MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

1B Satunnaismuuttujat ja todennäköisyysjakaumat

Pekka Pere

matematiikan ja systeemianalyysin laitos
perustieteiden korkeakoulu
Aalto-yliopisto

lukuvuosi 2023–2024
periodi I

Sisältö

Satunnaismuuttuja

Satunnaismuuttuja (lyh. sm) on suure, jonka arvo määräytyy satunnaisilmiön toteumasta:

- Sattuma määrää satunnaisilmiön toteuman $s \in S$
- Toteuma s määrää satunnaismuuttujan arvon $X(s)$
- Tapahtuma $\{X = a\} \ := \ \{s \in S \ : \ X(s) = a\}$
- Sm:n arvoja vastaavat tapahtumat ovat perusjoukon ositus

Esim (Kaksi nopanheittoa)

Perusjoukossa $S = \{(s_1, s_2) : s_1, s_2 = 1, \ldots, 6\}$

- Silmälukujen summa $N(s) = s_1 + s_2$ on satunnaismuuttuja
- Niiden maksimi $M(s) = \max(s_1, s_2)$ on satunnaismuuttuja
- Myös 1. silmäluku $X_1(s) = s_1$ on satunnaismuuttuja!

Usein sm:n arvoa $X(s)$ merkitään lyhyesti X.

Satunnaismuuttuja: Teoria ja käytäntö

Huomaa kaksi abstraktion askelta:

- Perusjoukko S ja todennäköisyysfunktio $\mathbb{P}$ kuvaavat kaiken *satunnaisuuden* satunnaisilmiössä.
- Kun (satunnainen) toteuma $s \in S$ toteutuu, niin se määrää kaikkien satunnaismuuttujien arvot.

Yhdessä satunnaisilmiössä (esim. "jaetaan yksi kortti", "jaetaan viisi korttia") voimme määritellä mielivaltaisen paljon satunnaismuuttujia (funktioita toteumasta).

Matemaattisesti satunnaismuuttuja on (deterministinen) *kuvaus* eli *funktio* toteumasta s esim. reaaliluvuksi $X(s)$.

Käytännössä satunnaismuuttujan arvoa $X(s)$ merkitään lyhyesti X, ja tapahtumaa, että X saa esim. arvon 5, merkitään $\{X = 5\}$ tai ilman sulkuja $X = 5$.

Eri tyyppisiä satunnaismuuttujia

Usein satunnaismuuttujan arvot $X(s)$ ovat reaalilukuja, mutta ne voivat olla jotain muutakin.

Nimitys	**Maalijoukko**	Esim.
Satunnaisluku	$\mathbb{R}$	
Satunnaisvektori	$\mathbb{R}^n$	(X_1, X_2, X_3) kolmesta nopanheitosta; tai noppatulosten (Minimi, Maksimi)
Satunnaismatriisi	$\mathbb{R}^{m \times n}$	
Satunnaismerkkijono	A^n	DNA-sekvenssi, $A = \{\text{A,C,T,G}\}$
Stokastinen prosessi	$\mathbb{R}^T$	aikavälin T funktiot
Satunnaiskenttä	$\mathbb{R}^U$	alueen U funktiot
Satunnaisverkko	$\{0, 1\}^{V \times V}$	solmujoukon V verkot

Tällä kurssilla käsitellään lähes yksinomaan satunnaislukuja (eli reaaliarvoisia satunnaismuuttujia) ja $\mathbb{R}^2$:n satunnaisvektoreita.

Satunnaismuuttuja on diskreetti jos sen arvojoukko on äärellinen kuten $\{1, 2, 3, 4\}$ tai numeroituvasti ääretön kuten $\mathbb{N}$.

Sisältö

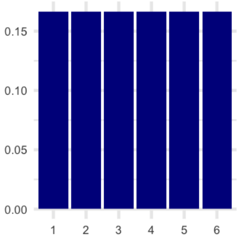
Satunnaismuuttujan jakauma

Satunnaismuuttujan X jakauma on taulukko tai funktio, josta voidaan määrittää X:n mahdolliset arvot ja niiden todennäköisyydet.

Esim (Kaksi nopanheittoa)

Nopan 1 silmäluvun X_1 jakauma on

k	1	2	3	4	5	6
$\mathbb{P}(X_1 = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$



eli lukujoukon $\{1, \dots, 6\}$ tasajakauma.

Nopan 2 silmäluvulla X_2 on sama jakauma.

$\implies$ Satunnaismuuttujat X_1 ja X_2 ovat samoin jakautuneita. Ne ovat kuitenkin eri satunnaismuuttujat. (Kun heität noppaa, voi hyvin sattua, että $X_1 \neq X_2$).

