



CHALMERS



GÖTEBORGS UNIVERSITET

Itererade slumpmässiga funktioner

*Examensarbete för kandidatexamen i matematik vid Göteborgs universitet
Kandidatarbete inom civilingenjörsutbildningen vid Chalmers*

Vilhelm Agdur
Daniel McKelvey

Institutionen för matematiska vetenskaper
Chalmers tekniska högskola
Göteborgs universitet
Göteborg 2018

Itererade slumpmässiga funktioner

Examensarbete för kandidatexamen i matematik vid Göteborgs universitet

Vilhelm Agdur

*Kandidatarbete i matematik inom civilingenjörsprogrammet Teknisk matematik
vid Chalmers*

Daniel McKelvey

Handledare: Serik Sagitov

Examinator: Maria Roginskaya och Marina Axelson-Fisk

Institutionen för matematiska vetenskaper

Chalmers tekniska högskola

Göteborgs universitet

Göteborg 2018

Populärvetenskaplig presentation

Många olika fenomen, även sådana som egentligen inte involverar någon slump, förstås bäst med hjälp av sannolikheteoretiska modeller. Ett exempel som är viktigt inom biologi och medicin är hur proteiner viker sig, antingen korrekt för att uppnå sin funktioner, eller felaktigt, så att proteinet inte har någon effekt eller till och med orsakar sjukdom.

Ett protein består av en eller flera långa kedjor av aminosyror som viker sig till en möjligen mycket komplicerad tredimensionell struktur, som ger proteinet de egenskaper det behöver för att fungera i kroppen.

Vikningen av proteinet drivs, såsom varje spontan fysikalisk process, av att minimera energin. Det vill alltså undvika till exempel spänningar i strukturen och att vattenavstötande aminosyror hamnar i kontakt med vattnet på utsidan. Den kan dock inte spontant hoppa direkt till det energiminimerande tillståndet, utan måste göra detta i steg. Om något steg ökar energin blir övergången omöjlig.

Alltså kan vi inte lösa problemet med hur proteinet viker sig enbart genom att leta upp vilket tillstånd som har lägst energi – det är inte säkert att det tillståndet alls är nåbart från utgångspositionen.

En ytterligare komplikation kommer från att proteinet befinner sig i vatten, så den kan få små "knuffar" av vattenmolekylerna som ger den lite extra energi, och ändrar hur den viker sig. Alltså kommer vi inte att ha att proteinet alltid är i det färdigvikta stadiet, utan ibland kommer den att få en liten knuff bakåt mot ett tillstånd med högre energi.

Dessa små knuffar, som väsentligen är slumpmässiga, och andra liknande effekter, gör att vikningen av proteinet inte är en helt deterministisk process. I varje skede kan det tänkas finnas flera möjliga sätt att vika, och vilken väg proteinet tar beror på slumpen i hur det blir knuffat.

Som tur är är det här en på vissa sätt lätthanterlig slump. Proteinets har inget minne, och det har, oavsett vad homeopater må anse, inte heller vattnet. Hur proteinet kommer att vika sig i nästa steg må alltså vara slumpmässigt, men det beror bara på proteinet ser ut just nu, inte på hur det kom att hamna där, eller hur länge det har hållit på att vika sig.

Detta betyder att hela processen i vilken proteinet viker sig kan, såsom [Pande et al., 2010] diskuterar, modelleras som en så kallad *Markovkedja*. Idén är att hela processen kommer vara helt bestämd av vilka tillstånd den kan ta, och sannolikheterna att gå från ett givet tillstånd till ett annat. Bägge dessa saker kan mätas upp experimentellt. Att göra detta kräver inte att vi observerar ett protein genomgå hela vikningen, utan det räcker att

ha en blandning av många olika tillstånd, och se hur de övergår i varandra, vilket sparar resurser.

När man väl har denna modell kan verktyg från sannolikheteorin användas för att besvara frågor om systemet. Föreställ dig att vi häller en stor mängd nyproducerat och ovikt protein i vatten. Precis i början kommer 100% vara i det helt ovikta tillståndet, men de börjar omedelbart vika sig, och efter någon viss tid uppnås ekvilibrium, och procentandelarna av olika tillstånd slutar ändra på sig.

Det visar sig att denna ekvilibriumfördelning väldigt ofta kommer vara deterministiskt bestämt, även om tillståndet hos varje enskilt protein är slumpmässigt så kommer vi alltid att hamna i samma ekvilibrium om vi har många proteiner.

Om detta ekvilibriumfördelning kan vi alltså ställa frågor som: Hur stor andel av proteinerna är i det önskade färdigvikta tillståndet? Hur många har fått en liten knuff av extra energi och är i något tidigare tillstånd? Hur många har blivit vikta på fel sätt, och är i något oönskat tillstånd?

Förutom frågor om vad detta ekvilibriumfördelning är, så kan vi också fråga hur lång tid det tar, från någon given fördelning, för systemet att jämna ut sig och uppnå ekvilibrium.

I vårt sannolikheteoretiska språk kallas denna ekvilibriumfördelning för den *stationära fördelningen*. Frågan vi har undersökt i detta arbete är hur man matematiskt kan studera tiden det tar att konvergera mot den stationära fördelningen, inte bara i fall som med proteiner där vi har ändligt många tänkbara tillstånd, utan också i mer komplicerade fall med oändligt många tänkbara tillstånd.

Sammanfattning

Vi ger en kort introduktion till hur itererade slumpmässiga funktioner inducerar en markovkedja, samt till konvergens av sannolikhetsmått. Vi presenterar sedan Letacs sats, som ger förutsättningar för existensen hos en stationär fördelning i termer av Lipschitzkonstanterna för funktionerna.

Vi studerar sedan till vilken grad satsen överlever utan Lipschitzkonstanter, och presenterar en generell sats som ger existens av en stationär fördelning, med andra förutsättningar. Vi studerar också huruvida satsen fortfarande håller om vi släpper på antaganden om oberoende och likafördelning, alltså släpper på att processen skall vara markovsk och tidshomogen. Vi ger en generalisering av Letacs sats som delvis innefattar även detta fall.

Abstract

We give a short introduction to how an iterated random function induces a Markov chain, and to convergence of probability measures. We then present Letacs theorem, which gives conditions for the existence of a stationary distribution in terms of the Lipschitz constants of the functions.

We then study to what degree the theorem survives without Lipschitz constants, and present a general theorem that gives existence of a stationary distribution, with different conditions. We also study whether the theorem still holds if we relax the assumption of independence and identical distribution, that is, relax the assumption that the process is markovian and time-homogeneous. We give a generalisation of Letacs theorem that partially covers also this case.

Innehåll

1	Bakgrund	6
2	Itererade slumpmässiga funktioner	7
3	Kopplingar av markovkedjor	7
4	Koppling mellan markovkedjor	9
5	Konvergenshastighet och avstånd mellan sannolikhetsmått	12
6	Letacs sats	14
7	Bortom Lipschitzkonstanterna	15
8	Generalisering av Letacs sats – bortom tidshomogeniteten och markoviteten	17
9	Ytterligare frågor	21
A	Tekniska definitioner och bevis	23

Förord

Med tack till vår handledare Serik Sagitov.

Arbetet på denna kandidatarbetesrapport var uppdelat så att Daniel skrev avsnitt fyra, och Vilhelm skrev resten. Varje person skrev de delar den ansvarat för att producera. Tidslogg och dagbok för projektet fördes. Inga relevanta etiska aspekter finns att behandla, eftersom arbetet är rent teoretiskt.

1 Bakgrund

En markovkedja är en stokastisk process som “inte minns det förflutna”, alltså, vad som händer beror enbart på det nuvarande tillståndet, och inte på hur vi hamnade där. De enklaste exemplen har ändligt många tänkbara tillstånd, eller heltalsvärda tillstånd.

Till exempel kan man tänka sig att vi singlar en slant, och räknar antalet gånger vi får krona. Våra tänkbara tillstånd blir då heltal. Eller så räknar vi om vi har fått ett udda eller jämnt antal krona – och får bara två tänkbara tillstånd.

För dessa enkla exempel kan vi skriva upp alla sannolikheterna för att övergå mellan två tillstånd. Om vi låter X_n vara antalet krona efter vi har singlar slanten n gånger, får vi att

$$\mathbb{P}(X_n = m | X_{n-1} = m) = \mathbb{P}(X_n = m + 1 | X_{n-1} = m) = \frac{1}{2}$$

och om vi låter

$$Y_n = \begin{cases} 1 & \text{om } X_n \equiv 1 \pmod{2} \\ 0 & \text{om } X_n \equiv 0 \pmod{2} \end{cases}$$

får vi att

$$\begin{aligned} \mathbb{P}(Y_n = 0 | Y_{n-1} = 0) &= \mathbb{P}(Y_n = 1 | Y_{n-1} = 0) \\ &= \mathbb{P}(Y_n = 1 | Y_{n-1} = 1) \\ &= \mathbb{P}(Y_n = 0 | Y_{n-1} = 1) = \frac{1}{2} \end{aligned} \tag{1}$$

och i bägge fallen är alla andra övergångssannolikheter noll.

Om vi nu ställer oss frågan om vad som händer efter väldigt stor tid framträder en tydlig skillnad mellan de två processerna. För den första väntar vi oss att $X_n \approx \frac{n}{2} \pm \sqrt{n}$, så vi behöver veta hur lång den väldigt långa tiden är för att kunna besvara frågan.

Däremot, för den andra, så väntar vi oss snabbt att bägge tillstånden skall vara lika sannolika, så vi kan svara $\mathbb{P}(Y_n = 0) \approx \mathbb{P}(Y_n = 1) \approx \frac{1}{2}$ utan att behöva veta den exakta längden på tiden.

Det koncept vi trevar efter här är den *stationära fördelningen* för våra processer. En tolkning av det är just hur processen ser ut efter lång tid. En annan snarlik tolkning är som det genomsnittliga tillståndet hos processen om den förlöper under en väldigt lång tid.

En aningen mer abstrakt men mycket användbar tolkning är som den unika invarianta fördelningen – där vi säger att π är invariant om, ifall $X_0 \sim \pi$, även $X_1 \sim \pi$. Sannolikheterna för de olika tillstånden ändras alltså inte med tiden om vi börjar i en invariant fördelning.

Som vi observerat i våra exempel existerar inte alltid en stationär fördelning. I vårt första exempel går processens tillstånd alltid mot oändligheten, så, informellt sagt, väntar vi oss att processen efter oändlig tid befinner sig i oändligheten, vilket inte är ett faktiskt tillstånd för processen.

Vi kan också få en process som saknar stationär fördelning om vi har någon form av periodicitet eller determinism i vår process. Som ett triviellt exempel har inte processen som är konstant någon stationär fördelning – om X_n alltid är lika med X_0 kan vi självklart inte förutspå värdet på X_n utan referens till X_0 .

Frågan om existensen hos en stationär fördelning är alltså inte helt enkel, och det är inte heller i allmänhet lätt att avgöra vad den är. I fallet med ändligt många möjliga tillstånd kan problemet i princip reduceras ner till att skriva upp en matris av övergångssannolikheter och räkna på den, men i allmännare fall behöver vi mer sofistikerade metoder.

För vårt första exempel finns det en väldigt naturlig tolkning i termer av slumpmässigt valda funktioner från \mathbb{N} till \mathbb{N} . Specifikt tar vi två funktioner $h : x \mapsto x + 1$ och $t : x \mapsto x$, och får att om vi får krona i tid n är $X_{n+1} = h(X_n)$, annars är $X_{n+1} = t(X_n)$. h är alltså funktionen som lägger till ett till vår räkning, och t ändrar inte vår räkning.

