Stokastisilla prosesseilla mallintaminen

Janika Kääriäinen

Sisältö

Luku 1

Johdanto

Stokastisilla prosesseilla tarkoitetaan satunnaisia ilmiöitä, jotka muuttuvat ajan kuluessa. Stokastinen prosessi on satunnaismuuttujajono (X_n ; n = 0, 1, 2, …). Satunnaismuuttujat ovat ajanhetkillä indeksöityjä kuvauksia todennäköisyysavaruudelta tila-avaruuteen, eli X_0 , X_1 , … , ϵ S. Tilajoukko $S \neq \emptyset$ on äärellinen tai numeroituva joukko. Tilat ovat tilajoukon alkioita i,j ϵ S. Ajanhetket ovat luonnollisia lukuja n, m = 0, 1, 2, … Toisinaan aika ja satunnaismuuttujat voivat olla diskreetin sijaan jatkuvia. Prosessi on stationaarinen, jos sen tilastolliset ominaisuudet (esim. odotusarvo, varianssi) eivät muutu ajan kuluessa.

Luku 2

Markovin ketju

Stationaarinen stokastinen prosessi (X_n) on <u>Markovin ketju</u> (Markov chain), jos kaikilla ajanhetkillä n, m ja tiloilla i, j ϵ S pätee

$$\mathrm{P}(X_{n+1}=j \mid X_0=i_0, \ldots, X_{n-1}=i_{n-1}, X_n=i) = \mathrm{P}(X_{n+1}=j \mid X_n=i)$$

Kun merkitään <u>siirtymätodennäköisyyttä</u> tilasta $X_n = i$ tilaan $X_{n+1} = j$ p_{ij} :llä, niin pätee myös

$$p_{ij} = \mathrm{P}(X_{n+1}=j \mid X_n=i) = \mathrm{P}(X_{m+1}=j \mid X_m=i)$$

Markovin ketjussa siis uusi tila riippuu vain edellisestä tilasta. Sillä on ns. ”unohtavaisuusominaisuus” eli uuteen tilaan ei vaikuta historia edellistä tilaa pitemmälle. Siirtymätodennäköisyydet riippuvat ainoastaan tiloista i ja j.

Markovin ketjun (X_n) <u>alkujakauma</u> on satunnaismuuttujan X_0 jakauma, eli todennäköisyydet

$$p_i = \mathrm{P}(X_0=i), \quad i \in S.$$

Merkintöjä: $$p_{ij}^{(m)} = \mathrm{P}(X_m=j \mid X_0=i),$$

siirtymämatriisi: $$P^{(m)} = (p_{ij}^{(m)})_{ij} = \begin{matrix} p_{00}^{(m)} & p_{01}^{(m)} & \cdots \\ p_{10}^{(m)} & p_{11}^{(m)} & \cdots \\ \vdots & \vdots & \ddots \end{matrix} \quad \text{jossa } P^{(1)} = P.$$

<u>Lause 2.1</u> *Kaikilla ajanhetkillä m on voimassa*

$$P^{(m)} = P^m, \quad \textit{jossa } P^m \textit{ on matriisipotenssi.}$$

Lisäksi kaikilla ajanhetkillä n ja m ja tiloilla i, j ϵ S pätee

$$\mathrm{P}(X_{n+m}=j \mid X_0=i_0, \ldots, X_{n-1}=i_{n-1}, X_n=i) = p_{ij}^{(m)} \quad \textit{(yleistetty Markov-ominaisuus)}$$

Todistus sivuutetaan.

<u>Lause 2.2</u> (*Chapmanin-Kolmogorovin yhtälöt*) *Kaikilla ajanhetkillä n ja m sekä tiloilla i, j ϵ S on voimassa*

$$p_{ij}^{(n+m)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}^{(m)}.$$

Todistus.

$$\sum_k p_{ik}^{(n)} p_{kj}^{(m)} = \sum_k (P^n)_{ik} (P^m)_{kj} = (P^n P^m)_{ij} = (P^{n+m})_{ij} = p_{ij}^{(n+m)}. \square$$

Esim 2.3 PRA:n periytyminen Hardy-Weinbergin mallilla.

PRA (Progressive Retinal Atrophy) eli etenevä verkkokalvon surkastuma on mm. Novascotiannoutajilla eli tollereilla esiintyvä perinnöllinen sairaus. Taudilla on peittyvä periytymistapa, eli sairastumiseen vaaditaan virheellinen geeni sekä isältä että emältä. Koira on terve kantaja, jos se on saanut yhden normaalin ja yhden virheellisen geenin.

Merkitään normaalia geeniä A:lla ja virheellistä a:lla. Nyt tilajoukko on S = {AA, Aa, aa}. Oletamme että jokainen tolleri saa yhden jälkeläisen. Merkitsemme X_n = "n:nnen sukupolven jälkeläisen geenitila", jolloin $X_n \in S$. Olkoon kaikkien tollereiden lukumäärä N, ja merkitään esimerkiksi N(AA):lla geeniparien AA lukumäärää. Merkitään geeniparien suhteellisia osuuksia seuraavasti: p = N(AA) / N, q = N(Aa) / N, r = N(aa) / N.