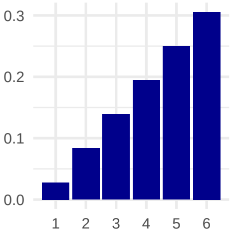
Esim. Kahden nopan maksimi

$M = \max(X_1, X_2)$, missä X_1 ja X_2 ovat kahden nopan tulokset.

$$
\begin{aligned}
\mathbb{P}(M = k) &= \mathbb{P}(M \leq k) - \mathbb{P}(M \leq k-1) \\
&= \mathbb{P}(X_1 \leq k \textbf{ ja } X_2 \leq k) - \mathbb{P}(X_1 \leq k-1 \textbf{ ja } X_2 \leq k-1) \\
&= \mathbb{P}(X_1 \leq k) \cdot \mathbb{P}(X_2 \leq k) - \mathbb{P}(X_1 \leq k-1) \cdot \mathbb{P}(X_2 \leq k-1) \\
&= \left(\frac{k}{6}\right)^2 - \left(\frac{k-1}{6}\right)^2 \\
&= \frac{2k-1}{36}
\end{aligned}
$$

Satunnaismuuttujan M jakauma:

k	1	2	3	4	5	6
$\mathbb{P}(M = k)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$



Uusia satunnaismuuttujia vanhoista

Yhdestä tai monesta satunnaismuuttujasta voi luonnollisesti laskea jollakin funktiolla uuden arvon, joka on taas satunnaismuuttuja.

Esim (Tunnusluvut)

Nopan tulokset X_1, X_2, X_3, niistä suurin $Y = \max(X_1, X_2, X_3)$, pienin $A = \min(X_1, X_2, X_3)$. Määritellään uusi sm $L = Y - A$, joka kuvaa, miten leveälle välille tulokset ovat asettuneet.

Jos tulokset ovat $(4, 2, 5)$, niin $Y = 5$, $A = 2$, ja edelleen $L = 5 - 2 = 3$.

- L on myös satunnaismuuttuja ja sillä on jokin jakauma, ts. eri arvojen todennäköisyydet.
- Tällaisia toteutunutta lukujonoa (dataa) kuvailevia tunnuslukuja käytetään paljon tilastotieteessä. Myös lukujonon keskiarvo on tunnusluku.

Uusia vanhoista: Yhden sm:n muunnos

Esim (Satunnaisen kokoinen neliö)

Kone tekee neliölaattoja, joiden sivu X määräytyy nopanheitolla.

k	1	2	3	4	5	6
$\mathbb{P}(X = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Laatan pinta-ala $A = X^2$ on myös sm. Mikä on sen jakauma? Otettava selville (1) mitä arvoja X^2 voi yleensäkään saada ja (2) millä tn:llä.

k	?	?	?	?	?	?
$\mathbb{P}(A = k)$	?	?	?	?	?	?

Uusia vanhoista: Yhden sm:n muunnos

Esim (Satunnaisen kokoinen neliö)

Kone tekee neliölaattoja, joiden sivu X määräytyy nopanheitolla.

k	1	2	3	4	5	6
$\mathbb{P}(X = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Laatan pinta-ala $A = X^2$ on myös sm. Mikä on sen jakauma? Otettava selville (1) mitä arvoja X^2 voi yleensäkään saada ja (2) millä tn:llä.

k	1	4	9	16	25	36
$\mathbb{P}(A = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

(Muunnoksista tulee lisää luennossa 2A.)

Esim. Metron odotusaika (reaaliluku)

X = seuraavan metron odotusaika (min) asemalla, jonne metroja saapuu 10 min välein. Mikä on satunnaismuuttujan X jakauma?

- $\mathbb{P}(2 \le X \le 3) \;=\; \frac{1}{10} \;=\; 0.1$
- $\mathbb{P}(2.9 \le X \le 3) \;=\; \frac{0.1}{10} \;=\; 0.01$
- $\mathbb{P}(2.999999 \le X \le 3) \;=\; \frac{0.000001}{10} \;=\; 0.0000001$
- $\mathbb{P}(X = 3) \;=\; 0$

Vastaavasti päätellen havaitaan, että $\mathbb{P}(X = t) \;=\; 0$ kaikilla t.

Menikö yllä olevassa päättelyssä jotain väärin?

Ei mennyt. Koska X:n arvojoukko on *jatkuva* väli $[0, 5]$, tarkoittaa $\{X = 3\}$ tapahtumaa, että X:n arvo on 3 äärettömän monen desimaalin tarkkuudella. Tällaisen tapahtuman todennäköisyys on nolla.

Tarvitaan vaihtoehtoinen tapa esittää odotusajan jakauma.

