0. Johdanto

0.1. **Satunnaisuudesta ja sen mallintamimisesta.** Tällä kurssilla käsittelemme kurssin nimen mukaisesti stokastisia prosesseja. Tulkitsemme sanan *stokastinen* tarkoittavan *satunnaista* ja sanan *prosessi* tarkoittavan ilmiötä, joka muuttuu ajan kuluessa eli *riippuu ajasta.*

Stokastisilla prosesseilla tarkoitamme siis satunnaisia ilmiöitä, jotka muuttuvat ajan kuluessa.[1] Kurssilla tulemme tarkastelemaan muutamia yksinkertaisia satunnaisia malleja. Nämä mallit ovat yksinkertaistuksia meitä kiinnostavista ilmiöistä, joita esiintyy esimerkiksi luonnossa, kaupankäynnissä tai vaikka uhkapeleissä.

Miksi haluamme tarkastella ilmiöitä satunnaisina, sillä varsin yleinen kysymys onkin, että onko satunnaisia ilmiöitä tai satunnaisuutta "oikeasti" olemassakaan. Tätä kysymystä väistämme käytännöllisistä syistä, sillä joka tapauksessa maailma on pullollaan tapahtumia, joiden tulevaa käytöstä on vaikuttaa olevan täysin mahdotonta ennakoida. Kuitenkin näitä tapahtumia tai niiden tiettyjä kiinnostavia piirteitä pyritään mallintamaan, jotta tapahtumista saataisiin edes jotain selkoa.

Tällaisia yksinkertaistuksia (eli malleja) voidaan luokitella niiden toteuttamistavan mukaan. On korkean tason *käsitteellisiä* malleja, joista tehtävät päätelmät ovat heuristisia. Yleensä tällaisten mallien pohjalta ei ole tarkoituksen mukaista tehdä päätelmiä, vaan malleja pyritään saamaan laskettavampaan muotoon. Tällöin voimme puhua esimerkiksi *matemaattisista, stokastisista* tai *tilastollisista* malleista.

Kuinka nämä viimeksi mainitut mallit sitten eroavat toisistaan? Vastaus tähän kysymykseen vaihtelee yleensä vastaajan mukaan. Stokastiset (eli satunnaiset) ja matemaattiset mallit ymmärrämme tällä kurssilla molemmat matemaattisiksi malleiksi, mutta erottavana tekijänä pidämme satunnaisuuden mukanaoloa tai sen puuttumista. Tilastollisessa mallissa mittaustapahtuma sekä mittausdata on yleensä mukana tarkastelussa, mutta muuten tilastollinen malli on myös satunnaismalli.

Tarkastellaan lyhyesti käsitteellisellä tasolla erään ilmiön matemaattista sekä satunnaista mallia ja niiden yhteyksiä. Lämpöyhtälö on(eräs) fysiikasta tuttu matemaattinen malli lämmön johtumiselle. Tämä malli osoittaa, että lämpötila kappaleessa pyrkii tasoittumaan siten, että lämpö virtaa kuumemmista kohdista kylmempiin.[2] Matemaattisesti lämpöyhtälö on toisen kertaluvun lineaarinen

[1] Yleisestikin tämä tulkinta on oikea, mutta "aika" ei yleisessä tilanteessa vastaa intuitiivista käsitystämme ajasta

[2] tai kylmyys virtaa kylmemmistä kohdista kuumempiin

osittaisdifferentiaaliyhtälö $\partial_t u = \partial_{x_1}^2 u + \ldots \partial_{x_n}^2 u$ ja mallin antamat ennusteet saadaan tarkastelemalla tätä yhtälöä ja sen ratkaisujen käyttäytymistä.