Vuonna 2003 tehdyssä tutkimuksessa, jossa testattiin maailmanlaajuisesti 898 tolleria, saatiin seuraavat suhteelliset osuudet: p = 48%, q = 45% ja r = 7%. Tällöin A-alleelien osuus on P = p + q / 2 = 48% + 45% / 2 = 70,5 % ja a-alleelien osuus on Q = r + q / 2 = 7% + 45% / 2 = 29,5 % .

Laskemme siirtymätodennäköisyydet edellä mainitun tilaston mukaan:

$P(X_{n+1} = AA \mid X_n = AA) = P = 70,5\%$

$P(X_{n+1} = AA \mid X_n = Aa) = (1 / 2)\, P = 35,25\%$

$P(X_{n+1} = AA \mid X_n = aa) = 0\%$

$P(X_{n+1} = Aa \mid X_n = AA) = Q = 29,5\%$

$P(X_{n+1} = Aa \mid X_n = Aa) = (1 / 2)\, Q + (1 / 2)\, P = 1 / 2 = 50\%$

$P(X_{n+1} = Aa \mid X_n = aa) = P = 70,5\%$

$P(X_{n+1} = aa \mid X_n = AA) = 0\%$

$P(X_{n+1} = aa \mid X_n = Aa) = (1 / 2)\, Q = 14,75\%$

$P(X_{n+1} = aa \mid X_n = aa) = Q = 29,5\%$

Kaavioissa pystygeeniparit ovat X_n ja vaakageeniparit ovat X_{n+1}

Kaavio 1	AA	Aa	aa
AA	P	Q	0
Aa	P / 2	50%	Q / 2
aa	0	P	Q

Kaavio 2	AA	Aa	aa
AA	70,5%	29,5%	0%
Aa	35,25%	50%	14,75%
aa	0%	70,5%	29,5%

Kaavio 3	AA	Aa	aa
AA	50%	50%	0%
Aa	25%	50%	25%
aa	0%	50%	50%

Kaaviossa 1 siirtymätodennäköisyydet on merkitty alleelien suhteellisilla osuuksilla P ja Q. Kaaviossa 2 siirtymätodennäköisyyksiin on sijoitettu tutkimuksen mukaiset suhteelliset alleeliosuudet. Kaaviossa 3 on käytetty vertailun vuoksi oletusta, että suhteelliset osuudet P ja Q ovat yhtäsuuret P = Q = 50%.

Luku 3

Poissonin prosessi

<u>Määritelmiä 3.1</u> Oletetaan, että $\lambda > 0$ on reaalinen. Satunnaismuuttuja $T \sim \text{Exp}(\lambda)$, eli T on <u>eksponenttijakautunut</u> parametrilla λ, jos

$$F(t) = P(T \leqslant t) = 1 - e^{-\lambda t} \qquad \text{kullakin} \quad t \in \mathbb{R}.$$

Oletetaan, että $\lambda > 0$ ja satunnaismuuttuja $N \in \mathbb{N}$, $N \sim \text{Poisson}(\lambda)$ eli N on <u>Poisson-jakautunut</u> parametrilla λ, jos

$$P(N=n) = (e^{-\lambda} \lambda^n)/n!, \quad \text{kun} \quad n \in \mathbb{N}$$

Oletetaan, että satunnaismuuttujat T_i ovat riippumattomia, i = 1, 2,…, ja $T_i \sim \text{Exp}(\lambda)$. Satunnaismuuttuja $S_i = T_1 + T_2 + \ldots + T_i$ on <u>i:nnen insidenssin hetki</u> (vertaa *incident*). Satunnaismuuttuja $N(t) \in \mathbb{N}$ on

$$N(t) = \begin{cases} 0, \; kun \;\; 0 \leqslant t \leqslant S_1 \\ 1, \; kun \;\; S_1 \leqslant t \leqslant S_2 \\ \quad \vdots \\ i, \; kun \;\; S_i \leqslant t \leqslant S_{i+1} \\ \quad \vdots \end{cases} .$$

Stokastinen prosessi $(N(t) \,;\, t \geq 0)$ on <u>Poissonin prosessi</u>, jonka parametri on λ .

<u>Lause 3.2</u> *Satunnaismuuttuja $N(t) \sim \text{Poisson}(\lambda t)$ jokaisella $t \geq 0$ eli*

$$p_k(t) = (e^{-\lambda t} (\lambda t)^k)/k!.$$

Todistuksen idea. Olkoon $\lambda > 0$ eräänlainen ”tapahtumaintensiteetti”, ja $h > 0$ hyvin pieni. Nyt

$P(N(t+h)=k+1 \parallel N(t)=k) \approx \lambda h$, joten

$$p_k(t) = P(N(t)=k) \approx P(N(t-h)=k-1, N(t)=k) + P(N(t-h)=k, N(t)=k)$$

$$= P(N(t-h)=k-1)\, P(N(t)=k \parallel N(t-h)=k-1) + P(N(t-h)=k)\, P(N(t)=k \parallel N(t-h)=k)$$

$$\approx p_{k-1}(t-h)\lambda h + p_k(t-h)(1-\lambda h).$$

Siis $\quad p_k(t) \approx p_{k-1}(t-h)\lambda h + p_k(t-h)(1-\lambda h)$.