Esim. Metron odotusaika

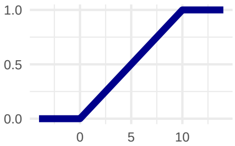
X = seuraavan metron odotusaika (min) asemalla, jonne metroja saapuu 10 min välein. Mikä on satunnaismuuttujan X jakauma?

Yksittäisten arvojen todennäköisyyksistä $\mathbb{P}(X = t) = 0$ ei ole hyötyä laskennassa, joten keskitymme välien todennäköisyyksiin.

$$\begin{aligned} \mathbb{P}(a \le X \le b) &= \mathbb{P}(a < X \le b) \\ &= \mathbb{P}(X \le b) - \mathbb{P}(X \le a) \\ &= F_X(b) - F_X(a), \end{aligned}$$

missä

$$F_X(t) = \begin{cases} 0, & t \le 0, \\ \frac{t}{10}, & 0 < t < 10, \\ 1, & t \ge 10. \end{cases}$$



on odotusajan jakauman kertymäfunktio.

Kertymäfunktio

Satunnaisluvun (eli reaaliarvoisen satunnaismuuttujan) jakauman **kertymäfunktio** on funktio $F_X(t) = \mathbb{P}(X \le t)$.

Fakta

Kertymäfunktio määrää jakauman: funktion $F_X(t)$ avulla voidaan laskea **kaikkien** *tapahtumien $\{X \in B\}$ todennäköisyydet.*

Esim (Metron odotusaika)

Millä todennäköisyydellä odotusaika osuu välille $(1, 2)$ tai $(3, 4)$?

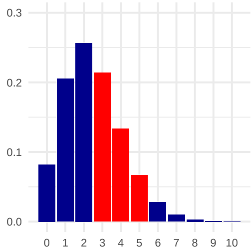
$$\begin{aligned}
\mathbb{P}(X \in (1,2) \text{ tai } X \in (3,4)) &= \mathbb{P}(X \in (1,2)) + \mathbb{P}(X \in (3,4)) \\
&= (F_X(2) - F_X(1)) + (F_X(4) - F_X(3)) \\
&= \left(\frac{2}{10} - \frac{1}{10}\right) + \left(\frac{4}{10} - \frac{3}{10}\right) \\
&= 0.2.
\end{aligned}$$

Sisältö

Tiheysfunktio

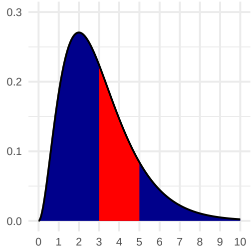
X on diskreetti, jos sen jakauma voidaan esittää funktion $f_X(x) \geq 0$ avulla muodossa

$$\mathbb{P}(X \in A) \;=\; \sum_{x \in A} f_X(x).$$



X on jatkuva, jos sen jakauma voidaan esittää funktion $f_X(x) \geq 0$ avulla muodossa

$$\mathbb{P}(X \in A) \;=\; \int_A f_X(x)\, dx.$$



Molemmissa tapauksissa funktiota f_X voidaan kutsua X:n tiheysfunktioksi (tf). Alaindeksi X voidaan jättää pois jos siitä ei tule sekaannusta.

Diskreetin jakauman tiheysfunktio

Diskreetin satunnaismuuttujan tiheysfunktio voidaan kirjoittaa muodossa $f_X(x) \;=\; \mathbb{P}(X = x)$ ja se toteuttaa ehdot

$$f_X(x) \geq 0 \qquad \text{ja} \qquad \sum_x f_X(x) \;=\; 1.$$

Vastaavasti mikä tahansa yo. ehdot toteuttava funktio on jonkin diskreetin jakauman tiheysfunktio.

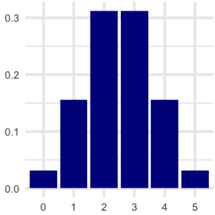
Diskreetin satunnaismuuttujan tiheysfunktion arvoja kutsutaan myös pistetodennäköisyyksiksi (= todennäköisyys, että sm:n arvo osuu tiettyyn pisteeseen).

Diskreetin jakauman tiheysfunktio

Kun satunnaismuuttujan arvojoukko on pieni, jakauman voi esittää taulukkona, jossa on sm:n kaikki mahdolliset arvot ja niiden tn:t.

Esim (Kruunien lukumäärä 5:llä kolikonheitolla)

k	0	1	2	3	4	5
$\mathbb{P}(X=k)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$



Suuren arvojoukon tapauksessa jakauma kannattaa esittää tiheysfunktion avulla jonkinlaisena lausekkeena.

Esim (Kruunien lukumäärä $n =$5 000 000:lla kolikonheitolla)

$$f_X(k) \;=\; \binom{n}{k}\left(\frac{1}{2}\right)^k\left(1-\frac{1}{2}\right)^{n-k}, \quad k=0,1,\ldots,n.$$

Tämä on binomijakauma parametreilla $n = 5\ 000\ 000$ ja $p = \frac{1}{2}$.