Om vi nu låter, för varje i , F_i vara en slumpmässig funktion, som är h eller t bägge med sannolikhet ett, kan vi tänka rekursivt och skriva att

$$X_n = F_n(F_{n-1}(\dots F_1(X_0)\dots))$$

Motsvarande kan vi även skriva vårt andra exempel som givet av två funktioner, en som lämnar tillstånden fixa och en som byter plats på dem.

Vi har alltså funnit ett sätt att skriva om processen i termer av en slumpmässigt vald funktion. Det kommer visa sig att detta kan, i de mer allmänna fallen, vara ett produktivt perspektiv, inte bara för att etablera existensen hos en stationär fördelning, utan också för att visa hur lång tid det tar för processens tillstånds fördelning att hamna nära den stationära fördelningen.

2 Itererade slumpmässiga funktioner

Vi övergår nu från våra konkreta exempel till att ge den allmänna konstruktionen av ett itererat funktionssystem på ett metriskt rum.

Låt $(\Omega, \mathcal{F}, \mu)$ vara ett sannolikhetsrum, och (S, d) vara ett fullständigt och separabelt metriskt rum. Låt vidare \mathfrak{W} vara en samling av funktioner från S in i sig självt, och låt $\{F_i\}_{i \in \mathbb{Z}}$ vara en samling av slumpvariabler definierade på Ω som tar värden i \mathfrak{W} .

Definition 1. För varje fixt $x \in S$ och $m \in \mathbb{Z}$ kan vi definiera framåtiterationen

$$X_n^{x,m} = F_n(F_{n-1}(\dots F_m(x) \dots))$$

för alla tider $n \geq m$. Om F_i är oberoende blir $X_n^{x,m}$ en markovkedja. Motsvarande kan vi, om Z är en slumpvariabel på S , definiera $X_n^{Z,m}$, vilket också blir en markovkedja om Z och F_i är oberoende. Med aningen missbrukad notation kan vi också, om π är något sannolikhetsmått på S , skriva $X_n^{\pi,m}$, med vilket vi menar $X_n^{Z,m}$, där $Z \sim \pi$ är en slumpvariabel som realiserar π oberoende av F_i .

Vi kan också, för $x \in S$ och $m \in \mathbb{Z}$, definiera bakåtiterationen

$$Y_n^{x,m} = F_m(F_{m+1}(\dots F_n(x) \dots))$$

för alla tider $n \geq m$. Denna blir i allmänhet inte en markovkedja.

Vi kan alltså, givet våra slumpmässiga funktioner, få ut en process. Det går även ofta att gå åt andra hållet:

Sats 1. För varje tidshomogen markovkedja som tar tillstånd i ett fullständigt och separabelt metriskt rum S existerar ett slumpmässigt element av mängden av mätbara funktioner på S som genererar markovkedjan.

Bevis. Detta är Sats 1.1 i [Kifer, 1986, s. 8], se där för de tekniska detaljerna i satsens formulering. \square

Det finns även vissa satser om när kedjan kan genereras av kontinuerliga funktioner, och i fallet när S är en orienterbar kompakt Riemannsk mångfald, när den kan genereras av glatta funktioner. Se [Blumenthal and Corson, 1970] och [Quas, 1991].

3 Kopplingar av markovkedjor

Från ett itererat funktionssystem kan vi alltså alltid få en process, som blir markovsk om funktionerna är oberoende av varandra. Men det här är inte allt vi får – vi får också en koppling mellan olika trajektorier av processen.

Om vi tar två slumpvariabler Z och W på S får vi två processer $X_n^{Z,m}$ och $X_n^{W,m}$. Om Z och W är oberoende och likafördelade kommer $X_n^{Z,m}$ och $X_n^{W,m}$ vara likafördelade för alla n och m , men de kommer inte vara oberoende eftersom de drivs av *samma* funktioner F_i . Detta är vad som kallas en koppling – vi har ett par av variabler med specificerad fördelning var för sig, men som interagerar på ett intressant sätt.

Specifikt för den här kopplingen har vi att om trajektorierna sammanfaller i någon tid kommer de följa varandra i varje tid framåt – det vill säga om $X_n^{Z,m} = X_n^{W,m}$ för något n är $X_k^{Z,m} = X_k^{W,m}$ för alla $k > n$.

Vi kan använda denna egenskap hos kopplingen för att analysera stationära fördelningen, och hur snabbt vi konvergerar mot den.

Om tillståndsrummet är ändligt, alltså $S = \{s_1, s_2, \dots, s_N\}$, kan vi skapa oss N stycken kopplade processer $X_n^{s_1,0}, X_n^{s_2,0}, \dots, X_n^{s_N,0}$. Uppenbarligen måste vi då ha, oavsett fördelningen hos W , att $X_n^{W,m} = X_n^{s_i,m}$ för något i . Så om vi sedan definierar en slumpmässig tid T som första tiden alla sammanfaller, alltså

$$T = \inf \left\{ n \geq 0 \mid \forall s, s' \in S; X_n^{s,0} = X_n^{s',0} \right\}$$

så måste vi, om $T < \infty$, ha för *alla* slumpvariabler Z och W att $X_T^{Z,0} = X_T^{W,0}$.¹

Denna observation ger en uppenbar hypotes, och idé för beviset till denna.

Hypotes 2. *Ifall $\mathbb{P}(T < \infty) = 1$ så existerar det som mest en invariant fördelning för Markovkedjan.*

Följande ser ut som ett trovärdigt bevis:

“*Bevis*”. Antag att W och Z bägge är fördelade enligt en invariant fördelning, så att $X_n^{W,0}$ har samma fördelning för alla n , och likaledes för $X_n^{Z,0}$. Men vi vet att i tiden T , som enligt hypotesen är ändlig, så är $X_T^{Z,0} = X_T^{W,0}$ nästan säkert, så de har samma fördelning i den tiden. Alltså måste $X_0^{Z,0} = Z$ och $X_0^{W,0} = W$ också ha samma fördelning, vilket visar lemmat. \square

Tyvärr är det här beviset fullkomligt felaktigt. Det gäller inte alls att $X_T^{Z,0}$ måste ha den stationära fördelningen ens om Z har det. Problemet är att, bara för att $X_n^{Z,0}$ har samma fördelning för alla n , så måste inte $X_T^{Z,0}$ ha den fördelningen, eftersom T inte är oberoende av $X_n^{Z,0}$.

Exempel 3. Låt $S = \{a, b\}$, och tag tre funktioner $id, g, h : S \rightarrow S$ givna av

$$\begin{aligned} id(a) &= a, & id(b) &= b \\ g(a) &= b, & g(b) &= a \\ h(a) &= b, & h(b) &= b \end{aligned}$$

och låt F vara likafördelad på dessa tre funktioner. En direkt beräkning ger då att $X_n^{x,0}$ har övergångsmatrix

$$\begin{pmatrix} 1/3 & 2/3 \\ 1/3 & 2/3 \end{pmatrix}$$

och att stationära fördelningen π ges av att $\pi(a) = 1/3$ och $\pi(b) = 2/3$.

Om vi i stället hade studerat T ser vi att trajektorierna uppenbarligen bara kan sammanfalla då $F = h$, så T är den minsta tiden sådan att $F_T = h$. Men då har vi att $X_T^{Z,0}$ deterministiskt måste vara lika med b , oavsett fördelningen på Z .

Vårt förra bevis fungerade alltså inte, och vi är inte garanterade stationär fördelning hos $X_T^{x,0}$.²

Vad vi kan göra är att slipa vidare på vårt misslyckade bevis, och få en lite annan förutsättning som faktiskt ger satsen. För att kunna göra detta behöver vi slutligen använda oss av bakåtiterationen vi introducerade. Specifikt definierar vi en ny snarlik tid \hat{T} som

$$\hat{T} = \inf \left\{ n \geq 0 \mid \forall s, s' \in S; Y_n^{s,0} = Y_n^{s',0} \right\}$$

Trots dess mycket snarlika utseende har den här slumpmässiga tiden \hat{T} ännu starkare egenskaper. Specifikt kommer vi att ha inte bara att $Y_{\hat{T}}^{x,0}$ är konstant som funktion av $x \in S$,

¹Det här fungerar så klart oavsett storleken på S , men är för det mesta inte intressant för oändliga S , eftersom T då oftast inte blir ändligt.

²Det är dock inte tydligt att Hypotes 2 inte är sann i alla fall i det ändliga fallet – alla försök att hitta ett motexempel har misslyckats, men tyvärr har också alla försök att bevisa hypotesen misslyckats. Förutsättningarna ger enkelt att det finns en unik sluten kommunicerande klass, men att visa att denna måste vara aperiodisk är svårt.

utan också att $Y_{\hat{T}+k}^{x,0}$ är konstant som funktion av både $x \in S$ och $k \in \mathbb{N}$. När vi väl har nått denna tid står alltså den bakåttitererade processen stilla.

Alltså så kommer $Y_{\hat{T}+k}^{x,0}$ ha samma fördelning för alla k , så om vi antar att processen har en stationär fördelning, så att $\mathcal{L}(Y_n^{x,0})$ har ett gränsvärde³, kan vi gå i gräns och säga att, ifall $\hat{T} < \infty$ nästan säkert

$$\begin{aligned} \mathcal{L}\left(Y_{\hat{T}}^{x,0}\right) &= \lim_{k \rightarrow \infty} \mathcal{L}\left(Y_{\hat{T}+k}^{x,0}\right) \\ &= \lim_{k \rightarrow \infty} \mathcal{L}\left(F_1(F_2(\dots F_T(F_{T+1}(\dots F_{T+k}(x)\dots))\dots))\right) \\ &= \lim_{n \rightarrow \infty} \mathcal{L}\left(F_1(F_2(\dots F_n(x)\dots))\right) \\ &= \lim_{n \rightarrow \infty} \mathcal{L}\left(Y_n^{x,0}\right) \end{aligned}$$

men vi vet att $Y_n^{x,0}$ har samma fördelning som $X_n^{x,0}$, eftersom alla F_i är oberoende likafördelade. Alltså har vi också

$$\mathcal{L}\left(Y_{\hat{T}}^{x,0}\right) = \lim_{n \rightarrow \infty} \mathcal{L}\left(Y_n^{x,0}\right) = \lim_{n \rightarrow \infty} \mathcal{L}\left(X_n^{x,0}\right)$$

och den allmänna teorin för stationära fördelningar för Markovkedjor säger oss nu att det sista gränsvärdet, om det existerar och är samma för alla x , är den stationära fördelningen för Markovkedjan. Alltså har vi visat:

Lemma 4. *Ifall $\mathbb{P}(\hat{T} < \infty) = 1$ och Markovkedjan har en stationär fördelning, så gäller det att $Y_{\hat{T}}^{x,0}$ har den stationära fördelningen.*

Det här visar sig vara mycket användbart i tillämpningar.

Exempel 5 (Propp-Wilson-algoritmen). Antag att S har en ordning $<$, och det finns ett maximalt element $\mathfrak{s} \in S$ och minimalt element $\mathfrak{m} \in S$, så att $\mathfrak{m} \leq x \leq \mathfrak{s}$ för alla $x \in S$. Antag också att F alltid är en ordningsbevarande funktion på S . Då gäller det att

$$\hat{T} = \inf \{n \in \mathbb{N} \mid Y_n^{\mathfrak{s},0} = Y_n^{\mathfrak{m},0}\}$$

Det räcker alltså att simulera två stycken bakåttitererade processer, en som startar i \mathfrak{s} och en som startar i \mathfrak{m} . Då de två har uppnått samma värde är det värdet exakt fördelat enligt stationära processen.

