

Otanta todennäköisyysjakaumista

1. Satunnaisotosten generoinnista

1.1. Yleistä sämplämisestä eli satunnaisten otosten tuottamisesta

Tämän osan yleinen teema on se, millä tavoin voimme tuottaa eli generoida satunnaisia olioita, joilla on haluttu jakauma. Esimerkiksi tavallisen napanheiton tarkoitus on tuottaa satunnaisluku, joka on tasaisesti jakautunut joukolla $\{1, 2, 3, 4, 5, 6\}$ — lauantai-iltaisen lottoarvonnan tarkoitus taas generoida (tasaisesti jakautunut) satunnainen 7 alkion osajoukko joukolle $\{1, 2, \dots, 39\}$. Korttipakan huolellisen sekoittamisen on tarkoitus tuottaa tasainen jakauma kaikkien korttien järjestysten joukossa.

Nopan ja kolikon heitot ovat perinteisiä tapoja tuottaa satunnaisuutta, mutta itseasiassa jo ennen näitä yksinkertaisia arvontavälineitä on käytetty luita tai luun paloja, jotka on heitetty, ja joiden asennosta on sitten tehty vaikkapa ennustuksia.¹ Tyypillisemmin matemaattisissa pohdinnoissamme ajatuksena on kuitenkin ennemminkin tuottaa satunnainen olio vaikkapa tietokoneella — tai ainakin algoritmisesti jostakin annetusta lähtökohdasta alkaen. Lähtökohtana voisi olla vaikkapa oletus siitä, että käytettävissämme on jo jokin (mieluiten alkeellinen) satunnainen olio, jonka jakauma tunnetaan: esimerkiksi ääretön jono $(\xi_n)_{n \in \mathbb{N}}$ riippumattomia $\{0, 1\}$ -arvoisia satunnaismuuttujia, joille $P[\xi_n = 0] = \frac{1}{2}$ ja $P[\xi_n = 1] = \frac{1}{2}$ (joka itsessään voisi olla kolikonheittoja toistamalla saatu). Tavoitteemme olisi sitten tätä käyttäen konstruoida satunnaisotos jostakin kiinnostavasta jakaumasta — vaikkapa Esimerkin ?? Gumbel-jakaumasta tai Ising-mallin Boltzmann jakaumasta hilalla \mathbb{Z}^3 annetuilla lämpötilan ja ulkoisen magneettikentän arvoilla.

Annetusta jakaumasta otosten tuottamista sanotaan hieman kontekstista riippuen myös sämplämiseksi tai simuloimiseksi (erityisesti jälkimmäinen viittaa yleensä tietokoneen käyttöön otoksen generoimisessa).

1.1.1. Miksi sämplätään?

Miksi sitten satunnaisten olioiden generoiminen on tärkeää? Tämän kurssin päätavoite ei ole järjestää lottoarvontoja tai muitakaan eksplisiittisesti sattumaan perustuvia pelejä, vaan melko ilmeiset vastauksemme lähtevät siitä, että satunnaismallien tavoite on mallintaa jotakin todellista ilmiötä.

Jos voimme tuottaa satunnaisia olioita, voimme sekä testata sitä, vastaako malli hyvin haluttua ilmiötä, että tutkia tätä ilmiötä turvallisesti vaikkapa kotitietokoneella (ja silloin kun ilmiö on esimerkiksi hurrikaani, ydinreaktio tai finanssikriisi, tarpeetonta määrää suoria empiirisiä kokeita on pyrittävä välttämään).

¹Nopan ja kolikon keskeinen etu hankalan muotoisiin luunpaloihin verrattuna on se, että käytettävissämme on valistunut arvaus siitä, miten satunnainen lopputulos on jakautunut.

Riittävän yksinkertaisissa satunnaismalleissa jotkin suureet saattavat olla analyytisestikin laskettavissa,² mutta usein realististen mallien ja kiinnostavien suureiden kohdalla tosiasia on, että suorat laskut ovat auttamatta liian monimutkaisia. Otoksia generoimalla voimme numeerisesti laskea (liikarvoja) kiinnostaville suureille.

Lisäksi otosten generointi auttaa usein merkittävästi hahmottamaan satunnaismallia kvalitatiivisesti — näissäkin luentomuistiinpanoissa on pyritty tukemaan visuaalista ymmärrystä esimerkiksi Ising-mallin faasitransitiosta (mm. Kuva II.5) ja itseään välttävän satunnaiskävelyn fraktaalista muodosta (mm. Kuvat II.13 ja II.14) simulaatioihin perustuvilla kuvilla.

1.1.2. *Sämpläämisen lähtökohdista*

Puutemme tuskin lainkaan filosofiseen kysymykseen siitä, mitä satunnaisuus todellisuudessa on, ja miten lähtökohtana käyttämämme satunnainen olio tuotetaan. Jätämme pelkäksi toteamukseksi myös sen matemaattisen tosiasian, että mikä tahansa puolalaisella avaruudella arvoja saava satunnaisotus on mahdollista toteuttaa mitallisella kuvauksella kolikonheittojonojoukosta $\{0, 1\}^{\mathbb{N}}$, joka varustetaan tavallisella painottamattoman kolikon toistuvaa heittelemistä mallintavalla todennäköisyysmitalla. Esimerkkinä konstruoimme tässä vain tasaisen todennäköisyysmitan yksikköväliä. Olkoon siis $(\xi_n)_{n \in \mathbb{N}}$ jono riippumattomia $\{0, 1\}$ -arvoisia satunnaismuuttujia, joille $P[\xi_n = 0] = \frac{1}{2}$ ja $P[\xi_n = 1] = \frac{1}{2}$. Asetetaan