Jatkuvan jakauman tiheysfunktio

Jatkuvan jakauman tiheysfunktio toteuttaa ehdot

$$f_X(x) \geq 0 \qquad \text{ja} \qquad \int_{-\infty}^{\infty} f_X(x)\, dx \;=\; 1,$$

Vastaavasti mikä tahansa yo. ehdot toteuttava funktio on jonkin jatkuvan jakauman tiheysfunktio.

Jatkuvan jakauman tiheysfunktiota *ei* voi kirjoittaa pistetodennäköisyyksien avulla, sillä $\mathbb{P}(X = x) \;=\; 0$.

Jatkuva tiheysfunktio ilmaisee, millä *tiheydellä* todennäköisyyttä ("massaa") on pisteen x ympäristössä. Jos f_X on jatkuva x:n ympärillä ja $h > 0$ on pieni, niin

$$\mathbb{P}(X = x \pm h/2) \approx f_X(x) \cdot h$$

(Huom. Tiheysfunktion arvo voi olla miten suuri tahansa, kunhan se on sitä vain lyhyellä välillä.)

Kertymäfunktio ja tiheysfunktio

Jatkuvan satunnaisluvun

- kertymäfunktio saadaan tiheysfunktion integraalina

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)\,dt$$

- tiheysfunktio saadaan kertymäfunktion derivaattana

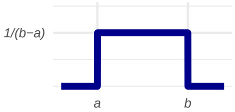
$$f_X(x) = F'_X(x)$$

kertymäfunktion derivoituvuuspisteissä.

Esim. Jatkuva tasajakauma

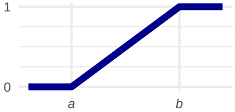
Funktio

$$f(t) = \begin{cases} \frac{1}{b-a}, & a < t < b, \\ 0, & \text{muuten.} \end{cases}$$



on erään jatkuvan jakauman tiheysfunktio: lukuvälin $[a, b]$ jatkuva tasajakauma.
Kertymäfunktio saadaan integraalina

$$F(t) = \int_{-\infty}^{t} f(s)\,ds = \begin{cases} 0, & t < a, \\ \frac{t-a}{b-a}, & a \le t \le b, \\ 1, & t > b. \end{cases}$$



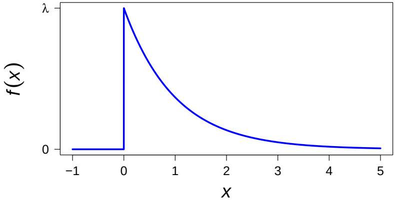
Sijoittamalla $a = 0$ ja $b = 10$ saadaan metron odotusajan jakauma.

Eksponenttijakauma

Tyypillinen käyttö: Ilmiö, jota tapahtuu toistuvasti satunnaisin väliajoin, keskimäärin λ kertaa aikayksikössä. Odotusaika on eksponenttijakautunut. (Esim. hyönteisiä tuulilasiin tai alkeishiukkasten hajoamisia.)

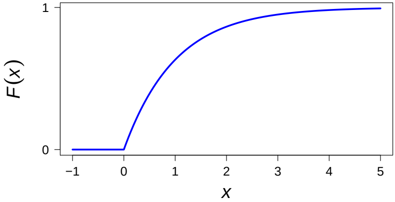
Eksponenttijakauman parametrilla $\lambda > 0$ tiheysfunktio on

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0. \end{cases}$$



Integroimalla tiheysfunktiota $\Longrightarrow$ kertymäfunktio on

$$F(x) = \begin{cases} 0, & x \leq 0, \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$



Eksponenttijakauman muistittomuus

$F(t) = 1 - e^{-\lambda t},\ t \geq 0$

$$\begin{aligned}
\mathbb{P}\big(X > s + t \,|\, X > s\big) &= \frac{\mathbb{P}(X > s + t \ \text{ ja } \ X > s)}{\mathbb{P}(X > s)} \\
&= \frac{\mathbb{P}(X > s + t)}{\mathbb{P}(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(X > t)
\end{aligned}$$

Siis $\mathbb{P}(X > s + t \,|\, X > s) = \mathbb{P}(X > t)$ kaikilla $s, t \geq 0$.

Tulkinta

Riippumatta siitä onko edellisen hyönteisen jälkeen ajettu 1 vai 5 minuuttia, on odotusaika seuraavaan hyönteiseen edelleen samalla tavoin eksponenttijakautunut.