Nyt $\quad p_k{}'(t) = \lambda \lim_{n \to 0+} (p_{k-1}(t-h) - p_k(t-h)) = \lambda(p_{k-1}(t) - p_k(t)).$

Jos k = 0 , niin $p_{-1}(t) = 0$. Käytetään generoivaa funktiota

$$f_s(t) = \sum_{k=0}^{\infty} p_k(t) s^k \quad , \; 0 \leq s \leq 1 ,$$

ja derivoidaan se, jolloin saadaan

$$f_s{}'(t) = \sum_{k=0}^{\infty} p_k{}'(t) s^k = \sum_{k=0}^{\infty} \lambda(p_{k-1}(t) - p_k(t)) s^k = \lambda \sum_{k=-1}^{\infty} p_k(t) s^{k+1} - \lambda f_s(t)$$

$$= s\lambda f_s(t) - \lambda f_s(t) = \lambda(s-1) f_s(t).$$

Pitämällä mielessä, että $e^x = \sum_{k=0}^{\infty} x^k / k!$, saadaan ratkaisu

$$f_s(t) = \sum_{k=0}^{\infty} p_k(t) s^k = e^{-\lambda t} e^{\lambda s} = e^{-\lambda t} \sum_{k=0}^{\infty} (\lambda^k s^k) / k! \quad \text{eli} \quad p_k(t) = (e^{-\lambda t} \lambda^k) / k!.$$

Lause 3.3 *Kun N(t) on Poissonin prosessi, niin*

$$P(N(s+t) - N(s) = i \parallel N(s) = k) = P(N(t) = k) = (e^{-\lambda t} (\lambda t)^i) / i!.$$

Todistus sivuutetaan.

Esimerkki 3.4 Neljän hengen ryhmämatkaa varten kerätään ilmoittautumisia. Oletetaan, että ilmoittautumisia tulee Poissonin prosessin mukaisesti. Kolmen tunnin aikana ilmoittautuu 15 ihmistä. Millä todennäköisyydellä ryhmä tuli täyteen 1. tunnin aikana?

Ratkaisu. Olkoon N(t) = ilmoittautuneiden lukumäärä hetkellä $t \geq 0$. Nyt $N(t) \sim$ Poisson (λt) . Merkitään A_i = ilmoittautuneita on i kpl , i = 0, 1, …, joten {ryhmä on täynnä} = $A_4 \cup A_5 \cup \ldots$. Lasketaan ensin todennäköisyydet p_0 = P(N(1) = 0 | N(3) = 15), p_1 = P(N(1) = 1 | N(3) = 15), p_2 = P(N(1) = 2 | N(3) = 15), p_3 = P(N(1) = 3 | N(3) = 15). Kun k = 0, 1, 2, 3, saadaan

$$p_k = \frac{P(N(1)=k) P(N(3) - N(1) = 15 - k)}{P(N(3)=15)} = \frac{e^{-\lambda} \lambda^k e^{-2\lambda} (2\lambda)^{15-k}}{k!(15-k)!} \Big/ \frac{e^{-3\lambda}(3\lambda)^{15}}{15!}$$

$$= \frac{2^{15-k} 15!}{3^{15} k!(15-k)!}.$$ Nyt

$p_0 = \dfrac{2^{15} 15!}{3^{15} 15!} = (\dfrac{2}{3})^{15} \approx 2{,}28366 * 10^{-3} \approx 0{,}0023$, $p_1 = \dfrac{2^{14} 15!}{3^{15} 14!} = 5 * (\dfrac{2}{3})^{14} \approx 0{,}01713$

$\approx 0{,}017$, $p_2 = \dfrac{2^{13} 15!}{3^{15} 2! 13!} = \dfrac{15*17}{9} (\dfrac{2}{3})^{13} \approx 0{,}05995 \approx 0{,}060$, $p_3 = \dfrac{2^{12} 15!}{3^{15} 3! 12!}$

$= \dfrac{455}{27} (\dfrac{2}{3})^{12} \approx 0{,}12988 \approx 0{,}13$. Nyt kysytty todennäköisyys on $1 - p_0 - p_1 - p_2 - p_3 = 1 - 2{,}28366*10^{-3} - 0{,}01713 - 0{,}05995 - 0{,}12988 = 0{,}79076 = 0{,}79$.

Luku 4

Tasapainojakauma ja absorptiotodennäköisyydet

Määritelmiä 4.1 Olkoon $\pi = (\pi_i)_{i \in S}$ todennäköisyysjakauma tilajoukolla S. Jakauma π on Markovin ketjun (X_n) tasapainojakauma, jos

$$\sum_{i \in S} \pi_i p_{ij} = \pi_j \qquad \text{jokaisella } j \in S.$$

Tämä voidaan kirjoittaa muodossa $\pi P = \pi$, kun π on vaakavektori. Tasapainojakauman avulla voidaan tarkastella, kuinka Markovin ketju käyttäytyy keskimäärin pitkillä aikaväleillä.

Oletetaan, että Markovin ketjulla (X_n) on yksikäsitteinen tasapainojakauma π . Merkitään

$$N_j(n) = \# \{0 \leqslant m < n : X_m = j\}$$

eli satunnaismuuttuja $N_j(n)$ on tilassa j käyntien kokonaislukumäärä aikavälillä $[0, n-1]$. Merkitään

$$\hat{\pi}_j(n) = \frac{N_j(n)}{n},$$

joka on siis tilassa j käyntien suhteellinen lukumäärä aikavälillä $[0, n-1]$.