Detta är användbart för att få exakta urval från fördelningar som vi annars inte kan ta exakta urval från, utan enbart har metoder som ger urval med approximativt rätt fördelning.

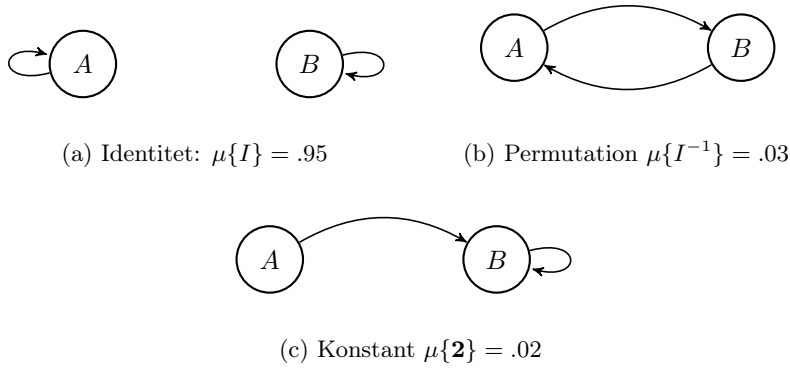
Notera här att vi, till skillnad från i fallet med framåtiterationen, behöver generera en hel lista av funktioner, eftersom vi lägger till vår nya funktion längst in. Det är alltså frestande att i stället simulera framåtiterationen, som ser ut att ha samma egenskap, men som vi har sett går detta inte.⁴

4 Koppling mellan markovkedjor

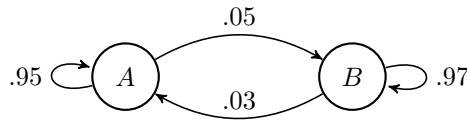
Som vi såg i det föregående avsnittet är det enkelt att gå från en itererat funktionssystem till en markovsk process. I detta avsnitt introducerar vi hur man går från en standard markovkedja till ett *kopplat* itererat funktionssystem.

³Detta antagande bör vara vad som behövs för att det tredje likhetstecknet skall vara korrekt. Alternativt är det någon egenskap hos T som behövs för att rättfärdiga den likheten. Lömskt nog gäller det nämligen inte i allmänhet för allmänna slumpvariabler T' – tänk att $S = \mathbb{Z}_2$, F är deterministiskt $f : x \mapsto x + 1$, och T är Bernoulli-0-1-fördelad. Då har $\mathcal{L}(Y_n^{x,0})$ inget gränsvärde, men alla $Y_{T'+n}^{x,0}$ har samma Bernoullifördelning och har alltså triviellt ett gränsvärde.

⁴Detta är något man varnar för i nästan alla artiklar om Propp-Wilson-algoritmen – man måste spara listan av funktioner, och algoritmen är ömtålig för missar med sådana saker. Varför detta är så är så klart inte uppenbart om man inte har sett teorin.



Låt oss nu introducera ett exempel på en koppling mellan markovkedjor. Låt Z_n vara markovkedjan som ges av Figur 1. Den har övergångsmatrisen $P = \begin{bmatrix} .95 & .05 \\ .03 & .97 \end{bmatrix}$. Tidsutvecklingen av markovkedjan ges av $Z_{n+1} = Z_n P$.



Figur 1: Markovkedja

Vi introducerar nu en markovkedja som är kopplad med Z_n . Låt nu S' representera tillstånden i S . $S' = \{[1, 0], [0, 1]\}$ ($A \sim [1, 0]$ $B \sim [0, 1]$). Låt $\boldsymbol{\psi}$ vara mängden av funktioner på S' till sig själv. Det vill säga: identitesfunktionen (I), permutationsfunktionen I^{-1} , och de konstanta funktionerna ($\mathbf{1}$, $\mathbf{2}$):

$$\boldsymbol{\psi} = \left\{ I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, I^{-1} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \mathbf{1} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \mathbf{2} = \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right\}$$

Definiera nu sannolikhetsrummet $(\boldsymbol{\psi}, \mathcal{2}^{\boldsymbol{\psi}}, \mu)$ med något sannolikhetsmått μ , och likafördelade, oberoende slumpvariabler $\{F_i\}_{i \in \mathbb{Z}}$ fördelade enligt μ . Den kopplade markovkedjan ges av $X_n^{s,0} = X_n^s = F_n(F_{n-1}(\dots F_1(x) \dots))$. För att visa att båda kedjorna är kopplade behöver vi visa att marginalfördelningarna av båda processerna är lika [Lindvall, 2002].

$$.95 = P(Z_{n+1} = A | Z_n = A) = P(X_{n+1} = [1, 0] | X_n = [1, 0])$$

Ur definitionen ser vi att högersidan är lika med sannolikheten att den $n+1$ funktionen är antingen identitesfunktionen (I) eller den konstanta funktionen ($\mathbf{1}$) det vill säga $P(Z_{n+1} = A | Z_n = \mu\{I\} + \mu\{\mathbf{1}\})$. Fortsätter vi på samma sätt får vi att

$$\begin{aligned} .97 &= P(Z_{n+1} = B | Z_n = B) = P(X_{n+1} = [0, 1] | X_n = [0, 1]) = \mu\{I\} + \mu\{\mathbf{2}\} \\ .05 &= P(Z_{n+1} = B | Z_n = A) = P(X_{n+1} = [0, 1] | X_n = [1, 0]) = \mu\{I^{-1}\} + \mu\{\mathbf{2}\} \quad (2) \\ .03 &= P(Z_{n+1} = A | Z_n = B) = P(X_{n+1} = [1, 0] | X_n = [0, 1]) = \mu\{I^{-1}\} + \mu\{\mathbf{1}\} \end{aligned}$$

Med ett lämpligt val av $\mu : \{\} \mapsto 0, \{I\} = .95, \{I^{-1}\} = .03, \{\mathbf{1}\} = 0, \{\mathbf{2}\} = .02$ får vi att $X_n^{x,m}$ är kopplad med markovkedjan Z_n .

Låt oss nu analysera vad som händer när vi startar samma markovkedja i två olika tillstånd, i.e. $X_n^{A,0}$ och $X_n^{B,0}$. I varje steg kan tillstånden göra en av tre saker: stå stilla (I), byta plats (I^{-1}) eller gå samman ($\mathbf{1}$, $\mathbf{2}$). I detta exempel är det enkelt att räkna ut fördelningen för första tiden alla trajektorier sammanfaller d.v.s. första gången funktion $\mathbf{2}$ väljs. Det är en typisk geometrisk fördelning d.v.s. $T \sim \text{Geo}(0.02)$.

Om vi tar ett steg tillbaka och analyserar sambandet mellan övergångsmatrisen P av Z_n och väntevärdet av slumpvariabeln F_i . Får vi i detta fallet

$$P = \begin{bmatrix} .95 & .05 \\ .03 & .97 \end{bmatrix} = 0.95 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 0.03 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + 0 \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} + 0.02 \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix}$$

Det är uppenbart att denna likhet är ekvivalent med Ekvation 2, samt att uppdelningen inte är unik. Alltså givet en markovkedja med ett begränsat antal tillstånd. Om man hittar en mängd funktioner ($\{0, 1\}$ -matriser) Φ_i sådana att övergångsmatrisen kan skrivas som $P = \sum_{i=0}^N a_i \Phi_i$. Kan man skapa en kopplad markovkedja som beskrivs av $X_n^x = \Phi_n(\Phi_{n-1}(\dots \Phi_1(x) \dots))$.

♣ (Figur 2a-2c)

Sats 6. Låt $P \in [0, 1]^{N \times N}$ vara en stokastisk matris. Då existerar en mängd unära stokastiska matriser $\Phi_i \in \{0, 1\}^{N \times N}$ och konstanter $a_i \in [0, 1]$ $\sum a_i = 1$ sådana att

$$P = \sum a_i \Phi_i$$

Bevisiden är att i upprepade steg sätta ett element till noll genom att dra ifrån en $\{0, 1\}$ -stokastisk matris med en skalfaktor. Det innebär att vi kan maximalt ha $N(N-1)+1$ olika $\{0, 1\}$ -matriser som då kan beskriva övergångsmatrisen P .

Sats 7. Låt Λ vara en matris med positiva element som uppfyller $\lambda \mathbf{1} = \mathbf{1} \Lambda$ $\lambda \in \mathbb{R} \setminus 0$. Då existerar det en $\{0, 1\}$ -stokastisk matris Φ ($\mathbf{1} = \mathbf{1} \Phi$) och en konstant a sådan att $\Lambda - a \Phi$ har ett element $\Lambda'_{i,j} = 0$ och $\Lambda_{i,j}$. Dessutom uppfyller $\lambda' \mathbf{1} = \mathbf{1} \Lambda'$

Bevis. Då Λ har ett nollskiljt egenvärde måste det finnas ett element i varje rad som är nollskiljd. Låt nu m_k vara det största elementet i rad k . Och låt a vara det minsta av mängden $a = \min_i m_i$. I varje rad k finns det nu en icke-tom mängd M_k bestående av de element som är större än eller lika med a . Vi kan nu skapa en $\{0, 1\}$ -stokastisk matris Φ genom att för varje rad k välja en position (k,j) där element $P_{k,j}$ är ett element i mängden M_k (t.ex. det minsta elementet i M_k). Kalla den nya matrisen $P' = P - a \Phi$. Alla element i P' är större än eller lika med noll enligt konstruktion. Dessutom måste det finnas ett element $P'_{k,j} = 0$ sådant att $P_{k,j} \neq 0$. Ty det finns en mängd M_k där det största elementet är a . Således är $P'_{k,j} = P_{k,j} - a \Phi_{k,j} = a - a * 1 = 0$.

För att visa den andra delen av satsen använder vi att $\Lambda - a \Phi = \lambda \mathbf{1} - a \mathbf{1} = (\lambda - a) \mathbf{1}$ \square

Bevis. Sats 3

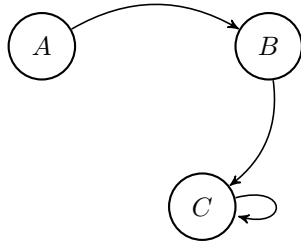
Enligt Sats 7 existerar det en matris Φ_1 och en konstant a_1 sådan att $P_1 = P - a_1 \Phi_1$, $(1 - a_1) \mathbf{1} = \mathbf{1} \Phi_1$. Dessutom har P_1 ett element $P_{i,j} = 0$ sådan att $P_{i,j} \neq 0$. Vi kan upprepa samma moment till något m sådant att P_m är lika med nollmatrisen. Det innebär att

$$\mathbf{0} = P - \sum_{k=1}^m a_k \Phi_k$$

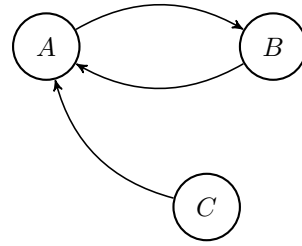
$$P = \sum_{k=1}^m a_k \Phi_k$$

\square

Vi återgår till att analysera första tiden då alla trajektorier sammanfaller. För vårt enkla fall var det enkelt att beräkna fördelningen av T . För stora matriser kan man bara få en övre begränsning. Det som gör att det inte går i dimensioner över 2 är att det inte bara finns permutationer och konstanta funktioner, utan funktioner som permuterar vissa tillstånd och är konstanta i andra exempel är



(a) Blandad funktion, Φ_A



(b) Blandad funktion Φ_B

Däremot kan man kombinera sekvenser av funktioner som som betar sig som konstanta funktioner. Till exempel är $\Phi_B(I(\Phi_A(x))) = \Phi_B(\Phi_A(x)) = x \mapsto A$, men $\Phi_A(\Phi_B(x))$ är inte en konstant funktion. Så antalet sekvenser är uppräknligt oändligt och man får nöja sig med en övre begränsning.