Satunnaisluvut — yhteenveto

Diskreetti jakauma	Jatkuva jakauma
X:n arvot sisältyvät äärelliseen tai numeroituvasti äärettömään arvojoukkoon	X:n arvot sisältyvät ylinumeroituvasti äärettömään lukujoukkoon
$\mathbb{P}(X = x) = f_X(x)$	$\mathbb{P}(X = x) = 0$ kaikilla x
Jakauma määräytyy tiheysfunktiosta kaavalla $$\mathbb{P}(X \in A) = \sum_{x \in A} f_X(x)$$	Jakauma määräytyy tiheysfunktiosta kaavalla $$\mathbb{P}(X \in A) = \int_A f_X(x)\, dx$$
Tiheysfunktion arvot ovat tarkkoja todennäköisyyksiä $$f_X(x) = \mathbb{P}(X = x)$$	Tiheysfunktion arvot ovat suhteellisia likiarvoisia todennäköisyyksiä $$f_X(x) \approx h^{-1}\mathbb{P}(X = x \pm h/2)$$

Sisältö

Satunnaismuuttujien yhteisjakauma 1 (Epäreilut nopat)

Samaan satunnaisilmiöön liittyvien satunnaismuuttujien X ja Y yhteisjakauma on taulukko tai funktio, josta voidaan määrittää parin (X, Y) mahdolliset arvot ja niiden todennäköisyydet.

Esim (Epäreilun nopan kaksi heittoa)

Heitto X_1 on 2 tn:llä 0.5 ja muut arvot 0.1 kukin.
Heitto X_2 on 2 tn:llä 0.5 ja muut arvot 0.1 kukin.
Silmälukujen X_1 ja X_2 yhteisjakauma on

	X_2					
X_1	1	2	3	4	5	6
1	0.01	0.05	0.01	0.01	0.01	0.01
2	0.05	0.25	0.05	0.05	0.05	0.05
3	0.01	0.05	0.01	0.01	0.01	0.01
4	0.01	0.05	0.01	0.01	0.01	0.01
5	0.01	0.05	0.01	0.01	0.01	0.01
6	0.01	0.05	0.01	0.01	0.01	0.01

Indikaattorifunktio

Joukon A indikaattorifunktio määritellään kaavalla

$$1_A(x) = \begin{cases} 1, & x \in A, \\ 0, & \text{muuten.} \end{cases}$$

Sen avulla voidaan yhden muuttujan jakaumien esityskaavat kirjoittaa muodossa

$$\mathbb{P}(X \in A) = \sum_{x \in S_X} 1_A(x) f_X(x)$$

ja

$$\mathbb{P}(X \in A) = \int_{-\infty}^{\infty} 1_A(x) f_X(x)\, dx.$$

Diskreetti ja jatkuva yhteisjakauma

Satunnaismuuttujilla X ja Y on diskreetti yhteisjakauma, jos niiden todennäköisyydet voidaan esittää funktion $f_{X,Y}(x,y) \geq 0$ avulla muodossa

$$\mathbb{P}((X,Y) \in A) \; = \; \sum_{x \in S_X} \sum_{y \in S_Y} 1_A(x,y) f_{X,Y}(x,y),$$

missä joukot S_X ja S_Y ovat numeroituvia, ja jatkuva yhteisjakauma, jos niiden todennäköisyydet voidaan esittää funktion $f_{X,Y}(x,y) \geq 0$ avulla muodossa

$$\mathbb{P}((X,Y) \in A) \; = \; \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 1_A(x,y) f_{X,Y}(x,y) \, dx \, dy.$$

Reunajakaumat 1 (Epäreilut nopat)

Yhteisjakauman rivi- ja sarakesummia kutsutaan reunajakaumiksi.

	X_2						
X_1	1	2	3	4	5	6	Sum
1	0.01	0.05	0.01	0.01	0.01	0.01	0.10
2	0.05	0.25	0.05	0.05	0.05	0.05	0.50
3	0.01	0.05	0.01	0.01	0.01	0.01	0.10
4	0.01	0.05	0.01	0.01	0.01	0.01	0.10
5	0.01	0.05	0.01	0.01	0.01	0.01	0.10
6	0.01	0.05	0.01	0.01	0.01	0.01	0.10
Sum	0.10	0.50	0.10	0.10	0.10	0.10	

Rivisummista saadaan X_1:n jakauma
Sarakesummista saadaan X_2:n jakauma

Reunajakaumat 2 (Asunnot)

Yhteisjakauman rivi- ja sarakesummia kutsutaan reunajakaumiksi.

