eleme 0 vagy 1. A mátrix egyes sorai megfelelnek a fa megfelelő szintjeinek, és pontosan annyi 1-est tartalmaznak, ahány csúcsot a megfelelő szint. Egy két csúcsú véletlen rekurzív fából indulunk ki: a nulladik szinten van a gyökér, az első szinten pedig a 2-es címkéjű csúcs. Ennek az állapotnak megfeleltetjük az

$$\begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

oszlopvektort. Ha már van egy $A \in \{0,1\}^{n \times m}$ mátrixunk, egy új csúcs csatlakoztatásakor eggyel növeljük a mátrixban levő 1-esek számát a következőképpen: Az i -edik ($i = 2, \dots, n + 1$) sorban eggyel nő az 1-esek száma, ha az új csúcs az $(i - 1)$ -edik szintre kerül (a gyökérnek megfelelő első sor a 0. szint). Az $(i - 1)$ -edik szint külső csúcsainak, azaz beszúrási pozícióinak száma éppen annyi, ahány csúcs van az $(i - 2)$ -edik és az $(i - 1)$ -edik szinteken összesen. Ezért az új 1-es i -edik sorba kerülésének valószínűségi aránya megegyezik a mátrix $(i - 1)$ -edik és i -edik soraiban levő elemek összegével. Ezt használja ki a program, elkerülve annak a problémánk szempontjából felesleges információnak a tárolását, hogy a szomszédos szinteket mely élek kötik össze. A mátrixban levő 1-esek számát addig növeljük, míg az el nem éri a fa méretének értékét, l -et. Ezután már egyszerű feladat a mátrix utolsó sorában levő elemek összegzése, melynek eredménye U_f . Egy 9 pontú fához egy új csúcs csatlakoztatása történhet például a következőképpen:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Példánkban az új csúcs $\frac{7}{17}$ valószínűséggel a fa 2. szintjére került, de azt nem tudjuk, és nem is érdekes, hogy e szint $7 = 3 + 4$ beszúrási pozíciója közül melyiket választotta. $f = 10$ esetén már készen is vagyunk: $U_{10} = 1$.

A MATLAB program

Elsőként következik a főprogram, melynek inputját és lehetséges outputjait fent megadtuk. Ez a program a fa nevű alprogram által létrehozott mátrixból számolja ki az outputokat:

```
function b=UtsoSzint(k,f);
for i=1:k
    A=fa(f);
    g=size(A,1);
    a(i)=sum(A(g,:));
end
b=zeros(1,max(a));
for i=1:max(a);
    for j=1:k
```

```

        if a(j)==i
            b(i)=b(i)+1;
        end
    end
end
end
b=b/sum(b);
bar(1:max(a),b,'c');

```

A következő program tartalmazza az algoritmus lényegét; létrehoz egy n pontú fához a fent leírt módon egy mátrixot (inputja a fa mérete, outputja a neki megfelelő mátrix), felhasználva két alprogramot:

```

function e=fa(n);
A=[1 0; 1 0];
for i=1:n-2
    sorszam=size(A,1);
    for j=1:sorszam-1
        v(j)=sum(sum(A(j:j+1,:)));
    end
    v(sorszam)=sum(A(sorszam,:));
    v=rulett(v);
    r=kivalaszt(v)+1;
    if r<sorszam+1
        k=1;
        while A(r,k)>0
            k=k+1;
        end
        A(r,k)=1;
        A(r,k+1)=0;
    else
        A(r,1)=1;
        A(r,2)=0;
    end
end
end
e=A;

```


A következő alprogram inputja egy tetszőleges nemnegatív n dimenziós \underline{v} vektor (melyre teljesül, hogy elemeinek összege pozitív), outputja pedig egy olyan 0 és 1 közötti elemű, ugyancsak n dimenziós \underline{u} vektor, melynek j -edik eleme

$$u_j = \frac{\sum_{i=1}^j v_i}{\sum_{i=1}^n v_i}.$$

```

function e=rulett(v)
s=sum(v);
u(1)=v(1)/s;
  for j=2:length(v)
    u(j)=v(j)/s;
    u(j)=u(j-1)+u(j);
  end
e=u;

```

Végül következzen az az alprogram, melynek inputja az előző, rulett nevű program outputvektorja. E vektor elemeivel felosztva a $(0, 1)$ intervallumot, egy $E(0, 1)$ eloszlású valószínűségi változó egyes intervallumokba esésének valószínűsége adja az új csúcstól választható szintek valószínűségét. Az output az inputvektor egy elemének sorszáma, mely meghatározza a választandó szintet.