Lause 4.2 *(Ergodilause) Olkoon Markovin ketjulla (X_n) yksikäsitteinen tasapainojakauma π . Tällöin*

$$\lim_{n \to \infty} \hat{\pi}_j(n) = \pi_j.$$

Todistus sivuutetaan.

Määritelmiä 4.3 Olkoon $A \subset S$. Joukko A on absoptiojoukko, jos $p_{ij} = 0$ aina kun $i \in A$ ja $j \in A^c$. Markovin ketju ei siis poistu absorptiojoukosta sinne saavuttuaan. Absorptiohetki T_A on satunnainen ajanhetki

$$T_A = min\{n \in \mathbb{N} : X_n \in A\}.$$

$T_A = \infty$, jos $X_n \in A^c$ jokaisella ajanhetkellä n. Merkitsemme {absorptio tapahtuu joskus} = $\{T_A < \infty\}$ ja {absorptio ei tapahtu koskaan} = $\{T_A = \infty\}$. Absorptiotodennäköisyys lähtötilasta i on $h_{iA} = P_i(T_A < \infty)$, jossa merkintä $P_i(A) = P(A \mid X_0 = i)$ tarkoittaa A:n todennäköisyyttä lähtötilasta i.

Olkoon h pystyvektori

$$h = \begin{pmatrix} h_0 \\ h_1 \\ h_2 \\ \vdots \end{pmatrix}$$

ja oletetaan, että $0 \le h_i \le 1$. Vektori h on <u>harmoninen</u> , jos

$$h_i = \sum_{j \in S} p_{ij} h_j \quad , i \in S ,$$

eli jos h = Ph .

<u>Lause 4.4</u> *Pystyvektori h_A on harmoninen. Lisäksi $h_{iA} = 1$ aina, kun $i \in A$.*

Todistuksen idea. Jos $i \in A$, niin $P_i(X_n \in A)$ jokaisella ajanhetkellä n, joten $h_{iA} = 1$.

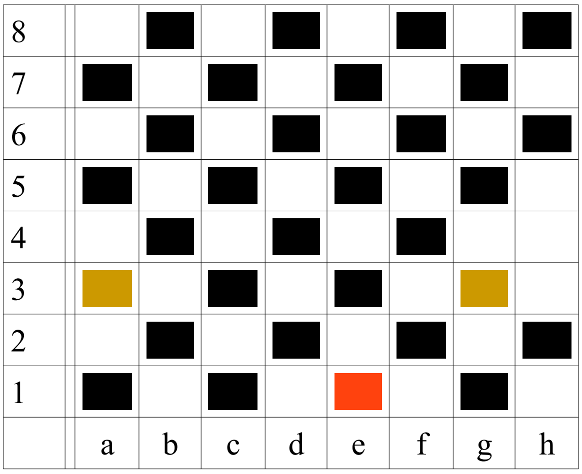
Osoitetaan $Ph_A = h_A$. Koska

$$h_{iA} = \lim_{n \to \infty} P_i(T_A \leqslant n) = \lim_{n \to \infty} P_i(X_n \in A) = \lim_{n \to \infty} \sum_{j \in A} p_{ij}^{(n)} ,$$

saamme

$$\sum_{j \in S} p_{ij} h_{jA} = \sum_j p_{ij} \lim_{n \to \infty} \sum_{k \in A} p_{jk}^{(n)} = \lim_{n \to \infty} \sum_j \sum_{k \in A} p_{ij} p_{jk}^{(n)} = \lim_{n \to \infty} \sum_{k \in A} \sum_j p_{ij} p_{jk}^{(n)} = \lim_{n \to \infty} \sum_{k \in A} p_{ik}^{(n+1)}$$
$$= \lim_{n \to \infty} \sum_{k \in A} p_{ik}^{(n)} = = h_{iA} .$$

<u>Esim 4.5</u>



Kuningas lähtee ruudusta e1, ja halutaan, että se liikkuu ruudun kerrallaan ja ainoastaan vain mustilla ruuduilla (liikkuu siis viistoon). Kuningas voi liikkua vasempaan tai oikeaan etu- tai takaviistoon. Poikkeuksena ovat laudan alimmat ruudut sekä reunaruudut. Jokaiseen sallittuun ruutuun kuningas liikkuu yhtä todennäköisesti.

Mikä on todennäköisyys, että kuningas pääsee ruudusta e1 ruutuihin a3 tai g3, eikä joudu ruutuihin c3 tai e3? Mistä alaruudusta on suurin todennäköisyys päästä ruutuihin a3 tai g3?

