Eötvös Loránd Tudományegyetem

Természettudományi Kar

Sulyán Tímea

Egy- és többdimenziós Lévy-folyamatok és alkalmazásaik

Szakdolgozat

Alkalmazott matematikus MSc, Sztochasztika szakirány

Témavezető:
Backhausz Ágnes
Valószínűségelméleti és Statisztika Tanszék

Budapest, 2019

Köszönetnyilvánítás

Szeretném megköszönni témavezetőmnek, Backhausz Ágnesnek a hasznos tanácsokat, észrevételeket, valamint köszönöm a végtelennek tűnő türelmét. Ezáltal is jelentősen segítette a szakdolgozatom elkészítését.

Továbbá köszönettel tartozom mindenkinek a környezetemben, aki bármilyen hatással volt rám az elmúlt időszakban, növelve ezzel kitartásom a szakdolgozat megírásához.

Tartalomjegyzék

1. Bevezetés **4**

2. Lévy-folyamatok **7**

2.1. Lévy-folyamatok tulajdonságai 10

2.1.1. Markov-folyamat 10

2.1.2. Önhasonlóság 10

2.2. Példák Lévy-folyamatra 11

3. Filogenetikai fák **15**

3.1. Brown-mozgás a filogenetikai fákon 15

3.2. Lévy-folyamat a filogenetikai fákon 20

4. Többdimenziós folyamatok **24**

5. Összefoglalás **34**

6. Függelék **35**

1. fejezet

Bevezetés

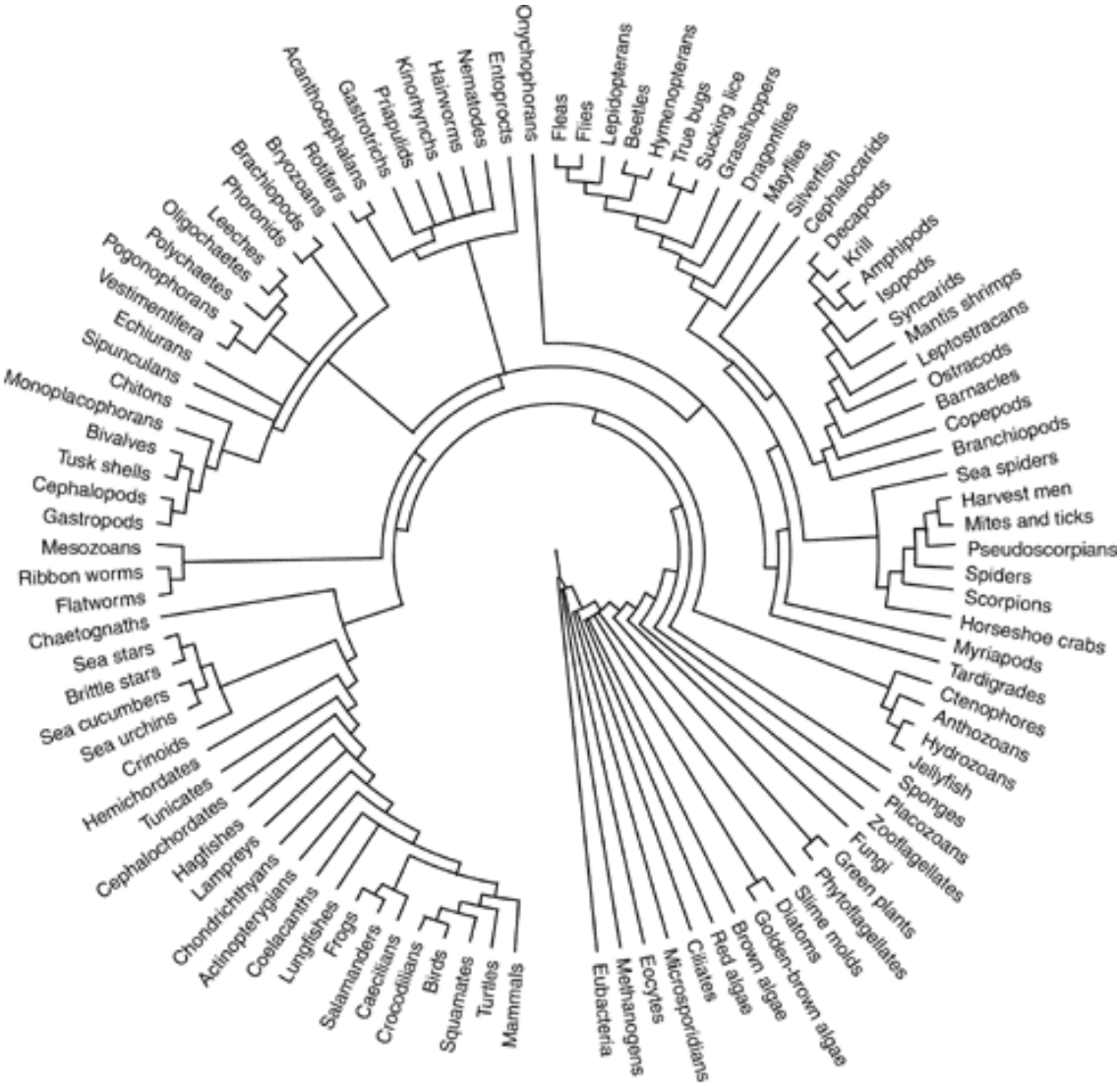
Az ember mindig kíváncsi volt arra, hogy hogyan is keletkezett az élővilágunk, a természet bármely egyedének eredetére, fejlődésének folyamatára. Gondoljunk csak a ma igen népszerű genetikai tesztekre, melyek választ ígérnek az emberek földrajzi, etnikai hovatartozásával kapcsolatban.

Elsőként *Carl von Linné* svéd természettudós, orvos és botanikus az élőlényeket természetes külső jellemzői alapján, illetve mérések mentén csoportosította. Őt tartják a modern biológiai rendszertan atyjának. Azonban volt egy igen fontos dolog, amit ő még nem vett figyelembe, ez pedig az evolúció. Az evolúciós leszármazási kapcsolatokkal foglalkozik a filogenetika. Az evolúció folyamatos, különböző tényezők hatására történik, generációról generációra történik a mutáció, ez alapján rajzolható a filogenetikai fa. Szakdolgozatomban ezt a fát és a rajta fejlődő evolúciós folyamatot mutatom be egy kicsit mélyebben, matematikai megközelítéssel. A fa értelmezése egyértelmű, így arra nem térek ki, ezzel szemben a folyamat egészen elbonyolódhat.

A dolgozat második fejezetében ezeknek a folyamatoknak a definíciójáról és tulajdonságairól olvashatunk, illetve néhány konkrét példát is láthatunk. ([1, 2, 3, 4, 9]) A harmadik fejezetében az előbb említett sztochasztikus folyamatok fejlődését láthatjuk majd a filogenetikai fán. ([6, 7]) Elsőként csak egy Brown-mozgás fejlődik a fán és megfigyelhetjük majd, hogy egyes egyedek között hogyan változik a korreláció a közös fejlődés függvényében. Ezt követően egy Brown-mozgás és egy összetett Poisson-folyamat összegét tekintjük, majd erre ugyanezt a korrelációt vizsgáljuk meg. Megnézzük azt is, hogy mi történik akkor, ha az egyedek jellemzőire egy nem folytonos függvény van hatással. A csoportok szétszakadása – amit leginkább külső tényezők befolyásolnak – okozza a jellemzők nem folytonos fejlődését, ezért lehet

indokolt a Lévy-folyamatok alkalmazása Brown-mozgás helyett. Természetes kérdésként vetődhet fel az, amikor például baktériumok elszaporodnak egy területen, akkor az össztömegük és korrelációjuk, hogyan alakul. Erre a fejezet végén láthatunk majd egy szimulációt. A negyedik fejezetben ugyanezen folyamatok többdimenziós változatával találkozunk és vizsgálataink, szimulációink is hasonlóak lesznek egy nagyobb fára, illetve összevetjük a két folyamat eredményeit. [10] Ennek a fejezetnek a végén egy érdekes algoritmusról olvashatunk, mely a Lévy-folyamat paraméter becsléseire, illetve ugrások detektálására ad egy módszert.

Az általam megírt kódokat, melyek ábráimat állították elő, a [11] segítségével írtam **R**-ben. Ebben található csomagok közül az „ape", „maps" és „phytools" csomagokat használtam, illetve a dokumentációkban található kódokat is segítségül hívtam. Az R kódot a függelék tartalmazza, azonban a hosszúsága miatt csak a legbonyolultabb modellre mellékeltem, ezt átalakítva a többire is adódik.



1.1. ábra. Evolúciós fa baktériumoktól az emlősökig [11]

2. fejezet

Lévy-folyamatok

Ebben a fejezetben arról fogok írni, hogy mi is az a Lévy-folyamat illetve ismertetem a kapcsolatát a korlátlanul osztható eloszlásokkal. Majd láthatunk néhány példát Lévy-folyamatra.[1, 2, 4]

2.1. Definíció. *Az* $(\Omega, \mathcal{F}, \mathbb{P})$ *valószínűségi mezőn értelmezett* $(B_t)_{t\geq 0}$ *folyamatot Brown-mozgásnak nevezzük, ha a következő tulajdonságok teljesülnek:*

(i) $B_0 = 0$,

(ii) független növekményű, azaz $B_{t_1} - B_{t_0}$, $B_{t_2} - B_{t_1}$, $\ldots$,$B_{t_n} - B_{t_{n-1}}$ *függetlenek tetszőleges* $t_0 = 0 < t_1 < t_2 < \ldots < t_n$ $(n \in \mathbb{N})$ *esetén,*

(iii) stacionárius, tehát $B_{t_n+h} - B_{t_{n-1}+h} \stackrel{d}{=} B_{t_n} - B_{t_{n-1}}$, $\forall\ h > 0$

(iv) minden $t > 0$*-ra,* $B_t \sim \mathcal{N}(0, t)$ *teljesül,*

(v) folytonos trajektóriájú.

2.2. Definíció. *Ugyanezen a valószínűségi mezőn értelmezett* $(N_t)_{t\geq 0}$ *folyamatot* λ*-intenzitású Poisson-folyamatnak nevezzük, ha az alábbiak igazak rá:*

(i) $N_0 = 0$,

(ii) független növekményű, azaz $N_{t_1} - N_{t_0}$, $N_{t_2} - N_{t_1}$, $\ldots$,$N_{t_n} - N_{t_{n-1}}$ *függetlenek tetszőleges* $t_0 = 0 < t_1 < t_2 < \ldots < t_n$ $(n \in \mathbb{N})$ *esetén,*

(iii) stacionárius, tehát $N_{t_n+h} - N_{t_{n-1}+h} \stackrel{d}{=} N_{t_n} - N_{t_{n-1}}$, $\forall\ h > 0$

(iv) minden $t > 0$-ra, $N_t \sim Poi(\lambda t)$ teljesül,

(v) jobbról folytonos trajektóriájú, balról pedig létezik határértéke.

Az előbbi két folyamatot már jól ismerjük, tulajdonságaikkal sokszor találkoztunk. E két folyamat közös tulajdonságai alapján definiáltak egy általánosabb folyamatot, a Lévy-folyamatot.

2.3. Definíció (Lévy-folyamat). *A fenti valószínűségi mezőn értelmezett $(L_t)_{t\geq 0}$ folyamatot Lévy-folyamatnak nevezzük, amennyiben ez a négy tulajdonság teljesül:*

(i) $L_0 = 0$,

(ii) független növekményű, azaz $L_{t_1} - L_{t_0}$, $L_{t_2} - L_{t_1}$, $\ldots$,$L_{t_n} - L_{t_{n-1}}$ függetlenek tetszőleges $t_0 = 0 < t_1 < t_2 < \ldots < t_n$ $(n \in \mathbb{N})$ esetén,

(iii) stacionárius, tehát $L_{t_n+h} - L_{t_{n-1}+h} \stackrel{d}{=} L_{t_n} - L_{t_{n-1}}$, $\forall\ h > 0$

(iv) jobbról folytonos trajektóriájú, balról pedig létezik határértéke,

(v) sztochasztikusan folytonos.