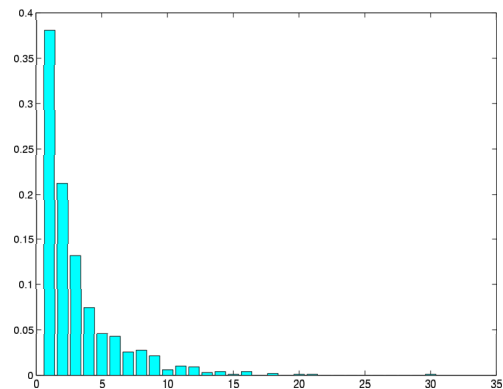
```

function e=kivalaszt(v)
a=rand;
i=1;
while v(i)<a
  i=i+1;
end
e=i;

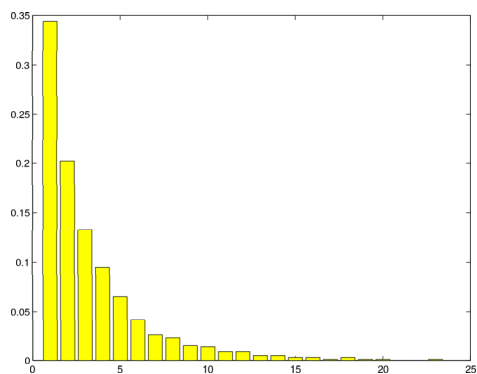
```

6.2. A futtatások eredményei

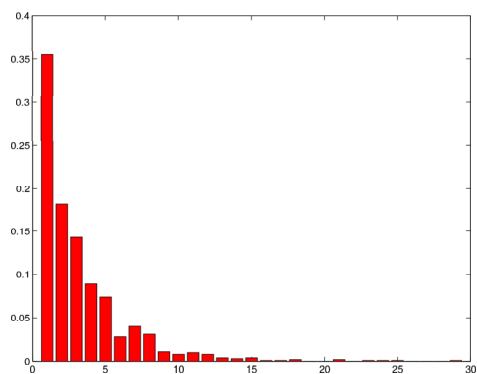
A 6.1. ábra az UtsoSzint(1000, 800), a 6.2. ábra az UtsoSzint(1000, 1000), a 6.3. ábra az UtsoSzint(1000, 1500), a 6.4. ábra pedig az UtsoSzint(1000, 2000) futtatásának eredményét mutatja be, azaz az U_{800} , U_{1000} , U_{1500} illetve U_{2000} mindegyikére kapott 1000 adat között előforduló értékek relatív gyakoriságait ábrázolják:



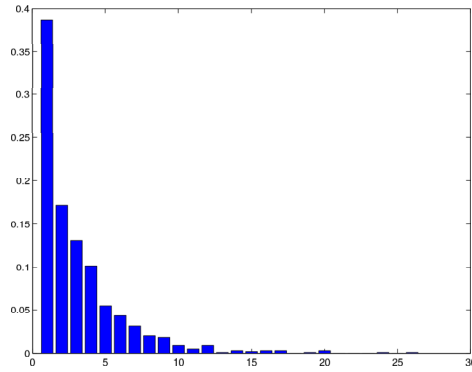
6.1. ábra. U_{800} értékei a hozzájuk tartozó relatív gyakoriságokkal