5 Konvergensthastighet och avstånd mellan sannolikhetsmått

Vår definition av T ger oss ett första konkret steg till att mäta hur lång tid det tar för markovkedjan att uppnå den stationära fördelningen – vi vet att den måste ha gjort det vid tid T , så vi hade kunnat studera saker som väntevärde och varians hos den variabeln för att få övre begränsningar på hur lång tid det tar att hamna i den stationära fördelningen.

Dock vore detta, så snart vi går över till fallet med oändliga eller kontinuerliga tillståndsrum, mycket restriktivt. Det är inte alls säkert att T skulle vara ändlig, även om processen har en stationär fördelning. Vi kan närma oss den stationära fördelningen utan att någonsin exakt uppnå den.

Vad vi behöver är först och främst ett koncept om vad det betyder för en följd av sannolikhetsmått att konvergera mot ett sannolikhetsmått – så vi kan få en matematisk betydelse i uttrycket “närma sig den stationära fördelningen” – och sedan också ett sätt att mäta avstånd mellan sannolikhetsmått. Dessa koncept måste också fungera i det allmänna fallet, inte bara när vi har ändligt många tänkbara tillstånd.

Det rätta konceptet av konvergens, som generaliserar konvergens i fördelning hos reellvärda slumpvariabler, visar sig vara *svag konvergens*. Idén vi letar efter är att konvergens hos fördelningarna skall betyda att väntevärdet av alla observationer vi kan göra konvergerar mot rätt sak. Formellt blir detta:

Definition 2. En följd av slumpvariabler Z_n på S konvergerar svagt mot en slumpvariabel Z på S ifall, för varje begränsad och kontinuerlig funktion $f : S \rightarrow \mathbb{R}$,

$$\mathbb{E}[f(Z_n)] \rightarrow \mathbb{E}[f(Z)]$$

Det finns en stor samling ekvivalenta definitioner av denna form av konvergens, som ger olika perspektiv på vad den innebär. De flesta av dem, inklusive den vi just gav, kommer vi aldrig att behöva befatta oss med.

En annan ekvivalent⁵ definition ger samlingen av alla sannolikhetsmått på S en metrik, och säger att konvergens i denna metrik är svag konvergens.

Definition 3. Låt S vara ett metriskt rum, med Borelsigmaalgebra \mathcal{B} . Vi definierar Prokhorov-avståndet mellan två sannolikhetsmått μ och ν på (S, \mathcal{B}) som

$$\rho(\mu, \nu) = \inf \{ \epsilon > 0 \mid \forall A \in \mathcal{B} : \mu(A) < \nu(A^\epsilon) + \epsilon, \nu(A) < \mu(A^\epsilon) + \epsilon \}$$

där A^ϵ är en *epsilonomgivning* av mängden A .

⁵Så länge S är ett fullständigt och separabelt metriskt rum.

Detta definierar en metrik på rummet av sannolikhetsmått på S som metrizerar svag konvergens av sannolikhetsmått. Om S är fullständigt och separabelt blir också rummet av sannolikhetsmått det.

Ifall S är diskret blir självklart $A^\epsilon = A$ för $\epsilon < 1$, så avståndet mellan två sannolikhetsmått blir precis den största skillnaden i sannolikheten de ger en mängd. I det kontinuerliga fallet tillåter vi också att vi gör mängden lite större, så den exakta platsen sannolikheten ligger på spelar mindre roll.

Även denna definition kommer vi inte behöva befatta oss med. Istället använder vi ett lemma, som ger en tredje ekvivalent definition.

Lemma 8. *Låt X och Y vara slumpvariabler på S med fördelningar μ och ν . Då gäller det att $\rho(\mu, \nu) \leq \delta$ om och endast om X och X' kan kopplas så att*

$$\mathbb{P}(d(X, X') > \delta) < \delta$$

Bevis. Detta är Sats 6.9 i [Billingsley, 2011, s. 74]. □

Detta lemma kan vi väldigt direkt använda i fallet med ändligt tillståndsrum. Samma observationer som innan Lemma 2 ger oss nämligen:

Korollarium 1. *Ifall S är ändligt gäller för alla slumpvariabler Z och W på S att*

$$\rho(X_n^{Z,0}, X_n^{W,0}) \leq \mathbb{P}(T > n)$$

Bevis. Såsom vanligt tar vi, när S är ändligt, metriken på S som att avståndet mellan alla distinkta punkter är ett. Vi har då att

$$\begin{aligned} \mathbb{P}((X_n^{Z,0}, X_n^{W,0}) > \mathbb{P}(T > n)) &= \mathbb{P}((X_n^{Z,0}, X_n^{W,0}) > 0) \\ &= \mathbb{P}(X_n^{Z,0} \neq X_n^{W,0}) \end{aligned}$$

och vi vet att om $n \geq T$ så måste vi ha $X_n^{Z,0} = X_n^{W,0}$, så den sista sannolikheten kan inte vara större än sannolikheten att $T > n$, vilket enligt Lemma 8 ger oss påståendet. □

Ett liknande argument som det i Korollarium 1 kommer, i det allmänna fallet, att ge oss en skattning av avståndet mellan fördelningen i tid n och den stationära fördelningen i termer av hur snabbt trajektorierna hamnar nära varandra.

En viktig sak att ha i åtanke är att fördelningen hos T beror inte bara på övergångssannolikheterna hos processen, utan också på hur vi har valt vår koppling. Korollarium 1 kan alltså ge drastiskt olika goda olikheter beroende på vårt val av koppling, och Lemma 2 ger ett tillräckligt men alls inte nödvändigt villkor.

Exempel 9. Vi återvänder till exemplet där vi singlar slant och räknar ifall vi har fått ett jämnt eller udda antal krona. Vi har alltså $S = \mathbb{Z}_2$, och övergångssannolikheter enligt (1).

Det finns en uppenbar koppling som vi får genom att projicera ner vår koppling h, t i första exemplet på \mathbb{Z}_2 . I denna har vi alltså två funktioner $h, t : \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$ givna av att $h : x \mapsto x + 1$, $t : x \mapsto x$, och bägge är lika sannolika. Tyvärr får vi här att trajektorierna *aldrig* konvergerar mot varandra, så $\mathbb{P}(T = \infty) = 1$, och varken Lemma 2 eller Korollarium 1 ger oss någon information alls.

Situationen förändras radikalt om vi gör ett klokare val av koppling. Tag istället $\alpha, \beta : \mathbb{Z}_2 \rightarrow \mathbb{Z}_2$ givna av $\alpha : x \mapsto 0$, $\beta : x \mapsto 1$, bägge lika sannolika. Även detta blir en koppling av samma markovkedja, men denna ger oss $\mathbb{P}(T = 1) = 1$, och vi får att vi uppnår den stationära fördelningen efter ett enda steg.

[Kazakevičius, 2012] diskuterar Exempel 9, och etablerar följande resultat:

Sats 10 (Kazakevičius, 2012). *Antag att S är uppräknligt, F ger en markovkedja som har stationär fördelning, och $\{F(s)\}_{s \in S}$ är en oberoende samling slumpvariabler. Då gäller det att $\mathbb{P}(T = \infty) = 0$.*

6 Letacs sats

Vi går nu slutligen helt över till det mest allmänna fallet, när S tillåts vara vilket fullständigt och separabelt metriskt rum som helst. Vi presenterar först en tidigare känd sats – Letacs sats – om existens hos och konvergens mot en stationär fördelning i termer av egenskaperna hos de slumpmässiga funktionerna, som antas vara oberoende och likafördelade, och presenterar sedan två olika generaliseringar. Dels går vi mot svagare antaganden om funktionerna, och dels går vi mot att inte anta att de är oberoende eller likafördelade.

Formuleringen av den vanliga varianten av Letacs sats är tagen direkt ur [Diaconis and Freedman, 1999], och beviset av vår generalisering av den till det tidsinhomogena fallet tar sin struktur och idé helt från deras bevis av den vanliga varianten.

Men för att kunna göra detta behöver vi först ett antal ytterligare definitioner:

Definition 4. *En samling av oberoende slumpvariabler $\{U_i\}_{i \in I}$ har likformigt algebraisk (höger-)svans ifall det existerar positiva konstanter α, β sådana att för varje $u > 0$ och varje $i \in I$,*

$$\mathbb{P}(U_i > u) \leq \frac{\alpha}{u^\beta}$$

En enskild slumpvariabel U som uppfyller detta sägs ha algebraisk svans.

Exempel 11. Om det finns en konstant C sådan att $\mathbb{E}[|U_i|] < C$ för alla $i \in I$ har U_i likformigt algebraisk svans. Detta följer av Markovs olikhet, som låter oss ta $\alpha = C$ och $\beta = 1$.

För varje $f \in \mathfrak{W}$ kan vi definiera K_f som den minsta Lipschitzkonstanten för funktionen f , eller ∞ om f inte är Lipschitzkontinuerlig. K_F blir en $[0, \infty]$ -värd slumpvariabel.

Om vi fixerar en referenspunkt $x_0 \in S$ så kan vi definiera en funktion ζ på \mathbb{F} , given av att $\zeta(f) = d(f(x_0), x_0)$. Även $\zeta(F)$ blir en reellvärd slumpvariabel.

Med dessa definitioner gjorda kan vi nu formulera den vanliga varianten av Letacs sats:

Sats 12 (Letac, 1986). *Antag att F_i är oberoende och likafördelade, så att $F_i \stackrel{d}{=} F$. Antag att K_F och $\zeta(F)$ bägge har algebraisk svans. Antag vidare att $\mathbb{E}[\log(K_F)] < 0$. Då har vi att*

1. *Det existerar ett unikt invariant mått π för Markovkedjan $X_n^{x,0}$.*
2. *Det existerar absoluta konstanter $r \in (0, 1)$, $C \in \mathbb{R}^+$, som inte beror på x eller n , sådana att för varje x och n*

$$\rho(\mathcal{L}(X_n^{x,0}), \pi) \leq Cr^n$$

där ρ är Prokhorovavståndet mellan sannolikhetsmått.

Sats 12 kommer att vara ett korollarium till vår generellare version av satsen, Sats 17, så vi ger enbart en grov skiss av beviset. Alla de tekniska detaljerna i en lite mer allmän kontext går att finna i beviset av den generellare varianten.

Idén är att vi kan använda Lipschitzkonstanterna och få en olikhet

$$d(X_n^{x,m}, X_n^{y,m}) \leq \left(\prod_{j=m}^n K_{F_j} \right) d(x, y)$$

Från denna observation får vi att vi kan gå över från att studera F , som vi vet väldigt lite om, till att studera K_F , som vi har ett visst grepp om, och ändå få mycket information om trajektorierna. Specifikt så vet vi att $\mathbb{E}[\log(K_F)] < 0$, och stora talens lag säger oss då för stora n att en summa av n oberoende kopior av $\log(K_F)$ med stor sannolikhet kommer ligga nära $n\mathbb{E}[\log(K_F)]$. Om vi sedan exponentierar bägge sidorna av detta får vi att

$$\sum_{i=m}^n \log(K_{F_i}) \approx (n-m)\mathbb{E}[\log(K_F)] \iff \prod_{i=m}^n K_{F_i} \approx e^{(n-m)\mathbb{E}[\log(K_F)]}$$

och eftersom m är fixt, $\mathbb{E}[\log(K_F)]$ är negativt och n stort kommer höger sida av denna skattningen vara väldigt nära noll.