Bruno de Finetti 1929-ben bevezette a korlátlanul osztható eloszlásokat és megmutatta a Lévy-folyamatokkal való kapcsolatát. Ez a kapcsolat rávilágít a Lévy-folyamatok családjának változatosságára. A korlátlanul oszthatóság definíciója tetszőleges valószínűségi változóra értelmes, ennek definíciója emlékeztetőül:

2.4. Definíció. *Egy valós értékű ξ valószínűségi változót korlátlanul oszthatónak nevezünk, ha minden $n = 1, 2, \ldots$ esetén létezik egy független, azonos eloszlású valószínűségi változókból álló $\xi_{1,n}, \ldots, \xi_{n,n}$ sorozat, melyre*

$$\xi \stackrel{d}{=} \xi_{1,n} + \cdots + \xi_{n,n}.$$

Egy valószínűségi változó korlátlanul osztható tulajdonságának megállapításához a karakterisztikus exponensét hívjuk segítségül, azonban ezt már csak a Lévy-folyamatokra definiálom. Tegyük fel, hogy L_t-nek létezik karakterisztikus exponense $\psi_t(u) := \log \mathbb{E}(e^{iuL_t})$. Ekkor L_t korlátlanul osztható, ha minden $n \geq 1$ esetén létezik az n valószínűségi változó karakterisztikus exponense $\psi_n(u)$, hogy

$$\psi(u) = n\psi_n(u), \quad \forall u \in \mathbb{R}.$$

A Lévy-folyamat definíciójából láthatjuk, az L_t egy korlátlanul osztható valószínűségi változó minden $t > 0$ esetén. Ez abból következik, hogy minden $n = 1, 2, \ldots$ esetén

$$L_t = L_{t/n} + (L_{2t/n} - L_{t/n}) + \cdots + (L_t - L_{(n-1)t/n});$$

valamint kihasználva, hogy az L folyamat független növekményű és $L_0 = 0$. Tegyük fel, hogy a karakterisztikus exponens minden $t \geq 0$-ra:

$$\psi_t(u) = \log E(e^{iuL_t}).$$

Ebbe az egyenletbe behelyettesítve az előző összefüggést és a logaritmus azonosságait felhasználva azt kapjuk, hogy

$$m\psi_1(u) = \psi_m(u) = n\psi_{m/n}(u).$$

A folyamat karakterisztikus függvénye felírható exponensének segítségével

$$\phi_{L_t}(u) = e^{t\psi(u)}.$$

Ha ezt a ψ-t ismerjük $t = 1$-re, akkor ismerjük az eloszlást minden t időpillanatban. Ugyanis ez meghatározza a karakterisztikus függvényt minden más t-re, az pedig egyértelműen meghatározza az eloszlást.

$$\phi_{L_1}(u) = e^{\psi_1(u)} \Rightarrow \psi_1(u) \Rightarrow \psi_t(u), \forall \text{ t-re} \Rightarrow \phi_{L_t}(u), \forall \text{ t-re} \Rightarrow \text{ eloszlás.}$$

A korlátlanul osztható eloszlások és a karakterisztikus exponens közötti kapcsolatot teljes mértékben a Lévy–Hincsin-formula írja le. Egy adott korlátlanul osztható eloszláshoz megadható egy L_t Lévy-folyamat úgy, hogy az L_1 eloszlása éppen ilyen lesz. Itt is csak a Lévy-folyamatra vonatkozó formulát használom. ([1] könyv 1.6 tétele)

2.5. Tétel (Lévy–Hincsin-formula Lévy-folyamatokra). *Tegyük fel, hogy $\sigma \in \mathbb{R}$, és μ az $\mathbb{R} \setminus \{0\}$-ra koncentrált mérték, úgy hogy eleget tesz*

$$\int_{\mathbb{R}} (1 \wedge x^2)\mu(dx) < \infty \quad \text{-nek.}$$

Ez a hármas definiálja minden $u \in \mathbb{R}$-re a

$$\psi(u) = iau + \frac{1}{2}\sigma^2 u^2 \int_{\mathbb{R}} (1 - e^{iux} + iux\mathbb{I}_{|x|<1})\mu(dx)$$

karakterisztikus exponenst. Ekkor egyértelműen létezik az $(\Omega, \mathcal{F}, \mathbb{P})$ valószínűségi mezőn Lévy-folyamat melynek ez a ψ függvény a karakterisztikus függvénye.

A Lévy-folyamat tehát három komponens összege. Ez a három komponens a drift (a), a Brown-mozgás (σ^2) és a tiszta ugró folyamat (μ).

2.1. Lévy-folyamatok tulajdonságai

Az alábbiakban kiemelek néhány tulajdonságot, [2] 6.9 és 6.11 fejezetéből, melyek fontosak lehetnek.

2.1.1. Markov-folyamat

Előbb láttuk, hogy a Lévy-folyamat független növekményű, ebből kifolyólag Markov-folyamat is. Legfontosabb jellemzője az átmenetvalószínűség:

$$P(s, x, t, B) = P(L(t) \in B | L(s) = x).$$

A független stacionárius növekményűségéből adódik, hogy az s időpontban az x állapotban lévő folyamat ugyanúgy fejlődik tovább, mintha a 0 időpontban 0-ból indulna (s-sel időben előre toljuk, és 0-ból x-be). Különlegessége, hogy az átmenetvalószínűség Lévy-folyamatra tér-idő homogén és megmutatható, hogy a Lévy-folyamatok az egyetlen ilyen jellemzővel rendelkező Markov-folyamatok:

$$P(s, x, t, B) = P(0, 0, t - s, B - x).$$

2.1.2. Önhasonlóság

A Brown-mozgás eloszlása megfelelően átskálázva nem változik meg. Ezt a tulajdonságot neveztük önhasonlóságnak, mely Lévy-folyamatokra kiterjesztve:

2.6. Definíció. *A Lévy-folyamat* ***önhasonló****, ha megfelelően átskálázva eloszlása nem változik, azaz létezik olyan $b(a)$ függvény, hogy minden $a > 0$ valós számra*

$$\frac{L(at)}{b(a)} \sim L(t).$$

2.7. Definíció (Stabilis eloszlás). *Az X valószínűségi változó* ***stabilis eloszlású****, ha minden n-re teljesül, hogy választva $X_1, X_2, \ldots, X_n$ független és X-szel azonos eloszlású valószínűségi változókat, ezek összegének eloszlására teljesül, hogy*

$$X_1 + \cdots + X_n \sim c_n X + d_n,$$

ahol c_n és d_n megfelelő konstansok.

2.8. Állítás. *Ha $L(t)$ önhasonló Lévy-folyamat, akkor $L(t)$ stabilis eloszlású $\forall t$-re.*

Bizonyítás. Az $L(t)$ egy Lévy-folyamat, ezért $L(nt) = Y_1 + \cdots + Y_n$, ahol Y_i a Lévy-folyamat növekménye $(i-1)t$ és it között, amik független, azonos eloszlásúak és L_t-vel megegyező eloszlású. Ha L_t önhasonló is, akkor

$$\frac{L(nt)}{b(n)} \sim L_t.$$

A két egyenletből azt kapjuk, hogy

$$\frac{Y_1 + \cdots + Y_n}{b(n)} \sim L_t,$$

ami a stabilis eloszlás definícióját elégíti ki $c_n = b(n)$ és $d_n = 0$ konstansok mellett. □

2.2. Példák Lévy-folyamatra

Összetett Poisson-folyamat

A biztosítók által alkalmazott folyamat, amikor az idő függvényében kívánják vizsgálni az összkár értékét. [3]

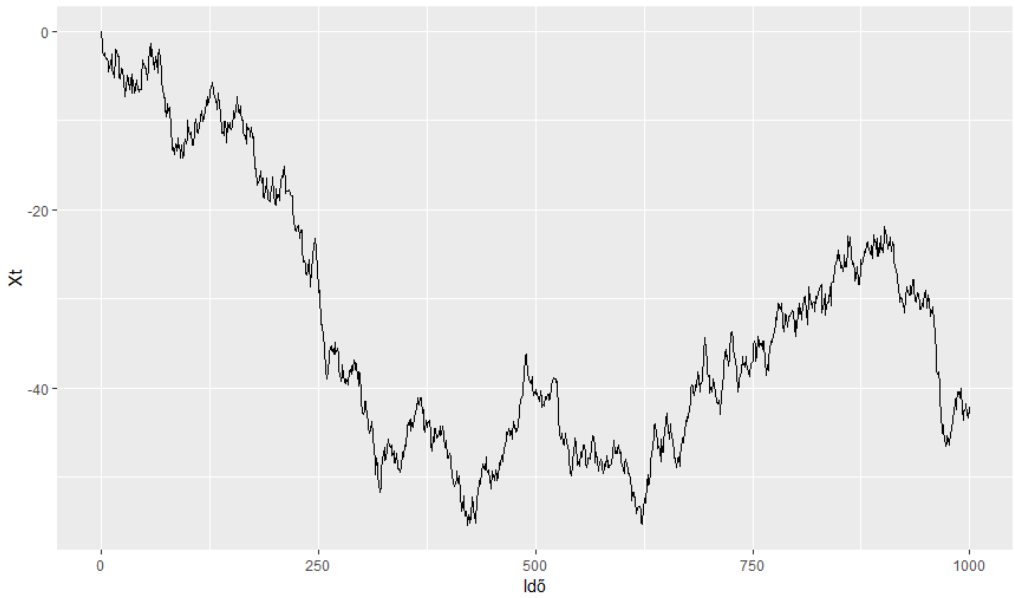
Legyen N_t egy $\lambda > 0$ intenzitású Poisson-folyamat és legyenek ξ_i-k egymástól és N_t-től is független, standard normális eloszlású valószínűségi változók. Ekkor az

$$X_t = \sum_{i=1}^{N_t} \xi_i, \quad t \geq 0$$

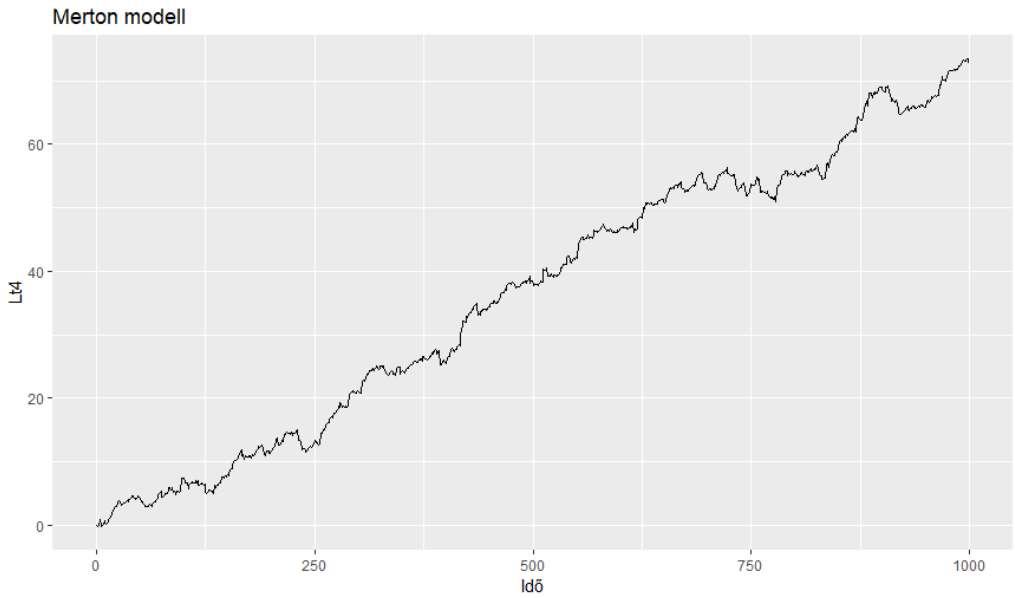
folyamatot összetett Poisson-folyamatnak nevezzük. Ez láthatjuk a 2.1 ábrán.