$$U = \sum_{n=1}^{\infty} \xi_n 2^{-n}.$$

Helposti nähdään, että $P[U < \frac{1}{2}] = P[\xi_1 = 0] = \frac{1}{2}$, ja yleisemmin kaikille dyadisille väleille $[(k-1)2^{-m}, k2^{-m}]$, $m \in \mathbb{N}$, $k = 1, 2, \dots, 2^m - 1, 2^m$, pätee

$$P[U \in [(k-1)2^{-m}, k2^{-m}]] = 2^{-m}$$

(nollamittaista poikkeusjoukkoa lukuunotamatta tämä tapahtuma vastaa sitä, että satunnaismuuttujat ξ_1, \dots, ξ_m saavat juuri tietyt arvot). Dyadiset välit ovat mitan määräävä kokoelma, joten tosiaankin $U \sim \text{Tas}([0, 1])$, ja olemme tuottaneet yksikkövälin tasaisen jakauman kolikonheittoista. Jatkossa otamme tavallisesti jo lähtökohdaksi jonon $(U_n)_{n \in \mathbb{N}}$ riippumattomia tasaisesti yksikkövälille jakautuneita satunnaismuuttujia.

1.1.3. *Hyvän sämpläämisen kriteerejä*

Kuinka sitten tulisi arvioida sitä, miten hyvä jokin tapa tuottaa otoksia jakaumasta on? Tähän liittyy ainakin kaksi kriteeriä, joiden suhteelliset tärkeydet määräytyvät kulloinkin käsillä olevasta ongelmasta ja tavoitteesta. Ensinnäkin on hyväksyttävä se, että täydellisiä otoksia on olemassa vain matemaattisessa idealisaatiossa — todellisuudessa voimme esimerkiksi heittää kolikkoa 40 000 kertaa³, mutta emme loputtomiin, ja samaan tapaan äärellisen hintainen tietokoneemme tuottaa äärellisessä

²Joitakin eksplisiittisesti ratkeavia esimerkkejä löytyy Osasta ???. Tällaiset harvinaiset ratkeavat esimerkit tosin enemmänkin vahvistavat sitä käsitystä, että numeerisia likimääräisiä ratkaisuja tarvitaan niiden lisäksi.

³Kaksi Berkeley yliopiston opiskelijaa, Priscilla Ku ja Janet Larwood, heittivät molemmat kevään 2009 aikana kolikkoa 20 000 kertaa. He kirjasivat tulokset ja koettivat muun muassa selvittää, pitääkö peräkkäisten heittojen riippumattomuus paikkaansa, vai olisiko jokin mahdollisista korrelaatioista tuottavista mekanismeista empiirisesti havaittavissa.

ajassa äärellisen tuloksen, joka voi olla korkeintaan likimääräisesti matemaattista mallia vastaava. Siksi sämpläämisen laatua on ilmeisesti arvioitava sillä, kuinka lähellä haluttua tuotettu jakauma on — tähän jakaumien läheisyyteen sopii esimerkiksi heikon suppenemisen topologia (muitakin vaihtoehtoja jakaumien etäisyyksien mittaamiselle on, ja tulemme sellaisiin myös törmäämään). Toinen huomioitava asia on, kuinka vähillä resursseilla otos saatiin tuotettua — esimerkiksi tietokoneohjelman käyttämä muisti ja prosessoriaika on käytännössä pyrittävä minimoimaan (samoin kuin opiskelijoiden kolikonheittoon käyttämien lukukausien määrä). Vaatimukset tuotetun jakauman tarkkuudesta ja laskennallisten resurssien vähyydestä sotivat osittain toisiaan vastaan. Toisaalta ne myös tarjoavat jälleen matemaattisen idealisaation sämpläämisestä: on annettava menetelmä, jonka tuottama satunnaisolio saadaan mielivaltaisen lähelle haluttua jakaumaa, kun käytettävissä olevat resurssit menevät äärettömään.

1.2. Reaalisten satunnaismuuttujien generoiminen

Ensimmäisenä varsinaisena aiheena tarkastelemme sitä, miten reaalisia satunnaismuuttujia voidaan generoida. Lähtökohdaksi otamme annetun tasaisesti yksikkövälille jakautuneen satunnaismuuttujan U .

Olkoon siis annettu todennäköisyysjakauma ν reaaliakselilla. Todennäköisyysjakauma voidaan spesifioida kertymäfunktion $F: \mathbb{R} \rightarrow [0, 1]$ avulla. Seuraava huomio johtaa yleiseen reaalisten satunnaismuuttujien sämpläämismenetelmään.

Lemma V.1. *Jos $F: \mathbb{R} \rightarrow [0, 1]$ on kertymäfunktio, niin kaava*

$$G(u) = \inf \{x \in \mathbb{R} \mid F(x) \geq u\}$$

määrittelee kasvavan funktion $G: (0, 1) \rightarrow \mathbb{R}$, jolle pätee

$$G(u) \leq x \iff u \leq F(x).$$

Huomautus V.2. Jos F on jatkuva ja aidosti kasvava, on yllä määritelty G täsmälleen kertymäfunktion käänteisfunktio $G = F^{-1}$.

Todistus. Suoraan määritelmästä seuraa, että G on kasvava: infimum yli pienemmän joukon on suurempi. Ominaisuudesta $F(x) \nearrow 1$ kun $x \nearrow +\infty$ seuraa, että millään $u < 1$ infimum ei ole yli tyhjän joukon, ja siten $G(u) < +\infty$. Ominaisuudesta $F(x) \searrow 0$ kun $x \searrow -\infty$ seuraa, että $G(u) > -\infty$ kaikilla $u > 0$. Selvästi määritelmän perusteella $G(F(x)) \leq x$, joten jos $u \leq F(x)$, niin soveltamalla kasvavaa funktiota G päätellään $G(u) \leq G(F(x)) \leq x$. Koska F on oikealta jatkuva, on $F(G(u)) \geq u$, joten jos $G(u) \leq x$, niin soveltamalla kasvavaa funktiota F päätellään $u \leq F(G(u)) \leq F(x)$. \square

Tästä saatavaa reaalisten satunnaismuuttujien sämpläämismenetelmää kutsutaan *käänteisen kertymäfunktion menetelmäksi*.