Alltså kommer vi ha att trajektorierna konvergerar mot varandra. Detta ger oss, via Lemma 14, att ett eventuellt stationärt mått måste vara unikt.

Som nästa steg studerar man bakåtiterationen, och bevisar med hjälp av en liknande olikhet att $Y_n^{x,m}$, för fixt $x \in S$ och $m \in \mathbb{Z}$, nästan säkert kommer bli en Cauchyföljd. Alltså kommer den ha ett gränsvärde. Eftersom vi antagit att alla F_i är likafördelade och oberoende kommer $Y_n^{x,m}$ ha precis samma fördelning som $X_n^{x,m}$, så när vi tar gränsvärdet när n går mot oändligheten är det rimligt att tro att vi får något som är fördelat som "processen efter oändlig tid", alltså enligt den stationära fördelningen. Detta visar sig också vara fallet.

7 Bortom Lipschitzkonstanterna

Ett sätt som Letacs sats är begränsad är att den mycket centralt beror på att vi har Lipschitzkonstanter som vi kan manipulera. Hela beviset bygger på att vi kan slippa resonera om våra faktiska funktioner genom att istället studera Lipschitzkonstanterna.

I fallet där $S = \mathbb{N}$ eller något annat rum där avstånden mellan punkter är heltal gör detta satsen i stort sett oanvändbar. Specifikt kommer då även K_F vara heltalsvärd, så kravet att $\mathbb{E}[\log(K_F)] < 0$ implicerar att det måste finnas ett utfall av $\log(K_F)$ som är mindre än 0. Men det enda möjliga sådana när K_F är heltalsvärdt är $-\infty$, och $K_F = 0$. Alltså blir ett av kraven att vi måste ha positiv sannolikhet att välja en konstant funktion.

Bara själva kravet att F är Lipschitz blir i kontexten av \mathbb{N} restriktivt – det blir till kravet att det existerar en konstant K_F beroende på F sådan att

$$\sup_{n \in \mathbb{N}} |F(n+1) - F(n)| < K_F$$

Det här kravet utesluter till exempel de flesta varianterna på en Galton-Watson-process, där $|F(n+1) - F(n)|$ är oberoende likafördelade.

Alltså är vi intresserade av hur mycket vi kan återhämta av satsen om vi släpper antagandena om Lipschitzfunktioner. Uppenbarligen kommer mycket gå förlorat, och vi kommer behöva nya förutsättningar.

En central del i beviset av Letacs sats är hur vi får nästan säker konvergens hos trajektorierna mot varandra. Alltså börjar vi från denna utgångspunkt och definierar:

Definition 5. *Låt S vara ett fullständigt och separabelt metriskt rum, $\{F_i\}_{i \in \mathbb{Z}}$ vara en samling av slumpmässiga funktioner från S in i sig självt, och $X_n^{x,m}$ vara den resulterande Markovkedjan startad i punkten $x \in S$ och tiden m . Vi säger att detta itererade funktionssystem är*

1. starkt kontraherande om vi har, för alla $x, y \in S$ och alla $m \in \mathbb{Z}$, att $d(X_n^{x,m}, X_n^{y,m}) \rightarrow 0$ nästan säkert
2. svagt kontraherande om vi, för alla $m \in \mathbb{Z}$ och alla par av slumpvariabler η, θ med värden i S som är oberoende av alla F_i , har att $d(X_n^{\eta,m}, X_n^{\theta,m}) \xrightarrow{w} 0$ svagt

Tanken i denna definition är alltså att ett kontraherande itererat funktionssystem glömmor sitt förflutna. Som vi kommer att se betyder det här att det inte kan ha två distinkta stationära fördelningar, eftersom den hade glömt vilken av dem den började i. Men först visar vi att vårt val av stark och svag som namn faktiskt är rimligt, i att den ena implicerar den andra på det väntade sättet.

Lemma 13. *Om ett itererat funktionssystem är starkt kontraherande är det svagt kontraherande.*

Bevis. Fixera något $m \in \mathbb{Z}$, och låt η och θ vara två slumpvariabler på S som är oberoende av alla F_i . Vi kan alltså specifikt låta (η, θ) realiserar av sannolikhetsmättet ν på S^2 , och låta F_i realiserar av μ på Ω .

Låt nu $A((x, y), \omega)$ vara indikatorfunktionen för händelsen att $d(X_n^{x,m}(\omega), X_n^{y,m}(\omega)) \rightarrow 0$. Detta blir en mätbar funktion. Påståendet att vårt itererade funktionssystem är starkt kontraherande säger oss nu att vi, för varje fixt utfall $(x, y) \in S^2$, har

$$\int_{\Omega} A((x, y), \omega) d\mu = 1$$

och om vi nu integrerar den här likheten över varje par (x, y) med avseende på måttet ν får vi att

$$\int_{S^2} \int_{\Omega} A((x, y), \omega) d\mu d\nu = 1$$

och eftersom A är en mätbar och begränsad funktion och vi integrerar mot ändliga mått ger oss nu Fubinis sats att

$$\int_{\Omega \times S^2} A((x, y), \omega) d(\mu \times \nu) = 1$$

det vill säga att vi har, nästan säkert med avseende på $\mu \times \nu$, att $d(X_n^{\eta,m}, X_n^{\theta,m}) \rightarrow 0$. Eftersom nästan säker konvergens implicerar svag konvergens har vi nu också att $d(X_n^{\eta}, X_n^{\theta}) \xrightarrow{w} 0$, vilket eftersom η och θ var godtyckliga ger oss att det itererade funktionssystemet är svagt kontraherande. \square

Lemma 14. *Ett svagt kontraherande itererat funktionssystem har som mest ett invariant sannolikhetsmått, det vill säga det finns som mest ett sannolikhetsmått π sådant att lagen av $F(\pi)$ är π .*

Bevis. Antag att π och π' bägge är invarianta. Låt $(\eta, \theta) \in S^2$ vara fördelat enligt $\pi \times \pi'$, oberoende av alla F_i . Från vårt antagande att vårt itererade funktionssystem är svagt kontraherande har vi nu att $d(X_n^{\eta}, X_n^{\theta}) \xrightarrow{w} 0$.

Vi har också, per definition av invariant mått, att de slumpmässiga vektorerna $(X_n^{\eta}, X_n^{\theta})$ alla har marginalfördelningar π och π' , för varje n . Vi har alltså skapat oss en följd av kopplingar mellan π och π' .

Notera nu att alla X_n^{η} har samma fördelning som η , så $X_n^{\eta} \xrightarrow{w} \eta$. Likaledes har vi att $X_n^{\theta} \xrightarrow{w} \theta$. Men vi har också $d(X_n^{\eta}, X_n^{\theta}) \xrightarrow{w} 0$, så Lemma 27 ger oss nu att $X_n^{\theta} \xrightarrow{w} \eta$. Eftersom svaga gränsvärden är unika, måste η och θ ha samma fördelning, det vill säga vi måste ha $\pi = \pi'$. \square

Här har vi alltså etablerat en första liten skärva av Letacs sats, nämligen unikheten hos ett invariant mått om det existerar. För att få fler slutsatser behöver vi fler förutsättningar, och idén är att vi vill lyfta problemet från S och studera det som ett problem på rummet av sannolikhetsmått på S , $\mathcal{M}(S)$.

Definition 6. *Om F är en slumpmässig funktion från S till S så definierar vi funktionen $\Phi : \mathcal{M}(S) \rightarrow \mathcal{M}(S)$, som tar sannolikhetsmått på S till sannolikhetsmått på S , som följer: Om X är en slumpvariabel på S fördelad enligt μ är $F(X)$ fördelad enligt $\Phi(\mu)$.⁶*

Notera nu att $\mathcal{M}(S)$ blir, under våra antaganden på S , inte bara ett topologiskt rum under topologin av svag konvergens, utan också ett fullständigt och separabelt metriskt rum $(\mathcal{M}(S), \rho)$ om vi ger det Prokhorovmetriken. Alltså kan vi tillämpa mycket analys även på den här nivån. Vi har också, i vår funktion Φ , sammanfattat hela verkan av F på S i en enda *deterministisk* funktion på $\mathcal{M}(S)$.

Det återstår nu att hitta rätt formulering av en sats som ger existens av ett stationärt mått. Kriteriet att det itererade funktionssystemet är svagt kontraherande motsvarar att Φ för olika mått närmre varandra, alltså att $\rho(\Phi^n(\mu), \Phi^n(\nu)) \rightarrow 0$ för alla mått μ och ν . Vad vi letar efter är en förutsättning för att Φ skall ha en fixpunkt.

Vi vet av varifrån problemet kom att vi inte kan låta trajektorierna divergera eller vara på för många olika ställen, vilket i termer av Φ betyder att Φ inte kan få smeta ut alla μ bredare över S . Slutligen vet vi att vår analys inte kommer att fungera om Φ inte är någorlunda snäll. Satsen vi får blir alltså som följer:

⁶Exakt vilket rum av funktioner F skall ta värden i, och vilken sigmaalgebra vi ger det rummet, kan tillåtas variera, och lämnas implicit.

Sats 15. Antag att det itererade funktionsystemet F är svagt kontraherande. Antag också att Φ är kontinuerlig i topologin av svag konvergens, samt att det existerar ett sannolikhetsmått μ sådant att följden

$$\mu, \Phi(\mu), \Phi(\Phi(\mu)), \Phi(\Phi(\Phi(\mu))), \dots$$

är stram⁷.

Då existerar ett unikt sannolikhetsmått $\pi \in \mathcal{M}(S)$ som är stationärt för markovkedjan driven av F .

Bevis. Att F är svagt kontraherande innebär att för varje fixt par μ, π så går $\rho(\Phi^n(\mu), \Phi^n(\pi))$ mot noll. Att följden $\{\Phi^n(\mu)\}_{n=0}^\infty$ är stram innebär enligt Prokhorovs sats att den är relativt kompakt, så det existerar en delföljd $\{\Phi^{n_k}(\mu)\}_{k=0}^\infty$ och ett sannolikhetsmått π sådant att

$$\Phi^{n_k}(\mu) \rightarrow \pi$$

Vad vi vill visa nu är att $\Phi(\pi) = \pi$. Att Φ är kontinuerlig ger oss att

$$\Phi(\pi) = \Phi\left(\lim_{k \rightarrow \infty} \Phi^{n_k}(\mu)\right) = \lim_{k \rightarrow \infty} \Phi^{n_k+1}(\mu)$$

så vi vill visa att det sistnämnda gränsvärdet också är π . Detta följer från att

$$\begin{aligned} \rho(\Phi^{n_k+1}(\mu), \pi) &\leq \rho(\Phi^{n_k+1}(\mu), \Phi^{n_k}(\mu)) + \rho(\Phi^{n_k}(\mu), \pi) \\ &= \rho(\Phi^{n_k}(\Phi(\mu)), \Phi^{n_k}(\mu)) + \rho(\Phi^{n_k}(\mu), \pi) \end{aligned}$$

och eftersom F är svagt kontraherande går första termen mot noll, och andra termen går mot noll per hur vi valde μ och π .

Således har vi visat att π är ett invariant mått. Att det är unikt följer av Lemma 14. \square

Man kan visa att Φ kommer vara kontinuerlig i topologin av svag konvergens ifall F nästan säkert är en kontinuerlig funktion på S . Detta antagande är alltså inte begränsande. Förutsättningen om en stram följd är i allmänhet en begränsning, men blir självklart triviell ifall S är kompakt, eftersom alla samlingar av mått då blir strama.