Lähemmin tarkasteltuna lämpötila on kuitenkin seurausta kappaleen pienimpien rakennusosasten lämpöliikkeestä, joka on rakennusosasten hyvin epäsäännöllistä liikettä kappaleen kidehilassa. Kappaleen kohta on sitä kuumempi, mitä voimakkaammin sen osaset liikkuvat. Lämmön johtuminen seuraa nyt osasten törmäyksistä naapureihinsa, sillä voimakkaimmin liikkuvat osat törmäävät vieressä oleviin osiin voimakkaammin kuin hitaammin liikkuvat ja luovuttavat vastaavasti enemmän liike-energiaansa törmäyksissä kuin sitä törmäyksessä saavat. Tämä saa keskiarvomielessä aikaan samanlaisen mallin kuin edellä ollut osittaisdifferentiaaliyhtälömalli.

Yksittäisen pienen rakennusosasen liikettä on käytännössä paras mallintaa puhtaasti satunnaisena liikkeenä. Voimme siis pitää tätä lämpömallia stokastisena mallina. Tämä on yksi esimerkki tilanteesta, milloin mallia on järkevää pitää satunnaisena. Seuraavassa esitämme muutaman tyypillisen satunnaisuuden lähteen.

Lähtötilanteen epävarmuus. Käytännön malleissa on usein hankalaa saada mitattua tarkasti mallinnettavan systeemin lähtötilanne. Joskus vain erilaisten lähtötilanteiden esiintymiskertojen suhteita voidaan mitata tai arvioida. Tällöin malli on satunnainen, sillä lähtötilanne voidaan mallintaa vain satunnaisena.

Herkkyys lähtötilanteen muutoksille. Jos tarkasteltava ilmiö on luonteeltaan *kaoottinen* eli pienet muutokset lähtötilanteessa voivat saada aikaan suuria muutoksia mittaustuloksissa, niin vaikka malli olisikin deterministinen, niin käytännössä se vaikuttaisi lähes satunnaiselta. Ilmiötä voi olla tällöin järkevää mallintaa yksinkertaisemmin analysoitavana stokastisena mallina.

Epätäydellinen mallinnus. Usein mallin teoreettinen pohja on hatara eli vain osia mallinnettavan ilmiön piirteistä kyetään kuvailemaan deterministisellä mallilla. Tällöin varsinaisen mallin käytös muuttuu ennustamattomaksi ja siksi sitä on parempi mallintaa satunnaisena ilmiönä.

Oleellinen mallin satunnaisuus. Viimeisenä kohtana voi pitää tilanteita, jotka vastaavat kysymykseen ”Onko satunnaisuutta olemassa?” myönteisesti. Esimerkiksi nykyaikainen kvanttiteoria perustuu ajatukseen ilmiöiden perinpohjaisesta satunnaisuudesta. Esimerkiksi yksittäisen fotonin ”päätöstä” heijastua tai olla heijastumatta lasin pinnasta ei voi mitenkään ennustaa, joten sitä on pidettävä puhtaan satunnaisena.

Tällä kursilla emme jatkossa ota enää kantaa mallinuksellisiin kysymyksiin vaan keskitymme yksinkertaisiin stokastisiin malleihin. Jotta saisimme niissä olevan satunnaisuuden kuriin, tulemme tarkastelemaan kaikkea satunnaisuutta yksinkertaisen todennäköisyyslaskennan keinoin.

0.2. **Todennäköisyyslaskennasta ja merkinnöistä.** Palautamme seuraavassa lyhyesti mieleen todennäköisyyslaskennan käsitteitä ja esittelemme myös muutamia kurssilla käytettäviä merkintätapoja.

Kaiken satunnaisuuden käsittelyn takana on (mahdollisesti suuri) musta laatikko, jota nimitetään *todennäköisyysavaruudeksi.* Tämä on kolmikko $(\Omega, \mathscr{F}, \mathbf{P})$.