Ratkaisu. Absorptiojoukko on A {a3, g3}. Nyt $h_{a3} = 1$, $h_{c3} = 0$, $h_{e3} = 0$, $h_{g3} = 1$, h = Ph.

$$
\begin{array}{c|cccccccccccc}
\square & a1 & c1 & e1 & g1 & b2 & d2 & f2 & h2 & a3 & c3 & e3 & g3 \\
a1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c1 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 \\
e1 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 & 0 \\
g1 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 & 0 & 0 & 0 & 0 \\
b2 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 0 \\
d2 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 & 0 \\
f2 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 & 1/4 & 1/4 \\
h2 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
a3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
c3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
e3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
g3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}
\;*\;
\begin{array}{c}
h_{a1} \\ h_{c1} \\ h_{e1} \\ h_{g1} \\ h_{b2} \\ h_{d2} \\ h_{f2} \\ h_{h2} \\ 1 \\ 0 \\ 0 \\ 1
\end{array}
\;=\;
\begin{array}{c}
h_{a1} \\ h_{c1} \\ h_{e1} \\ h_{g1} \\ h_{b2} \\ h_{d2} \\ h_{f2} \\ h_{h2} \\ 1 \\ 0 \\ 0 \\ 1
\end{array}
$$

$\Leftrightarrow$ $h_{b2} = h_{a1} = 103/239 \approx 0{,}431$, $h_{d2} = 37/239 \approx 0{,}155$, $h_{c1} = 70/239 \approx 0{,}293$,

$h_{f2} = 119/239 \approx 0{,}498$, $h_{e1} = 78/239 \approx 0{,}326$ (todennäköisyys ruudusta e1),

$h_{h2} = 199/239 \approx 0{,}833$, $h_{g1} = 159/239 \approx 0{,}665$ (suurin todennäköisyys alaruuduista).

Luku 5

Haarautumisketjut

<u>Määritelmiä 5.1</u> Oletetaan, että satunnaismuuttuja X_n on sukupolven n koko eli yksilöiden lukumäärä sukupolvessa n = 0, 1, 2, … Aluksi on yksi yksilö eli $X_0 = 1$. Kaikki yksilöt ovat riippumattomia, sekä identtisiä eli jälkeläisten lukumääräjakauma on jokaisella sama. Yksilöiden jälkeläisten lukumärää kuvaa satunnaismuuttuja ξ , jonka jakauma on $a_k = P(\xi = k)$, k = 0, 1, 2, …

Merkitään sukupuuton todennäköisyyttä

$$q \ = \ P(sukupuutto) \ = \ P(X_n \ jollakin \ n) \ = \ P(T_0 < \infty).$$

Oletamme, että $P(\ \xi = 0\ ) > 0$, koska jos $P(\ \xi = 0\ ) = 0$, niin

$$X_{n+1} \ = \ \sum_{i=0}^{X_n} \xi_i^{(n)} \geqslant \sum_{i=1}^{X_n} 1 = X_n,$$

eli $1 \ = \ X_0 \leqslant X_1 \leqslant ....,$ joten $X_n \neq 0,$ kun $n \geq 1$. Näin ollen sukupuuttoa ei tapahdu. Oletamme myös, että $a_0 + a_1 < 1$. Jos $a_0 + a_1 = 1$ sallittaisiin, niin $P(\ \xi > 1\ ) = 0$. Tällöin jälkeläisiä olisi joko 0 tai 1, joten pakostakin joskus tapahtuisi sukupuutto. Oletus $a_0 + a_1 < 1$ tarkoittaa, että $P\ (\ \xi \geq 2\ ) > 0$.

Merkitään jälkeläisten lukumäärää kuvaavan satunnaismuuttujan ξ odotusarvoa

$$m \ = \ \mathrm{E}\zeta \ = \ \sum_{k \geqslant 0} k a_k .$$

Saadaksemme selville sukupuuton todennäköisyyden, otamme käyttöön generoivan funktion φ,

$$\varphi : [0,1] \to \mathbb{R} \ \ \varphi(s) = \mathrm{E}\, s^k \ = \ \sum_{k=0}^{\infty} s^k a_k \ = \ a_0 + a_1 s + a_2 s^2 + \ldots \ .$$

Tämän funktion ominaisuuksien perusteella voidaan osoittaa, että kun m > 1, niin yhtälöllä $\varphi(s) = s$ on kaksi ratkaisua, $s_1 < 1$ ja triviaali ratkaisu $s_2 = 1$. Kun $m \leq 1$, yhtälöllä on vain triviaaliratkaisu.

<u>Lause 5.2</u> *Sukupuuton todennäköisyys q on yhtälön $\varphi(s) = s$ pienin juuri, eli*

$$P\ (\ sukupuutto\ ) = s_1 .$$

Lisäksi $q < 1 \ \Leftrightarrow \ m = \mathrm{E}\,\xi > 1$.

Todistus sivuutetaan.

<u>Esim 5.3</u> Oletetaan, että sairas yksilö tartuttaa erästä bakteeria k = 0, 1, 2, 3 yksilöä todennäköisyyksillä $a_0 = a$, $a_1 = (1 - a) / 2$, $a_2 = (1 - a) / 4$, $a_3 = (1 - a) / 4$, kun $a \geq 0$. Millä todennäköisyydellä puhkeaa epidemia? Kuinka suuri osa väestöstä on rokotettava, ettei epidemia puhkea?