Seuraus V.3. *Jos $U \sim \text{Tas}([0, 1])$, niin satunnaismuuttujan $G(U)$ jakauman kertymäfunktio on F .*

Todistus. Satunnaismuuttujan $G(U)$ kertymäfunktio saadaan suoralla laskulla edellisestä lemmaa käyttäen

$$\mathbb{P}[G(U) \leq x] = \mathbb{P}[U \leq F(x)] = F(x).$$

□

Tätä menetelmää käytetään usein reaalisten satunnaismuuttujien simuloimiseen tietokoneella. Lähes kaikissa ohjelmointikielissä on mahdollista pyytää konetta tuottamaan tasaisesti jakautunut (pseudo)satunnaisluku U . Halutusta kertymäfunktioista F lasketaan sitten vastaava G (muistutetaan, että sopivasti tulkiten $G = F^{-1}$), ja asetetaan $X = G(U)$, jolloin on saatu haluttua jakaumaa noudattava (pseudo)satunnaisluku X .

Esimerkki V.4. Eksponenttijakauman kertymäfunktio on $F(x) = 1 - e^{-\lambda x}$, kun jakauman parametri on λ . Silloin $G(u) = \frac{-1}{\lambda} \log(1 - u)$. Eksponenttijakautunut satunnaismuuttuja T voidaan siis konstruoida (sämplätä) seuraavasti. Otetaan on tasaisesti yksikköväliä jakautunut satunnaismuuttuja U , ja määritellään satunnaismuuttuja $\frac{-1}{\lambda} \log(1 - U)$. Toisaalta myös satunnaismuuttuja $1 - U$ tässä on tasaisesti jakautunut yksikköväliä, joten käytännössä hieman siistimpi konstruktio on

$$T = \frac{-1}{\lambda} \log(U).$$

1.2.1. Kätevä temppu Gaussisten satunnaislukujen sämpläämiseksi

Ylläolevalla menetelmällä voitaisiin toki sämplätä standardista normaalijakaumasta. Standardin normaalijakauman kertymäfunktio on

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{s^2}{2}} ds,$$

ja käänteisfunktio on $G = F^{-1}: (0, 1) \rightarrow \mathbb{R}$. Jos sitten $U \sim \text{Tas}([0, 1])$, niin $X = G(U) \sim N(0, 1)$. Huomataan kuitenkin, että F ja G eivät ole alkeisfunktioita, joten päällisin puolin siisti konstruktio $X = G(U)$ kätkee joitakin vaikeuksia funktion G evaluoimiseen.

Tässä tapauksessa toimii sama temppu kuin Gaussisen integraalin

$$\int_{-\infty}^{\infty} e^{-\frac{s^2}{2}} ds = \sqrt{2\pi}$$

laskemisessakin: kaksiulotteinen tapaus on yksiulotteista helpompi. Sämpläämme yhden reaalisesta Gaussisesta satunnaismuuttujan sijasta parin (X_1, X_2) riippumattomia standardia normaalijakaumaa noudattavia satunnaismuuttujia. Pyrimme siis jakaumaan

$$\mathbb{P}[(X_1, X_2) \in E \subset \mathbb{R}^2] = \frac{1}{2\pi} \int_E e^{-\frac{1}{2}\|\underline{s}\|^2} d^2\underline{s}.$$

Haluamme lausua parin (X_1, X_2) napakoordinaateissa

$$X_1 = R \cos(\Theta) \quad X_2 = R \sin(\Theta),$$

ja ilmeisesti silloin Θ olisi tasaisesti jakautunut esimerkiksi välillä $[0, 2\pi)$, ja satunnaismuuttujan $R = \sqrt{X_1^2 + X_2^2}$ kertymäfunktio olisi

$$\mathbb{P}[R \leq r] = \frac{1}{2\pi} \int_{\|\underline{s}\| \leq r} e^{-\frac{1}{2}\|\underline{s}\|^2} d^2\underline{s} = \int_0^r s e^{-\frac{1}{2}s^2} ds = 1 - e^{-\frac{1}{2}r^2}.$$

Lisäksi R ja Θ olisivat toisistaan riippumattomat. Tasaista jakaumaa noudattava Θ on helppo sämplätä, ja satunnaismuuttujan R kertymäfunktio on alkeisfunktio.

Huomaamme näin, että jos (U_1, U_2) on pari riippumattomia tasaisesti yksikköväliille jakautuneita satunnaismuuttujia, niin asettamalla

$$\Theta = 2\pi U_1 \quad R = \sqrt{-2 \log(1 - U_2)}$$

ja sitten

$$X_1 = R \cos(\Theta) \quad X_2 = R \sin(\Theta),$$

olemme onnistuneet sämpläämään kaksi riippumatonta standardia normaalijakaumaa noudattavaa satunnaismuuttujaa.

1.3. Koplaukset

Koplaukset ovat probabilistinen tekniikka, jossa kaksi (tai useampia) satunnaismuuttujaa toteutetaan samalla todennäköisyysavaruudella. Sitä käytetään abstrakteissakin probabilistisissä todistuksissa, mutta sämpläämisen yhteydessä se on luonnollinen ja konkreettinen asia.