Joukko Ω on kaikkien alkeistapahtumien muodostama joukko. Kurssin kannalta tällä joukolla ei ole juurikaan merkitystä eli suurimmaksi osaksi joukkoa Ω voi pitää äärellisenä tai numeroituvasti äärettömänä joukkona. Joukko $\mathscr{F}$ on alkeistapahtumien joukon osajoukkojen $\mathscr{P}(\Omega)$ osajoukko, eli niin sanottujen *tapahtumien* joukko. Käytännössä tällä kurssilla kaikki mahdolliset joukon Ω osajoukot ovat tapahtumia. Yleisessä tilanteessa alkeistapahtumia voi olla ”liikaa”, joten välttämättä kaikkien alkeistapahtumien osajoukkojen ei tarvitse olla tapahtumia, mutta ainakin Ω on aina tapahtuma. Yleisestikin tapahtumat on kuvailtavissa seuraavilla säännöillä.

0.1. **Määrittelevät ominaisuudet.**

- joukko Ω on *varma* tapahtuma
- jos A on tapahtuma, niin joukko $A^C := \Omega \setminus A$ on myös tapahtuma (ns. *komplementtitapahtuma*)
- jos $\{\ A_k :\ k = 0, 1, 2, \dots\ \}$ ovat tapahtumia, niin niiden yhdiste
$$\{A_k \text{ tapahtuu jollakin } k = 0, 1, 2, \dots\}$$
on tapahtuma
- jos $\{\ A_k :\ k = 0, 1, 2, \dots\ \}$ ovat tapahtumia, niin niiden leikkaus
$$\{A_k \text{ tapahtuu jokaisella } k = 0, 1, 2, \dots\}$$
on tapahtuma.

Koska tapahtumat $\{A_1, A_2, \dots \text{ ja } A_d\}$ ovat varsin yleisiä, niin käytämme näille lyhennysmerkintää

0.2. **Merkintä.** Kun $A_1, \dots, A_d$ ovat tapahtumia, niin käytämme merkintää
$$A_1 A_2 \dots A_d := \{A_1, A_2, \dots \text{ ja } A_d\}.$$

Kuvaus $\mathbf{P}$ liittää kuhunkin tapahtumaan sen *todennäköisyyden*, mikä on luku suljetulla välillä $[0, 1]$ ja se toteuttaa seuraavat ehdot:

0.3. **Määrittelevät ominaisuudet.**

– varman tapahtuman Ω todennäköisyys $\mathbf{P}(\Omega) = 1$
– jos A on tapahtuma, niin komplementtitapahtuman $A^C := \Omega \setminus A$ todennäköisyys on $\mathbf{P}\left(A^C\right) = 1 - \mathbf{P}(A)$ ja
– jos $(A_k)_{k\in\mathbb{N}}$ ovat pistevieraita tapahtumia, niin

$$\mathbf{P}(A_k \text{ tapahtuu jollakin } k \in \mathbb{N}) = \sum_{k\in\mathbb{N}} \mathbf{P}(A_k)$$

Mallintaaksemme stokastisia ilmiöitä tarvitsemme vielä satunnaismuuttujan sekä ehdollisen todennäköisyyden käsitteet. Palautamme ensin mieleen satunnaismuuttujat.

Satunnaismuuttuja X on (lähes) mielivaltainen kuvaus todennäköisyysavaruudesta *tilajoukkoon* S. Kurssilla S on yleensä jokin äärellinen tai numeroituvasti ääretön joukko. Tällöin satunnaismuuttuja X voidaan tulkita mielivaltaiseksi kuvauksi $\Omega \to S$. Yleisemmässä tapauksessa, meidän tulisi asettaa myös tilajoukkoon sen säännölliset eli mitattavat tapahtumat. Tällöin vaatimus olisi vain: jos $A \subset S$ on tilajoukon mikä tahansa säännöllinen tapahtuma, niin joukon $\{X \in A\}$ on oltava tapahtuma todennäköisyysavaruudessa Ω.

Olemme nyt saaneet kerrattua todennäköisyysavaruuden ja satunnaismuuttujan käsitteet. Jatkossa emme enää kirjoita allaolevaa todennäköisyysavaruutta Ω näkyviin lainkaan. Puhumme vain *tapahtumista* ja *todennäköisyyksistä*. Satunnaismuuttujien kohdalla tarvitsemme vain tiedon *tilajoukosta* S ja siten satunnaismuuttujaa X voimme pitää tilajoukon tuntemattomana alkiona $X \in S$ ja jota voimme käsitellä tarkalleen samoin kuin tilajoukon alkiota.