Ratkaisu.

$$P(epidemia) = 1 - q.$$

$$m = \mathrm{E}\zeta = \sum_{k\geqslant 0} k\,a_k = \frac{1-a}{2} + 2*\frac{1-a}{4} + 3*\frac{1-a}{4} = \frac{7-7a}{4} > 1 \Leftrightarrow \quad 0 \leqslant a \leqslant \frac{3}{7}.$$

$$\varphi(s) = \mathrm{E}\,s^k = \sum_{k=0}^{\infty} s^k a_k = a + \frac{1-a}{2}s + \frac{1-a}{4}s^2 + \frac{1-a}{4}s^3.$$

Olkoon $\varphi(s) = s$, jolloin

$$\frac{1-a}{4}s^3 + \frac{1-a}{4}s^2 + \frac{1-a}{2}s + a = s \Leftrightarrow \quad s^3 + s^2 - \frac{2(1+a)}{1-a}s + \frac{4a}{1-a} = 0 \Leftrightarrow$$

$$(s-1)(s^2 + 2s - \frac{4a}{1-a}) = 0 \Leftrightarrow \quad s_1 = 1,\ s_2 = -1 + \sqrt{\frac{1+3a}{1-a}},\ s_3 = -1 - \sqrt{\frac{1+3a}{1-a}}.$$

s_3 ei käy, ja koska $s_2 < s_1$ aina kun $a < 2$, niin

$$q = s_2 = -1 + \sqrt{\frac{1+3a}{1-a}}. \quad \text{Nyt} \quad P(epidemia) = 1 - (-1 + \sqrt{\frac{1+3a}{1-a}}) = 2 - \sqrt{\frac{1+3a}{1-a}}.$$

Oletetaan, että osuus $0 < \theta < 1$ rokotetaan. Merkitään ζ^* = tartuntojen lukumäärä rokotetussa populaatiossa.

$$a_0^* = P(\zeta^* = 0) = \theta + (1-\theta)a_0,\ a_k^* = (1-\theta)a_k, \quad \text{kun} \quad k \geqslant 1$$

$$m^* = \mathrm{E}\,\zeta^* = \sum_{k=0}^{\infty} ka_k^* = (1-\theta)\sum_{k=0}^{\infty} ka_k = (1-\theta)m \leqslant 1 \Leftrightarrow \quad \theta \geqslant 1 - 1/m = 1 - \frac{7-7a}{4} = \frac{7a-3}{4}.$$

Väestöstä on siis rokotettava osuus $\frac{7a-3}{4}$ jottei epidemia puhkea.

Luku 6

Syntymä- ja kuolemaprosessi

Määritelmiä 6.1 Nimitetään parametreja λ_0, λ_1, … ≥ 0 syntymäintensiteeteiksi ja parametreja μ_1, μ_2, … ≥ 0 kuolemaintensiteeteiksi.

Oletetaan, että

$P(X(t+h)=i+1 \| X(t)=i) \approx \lambda_i h$ ja $P(X(t+h)=i-1 \| X(t)=i) \approx \mu_i h,$ jolloin

$P(X(t+h)=i \pm 1 \| X(t)=i) \approx (\mu_i + \lambda_i)h.$

Olkoon T ~ Exp(λ_i) ja U ~ Exp(μ_i), sekä T ja U riippumattomia . Satunnaismuuttujat T ja U kuvaavat seuraavan syntymän ja kuoleman väliaikaa. Jos T < U , tapahtuu syntymä eli siirrymme tilasta i tilaan i+1, ja jos T > U, niin tapahtuu kuolema. Tilassa i viipymisaika on min(T,U) ~ Exp(λ_i +μ_i).

Johdetaan differentiaaliyhtälöt syntymä-kuolema-prosessin jakaumalle.

$$p_i(t) = P(X(t)=i)$$

$$\approx P(X(t-h)=i-1, X(t)=i) + P(X(t-h)=i+1, X(t)=i) + P(X(t-h)=i, X(t)=i)$$
$$P(X(t-h)=i-1)P(X(t)=i \| X(t-h)=i-1) + P(X(t-h)=i+1)P(X(t)=i \| X(t-h)=i+1) +$$
$$P(X(t-h)=i)P(X(t)=i \| X(t-h)=i),$$
$$p_i(t) \approx p_{i-1}(t-h)\lambda_{i-1}h + p_{i+1}(t-h)\mu_{i+1}h + p_i(t-h)(1-(\lambda_i+\mu_i)h).$$

Saadaan siis $p_i'(t) \approx \lim\left(\frac{p_i(t)-p_i(t-h)}{h}\right) = p_{i-1}(t)\lambda_{i-1} + p_{i+1}(t)\mu_{i+1} - p_i(t)(\lambda_i+\mu_i)$.

Todennäköisyysjakauma π on SK-prosessin tasapainojakauma, jos $p_i(t) = \pi_i$ kaikilla i ∊ S ja jokaisella t. Tästä saadaan $p_i'(t) = D\pi_i = 0$ eli $0 = p_i'(t) = \lambda_{i-1}\pi_{i-1} + \mu_{i+1}\pi_{i+1} - \pi_i(\lambda_i + \mu_i)$. Ratkaistaan tämä differenssiyhtälö. Nyt $\lambda_{-1}=0$ ja $\mu_0 = 0$.

i = 0: $\mu_1\pi_1 - \lambda_0\pi_0 = 0 \;\Rightarrow\; \pi_1 = (\lambda_0\pi_0) / \mu_1$

i = 1: $(\lambda_0\pi_0 - \mu_1\pi_1) + (\mu_2\pi_2 - \lambda_1\pi_1) = \mu_2\pi_2 - \lambda_1\pi_1 = 0 \;\Rightarrow\; \pi_2 = (\lambda_1\pi_1)/\mu_2 = [(\lambda_0\lambda_1)/(\mu_1\mu_2)]\pi_0$

Induktiolla saadaan

$\pi_i = \frac{\lambda_{i-1}\lambda_{i-2}\ldots\lambda_1\lambda_0}{\mu_i\mu_{i-1}\mu_1}\pi_0 =: \rho_i\pi_0.$ Valitaan $\rho_0 = 1$. $1 = \sum_{i\in S}\pi_i = \pi_0\sum_{i=0}^{\infty}\rho_i =: \pi_i C.$

Näin ollen tasapainojakauma π on olemassa, jos C = $\sum \rho_i < \infty$. Tällöin $\pi_i = \frac{\rho_i}{C}$.