Olkoon ν_1 ja ν_2 todennäköisyysmitat avaruuksilla \mathfrak{X}_1 ja \mathfrak{X}_2 , vastaavasti. Todennäköisyysmittaa ν tuloavaruudella $\mathfrak{X}_1 \times \mathfrak{X}_2$ sanotaan vastaavien satunnaismuuttujien *koplaukseksi*, jos pätee

$$\nu[\mathfrak{X}_1 \times E_2] = \nu_2[E_2] \quad \text{ja} \quad \nu[E_1 \times \mathfrak{X}_2] = \nu_1[E_1].$$

Toisin sanoen, kun satunnaismuuttujapari $(X_1, X_2) \in \mathfrak{X}_1 \times \mathfrak{X}_2$ noudattaa jakaumaa ν , niin komponentit X_1 ja X_2 noudattavat jakaumia ν_1 ja $X_2 \sim \nu_2$, vastaavasti.

Aina voitaisiin tietenkin vaikkapa määritellä ν tulomittana $\nu_1 \otimes \nu_2$, mutta silloin yhteisjakauma ei ole erityisen kiinnostava. Koplaukset ovat tyypillisesti hyödyllisiä silloin, kun satunnaismuuttujaparin (X_1, X_2) yhteisjakaumalla on jokin käyttökelpoinen relaatio. Ehkäpä parhaimmillaan koplaukset ovat tekniikkana silloin, kun toinen satunnaismuuttujista on vaikea käsitellä suoraan, ja toinen taas helppo, ja kun lisäksi sopivan relaation avulla voidaan saavuttaa johtopäätöksiä vaikeammasta tekemällä laskuja helpommalla.

1.3.1. Stokastinen dominointi

Annamme erään esimerkin koplauksesta, jossa saadaan voimaan järjestysrelaatio.

Tehtävä V.1. Olkoot X_1 ja X_2 reaalisia satunnaismuuttujia vastaavilla kertymäfunktioilla F_1 ja F_2 . Sanomme, että satunnaismuuttuja X_2 stokastisesti dominoi satunnaismuuttujaa X_1 , jos kaikilla $x \in \mathbb{R}$ pätee $F_1(x) \leq F_2(x)$. Osoita, käyttäen reaalisten satunnaismuuttujien sämpläämistä, että silloin on olemassa satunnaismuuttujien X_1 ja X_2 koplaukset (X_1, X_2) , jolle melkein varmasti $X_1 \leq X_2$.

1.4. Äärellisistä todennäköisyysavaruuksista sämpläämisestä

Statistisen fysiikan mallit on tyypillisesti määritelty äärellisillä todennäköisyysavaruuksilla.

Jos Ω on äärellinen, $\#\Omega = n < \infty$, olisi periaatteellisesti mahdollista sämplätä Ω -arvoista satunnaismuuttujaa ylläolevan menetelmän pienellä variantilla. Alkiot luettelaisiin

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\},$$

ja samoin vastaavat todennäköisyydet p_1, p_2, \dots, p_n , missä $p_j = \mathbf{P}[\{\omega_j\}]$. Kun sitten olisi annettu $U \sim \text{Tas}([0, 1])$, voitaisiin valita satunnainen indeksi

$$K = \min \left\{ k \in \llbracket 1, n \rrbracket \mid \sum_{j=1}^k p_j \geq U \right\},$$

ja silloin satunnainen alkio $X = \omega_K$ noudattaisi haluttua jakaumaa.

1.4.1. *Naivin sämpläämisen mahdottomuudesta*

Yllä kuvattu sämplääminen on kuitenkin tyypillisesti käytännössä mahdotonta (karkeasti ottaen ainakin silloin, kun tila-avaruuden Ω koko on eksponentiaalisen suuri jossakin luonnollisessa kokoparametrissa, ja haluttu todennäköisyysjakauma on monimutkainen, esimerkiksi statistisen fysiikan mallit ja niiden Boltzmann jakauma).

Ymmärtääksemme ylläolevan mahdottomuuden, tarkastelemme laskennallisen vaativuuden suuruusluokkaa Ising mallissa. Ottakaamme ajatuskokeeseen vaatimattomasti 2-ulotteinen Ising-malli ruutupaperiarkkia pienemmässä 10×10 -suorakaiteessa. Olettakaamme optimistisesti, että tietokoneemme olisi jo suorittanut tarvittavat laskut kaikkien tilojen Boltzmann painoista $e^{-\beta H(\sigma)}$. Todennäköisyydet olisivat silloin

$$p_\sigma = \frac{1}{Z} e^{-\beta H(\sigma)},$$

missä Z on partitiofunktio $Z = \sum_{\sigma} e^{-\beta H(\sigma)}$. Mutta pelkästään partitiofunktion Z määrittelevässä summassa on $\#\Omega = 2^{10^2} \approx 1.3 \times 10^{30}$ termiä. Vuonna 2013 maailman nopein tietokone, kiinalainen *Tianhe-2* pystyi suorittamaan noin 3.4×10^{16} laskutoimitusta sekunnissa — valitettavasti tällä tahdilla pelkän summan laskemiseen kuluisi yli satatuhatta vuotta. Jos tämä kuulostaa vielä järkevältä, otetaan vastaavaan tarkasteluun kolmiulotteisen Ising-malli vaikkapa haastavaa Rubikin kuution varianttia vastaavassa $5 \times 5 \times 5$ -suorakaiteessa. Hilapisteitä on silloin 125 äskeisen sadan sijasta, mikä ei kuulosta dramaattiselta — mutta tila-avaruuden koko on nyt $2^{5^3} \approx 4.3 \times 10^{37}$, ja partitiofunktion summan laskeminen veisi tässä tapauksessa useita kertalukuja enemmän kuin maailmankaikkeuden iän verran. Tällaiset pohdinnat ovat toki täysin epärelevantteja monestakin syystä, mutta ne auttavat hahmottamaan sitä, että naiivi lähestymistapa äärellistenkin todennäköisyysjakaumien simuloimiseen on kertakaikkiaan mahdoton.