	A						
H	1	2	3	4	5	6	Sum
1	0.1263	0.0129	0.0019	0.0009	0.0003	0.0001	0.1424
2	0.1961	0.0857	0.0124	0.0048	0.0014	0.0004	0.3010
3	0.0727	0.0966	0.0336	0.0190	0.0050	0.0013	0.2282
4	0.0382	0.0793	0.0306	0.0298	0.0100	0.0028	0.1908
5	0.0154	0.0414	0.0169	0.0206	0.0087	0.0023	0.1053
6	0.0042	0.0117	0.0055	0.0065	0.0034	0.0010	0.0324
Sum	0.4530	0.3277	0.1010	0.0816	0.0289	0.0078	

Rivisummista saadaan H:n jakauma
Sarakesummista saadaan A:n jakauma

Reunajakaumat

Diskreettiä yhteisjakaumaa noudattavien satunnaismuuttujien X ja Y tiheysfunktiot saadaan yhteisjakauman tiheysfunktiosta kaavoilla

$$\begin{aligned} f_X(x) &= \sum_{y \in S_Y} f_{X,Y}(x,y) \\ f_Y(y) &= \sum_{x \in S_X} f_{X,Y}(x,y). \end{aligned}$$

Jatkuvan yhteisjakauman tapauksessa

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dx. \end{aligned}$$

Esim. Yksikköneliön tasajakauma

Yksikköneliön $(0,1)^2$ tasajakaumaa noudattavan satunnaisvektorin (U_1, U_2) tiheysfunktio on

$$f_{U_1,U_2}(x_1,x_2) = \begin{cases} 1, & \text{kun } x_1 \in (0,1) \text{ ja } x_2 \in (0,1), \\ 0, & \text{muuten.} \end{cases}$$

Yhteisjakauman reunatiheysfunktiot ovat

$$f_{U_1}(x_1) = \int_{-\infty}^{\infty} f_{U_1,U_2}(x_1,x_2)\,dx_2 = \begin{cases} 1, & \text{kun } x_1 \in (0,1), \\ 0, & \text{muuten.} \end{cases}$$

$$f_{U_2}(x_2) = \int_{-\infty}^{\infty} f_{U_1,U_2}(x_1,x_2)\,dx_1 = \begin{cases} 1, & \text{kun } x_2 \in (0,1), \\ 0, & \text{muuten.} \end{cases}$$

U_1 ja U_2 noudattavat siis välin $(0,1)$ tasajakaumaa.

Sisältö

Ehdollinen jakauma (Asunnot)

Tarkastellaan asuntojen yhteisjakauman 1. riviä eli tapausta $H = 1$ (yksi huone; jaettu tapauksiin asukasmäärän mukaan).

	A						
H	1	2	3	4	5	6	Sum
1	0.1263	0.0129	0.0019	0.0009	0.0003	0.0001	0.1424

Rivin summa on $0.1424 \neq 1$, eli tämä rivi ei voi olla tn-jakauma. Rivillä on erilaisten yksiöiden osuuksia *koko asuntokannasta*.

Jaetaan rivisummalla $\longrightarrow$ saadaan erilaisten yksiöiden osuudet *yksiöistä*.

	A						
H	1	2	3	4	5	6	Sum
1	0.8870	0.0909	0.0131	0.0062	0.0021	0.0006	1.0000

Tämä on eräs tn-jakauma, nimittäin A:n ehdollinen jakauma kun $H = 1$.

Ehdollinen jakauma (Epäreilut nopat)

Kullakin rivillä (X_1:n arvo kiinnitetty) saadaan X_2:n *ehdollinen* jakauma jakamalla yhteisjakauman tn:t rivisummalla: esim.

$$P(X_1 = 1 \,|\, X_2 = 1) = P(X_1 = 1 \textbf{ ja } X_2 = 1)/P(X_2 = 1) = 0.01/0.1 = 0.1$$

Tässä X_1:n ehdollinen jakauma, kun $X_2 = 1$:

	X_2						
X_1	1	2	3	4	5	6	Sum
1	0.1	0.5	0.1	0.1	0.1	0.1	1.0

Toisaalta jokaisella muulla X_1:n arvolla saadaan X_2:lle sama ehdollinen jakauma (TAULU). Toisin sanoen X_1:n jakauma *ei riipu* X_2:sta.

Ehdolliset jakaumat

Y:n ehdollinen tiheysfunktio X:n suhteen määritellään kaavalla

$$f_{Y|X}(y|x) \;=\; \frac{f_{X,Y}(x,y)}{f_X(x)}.$$

Diskreetissä tapauksessa $f_{Y|X}(y|x) \geq 0$ ja $\sum_{y \in S_Y} f_{Y|X}(y|x) \;=\; 1$,

Jatkuvassa tapauksessa $f_{Y|X}(y|x) \geq 0$ ja $\int_{-\infty}^{\infty} f_{Y|X}(y|x)\, dy \;=\; 1$.
$\implies y \mapsto f_{Y|X}(y|x)$ on yhden muuttujan jakauman tiheysfunktio.