Merton-modell

A tőzsdei árfolyamatokban hirtelen nagy ugrásszerű változások jelenhetnek meg, melyek külső hatásra keletkeznek. Ezek nehezen modellezhetőek, nem elég a Brown-mozgás hozzá, mert az folytonos trajektóriájú folyamattal írja le a részvényárat. Az ugrások megjelenítéséhez a Poisson-mértéket is be kell építeni, tehát az árak folytonos mozgását hirtelen ugrások szakítják meg. Ezek az ugró diffúziós modellek, mint például a 2.2 képen látható Merton-modell.



2.1. ábra. Összetett Poisson-folyamat, $\lambda = 100$ paraméterrel, standard normális eloszlású ugrásokkal



2.2. ábra. Merton modell, $\mu = 0,5$, $\sigma = 0,75$, $\lambda = 1,5$

Amennyiben több különböző árfolyamatot tekintünk, melyek valamilyen értelemben kapcsolatban vannak egymással, akkor ezek a folyamatok nem lesznek egymástól függetlenek. Erre egy példa a sztochasztikus korreláció modellje, melynek egy egyszerűbb változatát az alábbiakban ismertetjük. [9]

Legyen X és Y két egymástól független, standard normális eloszlású valószínűségi változó. Ekkor

$$Z = rX + \sqrt{(1-r^2)}Y \sim \mathcal{N}(0;1)$$

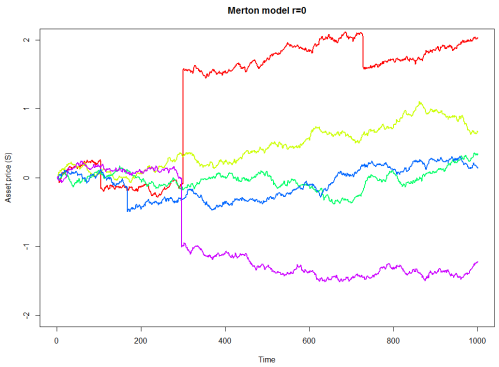
Ennek megfelelően, vegyünk 5 árfolyamatot és egy tőlük független Brown-mozgást $B_0(t)$-t. Melyet ha minden folyamathoz hozzáadunk, akkor a folyamatok között lesz egy r korreláció.

$$\begin{aligned}
X_1(t) &= rB_0(t) + \sqrt{(1-r^2)}B_1(t) \\
X_2(t) &= rB_0(t) + \sqrt{(1-r^2)}B_2(t) \\
X_3(t) &= rB_0(t) + \sqrt{(1-r^2)}B_3(t) \\
X_4(t) &= rB_0(t) + \sqrt{(1-r^2)}B_4(t) \\
X_5(t) &= rB_0(t) + \sqrt{(1-r^2)}B_5(t)
\end{aligned}$$

Ezt a 2.3 ábrán láthatjuk $r = 0; 0.25; 0.5; 0.75$ értékekre.

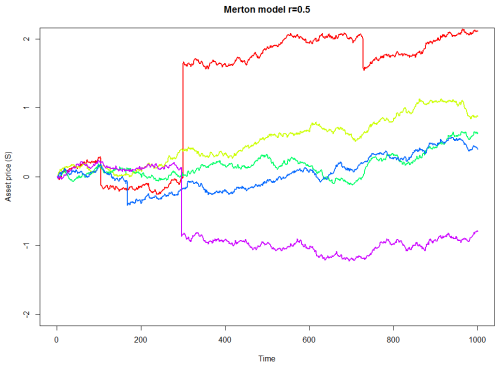
Az így létrejött folyamatok közötti korrelációt szeretnénk megfigyelni, melyhez 1000 darab folyamatpárt vizsgálunk és a köztük lévő korrelációt becsüljük pontonként, a beépített *cor()* függvénnyel. A folyamat egy konkrét t időpillanatában megfigyeljük a korreláció függését az $r = 0; 0.02; 0.04; \ldots; 1$ értékekre. A kapott értékeket kirajzolva a 2.4 ábrán látjuk.

Látható, hogy az r érték növekedésével a folyamatok közötti korreláció is egyre növekedik. Tehát, ha két termék közti kapcsolat növekedik, akkor az részvényárfolyamatuk közti kapcsolat is.

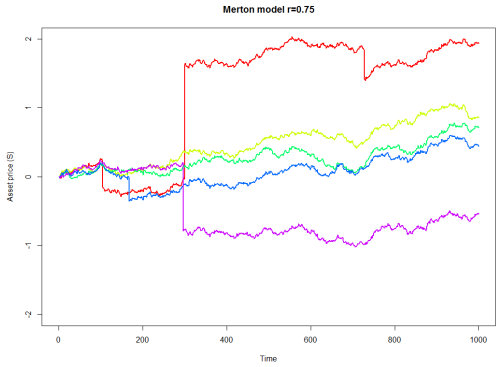


(a) r=0

(b) r=0.25

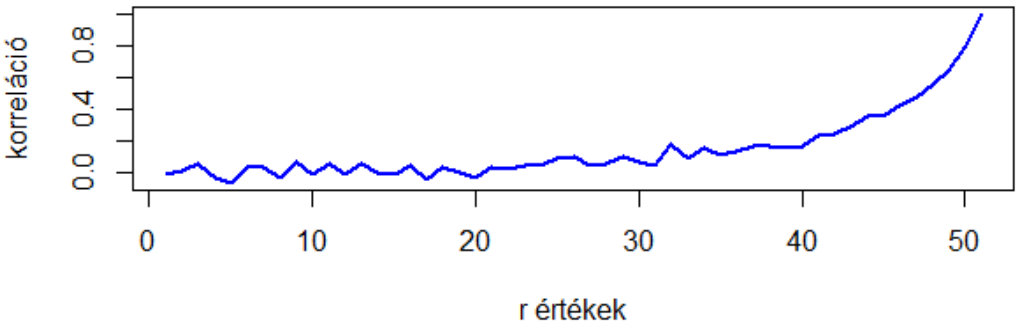


(c) r=0.5



(d) r=0.75

2.3. ábra. Korreláció az árfolyamok közt



2.4. ábra. Korreláció különböző r értékek mellett

3. fejezet

Filogenetikai fák

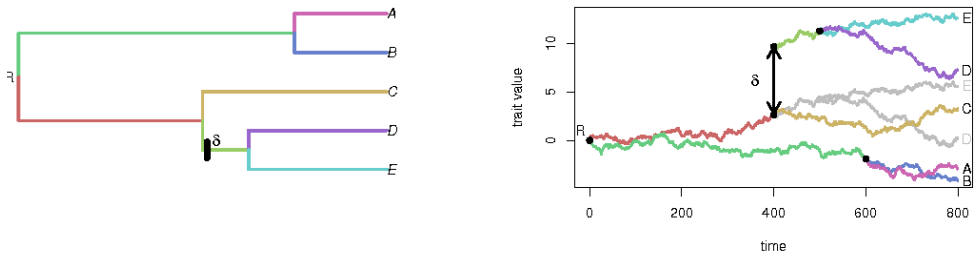
A bevezetésben már említett filogenetikai fák olyan fák, melyek élőlények evolúciós kapcsolatát mutatják be. Ez persze csupán hipotézisek alapján készül. A filogenetikai fa elágazásának mintázata tükrözi, hogy a fajok vagy más csoportok közös törzsekből alakultak ki. Természetesnek tűnik az, hogy a fák esetében két faj szorosan kapcsolódik egymáshoz, ha van egy közvetlen közös őse és kevésbé, ha csak egy korábbi közös ősük van. Felsenstein megállapította, hogy egymással kapcsolatban álló élőlények nem függetlenül fejlődnek egymástól. Közös evolúciójukat jól megfigyelhetjük a filogenetikai fák segítségével. [6, 7]

3.1. Brown-mozgás a filogenetikai fákon

A jellemvonások folyamatos fejlődésének modellezésére széles körben használják a Brown-mozgást, mint például a testtömeg-fejlődés vagy a gének információ öröklődése. Azt feltételezzük, hogy egy vagy több tulajdonság e szerint a sztochasztikus folyamat szerint fejlődik a fa minden ágán. Ezt szemlélteti jól a [8] cikkben található 3.1 ábra.

Elsőként tekintsük azt a modellt, melyben egyetlen tulajdonság szerint vizsgáljuk a fejlődést, azaz az egy változós esetet a 3.2 ábra segítségével.

Legyen $B(t)$ Brown-mozgás, melyre $B_{v_5}(0) \sim \mathcal{N}(0, \sigma^2)$ teljesül. A 0 időpontból kiindulva a folyamat t_1 idő alatt a v_7 pontba érkezve $B_{v_7}(t_1) \sim \mathcal{N}(0, \sigma^2(t_1 - t_0))$ eloszlású. Ennél a pontnál bekövetkezik egy mutáció és a folyamat elágazik. A v_1 és v_2 csúcsokban megfigyelt értékek már nem függetlenek egymástól, mert a fa gyökerétől

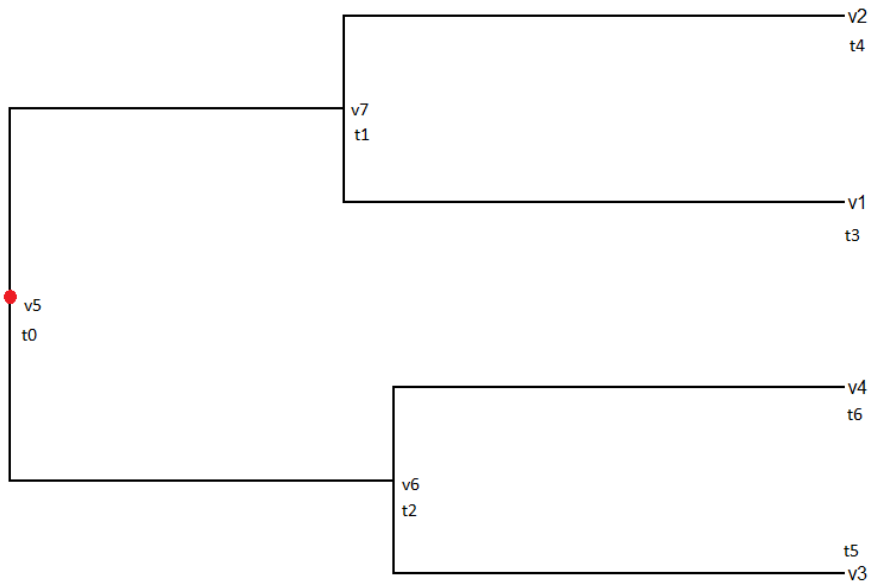


(a) A phylogenetic tree.

(b) Trait value evolution.

3.1. ábra. Jellemvonás fejlődése BM szerint [8]

a v_7 pontig közösen fejlődtek, ahogy az a 3.2 ábrán is látszik.



3.2. ábra. A filogenetikai fa

Az elágazás után t_3 idővel a folyamatunk $B_{v_1}(t_3) \sim \mathcal{N}(0, \sigma^2(t_1 + t_3))$, a v_2-ben pedig $B_{v_2}(t_4) \sim \mathcal{N}(0, \sigma^2(t_1 + t_4))$. Ez tehát tekinthető két folyamatnak, amiknek van közös részük, ezért nem független folyamatok. A folyamatok t_3 és t_4 időpontbeli kovarianciamátrixa a következő lesz:

$$\begin{bmatrix} \sigma^2(t_1 + t_3) & \sigma^2 t_1 \\ \sigma^2 t_1 & \sigma^2(t_1 + t_4) \end{bmatrix} = \sigma^2 \begin{bmatrix} t_1 + t_3 & t_1 \\ t_1 & t_1 + t_4 \end{bmatrix}$$

A C kovarianciamátrix n darab faj esetén $n \times n$-es, szimmetrikus és főátlójában azok az értékek szerepelnek, melyből kiolvasható, hogy az ahhoz tartozó levél (egyed)

milyen távol van a gyökértől. Ebből kitalálható, hogy a főátlón kívüli elemek azt az időhosszt jelölik, mely alatt a két egyed közösen fejlődött a kezdőponttól. Az előző fa teljes kovarianciamátrixa tehát:

	v_1	v_2	v_3	v_4
v_1	$t_1 + t_3$	t_1	0	0
v_2	t_1	$t_1 + t_4$	0	0
v_3	0	0	$t_2 + t_5$	t_2
v_4	0	0	t_2	$t_2 + t_6$

Ennek a szimulációjához az R-ben található „*ape*" és „*phytools*" csomagokat fogom használni.