Tarvitsemme vielä muutaman käsitteen sekä merkinnän. Kun satunnaismuuttujan X tilajoukko $S = \{i_0, i_1, \dots\}$ on jokin positiivisten reaalilukujen $\mathbb{R}_+$ numeroituva osajoukko, niin *satunnaismuuttujan* X *odotusarvo* $\mathbf{E}\,X$ on positiivinen reaaliluku (tai mahdollisesti ääretön ∞)

$$\mathbf{E}\,X := \sum_{k=0}^{\infty} i_k \mathbf{P}(X = i_k). \tag{0.4}$$

Jos tilajoukko $S \subset \mathbb{C}$ on äärellinen, niin sama määritelmä on voimassa, mutta jos tilajoukko on numeroituvasti ääretön kompleksilukujen osajoukko, niin satunnaismuuttujalla *on odotusarvo*, jos myös itseisarvolla $|X|$ on äärellinen odotusarvo. Käytännössä kurssilla satunnaismuuttujat ovat positiivisia[3] tai niillä on odotusarvo.

Yleisessä tapauksessa tilajoukko S voi olla ylinumeroituva kompleksilukujen osajoukko, ja tällöin tarvitsisimme hieman lisätietoja odotusarvosta. Tällaisia

[3] eli tilajoukko on $\mathbb{R}_+$:n osajoukko

tietoja käsitellään lähemmin todennäköisyysteorian kurssilla, mutta myös Mitta- ja integraali –kurssilla, sillä yleisesti odotusarvo on vain mittaintegraali todennäköisyysmitan $\mathbf{P}$ suhteen. Tällaista koneistoa emme kuitenkaan kurssilla tule tarvitsemaan, sillä rajoitumme kysymyksiin, joita pystymme tarkastelemaan yksinkertaisimmilla käsitteillä.

Odotusarvolla on seuraavia ominaisuuksia:

– odotusarvo on lineaarinen eli jos $\alpha, \beta \in \mathbb{C}$ ja X sekä Y ovat satunnaismuuttujia, niin

$$\mathbf{E}\,(\alpha X + \beta Y) = \alpha \mathbf{E}\,X + \beta \mathbf{E}\,Y$$

– jos $0 \leq X_0 \leq X_1, \ldots$ ovat satunnaismuuttujia ja $\lim X_n = X$, niin

$$\mathbf{E}\,X = \lim_{n \to \infty} \mathbf{E}\,X_n\,. \tag{0.5}$$

Näitä kahta ominaisuutta tulemme jatkossa tarvitsemaan usein. Tulemme myös käyttämään seuraavaa niin sanotun Iversonin[4] notaatiota tai *hakasulkumerkinnän*. Jotakin vastaavaa merkintää tarvitaan eri tilanteissa niin usein, että on järkevää käyttää mahdollisimman lyhyttä, selkeää sekä yhtenevää merkintää koko ajan.

0.6. **Merkintä.** Iversonin hakasulkumerkintä tarkoittaa kuvausta väitteiltä luvuille $\{0, 1\}$, joka määritellään seuraavasti:

$$[\,\text{väite}\,] := \begin{cases} 1, & \text{jos väite on tosi,} \\ 0, & \text{jos väite ei ole tosi.} \end{cases}$$

Tämän merkinnän erikoistapauksena saamme esimerkiksi Kroneckerin deltan, sillä $\delta_{ij} = [\,i = j\,]$. Tutustutaan lyhyesti tämän merkinnän "ominaisuuksiin". Voimme esimerkiksi kirjoittaa jokaisen satunnaismuuttujan X, jonka tilajoukko on jokin lukujoukko, yksinkertaisena summana

$$X = \sum_{k \in S} k[\,X = k\,].$$

Yleistämme merkinnän tapahtumille A seuraavasti

0.7. **Merkintä.** Jos A on tapahtuma, niin $[\,A\,]$ on satunnaismuuttuja, jolle

$$[\,A\,](\omega) := [\,\omega \in A\,] = \begin{cases} 1, & \text{jos } \omega \in A, \\ 0, & \text{jos } \omega \notin A, \end{cases}$$