Tulkinta diskreetille:

$$f_{Y|X}(y|x) \;=\; \mathbb{P}(Y = y \,|\, X = x).$$

Tulkinta jatkuvalle: yhteisjakauman tiheysfunktion jatkuvuuspisteissä pienillä $h > 0$ arvoilla pätee

$$f_{Y|X}(y|x) \;\approx\; \frac{\mathbb{P}(Y = y \pm h/2 \,|\, X = x \pm h/2)}{h}.$$

Satunnaismuuttujien riippuvuus ja riippumattomuus

Satunnaismuuttujat X ja Y ovat stokastisesti riippumattomat, jos kaikilla A, B pätee

$$\mathbb{P}(X \in A, Y \in B) \;=\; \mathbb{P}(X \in A)\mathbb{P}(Y \in B).$$

Yhtäpitävästi:

$$\mathbb{P}(Y \in B \,|\, X \in A) \;=\; \mathbb{P}(Y \in B)$$

tai

$$\mathbb{P}(X \in A \,|\, Y \in B) \;=\; \mathbb{P}(X \in A).$$

Tapahtuma $X \in A$ ei sisällä mitään informaatiota, josta olisi hyötyä Y:n arvon ennustamiseen.

Jos tapahtumat eivät ole riippumattomat, niin ne ovat (stokastisesti) riippuvat. Silloin Y:n ehdollinen jakauma on erilainen eri X:n arvoilla.

Satunnaismuuttujien riippuvuus ja riippumattomuus

Fakta

Diskreettiä tai jatkuvaa yhteisjakaumaa noudattavat satunnaismuuttujat X ja Y ovat riippumattomat jos ja vain niiden yhteisjakauman tiheysfunktio voidaan esittää muodossa

$$f_{X,Y}(x,y) \ = \ f_X(x)f_Y(y)$$

Tämä vastaa ehtoa

$$f_{Y|X}(y|x) \ = \ f_Y(y),$$

eli Y:n ehdollinen jakauma X:n suhteen on sama kuin Y:n jakauma sellaisenaan.

Esim. Satunnaisotanta

Kuinka moni opiskelijoista katsoi viime to *Salatut elämät*?

- $S =$ "Kaikki opiskelijat", $\#S = 80$
- $A =$ "Salkkarit katsoneet opiskelijat", $\#A = 3$.

($\#A$ olisi käytännön tilanteessa tuntematon)

Haastatellaan satunnaiset $n = 2$ opiskelijaa ja merkitään

$$X_1 = \begin{cases} 1, & \text{jos 1. haastateltu opiskelija} \in A \\ 0, & \text{muuten} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{jos 2. haastateltu opiskelija} \in A \\ 0, & \text{muuten} \end{cases}$$

Mikä on X_1:n ja X_2:n yhteisjakauma?
$\mathbb{P}(X_1 = 1, X_2 = 1) \; = ?$

Satunnaisotanta palauttaen ja palauttamatta

- Palauttaen = toinen opiskelija valitaan taas koko joukosta
- Palauttamatta = toinen opiskeilja valitaan jäljelläolevista

Palauttaen

	X_2		
X_1	0	1	Yht
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

Palauttamatta

	X_2		
X_1	0	1	Yht
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

Molemmissa tapauksissa saadaan samat reunajakaumat.

Yhteisjakaumaan vaikuttaa, miten otanta suoritetaan.

Satunnaisotanta palauttaen ja palauttamatta

Palauttaen

X_1	$X_2 = 0$	$X_2 = 1$	Yht
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_1,X_2}(i,j) = f_{X_1}(i) f_{X_2}(j)$$

Palauttamatta

X_1	$X_2 = 0$	$X_2 = 1$	Yht
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_1,X_2}(i,j) \neq f_{X_1}(i) f_{X_2}(j)$$

Molemmissa tapauksissa saadaan samat reunajakaumat.

Satunnaisotannassa palauttaen ovat X_1 ja X_2 riippumattomat.
Satunnaisotannassa palauttamatta X_1 ja X_2 ovat riippuvat.

Esim. Satunnaisotanta palauttaen

Mikä on satunnaismuuttujan X_2 ehdollinen jakauma tapahtuman $\{X_1 = 0\}$ sattuessa?

	X_2		
X_1	0	1	Yht
0	$\frac{77}{80} \times \frac{77}{80}$	$\frac{77}{80} \times \frac{3}{80}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{80}$	$\frac{3}{80} \times \frac{3}{80}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_2|X_1}(0|0) = \frac{\frac{77}{80} \times \frac{77}{80}}{\frac{77}{80}} = \frac{77}{80}.$$

$$f_{X_2|X_1}(1|0) = \frac{\frac{77}{80} \times \frac{3}{80}}{\frac{77}{80}} = \frac{3}{80}.$$

Tässä tapauksessa X_2:n ehdollinen jakauma tapahtuman $\{X_1 = 0\}$ sattuessa on sama kuin X_2:n ehdollistamaton jakauma.