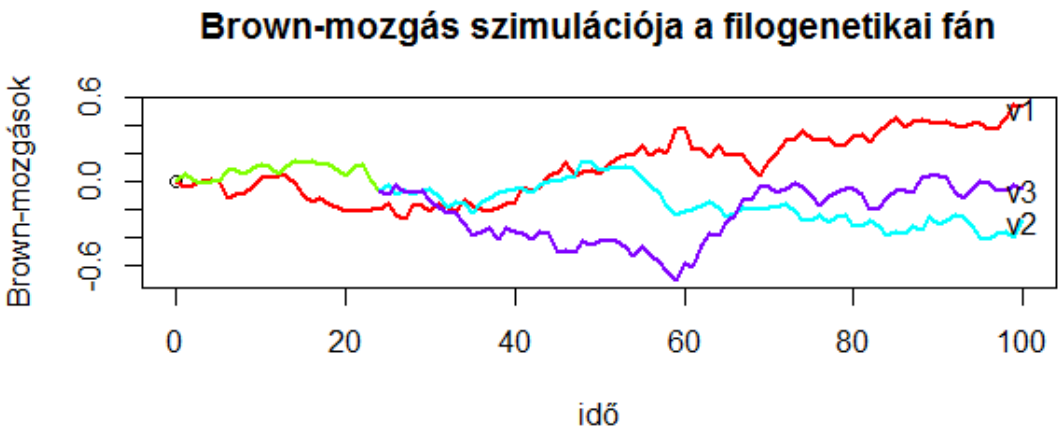
Az előzőtől eltérően most tekintsünk csak három egyedet. A *pbtree*() nevű függvénnyel épített 3 levelű fa, mely két belső pontot tartalmaz és $T = 100$ időpillanat alatt figyeljük meg a fejlődést. A továbbiakban a Brown-mozgás jelölésénél a t-től való függést nem írom ki az egyszerűség kedvéért.

A feladatunk tehát, hogy a fa végpontjaiban megfigyelt értékek korrelációs együtthatóját számoljuk ki. Azért erre a mérőszámra vagyunk kíváncsiak, mert legtöbb esetben nem ismerjük a filogenetikai fa alakját illetve azt sem, hogy mi az az időpillanat amikor egy mutáció következhetett be. Brown-mozgás esetén ezt egyszerűbb vizsgálni, azonban a későbbiekben ez nem lesz nyilvánvaló.

A 3.3 (b) ábrán lévő fa variancia- kovariancia mátrixának kiszámításához a *vcv()* függvényt hívom segítségül, mely az alábbi mátrixot adta eredményül:

	v_1	v_2	v_3
v_1	1.00	0	0
v_2	0	1.00	0.24
v_3	0	0.24	1.00

Itt láthatjuk, hogy a v_2 és v_3 csúcsok 24-edik időpillanatig fejlődtek közösen ezért a korrelációjuk 0,24. Ennek a beépített függvénynek a hátránya, hogy csak Brown-mozgásra számol korrelációt. Ahhoz, hogy a későbbiekben is tudjunk korrelációt számolni bonyolultabb Lévy-folyamatok esetén, tekintsük a 100 különböző futtatásból nyert végpontbeli értéket és ezek tapasztalati korrelációjából kapunk egy becslést az eredeti korrelációs együtthatóra. Ezzel a módszerrel kapott mátrixban az egymástól függetlenül fejlődő egyedek tapasztalati korrelációs együtthatója -0.104, 0.085,



(a)

(b)

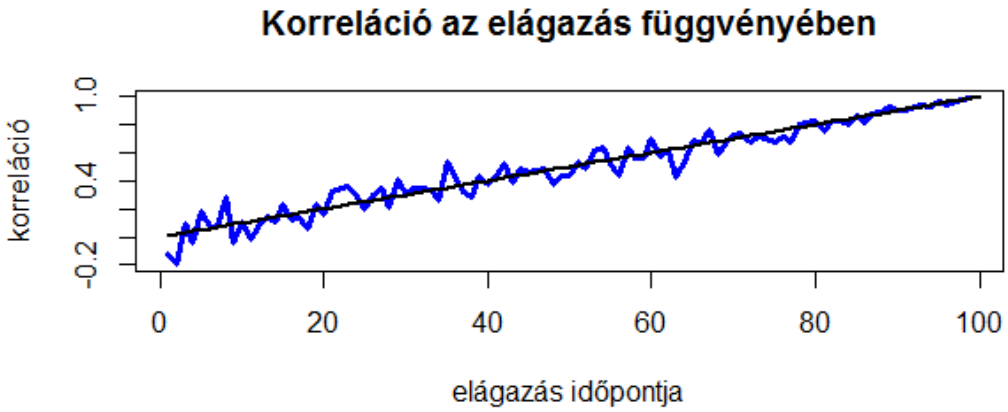
3.3. ábra. Brown mozgás fejlődése a filogenetikai fán

ami elfogadhatónak tekinthetünk. Itt 0.21 a korrelációs együttható, azaz a 21-edig időpontig fejlődött közösen, ami egészen jó becslés egy olyan fejlődési folyamatra, ahol nem lehet pontosan meghatározni egy elágazás pontos idejét.

	v_1	v_2	v_3
v_1	1.00	-0.104	0.085
v_2	-0.104	1.00	0.209
v_3	0.085	0.209	1.00

A kérdés, ami felvetődik, hogy a közösen fejlődött egyedek elágazási időpontjától,

hogyan függ a korrelációs együttható. Azt várjuk, hogy ha például 10-edik időpontig fejlődnek közösen, akkor 0.10 körüli, ha a 80-adik időpontig, akkor viszont körülbelül 0.8 lesz a korrelációs együttható. Ennek vizsgálatához készítettem a 3.4 ábrát. Ezen látszik, hogy a várakozásainknak megfelelően, egy lineáris trenddel rendelkező görbét kapunk. Az ábrán előforduló egyenestől való kilengések annak köszönhetőek, hogy amíg rövid kevés idő telik el az elágazásig, addig a tapasztalati korrelációban a véletlen kilengések játszanak nagyobb szerepet. Ugyanakkor azt is látjuk, hogy a görbe vége látszólag rásimul az egyenesre.

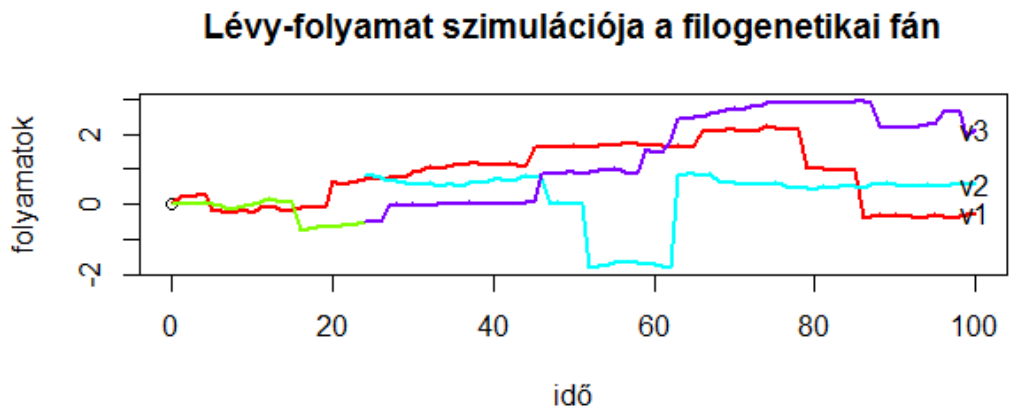


3.4. ábra. Brown-mozgásból szimulált korreláció az elágazás függvényében

3.2. Lévy-folyamat a filogenetikai fákon

A Lévy-folyamatok alkalmazása a biológiában számomra első hangzásra különösnek tűnt, azonban utána olvasva már teljesen észszerű. Bevezetésére azért került sor mert a fejlődési folyamatban előforduló hirtelen, ugrásszerű változások leírását jól követi a Lévy-folyamat. A folyamatban egy-egy ilyen ugrás megjelenése, a filogenetikai fában egy-egy elágazást jelent. A vizsgált jellemző lényeges eltérése a korábbi állapottól már indokolttá teszi, hogy ezen a különleges folyamat mentén vizsgálódjunk. Az ábráimon ez az ugrás nem minden ágon jelentkezik majd a későbbiekben sem. Ennek oka, hogy egy területen ugyanazon jellemző szerint fejlődő állatok közül néhány elvándorol élelemhiány miatt, vagy más területen is feltűnnek új baktériumok hordozók által. Ekkor az új környezeti tényezők másképpen hatnak rájuk, míg az eredeti környezetben maradt élőlények fejlődnek tovább az addigi folyamat szerint. A következő szimulációk során ezt a modellt fogom alapul venni. [7]

Az előző fejezet 3.3 (b) ábráján lévő fán nézzünk egy olyan Lévy-folyamatot, mely egy 0.05 szórásnégyzetű Brown-mozgás és egy $\lambda = 0.11$ paraméterű összetett Poisson-folyamat összege.



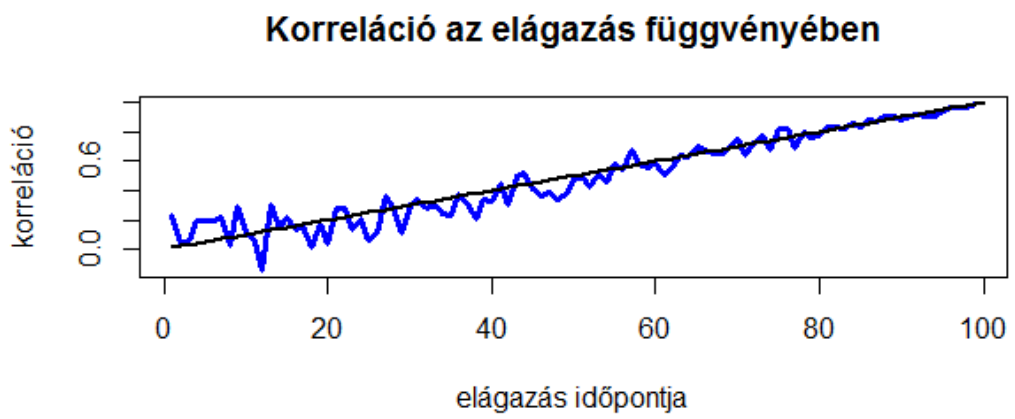
3.5. ábra. Lévy-folyamat fejlődése a fán, $\sigma = 0.05$, $\lambda = 0.11$

A 3.5 képen látható folyamat végpontbeli értékei közti korreláció kiszámításához az előző fejezetben megismert módszert fogom használni. Ezen módszer segítségével a következő variancia-kovariancia mátrixot kapjuk:

	v_1	v_2	v_3
v_1	1.00	-0.048	0.11
v_2	-0.048	1.00	0.21
v_3	0.11	0.21	1.00

Erre a folyamatra nincsen összehasonlítási alapunk, az egyetlen érték amit össze tudunk vetni az a v_2 és v_3 csúcsok korrelációja. Ezzel a modellel, ha egy kicsivel is de pontosabb értéket kaptunk, azonban ez az aktuális futtatási erdményeknek is „köszönhető".

Nézzük meg most azt, hogy mi történik akkor, ha a különböző elágazási időpontok függvényében vizsgáljuk a csúcsok közti összefüggést.