Lause 6.2 *(Ergodilause SK-prosesseille) Pitkällä aikavälillä π_i on likimain SK-prosessin tilassa i viettämä suhteellinen aika.*

Todistus sivuutetaan.

Esim 6.3 Linjastoon saapuu Poissonin prosessin mukaisesti asiakkaita keskimäärin 40 tunnissa. Linjastossa on aina vapaa palvelupiste. Palveluajat ovat eksponenttijakautuneita ja riippumattomia, ja asiakas viipyy linjastossa keskimäärin 15 minuuttia. Olkoon X(t) asiakkaiden lukumäärä linjastossa hetkellä $t \geq 0$.

Millä todennäköisyydellä asiakkaita on paikalla yhteensä 15? Entä paljonko asiakkaita on keskimäärin paikalla?

Ratkaisu. Oletetaan, että X(t) = i. Nyt

$$q_{i,i+1}=\lambda \quad ja \quad q_{i,j}=\begin{cases}\mu i, i\geqslant 0 \; ja \; j=i-1, \\ 0, \qquad muuten\end{cases}$$

Siis $\lambda_i = \lambda$ ja $\mu_i = \mu i$. Olkoon N(1) = asiakkaita saapuu hetkellä t =1, jolloin N(1) ~ Poisson(λ). N(1) = λ = 40. Olkoon S asiakkaan palvelusaika, jolloin S ~ Exp(μ). Nyt ES=1/μ= ¼ eli μ = 4.

$$\rho_i=\frac{\lambda_{i-1}\lambda_{i-2}\ldots\lambda_1\lambda_0}{\mu_i\mu_{i-1}\mu_1}=\frac{\lambda^i}{\mu^i i!}.$$

Kun käytetään Taylorin sarjaa, saadaan $C=\sum_i \rho_i=\sum_i \left(\frac{\lambda^i}{\mu^i i!}\right)=e^{\lambda/\mu}$, joten

$$\pi_i=\frac{\rho_i}{C}=\frac{\lambda^i}{\mu^i i!}:e^{\lambda/\mu}=\frac{\lambda^i e^{-\lambda/\mu}}{\mu^i i!}=\frac{e^{-\lambda/\mu}(\lambda/\mu)^i}{i!}$$

Nyt $\pi_{15}=\dfrac{e^{-10}10^{15}}{15!}\approx 0{,}034718\approx 0{,}035$,

$\mathrm{E}\,X(t)=\sum_{i=0}^{\infty} p_i(t) i=\sum_{i=0}^{\infty}\pi_i i$. Koska π_i ~ Poisson(λ / μ), niin $E\pi_i$ = λ / μ = 10. Eli todennäköisyys, että asiakkaita on 15, on 0,035 ja asiakkaita on keskimäärin paikalla 10.

Luku 7

Äärellinen Markovin prosessi

Määritelmiä 7.1   Parametreja $q_{ij} \geq 0$ nimitetään siirtymäintensiteeteiksi tilasta i tilaan j, kun i,j ϵ S = {1, 2, … , d} ja i≠j. Oletetaan, että

$$P(X(t+h)=j \| X(t)=i) \approx q_{ij} h.$$

Todennäköisyydet voidaan tulkita "kilpailuksi" d – 1 riippumattoman eksponenttijakautuneen satunnaismuuttujan välillä siten, että mikä ehtii tapahtua ensin. Eli jos i = 1, niin $T_{12} \sim \mathrm{Exp}(q_{12})$, … , $T_{1d} \sim \mathrm{Exp}(q_{1d})$ kilpailevat siitä, mikä on pienin eli $T_1 = \min\{T_{12}, T_{13}, \dots, T_{1d}\}$. Satunnainen aika $T_i = \min\{T_{ij} : j \neq i\}$ on tilassa i viipymisaika, ja se on $\mathrm{Exp}(q_i)$-jakautunut,

$$q_i = \sum_{j \neq i} q_{ij}.$$

Saadaan siis tilasta i siirtymätodennäköisyys aikavälillä (t, t+h]

$$P(X(t+h) \neq i \| X(t)=i) \approx q_i h$$

ja viipymistodennäköisyys tilassa i

$$P(X(t+h)=i \| X(t)=i) \approx 1-q_i h.$$

Kun laitamme siirtymäintensiteetit matriisimuotoon

$$Q = \begin{matrix} q_{11} & q_{12} & \cdots & q_{1d} \\ q_{21} & q_{22} & \cdots & q_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ q_{d1} & q_{d2} & \cdots & q_{dd} \end{matrix},$$

saadaan Markovin prosessin intensiteettimatriisi. Intensiteettimatriisissa $q_{ii} = -q_i$.