[4]Kenneth Eugene Iversonin mukaan lähteenä Donald Erwin Knuthin *The Art of Computer Progamming, Vol I*

Jatkossa emme tule kirjoittamaan alkeistapahtumaa ω näkyviin, joten jos A on tapahtuma, niin $[\,A\,]$ on satunnaismuuttuja, jonka tilajoukkona on kaksio $\{0,1\}$. Erityisesti havaitsemme, että odotusarvon määritelmän mukaan

$$\mathbf{E}\,[\,A\,] = 0 \times \mathbf{P}\,(\,[\,A\,] = 0\,) + 1 \times \mathbf{P}\,(\,[\,A\,] = 1\,) = \mathbf{P}\,(\,A\,)\,, \tag{0.8}$$

sillä $\{[\,A\,] = 1\} = A$. Siispä induktioilla voimme päätellä, että jos satunnaismuuttujan X tilajoukko $S = \{i_0, i_1, \ldots, i_d\}$, niin odotusarvon lineaarisuuden sekä identiteetin (0.8) avulla voimme johtaa esittämämme odotusarvon määritelmän, sillä

$$\mathbf{E}\,X \;= \mathbf{E}\Big(\sum_k i_k[\,X = i_k\,]\Big) \;= \sum_k i_k \mathbf{E}\,[\,X = i_k\,] \;= \sum_k i_k \mathbf{P}\,(\,X = i_k\,)\,.$$

Myös suorana sovelluksena Iversonin notaatiosta voimme laskea satunnaismuuttujan $f(X)$ odotusarvon, sillä

$$\mathbf{E}\,f(X) \;= \mathbf{E}\Big(\sum_k f(i_k)[\,X = i_k\,]\Big) \;= \sum_k f(i_k)\mathbf{P}\,(\,X = i_k\,)\,.$$

Summauksen ja odotusarvon järjestystä voi aina vaihtaa, kun tilajoukko S on äärellinen. Äärettömän tilajoukon tapauksessa voimme yleensä perustella summauksen ja odotusarvon järjestyksen vaihdon soveltamalla odotusarvon raja-arvo-ominaisuutta (0.5).

Todennäköisyyslaskennan pikakertauksessa tarvitsemme vielä *ehdollisen todennäköisyyden* käsitteen.

0.9. **Merkintä.** Merkitsemme tapahtuman A todennäköisyyttä ehdolla, että tapahtuma B on tapahtunut, seuraavasti

$$\mathbf{P}\,(\,A\,|\,B\,) := \frac{\mathbf{P}\,(\,AB\,)}{\mathbf{P}\,(\,B\,)}$$

Ehdollinen todennäköisyydellä on samat ominaisuudet kuin tavallisella todennäköisyydellä, joten sitä vastaa myös *ehdollinen odotusarvo*:

0.10. **Merkintä.** Merkitsemme satunnaismuuttujajan X ehdollista odotusarvo ehdolla, että tapahtuma B on tapahtunut, seuraavasti

$$\mathbf{E}\;(X\,|\,B)\;:= \sum_k i_k \mathbf{P}\,(\,X = i_k\,|\,B\,)\,.$$

Ehdollisen todennäköisyyden avulla voimme määritellä tapahtumien *riippumattomuuden.*

0.11. **Määritelmä.** Sanomme, että tapahtumajoukko $\{\;A_\lambda :\;\; \lambda \in I\;\}$ on *riippumaton*, jos jokaisella äärellisellä osajoukolla $\{\lambda_0, \ldots, \lambda_d\} \subset I$ on voimassa

$$\mathbf{P}\,\big(\,A_{\lambda_d}\,|\,A_{\lambda_0}A_{\lambda_1}\ldots A_{\lambda_{d-1}}\,\big) = \mathbf{P}\,(\,A_{\lambda_d}\,)\,.$$