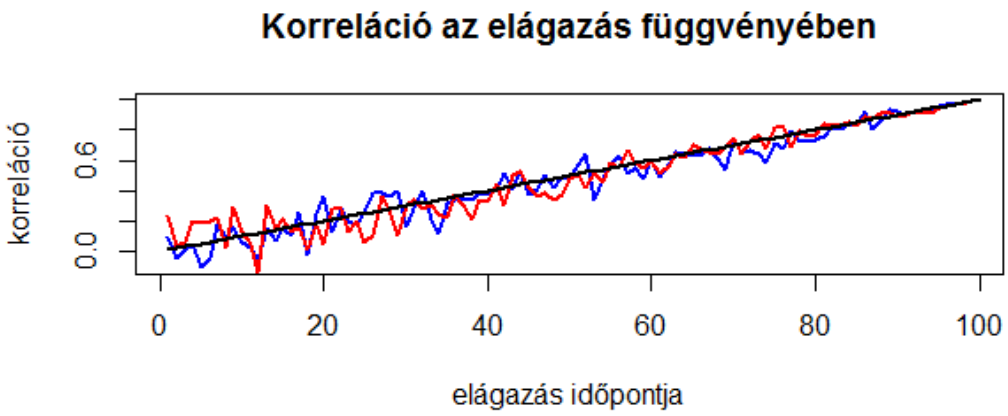


3.6. ábra. Korreláció függése az elágazás időpontjától Lévy-folyamatra

Itt is eredményül kapjuk a várt lineáris trenddel rendelkező görbét, mint az előző fejezetben. Az egyenestől való eltérések ugyanannak a véletlen sorsolási hibáknak tudhatók be. Arra a logikus következtetésre jutunk az ábrákból, hogy minél tovább fejlődnek közösen az egyedek, annál nagyobb az összefüggés a vizsgált tulajdonságaik között.

A két görbét összevetve a 3.7 ábrán, kékkel a Brown-mozgás futási eredménye látható, pirossal pedig a Lévy-folyamaté. Valójában egészen hasonló eredményeket kapunk, talán a Lévy-folyamat kilengései laposabbak a 20-adik időpillanattól kezdve.

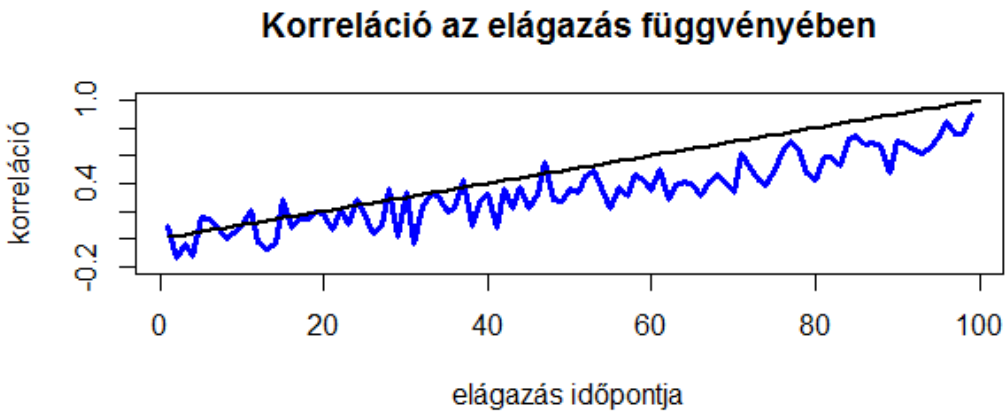
Az eddigi vizsgálataink akkor válnak igazán érdekessé, ha a folyamatoknak valamilyen függvényét nézzük. Ennek a függvénynek a jelentése lehet például az, hogy ha



3.7. ábra. Korreláció függése az elágazás időpontjától mindkét folyamatra

élőlények azonos helyen élnek, akkor a külső tényezők hogyan hatnak rájuk. Olyan függvényt is tekinthetünk, ami nem folytonos, mint például az indikátor függvény. Ez tulajdonképpen azt jelenti, hogy a vizsgált egyedek jellemzői elérnek-e egy adott szintet. Vajon ilyen függvények esetén is ugyanolyan nagyságú marad az összefüggés, vagy teljesen más eredményt kapunk?

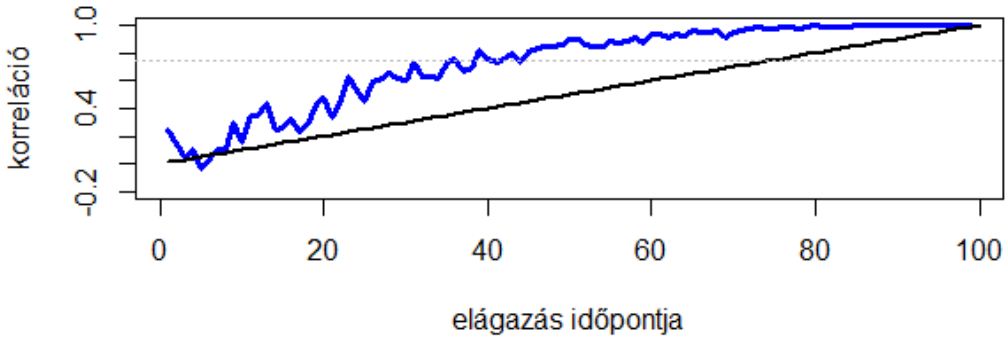
Most az utóbbi függvényt alkalmazzuk a folyamatainkra. A folyamatunk 0.05 szórású egy Brown-mozgás és egy $\lambda = 0.111$ paraméterű összetett folyamat összegéből áll most is, $a = 0.1$ pedig legyen az a szint, amit el kell érnie a jellemzőnek. A kapott eredményt ugyanúgy 100 futtatásból számítjuk.



3.8. ábra. Indikátor függvény hatása a korrelációra Lévy-folyamat esetén

Ebben az esetben az ingadozások sokkal nagyobbak, mert itt feltehetően nem minden egyes futás ugyanazon időpontjában éri el a megadott szintet. Kisebb korreláció figyelhető meg azoknál az időpontbeli elágazásoknál is, melyek majdnem végig közösen fejlődtek. (3.8 ábra) Így tehát, ha például egy egyed testtömege elér egy bizonyos szintet, akkor kevésbé valószínű, hogy a másik egyed testtömege is eléri azt a szintet.

Testtömeggel kapcsolatos vizsgálatunk folytatásához azt nézzük meg, hogy egy fejlődési szakasz alatt egy-egy egyed testtömegének kumulált összegét tekintjük a végpontokban, majd ezek korrelációját hasonlítjuk össze. Természetesen ennek emlősök esetén nem sok értelme van azonban, ha egy területen elszaporodó baktériumokat (akár bogarakat) vizsgálunk, akkor azok össztömegére vonatkozó következtetéseket vonhatunk le. Ezt szemlélteti a 3.9 ábra. Ezen jól kivehetően látszik, hogy az összefüggés jóval magasabb, mint amikor független növekményűek voltak a folyamataink. A kumulált összegben többszörösen megjelenik egy részösszeg, így persze nem várhatjuk a lineáris trenddel rendelkező növekedést. Az ábráról leolvasva, a 0.75-ös korrelációs együtthatós értéket már eredményül kapjuk akkor, amikor a folyamatunk csak a 30-adik időpillanatig fejlődik közösen.



3.9. ábra. Kumulált összegek korrelációja a folyamat elágazásának függvényében

A következő fejezetben hasonló elemzéseket fogunk látni egy nagyobb filogenetikai fán.

4. fejezet

Többdimenziós folyamatok

Legtöbbször nem csak egy karakter fejlődését szeretnénk vizsgálni, hanem több jellemzőét is egyszerre. Azonban nem csak az egyedek között állhat fenn korreláció, hanem a jellemzők között is. [10]

A többdimenziós folyamatainkat az alábbi paraméterekkel tudjuk leírni. Elsőként van egy **a** vektorunk, mely r darab karakter várható értékét tartalmazza és egyben ez a vektor a kezdőpontja az r dimenziós folyamatunknak

$$\mathbf{a} = [m_1(0), m_2(0), \ldots, m_r(0)].$$

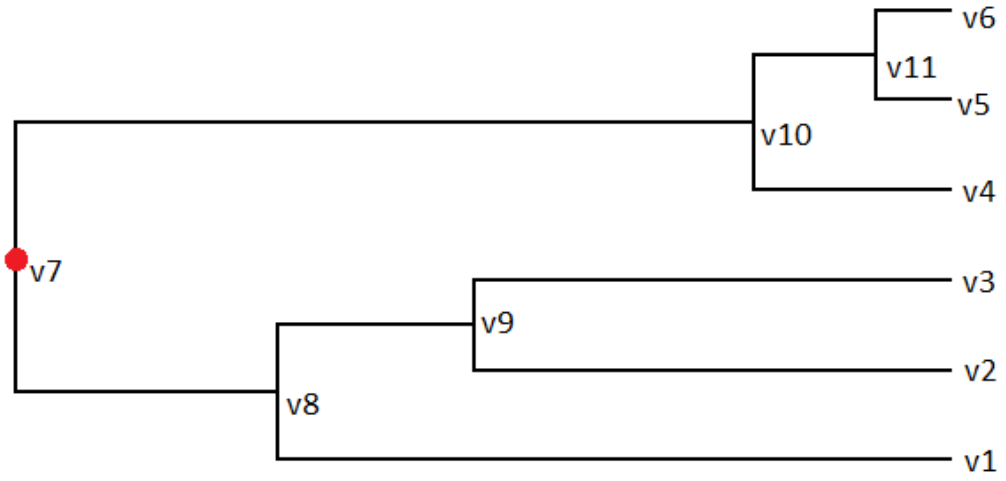
A másik paraméterünk az R variancia-kovariancia mátrix, melynek főátlójában a karakterek paraméterei, azon kívül pedig a jellemzők közti korrelációk vannak.

$$R = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1r} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{r1} & \ldots & & \sigma_r^2 \end{bmatrix}$$

A többdimenziós folyamatok esetében, amikor már a jellemzők közti korrelációt is vizsgáljuk, akkor ezt az R mátrixot az előző fejezetben bemutatott C variancia-kovariancia mátrixszal kell szorozni. Ez a $C \times R$ mátrix lesz a többdimenziós folyamat varianciakovariancia mátrixa.

Az előző fejezetben megismert Lévy-folyamat fejlődését fogom összevetni a Brown-mozgással egy 6 egyedből álló fán, azonban a Lévy-folyamatunk talán jobban lefedi a valóságot azzal, hogy több ugrás jelenik meg a fában. A 4.1 ábrán látható a fa, amin a folyamataink fejlődnek, előzőhöz hasonlóan $T = 100$ időpillanatig. A

folyamatok pedig most is hasonló paraméterűek, mint az előző fejezetben: a Brown-mozgás 0 várható értékű és 0.01 szórású, a Lévy-folyamatban 0.01 szórású Brown-mozgás és egy $\lambda = 0.111$ paraméterű összetett folyamat összege áll. Ezeket a 4.2-n látjuk kirajzolva.



4.1. ábra. Filogenetikai fa 6 levéllel, T=100 időpontig

Először a v_i $(i = 1 \dots 6)$ csúcsok variancia-kovariancia mátrixát tekintsük meghozzá úgy, hogy a fa mindkét „fő ága” a 32-edik időpillanatban ágazik el.

Így azonban elég nehéz összehasonlítani csak a mátrixok alapján a módszereket, ezért mérőszámként a mátrix elemeinek abszolútérték összegét veszem. A beépített függvénnyel számolt mátrixra ez a mérőszám 13.1, a tapasztalati módszerrel számolt mátrixra 12.2 értéket kaptam. Az eltérés abból is adódhat, hogy az eredeti fában nem ugyanabban az időpillanatban ágaznak el a folyamatok mindkét ágon. Ugyanakkor a végpontbeli korrelációt ez nem biztos, hogy befolyásolja, ezt lentebb látni fogjuk. Mivel az ***R*** csomagban található függvény csak Brown-mozgásra számol mátrixot, így a Lévy-folyamatra ezt nem tudjuk összehasonlítani, azonban ez az érték 10.92. Az, hogy kisebb értéket kaptunk nem csak a véletlen generálásból adódhat. Következhet ez abból is, hogy különböző eloszlásból generált értékeket adunk össze minden időpillanatra, így intuitívan kisebb korrelációt is várnánk.