Lause 7.2   *Yleisen äärellisen Markovin prosessin jakauma toteuttaa differentiaaliyhtälön*

$$p_{ij}'(t) = \sum_{k=1}^{d} p_{ik}(t) q_{kj}$$

jokaisella i, j ϵ {1,..., d}. Matriisimuodossa P'(t) = P(t)Q.

Todistus.

$$p_{ij}(t) = P_i(X(t)=j) = \sum_{k=1}^{d} P_i(X(t-h)=k, X(t)=j) = \sum_{k=1}^{d} P_i(X(t-h)=k),$$
$$P(X(t)=j \| X(0)=i, \dots, X(t-h)=k)$$
$$= \sum_{k \neq j} P_i(X(t-h)=k) P(X(t)=j \| X(t-h)=k) + P_i(X(t-h)=k) P(X(t)=j \| X(t-h)=j)$$

$$\approx \sum_{k \neq j} p_{ik}(t-h) q_{kj} h + p_{ij}(t-h)(1-q_j h) = \sum_{k=1}^{d} p_{ik}(t-h) q_{kj} h + p_{ij}(t-h).$$

Nyt saadaan lauseke muotoon

$$\frac{p_{ij}(t)-p_{ij}(t-h)}{h}=\sum_{k=1}^{d} p_{ik}(t-h)q_{kj}, \quad \text{joten} \quad p_{ij}'(t)=\sum_{k=1}^{d} p_{ik}(t)q_{kj}.$$

Oletetaan, että P(X(0) = i) = π_i , ja että jakauma π on tasapainojakauma, eli pätee πP(t) = π. Derivoimalla yhtälöä πP(t) = π saadaan πP'(t) = 0. Sijoittamalla tähän P'(t) = P(t)Q saadaan 0 = πP'(t) = πP(t)Q = πQ, koska π on tasapainojakauma. Siis

$$\pi Q=0, \quad \text{eli} \quad \sum_{i=1}^{d} \pi_i q_{ij}=0 \quad \text{kaikilla i = 1, … ,d.}$$

Esim 7.3 Tarkastellaan laitetta, joka on kunnossa keskimäärin 4kk, ja kunnossa ollessaan tuottaa 200e/vrk. Kun laite vikaantuu, paikalle kutsutaan korjaaja. Korjaajan saapuminen kestää keskimäärin 4 vuorokautta ja korjaus 10 vuorokautta. Korjaus maksaa 250e/vrk. Lopuksi on testaus, joka kestää keskimäärin 2 vuorokautta. Oletetaan, että kunnossaoloaika, korjaajan saapumiseen kuluva aika, korjaus sekä testaus ovat kaikki eksponentiaalisesti jakautuneita ja myös riippumattomia. Paljonko laite tuottaa pitkällä aikavälillä per vuorokausi?

Ratkaisu. Merkitään tiloja seuraavasti:

tila A = laite on kunnossa, A ~ Exp(λ_A);

tila B = odotetaan korjaajaa, B ~ Exp(λ_B);

tila C = korjaus, C ~ Exp(λ_C);

tila D = testaus, D ~ Exp(λ_D). Näistä saadaan odotusarvot

E(A) = 4 * 30 = 120 = 1/ λ_A ⇔ λ_A = 1/120,

E(B) = 4 = 1/ λ_B ⇔ λ_B = ¼,

E(C) = 10 = 1/ λ_C ⇔ λ_C = 1/10,

E(D) = 2 = 1/ λ_D ⇔ λ_D = ½,

q_{AB} = λ_A = 1/120 , q_{BC} = λ_B = ¼ , q_{CD} = λ_C = 1/10 , q_{DA} = λ_D = ½ .

Olkoon π = (π_A π_B π_C π_D) tämän äärellisen Markovin prosessin tasapainojakauma ja

$$Q = \begin{matrix} q_{AA} & q_{AB} & q_{AC} & q_{AD} \\ q_{BA} & q_{BB} & q_{BC} & q_{BD} \\ q_{CA} & q_{CB} & q_{CC} & q_{CD} \\ q_{DA} & q_{DB} & q_{DC} & q_{DD} \end{matrix}$$

intensiteettimatriisi. Nyt πQ = 0, eli

$$\pi_A \quad \pi_B \quad \pi_C \quad \pi_D \quad * \quad \begin{matrix} -(\frac{1}{120}) & \frac{1}{120} & 0 & 0 \\ 0 & -(\frac{1}{4}) & \frac{1}{4} & 0 \\ 0 & 0 & -(\frac{1}{10}) & \frac{1}{10} \\ \frac{1}{2} & 0 & 0 & -(\frac{1}{2}) \end{matrix} \quad = \quad 0 \ .$$

Kun ratkaistaan matriisitulosta saatavat yhtälöparit, saadaan arvot $\pi_A = 60/68 = 15/17$, $\pi_B = 2/68 = 1/34$, $\pi_C = 5/68$, $\pi_D = 1/68$. Kysytty tuotto on lopulta

$$(\frac{15}{17}) * 200 - (\frac{5}{68}) * 250 = \frac{91375}{578} \approx 158,08824 \approx 158,09\, e .$$

Luku 8

Viitteet

http://www.mytoller.net/optigen_pra.html

http://wiki.helsinki.fi/pages/viewpage.action?pageId=31432388 (Stokastiset prosessit, kesä 2008)

Karlin & Taylor : A First Course in Stochastic Processes