1.4.2. *Miten lopulta onnistumme sämpläämään statistisen fysiikan malleista?*

Tulemme käyttämään statistien fysiikan mallien sämpläämiseen nk. Markov-ketju Monte Carlo menetelmää. Menetelmän idea on tuttu vaikkapa korttipeleistä: saadaksemme korttipakan huolella sekaisin toistamme jotakin sekoitusoperaatiota, joka tuottaa korttien järjestykseen satunnaisuutta. Tarpeeksi monen sekoituskerran jälkeen korttipakka on pitkälti unohtanut alkuperäisen järjestyksensä ja on tarkasta alkutilasta riippumatta lähellä toivottua tasaisesti satunnaista järjestystä.

Statistisen fysiikan malleille keksimme kullekin sopivia sekoitusoperaatioita, joita toistamalla haluttuun jakaumaan päästään. Useat näistä sekoitusoperaatioista itseasiassa imitoivat sitä, mitä luonnossa tapahtuu: lämpöliike tuottaa pientä satunnaista sekoitusta vähän kerrallaan, ja ajan kuluessa tarpeeksi systeemi lähestyy lämpötasapainoa (Boltzmann-jakaumaansa).

2. Markov-ketjut

Tässä luvussa käsittelemme Markov-ketjujen perusteoriaa. Tarkastelemme aikahomogeenisia äärellistilaisia Markov-ketjuja, eli diskreettiaikaisia stokastisia prosesseja, jotka saavat arvoja jollakin äärellisellä joukolla, ja joilla on seuraavat ominaisuudet (matemaattinen muotoilu näille ominaisuuksille esitetään Luvussa 2.1)

- *Markov-ominaisuus*: Prosessi on muistiton siinä mielessä, että sen tila tulevalla ajanhetkellä voi riippua nykytilasta, mutta ei menneisyydestä.
- *Homogeenisuus*: Prosessin siirtymätodennäköisyydet tilasta toiseen ovat kaikilla ajanhetkellä samat.

On huomionarvoista, että Markov ketjujen pääosin alkeellisella teorialla on valtavasti sekä reaali maailmaa koskevia että teoreettisia sovelluksia. Emme tällä kursilla tietenkään pyri esittämään Markov ketjujen sovelluskohteita kattavasti. Tällä kurssilla tärkeimmät sovelluskohteet ovat seuraavat:

- Otosten tuottaminen (eli sämplääminen) monimutkaisista todennäköisyysjakaumista (kuten tasaisesta mitasta äärellisellä kombinatorisesti hankalalla joukolla, tai Boltzmannin jakaumasta valtavan suurella tila-avaruudella)
- Dynamiikka statistisen fysiikan malleissa

On kuitenkin hyvä pitää mielessä, että oleellisesti samalla teorialla on käyttöä hyvin erilaisilla aloilla, kuten:

- * Biologia: saalistaja-saaliseläin -mallit, vuorovaikuttavat molekyylit, solujen liike, ...
- * Lääketiede: epidemiologia, geenitransmissio, populaatiodynamiikka, ...
- * Talous: osakemarkkinat ja valuuttakurssit, vakuutusriskit, johdannaisten hinnoittelu, ...
- * Yhteiskuntatieteet: liikenne, mielipiteiden dynamiikka, laumakäyttäytyminen, ...
- * Tietojenkäsittelytiede: internetliikenne, algoritmit, ...
- * Huvi: vedonlyönti, pelit, ...

2.1. Markov ketjut äärellisillä tila-avaruuksilla

Tarkastelemme Markov-ketjuja äärellisellä tila-avaruudella⁴ \mathcal{S} , joka alkioden enumeroinnilla voitaisiin tarvittaessa samaistaa $\mathcal{S} = \llbracket 1, n \rrbracket$, missä $n = \#\mathcal{S}$. Markov-ketju on diskreettiaikainen stokastinen prosessi, formaalisti siis todennäköisyysmitta \mathbb{P} joukolla

$$\mathcal{S}^{\mathbb{Z}_{\geq 0}} = \{X: \mathbb{Z}_{\geq 0} \rightarrow \mathcal{S}\}. \quad (\text{V.1})$$

Käytännössä puhumme aina konkreettisesta satunnaisprosessista, eli satunnaisesta ajan t funktiosta X , joka saa arvoja joukolla \mathcal{S} ,

$$X = (X(t))_{t \in \mathbb{Z}_{\geq 0}} \in \mathcal{S}^{\mathbb{Z}_{\geq 0}}. \quad (\text{V.2})$$

⁴Useat tarkastelut toimisivat yleisemmässäkin kontekstissa, mutta esimerkiksi sämpläämistä varten äärellisen tila-avaruuden tapaus on luonnollinen valinta. Vältämme näin myös tarpeettomia teknisiä pohdintoja.

Kuten stokastisten prosessien teoriassa yleensäkin, tarkastelemme paljon prosessin arvon $X(t)$ jakaumia annetuilla ajanhetkillä t , jotka siis ovat todennäköisyysmittoja äärellisellä joukolla \mathcal{S} . Tällainen todennäköisyysmitta $\underline{\mu}$ on kätevää ajatella rivivektorina

$$\underline{\mu} = (\mu_x)_{x \in \mathcal{S}}, \quad \mu_x = \mathbb{P}[X(t) = x], \quad (\text{V.3})$$

jonka komponentit μ_x ovat yksiiöiden $\{x\} \subset \mathcal{S}$ todennäköisyydet.