Esim. Satunnaisotanta palauttamatta

Mikä on satunnaismuuttujan X_2 ehdollinen jakauma tapahtuman $\{X_1 = 0\}$ sattuessa?

	X_2		
X_1	0	1	Yht
0	$\frac{77}{80} \times \frac{76}{79}$	$\frac{77}{80} \times \frac{3}{79}$	$\frac{77}{80}$
1	$\frac{3}{80} \times \frac{77}{79}$	$\frac{3}{80} \times \frac{2}{79}$	$\frac{3}{80}$
Yht	$\frac{77}{80}$	$\frac{3}{80}$	

$$f_{X_2|X_1}(0|0) = \frac{\frac{77}{80} \times \frac{76}{79}}{\frac{77}{80}} = \frac{76}{79}.$$

$$f_{X_2|X_1}(1|0) = \frac{\frac{77}{80} \times \frac{3}{79}}{\frac{77}{80}} = \frac{3}{79}.$$

Tässä tapauksessa X_2:n ehdollinen jakauma tapahtuman $\{X_1 = 0\}$ sattuessa on eri kuin X_2:n ehdollistamaton jakauma.

Sisältö

Ääretön diskreetti arvojoukko

Diskreetin sm:n mahdollisten arvojen joukko voi olla ääretön.

Esim (Kimblen alkuvaihe)

N = nopanheittojen lukumäärä, kunnes saadaan kuutonen.

$$\begin{aligned}
\mathbb{P}(N = k) &= \mathbb{P}(X_1 \neq 6, \ldots, X_{k-1} \neq 6, X_k = 6) \\
&= \mathbb{P}(X_1 \neq 6) \cdots \mathbb{P}(X_{k-1} \neq 6)\mathbb{P}(X_k = 6) \\
&= \left(1 - \frac{1}{6}\right)^{k-1} \left(\frac{1}{6}\right)
\end{aligned}$$

Satunnaismuuttuja N noudattaa äärettömän arvojoukon $\{1, 2, \ldots\}$ geometrista jakaumaa onnistumistodennäköisyytenä $p = \frac{1}{6}$. Tämä on numeroituvasti äärettömän joukon diskreetti jakauma, tiheysfunktiona

$$f_N(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots$$

Esimerkki: Metron odotusaika

Y = odotusaika (min) asemalla, jonne metroja saapuu 10 min välein, ja jossa metrot pysähtyvät 1 min ajan. Y:n jakauma = ?

X = aika (min) edellisen metron saapumisesta noudattaa välin $[0, 10]$ tasajakaumaa. Kun $t \in [0, 9]$,

$$\begin{aligned}\mathbb{P}(Y \leq t) &= \mathbb{P}(Y = 0) + \mathbb{P}(0 < Y \leq t) \\ &= \mathbb{P}(X \leq 1) + \mathbb{P}(0 < 10 - X < t).\end{aligned}$$

$$\implies \quad F_Y(t) = \begin{cases} 0, & t < 0, \\ \frac{1}{10} + \frac{t}{10}, & 0 \leq t \leq 9, \\ 1, & t > 9. \end{cases}$$

Onko Y:n jakauma diskreetti vai jatkuva?

- Y saa arvoja jatkuvalla välillä $[0, 9] \implies$ ei diskreetti
- $\mathbb{P}(Y = 0) > 0 \implies$ ei jatkuva

Esimerkki: Metron odotusaika

Y = odotusaika (min) asemalla, jonne metroja saapuu 10 min välein, ja jossa metrot pysähtyvät 1 min ajan.

Kertymäfunktio voidaan kirjoittaa muodossa

$$F_Y(t) \; = \; \frac{1}{10} F_{Y_0}(t) + \frac{9}{10} F_{Y_1}(t),$$

missä

$$F_{Y_0}(t) \; = \; \begin{cases} 0, & t < 0, \\ 1, & t \geq 0, \end{cases} \qquad F_{Y_1}(t) \; = \; \begin{cases} 0, & t < 0, \\ \frac{t}{9}, & 0 \leq t < 9, \\ 1, & t \geq 9, \end{cases}$$

Y:n jakauma on *diskreetin ja jatkuvan jakauman sekoitus*:

- Y_0 on diskreetti sm, joka varmuudella saa arvon 0 (Y:n jakauma ehdolla, että metro on odottamassa asemalla)
- Y_1 on jatkuva sm, joka noudattaa välin $[0, 9]$ tasajakaumaa (Y:n jakauma ehdolla, että metroa joudutaan odottamaan)

Seuraavalla kerralla puhutaan satunnaismuuttujien odotusarvoista. . .