Sanomme, että satunnaismuuttujajoukko $\{\ X_\lambda :\ \ \lambda \in I\ \}$ on *riippumaton*, jos aina, kun $\{\ B_\lambda :\ \ \lambda \in I\ \}$ on perhe tilajoukon tapahtumia, niin vastaava tapahtumajoukko

$$\{\ \{X_\lambda \in B_\lambda\} :\ \lambda \in I\ \}$$

on riippumaton.

Olemme nyt käsitelleet lyhyesti tarvittavat todennäköisyyslaskennan käsitteet. Johdatus todennäköisyyslaskentaan –kurssilla esitettyjä malleja ja jakaumia emme tässä kertaa vaan palautamme ne mieleen tarpeen tullessa.

0.3. **Ensimmäinen stokastinen malli.** Voimme nyt aloittaa kurssin pääsisällön, eli stokastisten prosesien tarkastelun. Aloitamme esimerkillä, jossa mallinnetaan sadepäiviä ja niiden todennäköisyyksiä. Seuraava esimerkki on Guttorpin kirjasta. Mallin pohjana olevat havainnot on tehty Yhdysvaltoissa, paikkassa nimeltä Snoqualmie Falls, joka sijaitsee Länsi-Washingtonissa.

0.12**. Esimerkki** (Sadepäivämalli)**.** Jaamme päivät karkeasti kahteen luokkaan: *sadepäiviin* ja *poutapäiviin*. Päivä on *sadepäivä*, jos päivänä aikana sataa ennalta määrätty vähimmäismäärä vettä, muuten päivä on *poutapäivä*. Snoqualmie Fallsissa tehtyjen sademäärämittausten perusteella vuosien 1948–1983 tammikuun päivistä poutapäiviä oli 325 ja sadepäiviä 791. Koska sää on tunnetusti vaikeasti ennustettavissa oleva tapahtuma, yritämme mallintaa sitä stokastisena prosessina.

Merkitään

$$X_{ij} = [\,\text{tammikuun } i\text{. päivänä vuonna } j \text{ satoi}\,].$$

Tässä $i = 1, 2, \ldots, 31$, $j = 1948, \ldots, 1983$ ja kunkin satunnaismuuttujan X_{ij} tilajoukkona on $\{0, 1\} = \{\text{pouta, sade}\}$. Yksinkertaisin satunnaismalli, jolla sateensattumistodennäköisyyttä voisi selittää, on seuraava: oletamme, että kaikki päivät ovat samanlaisia ja täysin toisistaan riippumattomia ja $X_{ij} \sim \mathrm{Bin}(1, p)$, missä p on sateen todennäköisyys. Tätä mallia voitaisiin nimittää *Bernoullin malliksi*. Yksinkertainen lasku antaa mittaustuloksen *uskottavuusfunktion*

$$L(p) = \mathbf{P}\left(\Big(\sum_{ij} X_{ij} \Big) = 791 \right) \propto p^{791}(1-p)^{325}.$$

Tästä saamme suurimman uskottavuuden estimaatin $\hat{p}$ todennäköisyydelle p, joka on tässä tapauksessa

$$\hat{p} = \sum_{i,j} X_{ij}/n = 791/1116 \approx 0,709.$$

	Tänään pouta	Tänään sataa	Yhteensä
Eilen pouta	186 (91)	123 (223)	309
Eilen satoi	128 (223)	643 (543)	771
Yhteensä	314	766	1080

KUVA 1. Sademääräjakaumahavainnot

Tämän avulla voimme arvioida otoskeskihajontaa $\sqrt{p(1-p)/n}$ ja saamme

$$\sqrt{\hat{p}(1-\hat{p})/n} \approx \sqrt{0,709 \times 0,291/1116} \approx 0,014.$$

Kuinka hyvin tämä Bernoullin malli sopii mittausdataan? Tämän selvittämiseksi mallin avulla voi laskea ennusteita tapahtumille ja verrata näitä todellisiin tuloksiin. Eräs on määrätä ennuste niiden päivien lukumäärälle, jolloin päivänä i on pouta ja myös päivänä $i+1$ on pouta, päivänä i sataa ja myös päivänä $i+1$ sataa.