6.2. ábra. U_{1000} értékei a hozzájuk tartozó relatív gyakoriságokkal

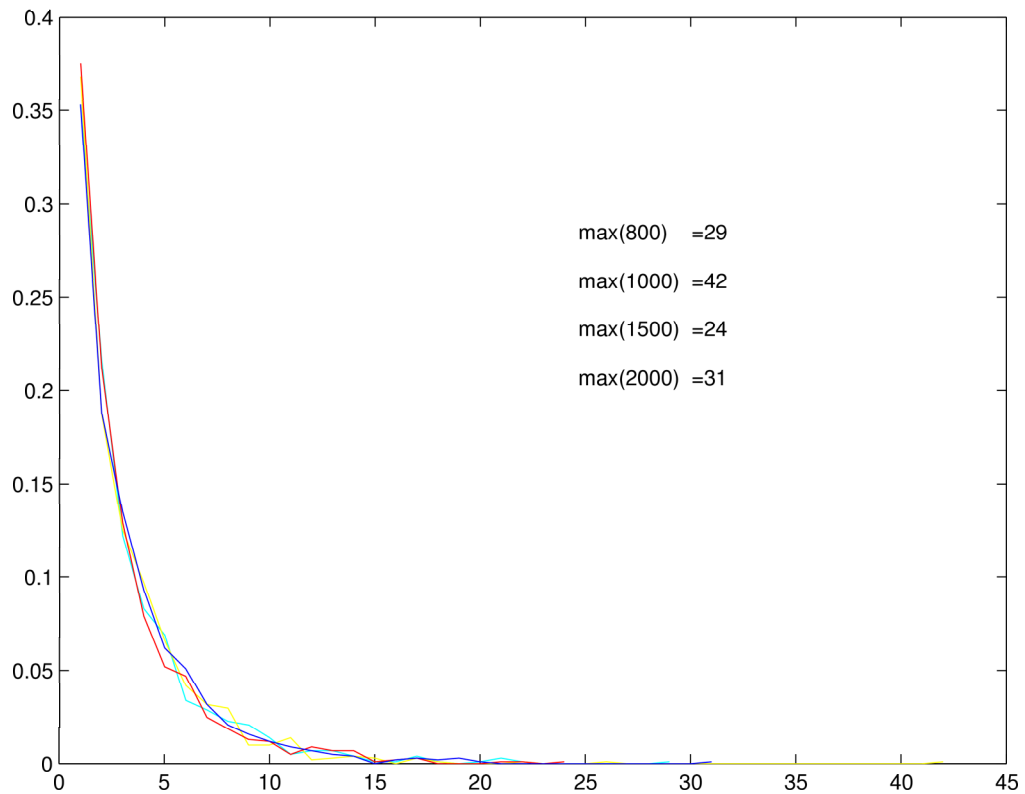


6.3. ábra. U_{1500} értékei a hozzájuk tartozó relatív gyakoriságokkal



6.4. ábra. U_{2000} értékei a hozzájuk tartozó relatív gyakoriságokkal

A 6.5. ábra a fenti eredményeket ábrázolja egy koordinátarendszerben. A kapott eredményekből nyilván nem következtethetünk a $\lim_{n \rightarrow \infty} P(U_n = k)$ aszimptotikus eloszlás logaritmikus periodicitására, de nem tűnik kizártnak e feltételezés.



6.5. ábra. $U_{800}, U_{1000}, U_{1500}$ és U_{2000} értékei a hozzájuk tartozó relatív gyakoriságokkal

Irodalomjegyzék

- [1] ATHREYA, K., NEY, P.: Branching Processes, Springer-Verlag, New York, 1972.
- [2] BAGCHI, A., PAL, A.: Asymptotic normality in the generalized Pólya-Eggenberger urn model, with an application to computer data structures, *SIAM J. Algebr. Discrete Methods*, Vol. 6, 1985.
- [3] COOPER, C., FRIEZE, A.: A general model of web graphs (English summery), *Random Structures Algorithms*, Vol. 22, 2003.
- [4] DEVROYE, L.: Universal limit laws for depths in random trees, *SIAM J. Comput.*, Vol. 28, 1998.
- [5] GASTWIRTH, J.: A probability model of a pyramid scheme, *Am. Stat.*, Vol. 31, 1977.
- [6] GRAHAM, R. L., KNUTH, D. E., PATASHNIK, O.: Konkrét Matematika, Műszaki Könyvkiadó, Budapest, 1998.
- [7] HALL, P., HEYDE, C. C.: Martingale Limit Theory and Its Application, Academic Press, New York, 1980.
- [8] JOHNSON, N., KOTZ, S.: Urn Models and Their Applications, Wiley, New York, 1977.
- [9] KARLIN, S., TAYLOR, H. M.: Sztochasztikus folyamatok, Gondolat, Budapest, 1985.
- [10] MAHMOUD, H.: Distances in plane-oriented recursive trees, *Journal of Computational and Applied Mathematics*, Vol. 41, 1992.
- [11] MAHMOUD, H., SMYTHE, R.: On the distribution of leaves in rooted subtrees of recursive trees, *Ann. Appl. Probab.*, Vol. 1, 1991.

- [12] MAHMOUD, H., SMYTHE, R. T., SZYMAŃSKI, J.: On the structure of random plane-oriented recursive trees and their branches, *Random Structures and Algorithms*, Vol. 4, 1993.
- [13] MEIR, A., MOON, J.: Recursive trees with no nodes of outdegree one, *Congr. Numer.*, Vol. 66, 1974.
- [14] MOON, J.: The distance between nodes in recursive trees, *London Math. Soc. Lecture Notes*, Ser. 13, Cambridge University Press, London, 1974.
- [15] NA, H., RAPAPORT, A.: Distribution of nodes of a tree by degree, *Math. Biosci.*, Vol. 6, 1970.
- [16] NAJOCK, D., HEYDE, C.: On the number of terminal vertices in certain random trees with an application to stemma construction in Philology, *J. Appl. Probab.*, Vol. 19, 1982.
- [17] NEVEU, J.: Discrete Parameter Martingales, North-Holland, Amsterdam, 1975.
- [18] PITTEL, B.: Note on the heights of random recursive trees and random m-ary search trees, *Random Structures and Algorithms*, Vol. 5, 1994.
- [19] SZYMAŃSKI, J.: On a nonuniform random recursive tree, *Annals of Discrete Mathematics*, Vol. 33, 1987.