Dynaaminen tapa kuvailla stokastinen prosessi on kertoa, mitkä ovat todennäköisyydet tuleville arvoille, jos menneet arvot tunnetaan. Diskreettiaikaisen äärellisen tila-avaruuden prosessin täydelliseen kuvailemiseen riittää esimerkiksi alkujakauma $\underline{\nu}$

$$\nu_x = \mathbb{P}[X(0) = x] \quad (\text{V.4})$$

ja siirtymätodennäköisyydet

$$\mathbb{P}\left[X(t+1) = y \mid X(0) = x_0, X(1) = x_1, \dots, X(t) = x_t\right]. \quad (\text{V.5})$$

Tästä lähtien tarkastelemme homogeenisia Markov-ketjuja, mikä tarkoittaa, että teemme siirtymätodennäköisyyksistä seuraavat kaksi oletusta:

- *Markov-ominaisuus:*

$$\begin{aligned} & \mathbb{P}\left[X(t+1) = y \mid X(0) = x_0, X(1) = x_1, \dots, X(t) = x_t\right] \\ &= \mathbb{P}\left[X(t+1) = y \mid X(t) = x_t\right]. \end{aligned} \quad (\text{V.6})$$

- *Homogeenisuus:*

$$\mathbb{P}\left[X(t+1) = y \mid X(t) = x\right] = P_{x,y}, \quad (\text{V.7})$$

missä $P_{x,y}$ on sama kaikilla t .

Siirtymätodennäköisyydet on kätevä koota nk. *siirtymätodennäköisyysmatriisiksi*

$$P \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}, \quad P_{x,y} = \mathbb{P}\left[X(t+1) = y \mid X(t) = x\right]. \quad (\text{V.8})$$

Aivan ensimmäisenä toteamme, että kun annetun ajanhetken jakauma tulkitaan rivivektorina, niin seuraavan ajanhetken jakauma saadaan kertomalla tätä vektoria oikealta siirtymätodennäköisyysmatriisilla.

Lemma V.5. *Jos $\underline{\mu} = (\mu_x)_{x \in \mathcal{S}}$ on prosessin tilan $X(t)$ jakauma ajanhetkellä t , $\mu_x = \mathbb{P}[X(t) = x]$, niin prosessin tilan jakauma ajanhetkellä $t+1$ on $\underline{\mu}P$.*

Todistus. Kirjoitetaan

$$\begin{aligned} \mathbb{P}[X(t+1) = y] &= \sum_{x \in \mathcal{S}} \mathbb{P}[X(t+1) = y \text{ ja } X(t) = x] \\ &= \sum_{x \in \mathcal{S}} \underbrace{\mathbb{P}[X(t) = x]}_{=\mu_x} \underbrace{\mathbb{P}[X(t+1) = y \mid X(t) = x]}_{=P_{x,y}} = (\underline{\mu}P)_y. \end{aligned}$$

□

Samaan tapaan lausumme prosessin äärellisulotteiset jakaumat alkujakauman (V.4) ja siirtymätodennäköisyyksien (V.8) avulla.

Propositio V.6. *Markov-prosessille, jonka alkujakauma on $\underline{\nu}$ ja siirtymätodennäköisyysmatriisi P , pätee*

$$\mathbb{P}\left[X(0) = x_0, X(1) = x_1, \dots, X(t) = x_t\right] = \nu_{x_0} P_{x_0, x_1} P_{x_1, x_2} \cdots P_{x_{t-1}, x_t}.$$

Todistus. Väite todistetaan soveltamalla edellisen lemmän laskun ideaa toistuvasti: kirjoitetaan ensin

$$\begin{aligned} & \mathbb{P}\left[X(0) = x_0, \dots, X(t-1) = x_{t-1}, X(t) = x_t\right] \\ = & \mathbb{P}\left[X(0) = x_0, \dots, X(t-1) = x_{t-1}\right] \mathbb{P}\left[X(t) = x_t \mid X(0) = x_0, \dots, X(t-1) = x_{t-1}\right], \end{aligned}$$

tunnistetaan jälkimmäinen tekijä oikealla puolella matriisielementiksi P_{x_{t-1}, x_t} , ja jatketaan rekursiivisesti ensimmäisen tekijän purkamista, kunnes lauseke on muotoa matriisielementtien tulo kertaa $\mathbb{P}[X(0) = x_0] = \nu_{x_0}$. \square

Suorana seurauksena näemme muun muassa, että prosessin tilan $X(t)$ jakauma ajanhetkellä $t > 0$ saadaan kertomalla rivivektoria $\underline{\nu}$ oikealta matriisiin P potenssilla t .

Seuraus V.7. *Markov-prosessille, jonka siirtymätodennäköisyysmatriisi on P , pätee*

- (a) *Jos prosessin tilan jakauma ajanhetkellä t on $\underline{\mu}$, eli $\mu_x = \mathbb{P}[X(t) = x]$, niin prosessin tilan jakauma ajanhetkellä $t + s$ on $\underline{\mu}P^s$.*
- (b) *Jos prosessin alkujakauma on $\underline{\nu}$, niin prosessin tilan jakauma ajanhetkellä t on $\underline{\nu}P^t$.*

On ilmeistä, että siirtymätodennäköisyyksillä on seuraavat ominaisuudet.

$$P_{x,y} \geq 0 \quad \forall x, y \in \mathcal{S} \tag{V.9}$$

$$\sum_{y \in \mathcal{S}} P_{x,y} = 1 \quad \forall x \in \mathcal{S}. \tag{V.10}$$

Ehdot (V.9) ja (V.10) toteuttavaa matriisia $P \in \mathbb{R}^{\mathcal{S} \times \mathcal{S}}$ sanotaan *stokastiseksi matriisiksi*.