8 Generalisering av Letacs sats – bortom tidshomogeniteten och markoviteten

En annan begränsning i förutsättningarna i Letacs sats är att vi antar att processen är tidshomogen, det vill säga att alla F_i är likafördelade, och att processen är markovsk, i detta fall alltså att F_i är oberoende. Detta antagande används sedan på två sätt – dels för att få att våra K_{F_i} och $\zeta(F_i)$ är oberoende och likafördelade, och dels i sista steget när vi får att fördelningen för Y_∞ är stationär för markovkedjan.

Om man studerar beviset ser man snart att vi, för att få de begränsningar vi använder för att etablera existensen hos Y_∞ , absolut inte behöver att F_i är oberoende likafördelade. Vi behöver egentligen inte ens att K_{F_i} eller $\zeta(F_i)$ är det, utan bara likformiga begränsningar på deras svansar och kontroll över medelvärden av $\log(K_{F_i})$ för stora samlingar av i .

Det är alltså naturligt att fråga sig ifall vi kan klara oss utan det antagandet på F_i och få satsen även för tidsinhomogena processer som inte nödvändigtvis är markovska. Antagandet om att F_i är likafördelade oberoende är nödvändigt för sista steget, så vi kommer få någon annorlunda slutsats, och annorlunda förutsättningar.

Alltså har vi tre problem att reda ut: hur långt beviset överlever om vi tar bort antagandena om likafördelning, vad vi behöver lägga till i slutet för att få ett meningsfullt resultat, och vad ett meningsfullt resultat egentligen blir.

Vi kan inte få ett invariant mått för en sådan process – det är inte ens helt tydligt vad det skulle betyda. Däremot kan vi få något analogt, om vi tolkar om konceptet av stationär fördelning på rätt sätt.

⁷Se Definition 10 i appendix för definition av vad det innebär för en samling av sannolikhetsmått att vara stram.

Specifikt vet vi att den stationära fördelningen informellt “är fördelningen av processen efter oändlig tid”. I det tidshomogena fallet är naturligtvis detta helt symmetriskt, men i det tidsinhomogena fallet spelar det roll i vilken riktning och varifrån vi låter tiden gå mot oändligheten. Vad vi specifikt får som resultat är att det, under vissa förutsättningar, för varje n finns en fördelning som kan tolkas som “fördelningen av processen i tid n om den har förlöpt oändligt länge”.

En av förutsättningarna i satsen kommer vara att vårt argument med stora talens lag fortfarande fungerar, samt att det itererade funktionssystemet i övrigt beter sig liknande ett som faktiskt är oberoende likafördelat. Vi ger alltså först tre definitioner.

Definition 7. *En samling av slumpvariabler $\{\xi_i\}_{i \in \mathbb{Z}}$ sägs uppfylla stora talens lag med väntevärde C ifall det existerar en konstant $C \in \mathbb{R}$ sådan att om c är något reellt tal större än C så finns det konstanter $A \in \mathbb{R}^+$ och $r \in (0, 1)$ sådana att*

$$\mathbb{P} \left(\frac{\sum_{i=m}^{m+n} \xi_i}{n} > c \right) \leq Ar^n$$

för alla $m \in \mathbb{Z}$ och $n > 0$.

Typexemplet på när en samling av slumpvariabler uppfyller detta ges av en variant på den eponyma stora talens lag:

Sats 16 (Variant av stora talens lag). *Låt ξ_i vara oberoende och likafördelade, där vi tillåter att $\mathbb{P}(\xi_i = -\infty) > 0$. Antag att det finns positiva ändliga konstanter α och β sådana att*

$$\mathbb{P}(\xi_i > v) < \alpha e^{-\beta v}$$

för alla $v > 0$.

Låt ξ vara fördelad som ξ_i . Vi har då att

1. $-\infty \leq \mathbb{E}[\xi] < \infty$
2. Om c är något reellt tal större än $\mathbb{E}[\xi]$ så finns det positiva ändliga konstanter A och $r \in (0, 1)$ sådana att

$$\mathbb{P} \left(\frac{\xi_1 + \xi_2 + \dots + \xi_n}{n} > c \right) \leq Ar^n$$

för alla $n > 0$. Konstanterna A och r tillåts bero på c och på lagen för ξ , men inte på n .

Bevis. Detta är Lemma 5.2 i [Diaconis and Freedman, 1999, s. 59]. □

Definition 8. *Vi säger att ett sannolikhetsmått π är fullständigt invariant för processen genererad av F_i ifall $\mathcal{L}(X_n^{\pi, m}) = \mathcal{L}(X_{n'}^{\pi, m})$ för alla $n, n' \geq m$.*

Definition 9. *Det itererade funktionssystemet $\{F_i\}_{i \in \mathbb{Z}}$ sägs vara långsiktigt tidssymmetriskt om det för varje $n \in \mathbb{Z}$ existerar konstanter $A \in \mathbb{R}^+$ och $r \in (0, 1)$ sådana att det gäller för varje $x \in S$ och $k \in \mathbb{N}$ att*

$$\rho(\mathcal{L}(X_n^{x, n-k}), \mathcal{L}(Y_{n+k}^{x, n})) < Ar^k$$

Med dessa definitioner gjorda kan vi nu formulera vår allmännare version av Letacs sats:

Sats 17 (Generellare version av Letacs sats). *Antag att K_{F_i} och $\zeta(F_i)$ bägge har likformigt algebraisk svans. Antag vidare att samlingen $\log(K_{F_i})$ uppfyller stora talens lag med väntevärde C för något $C < 0$. Då existerar det, för varje m , som mest ett fullständigt invariant sannolikhetsmått π .*

Antag också att det itererade funktionssystemet är långsiktigt tidssymmetriskt. Då existerar det för varje $n \in \mathbb{Z}$ och $x \in S$ en slumpvariabel $X_n^{-\infty}$, som inte beror på x , och konstanter $A \in \mathbb{R}^+$ och $r \in (0, 1)$ sådana att för varje $k \in \mathbb{N}$

$$\rho(\mathcal{L}(X_n^{x, n-k}), \mathcal{L}(X_n^{-\infty})) < Ar^k$$

$X_n^{-\infty}$ har alltså “fördelningen för processens tillstånd i tid n om den började för oändligt länge sedan”, och denna beror inte på punkten i vilken processen startade.

För att förenkla våra argument senare börjar vi med att konstatera att valet av x_0 med avseende på vilket $\zeta(F_i)$ definieras är irrelevant under satsens förutsättningar.

Lemma 18. *Antag att K_{F_i} har likformigt algebraisk svans. Om $\zeta(F_i)$ har likformigt algebraisk svans för något val av x_0 har dem det för alla val av x_0 .*

Bevis. Se appendix. □

Vårt första steg i att bevisa satsen är att begränsa avstånden mellan trajektorierna i termer av Lipschitzkonstanterna.

Lemma 19. *För varje fixt par $x, y \in S$ och varje $m \in \mathbb{Z}$ gäller deterministiskt för varje $n \geq m$ att*

$$d(X_n^{x,m}, X_n^{y,m}) \leq \left(\prod_{j=m}^n K_{F_j} \right) d(x, y)$$

Bevis. Följer triviellt med induktion i n . □

Så om vi kan få den produkten att vara liten kommer vi få att trajektorierna, för fixt ω och m , konvergerar mot varandra i n , det vill säga för stora tider n beror $X_n^{x,m}$ väsentligen bara på ω , m och n , inte på x . Från detta kommer det omedelbart att följa att ett fullständigt invariant mått nödvändigtvis är unikt.

Det är här vi utnyttjar vårt antagande om att $\log(K_{F_i})$ uppfyller stora talens lag med negativt väntevärde, och begränsar produkten genom att begränsa summan av logaritmerna av dess termer. Vi får då följande påstående:

Lemma 20. *Antag att $\log(K_{F_i})$ uppfyller stora talens lag med väntevärde $C < 0$. Då existerar, för $\epsilon \in (0, |C|)$, konstanter $A > 0$, $r \in (0, 1)$ sådana att det för alla $x, y \in S$ och alla $m \in \mathbb{Z}$ gäller för alla $n \geq m$*

$$\mathbb{P} \left(d(X_n^{x,m}, X_n^{y,m}) > e^{-(n-m)\epsilon} d(x, y) \right) \leq Ar^{n-m}$$

Bevis. Vi från Lemma 19 att

$$\begin{aligned} \mathbb{P} \left(d(X_n^{x,m}, X_n^{y,m}) > e^{-(n-m)\epsilon} d(x, y) \right) &\leq \mathbb{P} \left(\left(\prod_{j=m}^n K_{F_j} \right) d(x, y) > e^{-(n-m)\epsilon} d(x, y) \right) \\ &= \mathbb{P} \left(\frac{\sum_{j=m}^n \log K_{F_j}}{n-m} > -\epsilon \right) \end{aligned}$$

och existensen av konstanterna A och r är nu precis antagandet att K_F uppfyller stora talens lag med väntevärde C . □

Vilket, som ett korollarium, ger oss att det itererade funktionssystemet är starkt kontraherande.

Korollarium 2. *För varje fixt par av $x, y \in S$ och varje $m \in \mathbb{Z}$ har vi att $d(X_n^{x,m}, X_n^{y,m}) \rightarrow 0$ nästan säkert då $n \rightarrow \infty$.*

Bevis. Följer från Lemma 20 och Borel-Cantelli. För detaljerna, se appendix. □

Det är tydligt att Korollarium 2, Lemma 13, och Lemma 14 tillsammans ger oss det första påståendet i Sats 17.

För att studera existensen av en stationär fördelning, och för att bestämma konvergenstaktheten mot den om den existerar, behöver vi nu också bakåtiterationen. Till skillnad mot hur iterationen framåt beter sig så kommer bakåtiterationen snabbt att stabilisera sig, under förutsättningarna i satsen.

Lemma 21. Låt $\{g_i\}$ vara Lipschitzavbildningar från S in i sig självt, och x vara en punkt i S . Vi har då att

$$d(g_1(g_2(\dots g_n(x)\dots)), x) \leq \sum_{k=0}^{n-1} \left(\prod_{j=1}^k K_{g_j} \right) d(g_{k+1}(x), x)$$

Bevis. Vi har från triangelolikheten och definitionen av K_f att

$$d(f(g(x)), x) \leq d(f(g(x)), f(x)) + d(f(x), x) \leq K_f d(g(x), x) + d(f(x), x)$$

och lemmat följer från detta med induktion, om man skriver ut vad summan faktiskt är. \square

Det är nu vi utnyttjar vårt antagande om att S är fullständigt, och bevisar att följderna $Y_n^{x,m}$ nästan säkert är Cauchy för att få att den konvergerar. Vad vi vill analysera är alltså uttrycket $d(Y_{n+k}^{x,m}, Y_n^{x,m})$.

Lemma 22. Antag att $\zeta(F_i)$ och K_{F_i} bägge har likformigt algebraisk svans, och antag att $\log(K_{F_i})$ uppfyller stora talens lag med negativt väntevärde. Då existerar positiva konstanter B och C samt konstanter $r_0, r_1 \in (0, 1)$ som bara beror på x sådana att för varje $n_0 \geq m$

$$\mathbb{P}(\forall n > n_0, k \in \mathbb{N}; d(Y_{n+k}^{x,m}, Y_n^{x,m}) < Cr_1^{n_0-m}) \geq 1 - Br_0^{n_0-m} \quad (3)$$

Bevis. Beviset är långt och använder bara att geometriska summor konvergerar utan nya idéer, så det lämnas till appendix. \square

Proposition 23. Antag att $\zeta(F_i)$ och K_{F_i} bägge har likformigt algebraisk svans, och antag att $\log(K_{F_i})$ uppfyller stora talens lag med negativt väntevärde. Då existerar, för varje tid m , en slumpvariabel Y_∞^m sådan att $Y_n^{x,m} \rightarrow Y_\infty^m$ geometriskt snabbt nästan säkert, och Y_∞^m beror inte på startpunkten x .