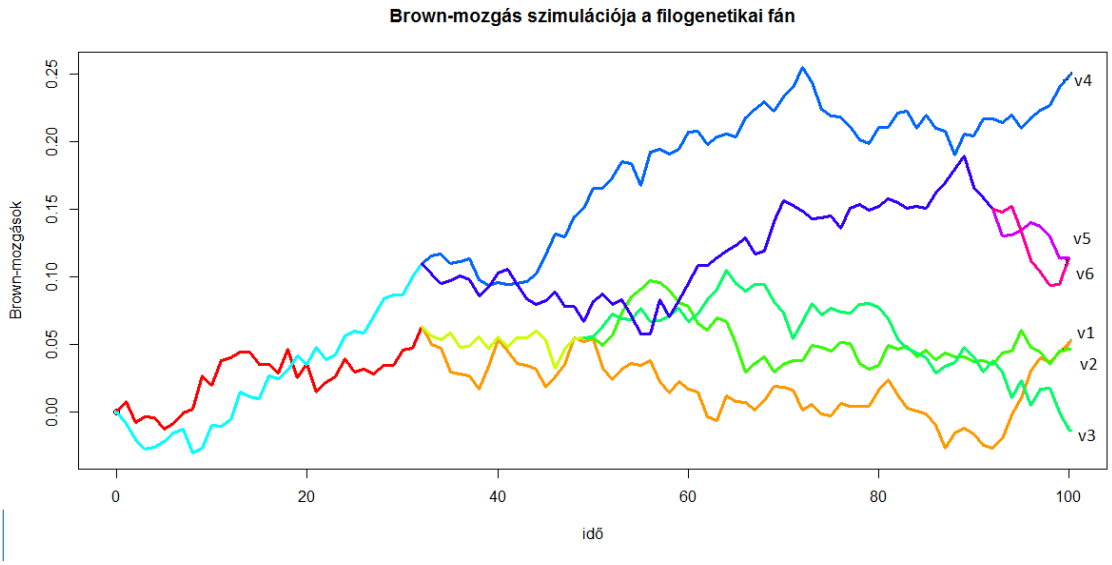
Az előbb látott folyamatokra most nézzük meg azt, hogy hogyan viselkedik a végpontokban megfigyelt értékek korrelációja az elágazások időpontjától függően. A 4.3 ábrán a v_2 és v_3 csúcsok korrelációját láthatjuk. Kék színnel a Brown-mozgásból számolt korreláció, pirossal pedig a Lévy-folyamatra vonatkozó korreláció van fel-

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	1.00	0.41	0.26	-0.05	0.11	0.01
v_2	0.41	1.00	0.39	-0.08	0.16	0.01
v_3	0.26	0.39	1.00	0.10	0.16	0.001
v_4	-0.05	-0.08	0.10	1.00	0.24	0.41
v_5	0.11	0.16	0.16	0.24	1.00	0.91
v_6	0.01	0.01	0.001	0.41	0.91	1.00

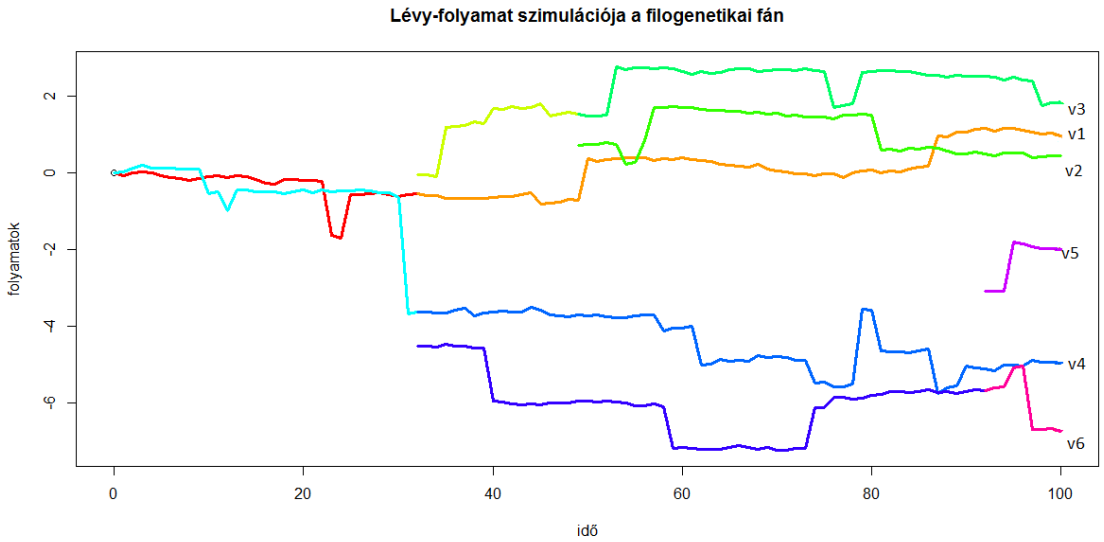
Brown-mozgás variancia-kovariancia mátrixa

	v_1	v_2	v_3	v_4	v_5	v_6
v_1	1.00	0.31	0.36	0.03	0.04	0.07
v_2	0.31	1.00	0.48	0.03	-0.02	-0.01
v_3	0.36	0.48	1.00	-0.05	-0.01	-0.03
v_4	0.03	0.03	-0.05	1.00	0.24	0.19
v_5	0.04	-0.02	-0.01	0.24	1.00	0.88
v_6	0.07	-0.01	-0.03	0.19	0.88	1.00

Lévy-folyamat variancia-kovariancia mátrixa



(a) Brown-mozgás fejlődése a filogenetikai fán

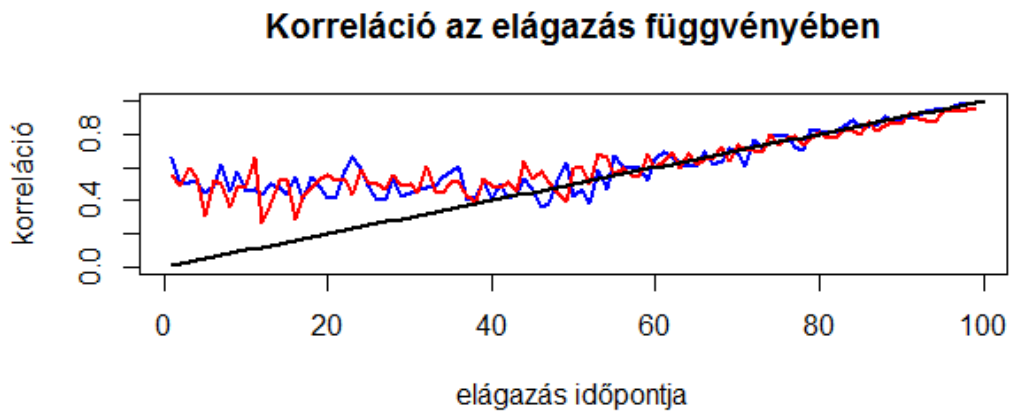


(b) Lévy-folyamat fejlődése a filogenetikai fán

4.2. ábra

tüntetve. Itt jól látható, hogy Lévy-folyamattal szimulált korrelációs együtthatókra végig kisebb értékeket kapunk, tehát ha az előbbi mátrix mérőszámot minden egyes időpontra kiszámolnánk kisebb értékeket kapnánk.

Az ábra első felén magas korrelációt láthatunk, amit azért kapunk mindkét folyamat esetén, mert a folyamat elágazási időpontjai az 49-edik időpillanatig végig

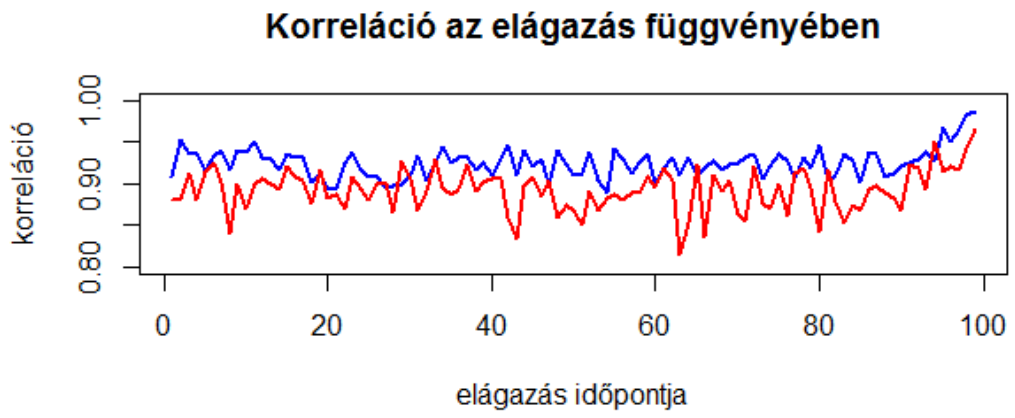


4.3. ábra. Brown-mozgásból és Lévy-folyamatból szimulált korreláció az elágazás függvényében a v_2 és v_3 csúcsokon. Kékkel a Brown-mozgás, pirossal a Lévy-folyamat

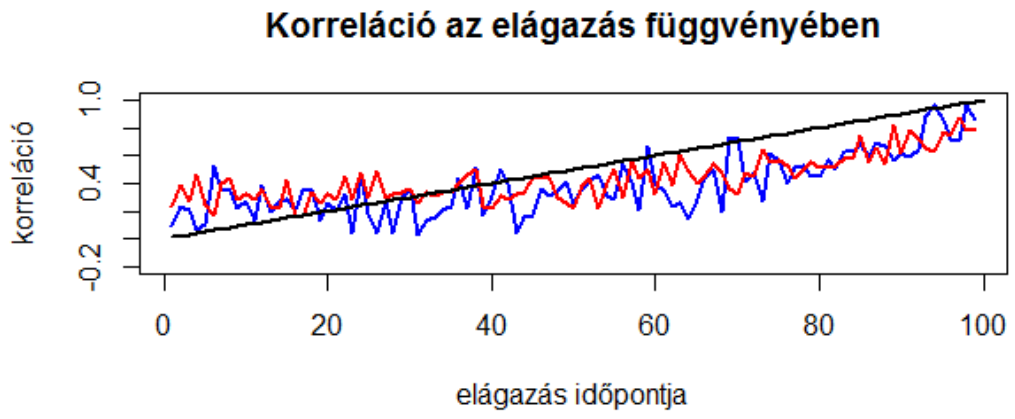
iterálnak 1-től egészen 49-ig. (4.3 ábra) Amint ezt az időpontot elértük, onnantól ugyanúgy viselkedik a fa és a folyamataink is, mint a kisebb fa esetében. Ebből adódóan érthető, hogy amíg nem érjük el azt az időpillanatot addig a korreláció közel ugyanakkora lesz. Ez megmagyarázza az előbbi gyanúnkat, hogy az első elágazás időpontjában lévő változás nincs hatással a különbségre. Az eltérés csökkentése leginkább a minta elemszám növelésével érhető el.

Hasonló a helyzet a v_5 és v_6 csúcsok kapcsolatánál annyi eltéréssel, hogy az eredeti fában a 92-edik időpontig fejlődtek közösen. A 4.4 képen láthatjuk ezt, szinte végig 0.9 körüli értékeket kapunk, csak a végén indul el némi növekedés. Azt is megfigyelhetjük itt, hogy az előző fejezetben a 3.7 ábrán, ahogy közeledtünk a 100-adik időpillanathoz egyre „simábbnak” láttuk a görbét, azonban hasonló kiugrások fordulnának elő ott is, ha ránagyítanánk.

Ahogyan azt előbb is tettük, következzen most az a szimuláció, amikor az indikátor függvényt alkalmazzuk a $T = 100$-adik pillanatban megfigyelt értékekre és úgy tekintjük a korrelációt. A 4.5 ábrán hasonló hatás figyelhető meg a v_2 és v_3 csúcsok esetében, mint korábban a kisebb fánál. Amíg nem érjük el a második elágazás időpontját, addig a korreláció közel ugyanakkora, mintha nem alkalmaztuk volna a függvényt, azt követően pedig kisebb korrelációt kapunk ugyanazon okokból kifolyólag.