Myös käänteinen pätee: jos on annettu alkujakauma ja stokastinen matriisi, voidaan niiden avulla konstruoida satunnaisprosessi X , nimittäin Proposition V.6 äärellisulotteiset jakaumat määrittävät yksikäsitteisen todennäköisyysmitan joukolla $\mathcal{S}^{\mathbb{Z}_{\geq 0}}$.

Annamme lopulta muutamia esimerkkejä homogeenisista Markov-ketjuista.

Esimerkki V.8. Jos tavallista noppaa heitetään toistuvasti, ja ensimmäisten t heiton silmäluvuista suurinta merkitään $X(t)$, niin $X = (X(t))_{t \in \mathbb{Z}_{\geq 0}}$ on Markov-prosessi, jonka siirtymätodennäköisyysmatriisi on

$$P = \begin{bmatrix} \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{2}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & \frac{3}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & \frac{4}{6} & \frac{1}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & \frac{5}{6} & \frac{1}{6} \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Esimerkki V.9 (Ehrenfestin uurna Markov-ketjuna). Osan I esimerkissä tarkasteltu naiivi kaksiosaisen kaasusäiliön malli, Ehrenfestin uurna, on Markov prosessi: tila-avaruus on joukko $\llbracket 0, N \rrbracket$ (tila on osassa A olevien hiukkasten lukumäärä), ja dynaaminen askel määrittelee

siirtymätodennäköisyydet

$$P_{m,m'} = \begin{cases} (1-q)\frac{m}{N} & \text{jos } m' = m-1 \\ 1 - (1-q)\frac{m}{N} - q\frac{N-m}{N} & \text{jos } m' = m \\ q\frac{N-m}{N} & \text{jos } m' = m+1 \\ 0 & \text{muuten} \end{cases}.$$

Esimerkki V.10 (Satunnaiskävely äärellisellä graafilla). Olkoon $G = (V, E)$ äärellinen graafi. Pisteen $x \in V$ asteeksi $\deg(x)$ sanotaan sen naapurien lukumäärää, eli sellaisten viivojen lukumäärää, joihin piste kuuluu $\deg(x) = \#\{e \in E \mid x \in e\}$. Satunnaiskävely tällä graafilla on diskreettiaikainen stokastinen prosessi, jonka tila-avaruus on graafin pisteiden joukko, $\mathcal{S} = V$. Yksinkertainen satunnaiskävely on sellainen Markov-ketju, jossa seuraava tila valitaan umpimähkään edellisen tilan naapurien joukosta (oletetaan tässä, että jokaisella tilalla on naapureita, $\deg(x) > 0$ kaikilla $x \in V$) — siirtymätodennäköisyydet ovat siis

$$P_{x,y} = \begin{cases} \frac{1}{\deg(x)} & \text{jos } \{x,y\} \in E \\ 0 & \text{muuten} \end{cases},$$

missä pisteen x aste $\deg(x) = \#\{e \in E \mid x \in e\}$ on sellaisten viivojen lukumäärä, joihin x kuuluu.

Esimerkki V.11 (Satunnaiskävelyt lokaalisti äärellisillä graafeilla). Edellisen esimerkin satunnaisprosessi on mielekäs myös silloin, kun graafi $G = (V, E)$ on vain lokaalisti äärellinen eli kun kaikilla pisteillä on äärellisen monta naapuria, $\deg(x) < \infty \forall x \in V$. Esimerkiksi aiemmin tarkastelemamme satunnaiskävelyt \mathbb{Z} :lla ja \mathbb{Z}^d :llä ovat tällaisia prosesseja.

Huomautus V.12 (Googlen verkkosivuhaku). Googlen verkkosivuhakupalvelu nousi johtavaksi internetin hakukoneeksi 2000-luvun alussa. Haku perustuu PageRank-algoritmiin, jolla hakutulosten relevanssia vertaillaan — tämä algoritmi teki Googlen hakukoneesta kilpailijoita tehokkaamman. Karkeasti ottaen, algoritmi simuloi satunnaista nettisurffaajaa, ja arvioi verkkosivun relevanssia todennäköisyydellä, että surffaaja päätyy kyseiselle sivulle. Satunnaisen surffaajan malli on (karkeasti) satunnaiskävely sillä suunnistetulla graafilla, jonka pisteet ovat eri verkkosivut ja verkkosivujen A ja B välillä on (suunnistettu) viiva, jos sivulla A on linkki sivuun B .

2.2. Stationaariset jakaumat

Lemmasta V.5 nähdään suoraan, milloin todennäköisyysjakauma Markov-prosessin peräkkäisille tiloille $X(t)$ ja $X(t+1)$ on sama. Näin käy, jos tilan $X(t)$ jakauma $\underline{\pi} = (\pi_x)_{x \in \mathcal{S}}$, $\pi_x = \mathbb{P}[X(t) = x]$, on sellainen, että

$$\underline{\pi}P = \underline{\pi}, \tag{V.11}$$

toisin sanoen jos $\underline{\pi}$ on siirtymätodennäköisyysmatriisin P vasemmanpuoleinen ominaisvektori ominaisarvolla 1. Sanomme tällaista $\underline{\pi}$ Markov-ketjun *stationaariseksi jakaumaksi*.

Esimerkki V.13. Tarkastellaan Esimerkin V.8 Markov-ketjua, maksimia nopanheittojen silmäluvuista. Rivivektori $\underline{\pi} = [0 \ 0 \ 0 \ 0 \ 0 \ 1]$ toteuttaa $\underline{\pi}P = \underline{\pi}$, ja määrittelee siis stationaarisen jakauman. Nopanheittojen silmäluvuista suurin ei muutu enää sen jälkeen, kun silmäluvu on kerran saavutettu.