Bernoullin mallin mukaan

$$\mathbf{P}(\text{ tänään ja huomenna poutaa }) = (1-p)^2,$$

joten arviolta $36 \times 30 \times (1-\hat{p})^2 \approx 91$ päivää 1080:sta toteuttaisi tämän. Kuitenkin havaintotaulukon mukaan on helppo havaita, että metsään menee. Taulukossa sulkeissa oleva arvo on Bernoullin mallin antama ennuste. Havaitsemme, että havainnot tukevat väitettä, että sadepäivän jälkeen seuraava päivä on useammin myös sadepäivä, kuin mitä riippumattomuusoletus ehdottaisi. Sama pätee myös poutapäiville. Voimme siis pitää riippumattomuusoletusta varsin epäilyttävänä tässä tapauksessa.

Voikin kysyä, että kuinka tätä mallia voisi muuttaa, jotta se selittäisi sadejakaumahavainnoissa esiintyvän korrelaation paremmin? Jos luovumme riippumattomuusoletuksesta, niin emme voi käyttää kaavaa

$$\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbf{P}(X_0 = i_1) \times \cdots \times \mathbf{P}(X_n = i_n)$$

satunnaismuuttujien yhteisjakauman määräämiseksi. Ehdollisen todennäköisyyden määritelmän mukaan voisimme kyllä kirjoittaa

$$\begin{aligned}
&\mathbf{P}(X_0 = i_0, X_1 = i_1, \dots, X_{n+1} = i_{n+1}) \\
&= \mathbf{P}(X_0 = i_1, \dots, X_n = i_n)\, \mathbf{P}(X_{n+1} = i_{n+1} \mid X_0 = i_1, \dots, X_n = i_n),
\end{aligned}$$

eli meidän tulisi määrätä ehdolliset todennäköisyydet

$$\mathbf{P}(X_{n+1} = i_{n+1} \mid X_0 = i_1, \dots, X_n = i_n). \tag{0.13}$$

Tulemmekin koko kurssin ajan tarkastelemaan sademallia vastaavaa tilanetta, kun oletamme, että ehdollinen todennäköisyys (0.13) riippuukin vain tiedosta $\{X_n = i_n\}$.

Esimerkissämme tämä oletus tarkoittaa seuraavaa: sen sijaan, että olettaisimme satunnaismuuttujien $X_{ij} \sim \mathrm{Bin}(1, p)$ olevan riippumattomia, oletamme että

$$\begin{aligned} p_{00} &:= \mathbf{P}\left(X_{(n+1),j} = 0 \,|\, X_{nj} = 0 \right) = 186/309 \approx 0,602 \\ p_{01} &:= \mathbf{P}\left(X_{(n+1),j} = 1 \,|\, X_{nj} = 0 \right) = 123/309 \approx 0,398 \\ p_{10} &:= \mathbf{P}\left(X_{(n+1),j} = 0 \,|\, X_{nj} = 1 \right) = 128/771 \approx 0,166 \\ p_{11} &:= \mathbf{P}\left(X_{(n+1),j} = 1 \,|\, X_{nj} = 1 \right) = 643/771 \approx 0,834 \end{aligned}$$

kun $n = 1, \dots, 30$. Edelleen oletamme, että

$$\mathbf{P}\left(X_{(n+1),j} = i_{n+1} \,|\, X_{1j} = i_1, \dots, X_{nj} = i_n \right) = \mathbf{P}\left(X_{(n+1),j} = i_{n+1} \,|\, X_{nj} = i_n \right).$$

Vuosien ajattelemme (ainakin vielä) olevan riippumattomat.