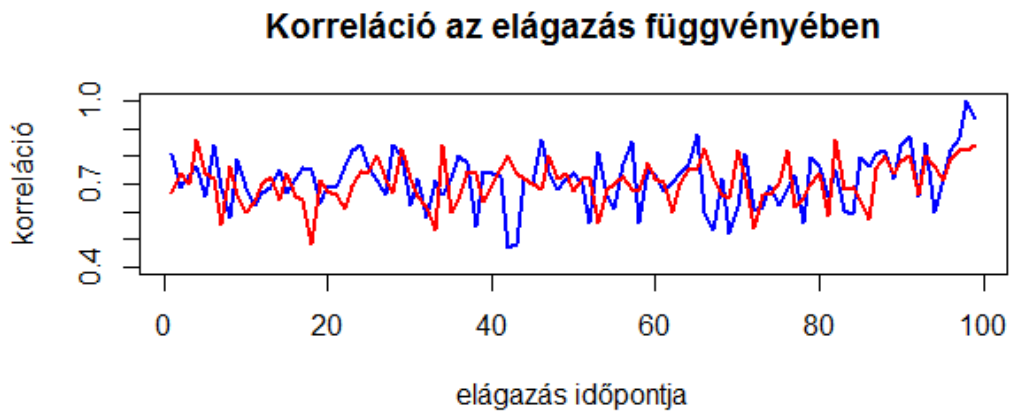
4.4. ábra. Kékkel a Brown-mozgás, pirossal a Lévy-folyamat látható. Korreláció a v_5 és v_6 csúcsokra.



4.5. ábra. Indikátor függvény hatása a korrelációra v_2 és v_3 csúcsokon

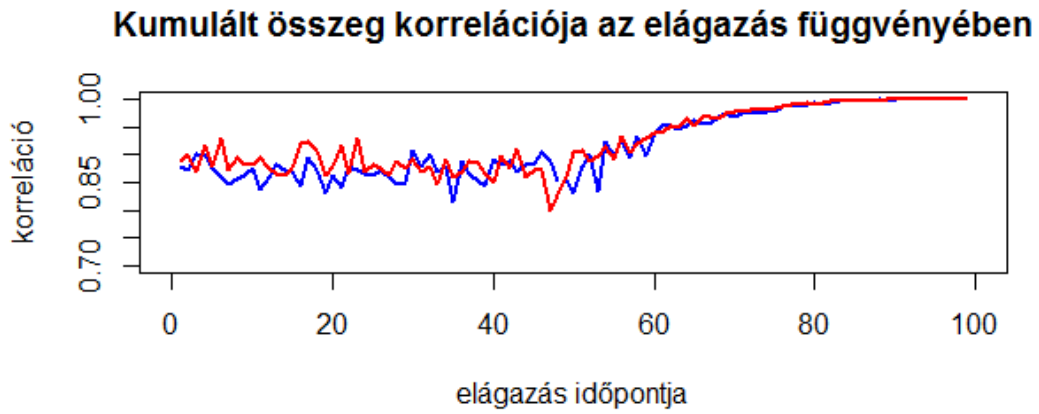
Kíváncsiságból nézzük meg, hogy más történik-e, ha a v_5 és v_6 csúcsokban lévő értékekre alkalmazom a függvényt. A 4.6 képen látjuk, hogy a görbe lefelé tolódott, kisebb a korreláció és az ingadozások is nagyobbak a 4.4-hez képest.

Utolsó szimulációként tekintsük meg a kumulált összegek korrelációját ugyanezen csúcsokra. Ebben az esetben azt várjuk, hogy a folyamatok amíg közösen fejlődtek, addig különösebb változás nem történik a görbe trendjében. Azt követően pedig a véletlen hibák által okozott kilengésektől eltekintve, egy monoton növekedő görbét



4.6. ábra. Indikátor függvény hatása a korrelációra v_5 és v_6 csúcsokon

kellene kapnunk.



4.7. ábra. Kumulált összeg vizsgálata a v_2 és v_3 csúcsokra

A 4.7 ábrán a várt eredményt kapjuk, azonban érdemes figyelni arra, hogy az y tengely 0.7-nél kezdődik. Tehát magasabb korrelációt kaptunk, mint 4.4-nél a részösszegek többszöri megjelenése miatt. Eddigi megfigyeléseink alapján, ugyanezt a vizsgálatot a v_5 és v_6 csúcsokra nem érdemes megfigyelnünk, ugyanis abban az esetben folyamataink $t = 92$ időpontig azonosak. Így már a kumulált összeg nélkül is magas korrelációt kaptunk, azzal együtt még nagyobb összefüggést kapnánk és az utolsó 8 megfigyelt időpontban jelenne meg további növekedés a görbénken.

Kitekintés

Az eddigi vizsgálataim során a folyamatokat tetszőleges paraméterek mellett futtattam. Illetve a felhasznált irodalomban található paraméterektől nem vagy kis mértékben tértem el. Ennek hátránya azonban meg is mutatkozik a 4.2 (b) ábráján, mert a folyamat megfigyelt ideje alatt láttunk többször is nagyobb ugrásokat. Azonban a fa elágazásai ezt nem követik, nem ugyanazokban az időpillanatokban ágazik el a fa, ahol nagyobb ugrások láthatóak. Természetesen felvetődik a kérdés, hogy mikor tekintünk egy változást akkora volumenűnek, hogy azt egy evolúciós lépésnek vehessük és a fa abban a pillanatban elágazzon. Landis a [7] cikkében ezzel külön foglalkozik, ugyanis különböző Lévy-folyamatokkal modellezi a megfigyelt adatokat és a folyamatok paramétereire próbál egy becslést készíteni. [10]

Az eljárás a Bayesi megközelítést alkalmazza arra, hogy a Lévy-folyamat fejlődését vizsgálja a filogenetikai fán. Legyen $p(\theta)$ az a priori eloszlása a Lévy-folyamat paramétereinek, $L(D|\theta)$ pedig a likelihood-függvénye a megfigyelt adatainknak a θ paraméter mellett. Kíváncsiak vagyunk az a posteriori eloszlásra, azaz

$$p(\theta|D) \propto L(D|\theta)p(\theta).$$

A likelihood függvény kiszámolásához Felsenstein „pruning algoritmusát” használja föl. Melyet, egy példán keresztül könnyebb lesz megérteni. Tekintsünk egy hat egyeddel rendelkező fát, melynek minden csúcsán a vizsgált karakter a 0, 1 vagy 2 értékeket veheti föl, mint a 4.8 ábrán.

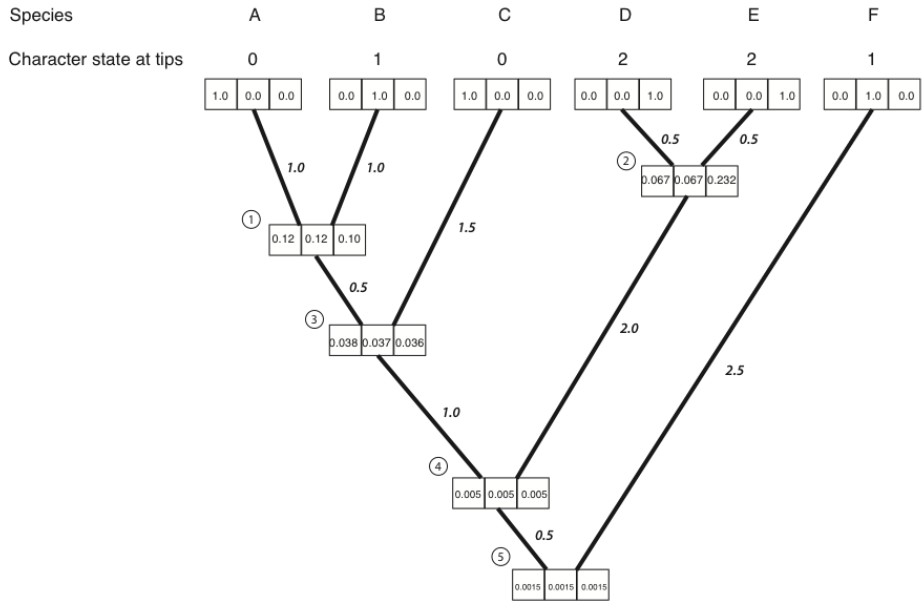
Az algoritmus a következő:

1. Azt töltjük ki, hogy milyen valószínűséggel veszi föl az adott értéket a leveleken a folyamat.

2. Megkeressük azt a következő csúcsot, amelyiknek minden leszármazottja levél. Ha több ilyen van egyet választunk.

3. A 2. pontban kiválasztott csúcshoz a leszármazottainak segítségével, a feltételes likelihood módszerével számoljuk ki a karakter értékek valószínűségét.

4. Az előző számolást ismételjük, olyan sorrendben haladva végig a csúcsokon, hogy a leszármazottakhoz tartozó valószínűségek már ismertek legyenek.



4.8. ábra. Minden csúcs, három négyzetében a valószínűségekkel [10] 8.4A ábra

5. Kiszámítjuk a valószínűséget az egész fán, a fa gyökerében lévő állapotok feltételes valószínűségével.

A 3. pontban található feltételes likelihood az, ami az algoritmus gerincét alkotja és az alábbi képlettel számol, ha az i csúcsban szeretnénk meghatározni a valószínűségeket annak már ismert j és k leszármazottai alapján.

Tekintsük az $(X_t)_{t>0}$ Lévy-folyamatot és az i-edik pillanatban a folyamat értéke legyen y_i, ekkor

$$L_i(y_i|\theta) = \left(\int \mathbb{P}(X_{t_j} = y_j|X_0 = y_i)L_j(y_j|\theta)dy_j\right) \times \left(\int \mathbb{P}(X_{t_k} = y_k|X_0 = y_i)L_k(y_k|\theta)dy_k\right). \tag{4.1}$$

Itt a t_j és t_k annak az ágnak a hossza, ami i-ből j-be és k-ba vezet. Az $L_i(y_i)$ pedig azt mondja meg, hogy mennyi annak a valószínűsége, hogy a hátralévő részben a j és k csúcsokig azt a folyamatot kapjuk, ami a megfigyelt adatunk. A j és k pontbeli állapotra írjuk fel a teljes valószínűség tételt, majd ezek szorzatával megkapjuk a kívánt valószínűséget. A szorzás azért alkalmazható, mert i után elágazik a folyamat, tehát függetlenül fejlődnek. Mivel folytonos karakterváltozást is vizsgálhatunk, ezért végtelen sok kimenetele lehet az y_j és y_k értékeknek, tehát összegzés

helyett integrálunk a megfelelő mérték szerint.

Az algoritmust folytatva eljutunk az $L_0(y_0|\theta)$ valószínűségekig. A likelihood-függvény meghatározásához a gyökérnél feltesszük, hogy egyenletes eloszlású a priori eloszlása van az y_0 értéknek, tehát az összes lehetséges értékre azonos súlyt helyezünk:

$$L(D|\theta) = \int_{-\infty}^{\infty} L_0(y_0|\theta) dy_0. \tag{4.2}$$

Azonban a 4.1 és a 4.2 integrálok a legtöbb Lévy-folyamatra kiszámíthatatlanok, ezért csak közelítéseket használnak.

5. fejezet

Összefoglalás

Az egyik legismertebb evolúciós elmélet, melyet talán mindannyian ismerünk Charles Darwintól származik, amit már az 1800-as évek közepén publikált. Azóta folyamatos fejlődés alatt volt a biológia ezen ága is, melyet az motivált, hogy karakterek és egyedek (együttes) változását kövessük matematikai eszközök segítségével is.

Az összehasonlító biológia egy kis keretein belül megismerkedhettünk a Brown-mozgás és egy bonyolultabb Lévy-folyamat változásaival. Vizsgálataim legfőképp az egyedek közt fennálló kapcsolaton alapultak. Láttuk, hogy az egyedek közti korreláció lineárisan változik az elágazás időpontjától függően. Tehát a szimulált adatokból is azt az információt kaptuk, hogy minél tovább fejlődnek az egyedek közösen annál inkább hasonlítanak egymásra. Ez a linearitás fennállt akkor is, amikor az indikátor függvény hatását vizsgáltuk az adott jellemzőre. Abban az esetben is megtartotta ezt a tulajdonságot, ha a kumulált összeget tekintettük, annyi változással, hogy magasabb korrelációt kaptunk.

Az én előzetes várakozásaimnak megfelelt az az eredmény is, amikor a nagyobb fán vizsgáltuk ugyanezeket a folyamatokat és az összetettebb Lévy-folyamat esetében kisebb tapasztalati korrelációs együtthatókat kaptunk.