Bevis. Fixera något m . Från Lemma 22 får vi att det existerar positiva konstanter B och C samt konstanter $r_0, r_1 \in (0, 1)$ som bara beror på x sådana att för varje $n_0 \geq m$

$$\mathbb{P}(\forall n > n_0, k \in \mathbb{N}; d(Y_{n+k}^{x,m}, Y_n^{x,m}) < Cr_1^{n_0-m}) \geq 1 - Br_0^{n_0-m} \quad (4)$$

Från detta kan man få väldigt direkt att följderna är Cauchy.

Så, tag ett godtyckligt $\epsilon > 0$, och välj n_0 stort nog att $Cr_1^{n_0-m} < \epsilon$. Definiera nu händelsen E_n som

$$E_n = \left\{ \forall k > n, j \in \mathbb{N}; d(Y_{k+j}^{x,m}, Y_k^{x,m}) < Cr_1^{n-m} \right\}$$

Vad vi vill visa är alltså att det nästan säkert finns något $n \geq n_0$ sådant att E_n inträffar. Men vi har att

$$\sum_{n \geq n_0} \mathbb{P}(E_n^c) \leq \sum_{n \geq n_0} Br_0^{n-m} < \infty$$

och enligt Borel-Cantelli inträffar alltså E_n nästan säkert bara ändligt många gånger, så E_n måste inträffa för något $n > n_0$. Alltså har vi visat att $Y_n^{x,m}$ är Cauchy, så det existerar ett gränsvärde $Y_\infty^{x,m}$.

Vad som återstår att bevisa är att $Y_\infty^{x,m}$ i själva verket är oberoende av x , och att $Y_n^{x,m} \rightarrow Y_\infty$ geometriskt snabbt.

Det första påståendet följer från att Lemma 19 fungerar även för Y_n^x , så det tillsammans med att $\log(K_{F_i})$ uppfyller stora talens lag med negativt väntevärde ger för varje par x, y , precis som för $X_n^{x,m}$, att $d(Y_n^{x,m}, Y_n^{y,m}) \rightarrow 0$ nästan säkert. Men vi har så klart också att $\lim_n d(Y_n^{x,m}, Y_n^{y,m}) = d(Y_\infty^{x,m}, Y_\infty^{y,m})$, så $Y_\infty^{y,m} = Y_\infty^{x,m}$.

Det andra påståendet följer av att observera att om E_n inträffar kan vi ta ett gränsvärde när $k \rightarrow \infty$ och även få att händelsen

$$\tilde{E}_n = \left\{ \forall k > n; d(Y_\infty, Y_k^{x,m}) < Cr_1^{n-m} \right\}$$

inträffar. Så enligt vårt föregående resonemang inträffar även \tilde{E}_n oändligt ofta, vilket betyder att $Y_n^{x,m}$ går mot Y_∞^m geometriskt snabbt. \square

När vi nu har bevisat detta är idén att använda Lemma 8 för att omvandla vårt påstående om konvergens av följderna $Y_n^{x,m}$ mot Y_∞^m till ett påstående om konvergens av fördelningen för $Y_n^{x,m}$ till fördelningen för Y_∞^m .

Lemma 24. *För varje $m \in \mathbb{Z}$ existerar det konstanter $A \in \mathbb{R}^+$ och $r \in (0, 1)$ sådana att för varje $n > m$ och varje $x \in S$*

$$\rho(\mathcal{L}(Y_\infty^m), \mathcal{L}(Y_n^{x,m})) < Ar^{n-m}$$

Bevis. Följer omedelbart från Lemma 8, eftersom Proposition 23 ger oss att vi kan koppla $Y_n^{x,m}$ och Y_∞^m så att avståndet mellan dem går mot noll geometriskt snabbt, med motsvarande begränsning på sannolikheten att de är långt ifrån varandra. \square

Med detta i handen är nu beviset av satsen enkelt.

Bevis av vår generellare Letacs sats, Sats 17. Vad vi önskar göra nu är att definiera en slumpvariabel som ”var processen befinner sig i tid n om den började i tid $-\infty$ i punkten x ”, och sedan visa att denna är oberoende av x . För att detta gränsvärde faktiskt skall existera behöver vi nu slutligen våra antaganden om långsiktig tidssymmetri.

Specifikt så har vi att

$$\rho(\mathcal{L}(Y_\infty^n), \mathcal{L}(X_n^{x,n-k})) \leq \rho(\mathcal{L}(Y_\infty^n), \mathcal{L}(Y_{n+k}^{x,n})) + \rho(\mathcal{L}(Y_{n+k}^{x,n}), \mathcal{L}(X_n^{x,n-k}))$$

och det första uttrycket går mot noll geometriskt snabbt enligt Lemma 24 och det andra går mot noll geometriskt snabbt enligt vårt antagande om långsiktig tidssymmetri. Alltså har vi att $X_n^{x,n-k}$ går mot Y_∞^n i fördelning geometriskt snabbt med k . Alltså kan vi låta det sökta $X_n^{-\infty}$ vara precis Y_∞^n . \square

Korollarium 3. *Den vanliga varianten av Letacs sats, Sats 12, håller.*

Bevis. Att $\log(K_{F_i})$ uppfyller stora talens lag med negativt väntevärde följer, tack vare att de är oberoende, av stora talens lag, Sats 16. Att de har likformigt algebraisk svans följer av att de enskilt har algebraisk svans och är likafördelade. Att systemet är långsiktigt tidssymmetriskt är triviellt, eftersom alla F_i har samma fördelning.

Likafördelningen ger också att fördelningen hos Y_∞^n inte beror på n , och denna kan vi ta som π , samt att $\mathcal{L}(X_n^{x,0}) = \mathcal{L}(X_0^{x,-n})$, vilket ger begränsningen. \square

9 Ytterligare frågor

Vad vi har gjort ger en del intressanta resultat, men det lämnar också en del otillfredställande hål i resultaten. Huvudsakligen så är Definition 9, av långsiktig tidssymmetri, mycket svårmotiverad av annat än att den får argumentet i vårt bevis att fungera. Det vore önskvärt att finna en mer naturlig formulering, som kan tolkas och bevisas. Att oberoende hos F_i implicerar den långsiktiga tidssymmetrin är triviellt, men det saknas mindre triviella implikationer. Ger till exempel någon form av stationaritet eller mixning också långsiktig tidssymmetri?

Även Sats 15 har aningen komplicerade förutsättningar, även om de är mer lättmotiverade. Förutsättningen att det existerar ett mått μ sådant att följderna $\{\Phi^n(\mu)\}_{n \in \mathbb{N}}$ är stram är visserligen triviell ifall S är kompakt eller om Markovkedjan har ett uppenbart absorberande tillstånd, men hjälp med att finna ett klokt val av μ i det allmänna fallet vore ändå bra. Hela det avsnittet använder dessutom inte bakåtiterationen alls, så den hade kunnat vara ett utforskat verktyg.

Det finns även andra spår att arbeta vidare på. Till exempel kan man visa, i det tidshomogena markovska fallet, att ifall bakåtiterationen konvergerar nästan säkert så konvergerar framåtiterationen i fördelning. [Kazakevičius, 2012] studerar den motsatta frågan i det diskreta fallet – ifall vi, givet att framåtiterationen konvergerar i fördelning får att bakåtiterationen konvergerar. Han får följande sats:

Sats 25 (Kazakevičius, 2012). *Antag att S är diskret⁸ och F_i är oberoende likafördelade och inducerar en Markovkedja som har en stationär fördelning. Då är följande påståenden ekvivalenta*

1. *Följden $Y_n^{x_0,0}$ konvergerar i sannolikhet för något $x_0 \in \mathfrak{S}$*
2. *Följden $Y_n^{x,0}$ konvergerar, som funktioner av x , mot någon slumpmässig konstant funktion på S*
3. *För alla $x, y \in S$ existerar det något n sådant att $\mathbb{P}(X_n^{x,0} \neq X_n^{y,0}) < 1$*
4. *För alla $x, y \in S$ gäller att $\mathbb{P}(X_n^{x,0} \neq X_n^{y,0}) \rightarrow 0$ då $n \rightarrow \infty$.*

Det sista påståendet är mycket snarlikt våra definitioner av starkt och svagt kontraherande, Definition 5, fast i sannolikhet och bara i det diskreta fallet. Det vore alltså intressant att studera ifall resultatet går att utvidga till det allmänna fallet, och hur det relaterar till våra resultat.

En annan tanke är att beviset av Letacs sats, och vår generalisering, moraliskt sett är en tillämpning av Banachs fixpunktssats. Detta döljs dock av hur vi behöver hantera våra olikheter. Det vore alltså intressant att skapa ett renare bevis som gör detta tydligt. En specifik idé vore att analysera under vilka förutsättningar på F vår Φ blir en kontraktion på $\mathcal{M}(S)$ – argumentet vore därefter en direkt tillämpning av Banachs fixpunktssats för att få att en stationär fördelning existerade.

Referenser

- [Billingsley, 2011] Billingsley, P. (2011). *Convergence of probability measures*. Wiley.
- [Blumenthal and Corson, 1970] Blumenthal, R. M. and Corson, H. H. (1970). On continuous collections of measures. *Annales de l'institut Fourier*, 20(2):193–199.
- [Diaconis and Freedman, 1999] Diaconis, P. and Freedman, D. (1999). Iterated random functions. *SIAM Review*, 41(1):45–76.
- [Grimmett and Stirzaker, 2001] Grimmett, G. and Stirzaker, D. (2001). *Probability and random processes*. Oxford university press.
- [Kazakevičius, 2012] Kazakevičius, V. (2012). Limits of the Letac principle. The discrete case. *Lithuanian Mathematical Journal*, 52(4):390–399.
- [Kifer, 1986] Kifer, Y. (1986). *Ergodic theory of random transformations*. Birkhauser.
- [Lindvall, 2002] Lindvall, T. (2002). *Lectures on the coupling method*. Courier Corporation.
- [Pande et al., 2010] Pande, V. S., Beauchamp, K., and Bowman, G. R. (2010). Everything you wanted to know about Markov State Models but were afraid to ask. *Methods*, 52(1):99–105.
- [Quas, 1991] Quas, A. N. (1991). On representations of Markov chains by random smooth maps. *Bulletin of the London Mathematical Society*, 23(5):487–492.
- [Strang et al., 1993] Strang, G., Strang, G., Strang, G., and Strang, G. (1993). *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA.

⁸Och således uppräknelig, eftersom vi som alltid antar S separabel.

A Tekniska definitioner och bevis

I detta appendix samlar vi de mer tekniska och perifera definitionerna, och bevis som inte är upplysande för argumentet.

Definition 10. Låt X vara något topologiskt rum med Borelsigmaalgebra \mathcal{B} . En mängd av sannolikhetsmått $\{\mu_\lambda\}_{\lambda \in \Lambda}$ på (X, \mathcal{B}) sägs vara stram ifall det för varje $\epsilon > 0$ existerar en kompakt delmängd $K_\epsilon \subset X$ sådan att för varje $\lambda \in \Lambda$, $\mu_\lambda(K_\epsilon) > 1 - \epsilon$.