Esimerkki V.14. Tarkastellaan Esimerkin V.10 satunnaiskävelyä äärellisellä graafilla $G = (V, E)$. Asetetaan $\pi_x = \frac{\deg(x)}{2\#E}$ ja huomataan ensinnäkin, että $\underline{\pi}$ on todennäköisyysjakauma joukolla

V (kts. Osan I Esimerkki ??). Lisäksi pätee

$$(\underline{\pi}P)_x = \sum_{y \in V} \pi_y P_{y,x} = \sum_{\substack{y \in V \\ \{x,y\} \in E}} \frac{\deg(y)}{2\#E} \frac{1}{\deg(y)} = \sum_{\substack{y \in V \\ \{x,y\} \in E}} \frac{1}{2\#E} = \frac{\deg(x)}{e\#E} = \pi_x,$$

joten tämä on satunnaiskävelyn stationaarinen jakauma.

Esimerkki V.15. Tarkastellaan Ehrenfestin uurnaa, eli Esimerkin V.9 Markov-ketjua. Osan I harjoitustehtävän laskusta seuraa, että binomijakauma $\underline{\pi}$ komponentein

$$\pi_m = q^m (1-q)^{N-m} \binom{N}{m}$$

on stationaarinen jakauma tälle Markov-ketjulle.

On luontevaa kysyä:

- Onko stationaarisia jakaumia aina olemassa?
- Onko stationaarinen jakauma yksikäsitteinen?

Vastaus ensimmäiseen kysymykseen osoittautuu myönteiseksi.⁵ Jälkimmäinen sen sijaan ei päde ilman lisäoletuksia — Luvussa 2.3 määrittelemämme Markov-ketjun redusoitumattomuus oletetaan tätä varten.

Lähestykäämme kuitenkin aihetta aluksi hieman toisesta näkökulmasta. Kysymme stabiloituuko tilojen $X(t)$ jakauma, kun $t \rightarrow \infty$.⁶ Seuraavan apulauseen mukaan jos näin on, on rajajakauma Markov-ketjun stationaarinen mitta.

Lemma V.16. *Jos $P[X(t) = x] \rightarrow \pi_x$ kaikilla $x \in \mathcal{S}$, niin $\underline{\pi} = (\pi_x)_{x \in \mathcal{S}}$ on todennäköisyysjakauma joukolla \mathcal{S} , joka on prosessin X stationaarinen jakauma, eli $\underline{\pi}P = \underline{\pi}$.*

Todistus. Ensinnäkin $P[X(t) = x] \geq 0$, joten rajalla $t \rightarrow \infty$ saamme $\pi_x \geq 0$. Samaan tapaan, $\sum_x P[X(t) = x] = 1$, joten rajalla päätelemme $\sum_x \pi_x = 1$. Siis raja-arvot $\underline{\pi}$ määrittelevät tosiaankin todennäköisyysmitan.

Muistetaan Lemmasta V.5, että

$$P[X(t+1) = y] = \sum_{x \in \mathcal{S}} P[X(t) = x] P_{x,y}.$$

Koska $P[X(t) = x] \rightarrow \pi_x$ kun $t \rightarrow \infty$, ja $P[X(t+1) = y] \rightarrow \pi_y$ kun $t \rightarrow \infty$, saadaan rajalla

$$\pi_y = \sum_{x \in \mathcal{S}} \pi_x P_{x,y}.$$

□

Luonteva jatkokysymys on nyt:

- Millä ehdoilla tilojen $X(t)$ jakaumat suppenevat kohti jakaumaa $\underline{\pi}$?

Huomataan ainakin, että joukon \mathcal{S} äärellisyydestä (kompaktisuudesta) seuraa jonon $(X(t))_{t \in \mathbb{Z}_{\geq 0}}$ tiukkuus, ja siten edelleen se, että on olemassa suppenevia osajonoja.

⁵Jos sallisimme numeroituvat tila-avaruudet \mathcal{S} , olisi kaikilla Markov-ketjuilla stationaarisia mittoja, jotka eivät välttämättä kuitenkaan olisi todennäköisyysmittoja. Tästä ilmiöstä kiinnostunut lukija keksii esimerkin pohtimalla vaikkapa yksinkertaista satunnaiskävelyä \mathbb{Z} :lla.

⁶Formaalimmin, konvergoivatko satunnaismuuttujat $X(t)$ heikosti kohti jotakin \mathcal{S} -arvoista satunnaismuuttujaa, kun $t \rightarrow \infty$?

Luvussa 2.3 määrittelemämme Markov-ketjun jaksottomuus oletetaan muun muassa siksi, ettei osajonoon siirtymistä tarvittaisi.

2.2.1. *Yhteys sämpläämiseen*

Miten Markov-ketjut ja niiden stationaariset jakaumat sitten liittyvät tämän osan pääteemaan, eli sämpläämiseen? Ideana on, että sämplätäksemme todennäköisyysmitasta $\underline{\pi}$ joukolla \mathcal{S} , konstruoinne Markov-ketjun X , joka konvergoi (parhaassa tapauksessa mistä tahansa alkutilasta lähtien) kohti tätä jakaumaa. Silloin kun Markov-ketjun siirtymien sämplääminen on helpompaa kuin stationaarisen jakauman sämplääminen suoraan, riittää siis simuloida Markov ketjua $X = (X(t))_{t \in \mathbb{Z}_{\geq 0}}$ suureen t :n arvoon asti, ja näin saatu satunnainen $X(t)$ noudattaa likimain haluttua (stationaarista) jakaumaa $\underline{\pi}$. Tällaisen Markov-ketju Monte Carlo -menetelmän käyttökelpoisuuden ymmärtää parhaiten esimerkkien valossa — palaamme sellaisiin Luvussa 3, kun olemme ensin käyneet läpi tarvittavan teorian.