Végül megismertünk egy bonyolult de annál hasznosabb algoritmust, a Felseinstein-féle „pruning” algoritmust. Ez akkor lehet hasznunkra, ha vannak adataink élőlényekről vagy gének mutációjával kapcsolatban tárolunk adatot. Ezekre az adatokra igen nehéz illeszteni folyamatot, ami a rövid időn belül lezajló, nagy ugrásszerű változásokat követi. A folyamatok bonyolultsága miatt azonban nem lehet pontos választ adni, hogy mekkora paramétereket ad az algoritmus válaszul, ezt is csak közelítéssel kapjuk meg.

6. fejezet

Függelék

```
##fuggvenyek, melyek a folyamataimat allitjak elo
##Brown-mozgas
set.seed(123)
BM=function(mu,sigma,T,N){
  h=T/N
  t=(0:T)/N
  X=rep(0,N+1)
  X[1]=0
  for(i in 1:N){X[i+1]=X[i]+mu*h+sigma*sqrt(h)*rnorm(1)}
  return(X)

}
##Osszetett Poisson-folyamat
PPgen=function(lambda){
  Y=0
  Sum=0
  flag=0
  while (flag==0){
    E=-log(runif(1))
    Sum=Sum+E
    if (Sum < lambda) {Y=Y+1} else {flag=1}
  }
```

```
  return(Y)
}

CPP=function(lambda, mu_xi, sigma_xi, T, N) {
  h=T/N
  t=(0:T)/N
  Z=rep(0, N+1)
  F=rep(0, N+1)
  I=rep(0, N)
  Z[1]=0
  for(i in 1:N) {
    I[i]=PPgen(h*lambda)
    if(I[i]==0){F[i]=0} else {F[i]=mu_xi*I[i]+
                sqrt(sigma_xi)*sqrt(I[i])*rnorm(1)}
    Z[i+1]=Z[i]+F[i]}
  return(Z)
}

##Merton-modell
Merton=function(mu,sigma,lambda,mu_xi,sigma_xi,N, T) {
  h=T/N
  t=(0:T)/N
  X=rep(0, N+1)
  F=rep(0, N+1)
  I=rep(0, N)
  X[1]=0
  for(i in 1:N) {
    I[i]=PPgen(h*lambda)
    if(I[i]==0){F[i]=0} else {F[i]=mu_xi*I[i]+
                sqrt(sigma_xi)*sqrt(I[i])*rnorm(1)}
    X[i+1]=X[i]+mu*h+sigma*sqrt(h)*rnorm(1)+F[i]}
  return(X)

##szimulaciok a kisebb meretu fahoz
```

```
##nagyobb fara a folyamataim
set.seed(1234)
t2 <- 100  # total time
n2 <- 6# total taxa
b2 <- (log(n2) - log(2))/t2

fafa<-pbtree(b = b2, n = n2, t = t2, type = "continuous")
plotTree(fafa)
fafa$edge
fafa$edge.length


Mt<-fafa$edge
Mt<-cbind(Mt, fafa$edge.length)
gyt<-fafa$edge[, 1] %in% fafa$edge[, 2]*1 ##ez 0 ha gyoker
Mt<-cbind(Mt, gyt)
colnames(Mt)<-c('szulo', 'gyerek', 'hossz', 'gyoke')
sam<-matrix()
a<-c()
b<-c()
h<-c()
for (i in 1:nrow(Mt)){
  if (Mt[i,'gyoke']==0){(a<-c(a,Mt[i,'gyerek'])) &
        (b<-c(b,Mt[i,'hossz']))}
 }
sam<-rbind(a,b)
for (i in 1:length(fafa$tip.label)){
  g[i]<-length(nodepath(fafa)[[i]])
  if (length(nodepath(fafa)[[i]])==max(g)){h<-
                c(h,nodepath(fafa)[[i]][max(g)-1])}

}
h<-unique(h)
l<-c()
```

```
for (i in 1:nrow(Mt)){
  if (Mt[i,'gyerek'] %in% h){l<-c(l,Mt[i, 'hossz'])}
}
sam<-rbind(sam,l)
rownames(sam)<-c('gyerek', 'hossz','l')
colnames(sam)<-c(1:ncol(sam))
Dhossz2<-matrix(0, 100, nrow(fafa$edge))
colnames(Dhossz2)<-Mt[,'hossz']
for (j in 1:nrow(Dhossz2)){
  for (i in 1:nrow(Mt)){
    if (Mt[i,'gyoke']==0 & Mt[i,'hossz']!=100) {
                Dhossz2[j,i]=j}
    else if (Mt[i,'gyoke']==0 &
                Mt[i,'hossz']==100) {Dhossz2[j,i]=100}
    else if (Mt[i,'szulo']== sam['gyerek',1] &
        j<sam['hossz',1]+Mt[i,'hossz']){Dhossz2[j,i]<-
        (Mt[i,'hossz']+sam['hossz',1]-j)}
    else if (Mt[i,'szulo']== sam['gyerek',2] &
        j<sam['hossz',2]+Mt[i,'hossz']){Dhossz2[j,i]<-
        (Mt[i,'hossz']+sam['hossz',2]-j)}
    else if (Mt[i,'szulo']== sam['gyerek',1] &
        j>=sam['hossz',1]+Mt[i,'hossz']){Dhossz2[j,i]<-1}
    else if (Mt[i,'szulo']== sam['gyerek',2] &
        j>=sam['hossz',2]+Mt[i,'hossz']){Dhossz2[j,i]<-1}
    else if (Mt[i, 'szulo']==h[1] & Mt[i,'szulo']!=
        sam['gyerek',1] & Mt[i,'szulo']!= sam['gyerek',2] &
        Mt[i,'gyoke']!=0 & j>sam['hossz',1]+sam['l',1]){
        Dhossz2[j,i]<-100-j}
    else if (Mt[i, 'szulo']==h[2] & Mt[i,'szulo']!=
        sam['gyerek',1] & Mt[i,'szulo']!= sam['gyerek',2] &
        Mt[i,'gyoke']!=0 & j>sam['hossz',2]+sam['l',2]){
        Dhossz2[j,i]<-100-j}
    else {Dhossz2[j,i]<-Mt[i,'hossz']}
    }
```

```
}
m<-list()
for (i in 1:length(fafa$tip.label)){
  g[i]<-length(nodepath(fafa)[[i]])
  if (length(nodepath(fafa)[[i]])==max(g)){m<-
        rbind(m,nodepath(fafa)[[i]])}
  else {m<-rbind(m,rep(0,max(g)))}
}
as.numeric(c(m))
m<-matrix(as.numeric(c(m)),nrow(m),ncol(m))
m
o<-c()
for (i in 1:nrow(m)){
  if (i==nrow(m)){}
  else if (i<nrow(m) & m[i,2]==m[i+1,2] &
        sum((m[i,]<m[i+1,]))){o<-c(o,m[i,ncol(m)])}
}
Foly<-matrix()
path<-nodepath(fafa)
folyamatok<-list(c(),c(),c(),c(),c(),c())


tobbs_rv1v2=function(t, v1, v2, phy){
  cwe3 <- reorder(fafa)
  my_tobbs=function(t){
    for (i in 1:nrow(cwe3$edge)){
      D[[i]] <- Merton(0,0.01,0,0,0,Dhossz2[t,i],
                        Dhossz2[t,i])
    }
    #for (i in 1:length(D)){
    #  if (M[i,'gyoke']==0) {}
    #  else if (M[i,'szulo']%in%sam['gyerek',] &
              M[i,'gyerek']%in%h){D[[i]]<-D[[i]]+
              rnorm(1, mean = 0, sd=0.05)}
```

```
  #   else  if  (M[i, 'szulo']%in%h & M[i, 'gyerek']%in%o){
                D[[i]]<-D[[i]]+rnorm(1, mean = 0, sd=0.05)}
  #}
  #szimulaio a fan
  llt <- Dhossz2[t,] + 1
  for (i in 1:nrow(cwe3$edge)) {
    ppt <- which(cwe3$edge[, 2] == cwe3$edge[i, 1])
    if (length(ppt) > 0)
      D[[i]] <- D[[i]] + D[[ppt]][llt[ppt]]
                        else D[[i]] <- D[[i]] + D[[1]][1]
  }
  return(D)
}

N<-replicate(100, my_tobbs(t))
nev<-fafa$edge[,2]
names(N)<-rep(nev,100)


NN<-split(N, rep(1:100, each=10)[1:1000])
path<-nodepath(fafa)
Folyamatok<-matrix(0,600,101)

for (k in 1:100){
  folyamatok<-list(c(),c(),c(),c(),c(),c())
 for (i in 1:length(path)){
 for (j in 2:length(path[[i]])){
   q<-which((path[[i]][j]==
                as.numeric(names(NN[[k]])))==TRUE)
  if (as.numeric(names(NN[[k]])[q])%in%
                sam['gyerek',]){folyamatok[[i]]<-
                c(folyamatok[[i]],NN[[k]][[q]])}
  else{folyamatok[[i]]<-c(folyamatok[[i]],
                NN[[k]][[q]][2:length(NN[[k]][[q]])])}
```

```
      }
    }
    foly<-do.call(rbind,folyamatok)
    Folyamatok<-rbind(Folyamatok,foly)
  }
  SUM<-matrix(0,600,101)
  for(i in 1:nrow(Folyamatok)){
    SUM[i,]<-cumsum(Folyamatok[i,])
  }
  SUMkor<-SUM[,101]
  SUMcor<-matrix(SUMkor,length(fafa$tip.label),100)

  #tapasztalati korrelacio eredeti

  ind<-c(1:length(fafa$tip.label))
  c<-matrix(0, length(N), 2)
  for (i in 1:length(N)){
    if (as.numeric(names(N)[i]) %in% ind){(c[i,1]<-
                  as.numeric(names(N)[i])) & (c[i,2]<-
                  N[i][[1]][length(N[i][[1]])])}
  }
  X<-matrix(0, nrow(fafa$edge), 100)#t=100-ban levo elemek
  for (i in 1:nrow(fafa$edge)){
    op[i]<-names(N)[i]
  }
  cv<-c(c[,2])
  X<-matrix(cv,nrow(fafa$edge),100)
  row.names(X)<-op
  ind<-rep(1:length(fafa$tip.label))
  for (i in 1:nrow(fafa$edge)){
    if(as.numeric(rownames(X)[i])%in%ind){}
    else {X<-X[-i,]}

  }
```

```
  #fv=function(x){
  #    if  (x>0.1){x<-1}
  #    else  {x<-0}
  #}
  #fXv1<-sapply(X[v1,],fv)
  #fXv2<-sapply(X[v2,],fv)

  #return(cor(fXv1,  fXv2))
  return(cor(X[v1,],  X[v2,]))
  #return(cor(SUMcor[v1,],SUMcor[v2,]))
}
```

Irodalomjegyzék

[1] Andreas E. Kyprianou, *Fluctuations of Lévy Processes with Applications*, Springer, 2013

[2] Márkus László, *Pénzügyi folyamatok matematikája*, előadás jegyzet

[3] Michaletzky György, *Kockázati folyamatok jegyzet*, 2001

[4] Prokaj Vilmos, *Sztochasztikus folyamatok*, előadás jegyzet

[5] https://evolution.berkeley.edu/evolibrary/article/0_0_0/evo_08

[6] Pénzes Zsolt *Bevezetés a filogenetikába*, egyetemi jegyzet

[7] Michael J. Landis, Joshua G. Schraiber, Mason Liang *Phylogenetic Analysis Using Lévy Processes: Finding jumps in the Evolution of continuous traits*, 2012, 193-204. oldal

[8] Paul Bastide, Mahendra Mariadassou, Stéphane Robin *Detection of adaptive shifts on phylogenies using shifted stochastic processes on a tree*, 2016,

[9] Jun Ma, *Pricing Foreign Equity Options with Stochastic Correlation and Volatility*, Annals of economics and finance 10-2, 303–327 (2009)

[10] Luke J. Harmon *Phylogenetic Comparative Methods*,2018

[11] http://phytools.org/eqg/Exercise_4.1

[12] http://www.thinksciencemaurer.com/evidence-for-evolution/