Lemma 26. 1. Om U_i är ickenegativa och likformigt övre begränsade har U_i likformigt algebraisk svans.

2. Om U_i har likformigt algebraisk svans och $c \geq 0$ så har cU_i likformigt algebraisk svans.

3. Om U_i och V_i har likformigt algebraisk svans så har $U_i + V_i$ det, även om U_i och V_i är beroende.

Bevis. Första och andra påståendet är uppenbara. För påstående tre, notera att $U_i + V_i$ bara kan vara större än t om åtminstone en av U_i och V_i är större än $t/2$. Så låt α, β vara konstanter för U_i och α', β' för V_i . Då har vi, med en unionsbegränsning

$$\begin{aligned} \mathbb{P}(U_i + V_i > t) &\leq \mathbb{P}(\{U_i > t/2\} \cup \{V_i > t/2\}) \\ &\leq \mathbb{P}(U_i > t/2) + \mathbb{P}(V_i > t/2) \\ &\leq \frac{\alpha}{(t/2)^\beta} + \frac{\alpha'}{(t/2)^{\beta'}} \leq \frac{\max(\alpha, \alpha') 2^{1+\min(\beta, \beta')}}{t^{\min(\beta, \beta')}} \end{aligned}$$

och vi har funnit lämpliga konstanter för $U_i + V_i$. □

Lemma 27. Låt S vara något metriskt rum. Antag att (X_n, Y_n) är slumpmässiga element i S^2 . Om $Y_n \xrightarrow{w} Y$ och $d(X_n, Y_n) \xrightarrow{w} 0$, så har vi också att $X_n \xrightarrow{w} Y$.

Bevis. Detta är Sats 3.1 i [Billingsley, 2011, s. 27]. □

Bevis av Lemma 18. Antag att $\zeta(F_i)$ har likformigt algebraisk svans för punkten x_0 , och låt $x'_0 \in S$ vara godtycklig. Vi har då att

$$\begin{aligned} d(F_i(x'_0), x'_0) &\leq d(F_i(x'_0), F_i(x_0)) + d(F_i(x_0), x_0) + d(x_0, x'_0) \\ &\leq (K_{F_i} + 1)d(x'_0, x_0) + \zeta(F_i) \end{aligned}$$

Nu kan Lemma 26 enkelt tillämpas på denna olikhet. □

Bevis av Lemma 2. Tag något $\epsilon > 0$. Vad vi önskar visa är att det existerar en slumpvariabel N som nästan säkert är ändlig sådan att för varje $n > N$, $d(X_n^{x,m}, X_n^{y,m}) < \epsilon$. Vi kan alltså definiera en samling händelser E_k som

$$E_k = \{\omega \in \Omega \mid \forall n > k; d(X_n^{x,m}(\omega), X_n^{y,m}(\omega)) < \epsilon\}$$

och definiera vår slumpvariabel N som

$$N = \inf(k \in \mathbb{Z} \mid \mathbb{1}_{E_k} = 1)$$

så att vad vi vill visa är att E_k nästan säkert inträffar för något k .

För att visa detta vill vi använda Borel-Cantelli och Lemma 20. Lemmat ger oss konstanter $A, B > 0$ och $r \in (0, 1)$, som tillåts bero på m , sådana att

$$\mathbb{P}(d(X_n^{x,m}, X_n^{y,m}) > e^{-n} B) \leq Ar^n$$

för alla n .⁹ Tag nu ett $K \in \mathbb{N}$ stort nog att $\epsilon > e^{-K}B$. För varje $k > K$ har vi nu, med en unionsbegränsning, att

$$\begin{aligned}
\mathbb{P}(E_k^c) &= \mathbb{P}(\{\omega \in \Omega \mid \neg(\forall n > k; d(X_n^{x,m}(\omega), X_n^{y,m}(\omega)) < \epsilon)\}) \\
&= \mathbb{P}(\{\omega \in \Omega \mid \exists n > k; d(X_n^{x,m}(\omega), X_n^{y,m}(\omega)) > \epsilon\}) \\
&= \mathbb{P}\left(\bigcup_{n>k} \{\omega \in \Omega \mid d(X_n^{x,m}(\omega), X_n^{y,m}(\omega)) > \epsilon\}\right) \\
&\leq \sum_{n>k} \mathbb{P}(d(X_n^{x,m}, X_n^{y,m}) > \epsilon) \\
&\leq \sum_{n>k} \mathbb{P}(d(X_n^{x,m}, X_n^{y,m}) > e^{-n}B) \\
&\leq \sum_{n>k} Ar^n = \frac{Ar}{1-r} r^k
\end{aligned}$$

och således har vi att

$$\sum_{k>K} \mathbb{P}(E_k^c) \leq \sum_{k>K} \frac{Ar}{1-r} r^k = \frac{Ar^2}{(1-r)^2} r^K < \infty$$

och Borel-Cantelli ger oss nu att E_k^c nästan säkert inträffar enbart ändligt ofta, vilket ger oss att det nästan säkert existerar ett k så att E_k inträffar, vilket är precis vad vi behövde för att visa lemmat. \square

Bevis av Lemma 22. Enligt samma resonemang som i Lemma 19 har vi att

$$d(Y_{n+k}^{x,m}, Y_n^{x,m}) \leq K_{F_m} K_{F_{m+1}} \dots K_{F_n} d(F_{n+1}(F_{n+2}(\dots F_{n+k}(x) \dots)), x)$$

Från Lemma 21 får vi från detta att

$$\begin{aligned}
d(Y_{n+k}^{x,m}, Y_n^{x,m}) &\leq K_{F_m} K_{F_{m+1}} \dots K_{F_n} \sum_{l=0}^{k-1} \left(\prod_{j=1}^l K_{F_{n+j}} \right) d(F_{n+l+1}(x), x) \\
&= \sum_{l=0}^{k-1} \left(\prod_{j=m}^{n+l} K_{F_j} \right) d(F_{n+l+1}(x), x)
\end{aligned} \tag{5}$$

Från antagandet om att $\log(K_{F_i})$ uppfyller stora talens lag med väntevärde $C < 0$ får vi, precis som i beviset av Lemma 20, om vi tar ett $\epsilon \in (0, |C|)$, att det existerar konstanter $A > 0$ och $r \in (0, 1)$ sådana att

$$\mathbb{P}\left(\prod_{j=m}^n K_{F_j} > e^{-(n-m)\epsilon}\right) < Ar^{n-m}$$

och en unionsbegränsning ger oss nu för varje $n_0 \geq m$ att

$$\begin{aligned}
\mathbb{P}\left(\forall n \geq n_0; \prod_{j=m}^n K_{F_j} \leq e^{-(n-m)\epsilon}\right) &= 1 - \mathbb{P}\left(\bigcup_{n \geq n_0} \left\{ \prod_{j=m}^n K_{F_j} > e^{-(n-m)\epsilon} \right\}\right) \\
&\geq 1 - \sum_{n \geq n_0} \mathbb{P}\left(\prod_{j=m}^n K_{F_j} > e^{-(n-m)\epsilon}\right) \\
&\geq 1 - \sum_{n \geq n_0} Ar^{n-m} = 1 - \frac{A}{1-r} r^{n_0-m} = 1 - A'r^{n_0-m}
\end{aligned}$$

⁹Vi har bakat ihop ett par delar i olikheten vi inte behöver, eftersom vi håller m fixt.

för en konstant $A' \in \mathbb{R}^+$ och $r \in (0, 1)$.

Nu behöver vi vårt antagande om att $\zeta(F_i)$ har likformigt algebraisk svans, och att vi inte förlorar någon generalitet i att anta att $x = x_0$. Skriv, för enklare notation, $\zeta_j = \zeta(F_j) = d(F_j(x), x)$. Från vårt antagande får vi att det existerar konstanter $\alpha, \beta \in \mathbb{R}^+$ sådana att

$$\mathbb{P}(\zeta_j > s^{j-m}) < \frac{\alpha}{s^{\beta(j-m)}}$$

för alla $j \in \mathbb{Z}$.

Välj nu ett $s \in (1, e^\epsilon)$, så att $se^{-\epsilon} < 1$. Vi har då, likt vår tidigare beräkning, för alla $n_0 \geq m$, att

$$\begin{aligned} \mathbb{P}(\forall n > n_0; \zeta_n \leq s^{n-m}) &= 1 - \mathbb{P}\left(\bigcup_{n > n_0} \{\zeta_n > s^{n-m}\}\right) \\ &\geq 1 - \sum_{n > n_0} \mathbb{P}(\zeta_n > s^{n-m}) \\ &\geq 1 - \sum_{n > n_0} \frac{\alpha}{s^{\beta(n-m)}} \\ &= 1 - \frac{\alpha}{s^\beta - 1} s^{-\beta(n_0-m)} \end{aligned}$$

Slår vi ihop dessa beräkningar får vi nu att

$$\begin{aligned} &\mathbb{P}\left(\forall n, k > n_0; \prod_{j=m}^n K_{F_j} \leq e^{-(n-m)\epsilon}, \zeta_k \leq s^{k-m}\right) = \\ &= 1 - \mathbb{P}\left(\left\{\exists n > n_0; \prod_{j=m}^n K_{F_j} > e^{-(n-m)\epsilon}\right\} \cup \{\exists n > n_0; \zeta_k > s^{k-m}\}\right) \\ &\geq 1 - \mathbb{P}\left(\exists n > n_0; \prod_{j=m}^n K_{F_j} > e^{-(n-m)\epsilon}\right) - \mathbb{P}(\exists n > n_0; \zeta_k > s^{k-m}) \\ &\geq 1 - A' r^{n_0-m} - \frac{\alpha}{s^\beta - 1} s^{-\beta(n_0-m)} \geq 1 - \left(A' + \frac{\alpha}{s^\beta - 1}\right) \max(r, s^{-\beta})^{n_0-m} \end{aligned}$$

så, eftersom $s^{-\beta} \in (0, 1)$ och $s^\beta > 1$, så kan vi låta $B = \left(A' + \frac{\alpha}{s^\beta - 1}\right) \in \mathbb{R}^+$ och $r_0 = \max(r, s^{-\beta}) \in (0, 1)$.

Vi har nu, om vi använder vår olikhet (5), för varje n_0 , att med sannolikhet större än $1 - Br_0^{n_0}$, för alla $n > n_0$ och $k \in \mathbb{N}$

$$\begin{aligned} d(Y_{n+k}^{x,m}, Y_n^{x,m}) &\leq \sum_{l=0}^{k-1} \left(\prod_{j=m}^{n+l} K_{F_j}\right) d(F_{n+l+1}(x), x) \\ &\leq \sum_{l=0}^{k-1} e^{-(n+l-m)\epsilon} s^{n+l+1-m} = e^{-(n-m)\epsilon} s^{n-m} s \sum_{l=0}^{k-1} (e^{-\epsilon} s)^l \\ &= (e^{-\epsilon} s)^{n-m} s \frac{1 - (e^{-\epsilon} s)^k}{1 - e^{-\epsilon} s} \leq (e^{-\epsilon} s)^{n_0-m} \frac{s}{1 - e^{-\epsilon} s} \end{aligned}$$

där beräkningarna håller eftersom $se^{-\epsilon} \in (0, 1)$. Vi kan nu låta $C = \frac{s}{1 - e^{-\epsilon} s} \in \mathbb{R}^+$ och $r_1 = se^{-\epsilon} \in (0, 1)$. Vi har alltså funnit de sökta konstanterna. \square