MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

1B Satunnaismuuttujat ja todennäköisyysjakaumat

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

Sisältö

Satunnaismuuttuja

Satunnaismuuttuja on suure, jonka arvo määräytyy satunnaisilmiön toteumasta:

- Sattuma määrää satunnaisilmiön toteuman $s \in S$
- Toteuma s määrää satunnaismuuttujan arvon $X(s)$
- Tapahtuma $\{X = a\} := \{s \in S : X(s) = a\}$

Esim (Kaksi nopanheittoa)

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

- Silmälukujen maksimi $M(s) = \max\{s_1, s_2\}$
- Silmälukujen summa $N(s) = s_1 + s_2$

Satunnaismuuttuja: Tulkinta

Satunnaismuuttuja X on suure, jonka arvo $X(s)$ määräytyy satunnaisilmiön toteumasta $s \in S$:

- Yhteen satunnaisilmiöön liittyy useita satunnaismuuttujia. Toteuma $s \in S$ määrää niiden kaikkien arvot.
- Todennäköisyysteoriassa tutkitaan satunnaismuuttujien arvojen todennäköisyyksiä, kun satunnaisilmiötä kuvaavan perusjoukon S todennäköisyysjakauma P tunnetaan.
- Tilastotieteessä pyritään havaittujen satunnaismuuttujien arvojen perusteella tekemään johtopäätöksiä perusjoukon S tuntemattomasta todennäköisyysjakaumasta P.

Matemaattisesti (**MS-E1600 Probability theory**):

- Satunnaismuuttuja on mitallinen kuvaus $X : S \to S'$
- Tapahtuman $\{X \in B\}$ todennäköisyys on luku $\mathrm{P}(A)$, missä $A = X^{-1}(B)$ on joukon B alkukuva

Eri tyyppisiä satunnaismuuttujia

Satunnaismuuttujasta $X : S \to S'$ saatetaan käyttää nimitystä

Nimitys	**Arvojoukko**	
Satunnaisluku	$S' \subset \mathbb{R}$	
Satunnaisvektori	$S' \subset \mathbb{R}^n$	
Satunnaismatriisi	$S' \subset \mathbb{R}^{m \times n}$	
Stokastinen prosessi	$S' \subset \mathbb{R}^T$	(aikavälin T funktiot)
Satunnaiskenttä	$S' \subset \mathbb{R}^U$	(alueen U funktiot)
Satunnaisverkko	$S' \subset \{0,1\}^{V \times V}$	(solmujoukon V verkot)

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

Sisältö

Satunnaismuuttujan jakauma

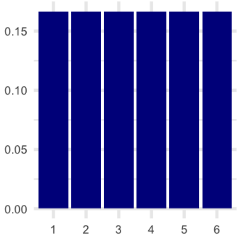
Satunnaismuuttujan X jakauma on taulukko tai funktio, josta voidaan määrittää X:n mahdolliset arvot ja niiden todennäköisyydet.

Esim (Kaksi nopanheittoa)

Nopan 1 silmäluvun X_1 jakauma on

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

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



Nopan 2 silmäluvulla X_2 on sama jakauma.

$\implies$ Satunnaismuuttujat X_1 ja X_2 ovat samoin jakautuneita.

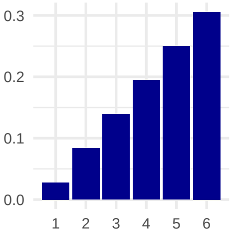
Esim. Kahden nopan maksimi

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

$$
\begin{aligned}
P(M = k) &= P(M \le k) - P(M \le k-1) \\
&= P(X_1 \le k, X_2 \le k) - P(X_1 \le k-1, X_2 \le k-1) \\
&= P(X_1 \le k)\,P(X_2 \le k) - P(X_1 \le k-1)\,P(X_2 \le k-1) \\
&= \left(\frac{k}{6}\right)^2 - \left(\frac{k-1}{6}\right)^2 \\
&= \frac{2k-1}{36}
\end{aligned}
$$

Satunnaismuuttujan M jakauma:

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



Esim. Metron odotusaika

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

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

Vastaavasti päätelleen havaitaan, että $\mathrm{P}(X = t) = 0$ kaikilla t.

Menikö yo. 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

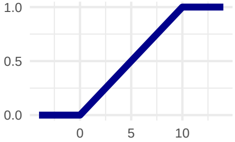
X = seuraavan metron odotusaika (min) asemalla, jonne metroja saapuu 10 min välein. Mikä on satunnaismuuttujan X jakauma?

Koska $\mathrm{P}(X=t) = 0$ kaikilla t, ei X:n jakaumaa voi määrittää pistetodennäköisyyksien avulla. Tarvitaan muita keinoja.

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

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 $F_X(t) = \mathrm{P}(X \le t)$.

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 $(3,4)$?

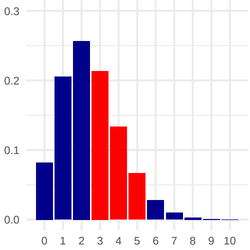
$$
\begin{aligned}
\mathrm{P}(X \in (1,2) \text{ tai } X \in (3,4)) &= \mathrm{P}(X \in (1,2)) + \mathrm{P}(X \in (3,4)) \\
&= (F_X(2) - F_X(1)) + (F_X(4) - F_X(3)) \\
&= \left(\frac{2}{10} - \frac{1}{10}\right) + \left(\frac{4}{10} - \frac{3}{10}\right) \\
&= 0.2.
\end{aligned}
$$

Sisältö

Tiheysfunktio

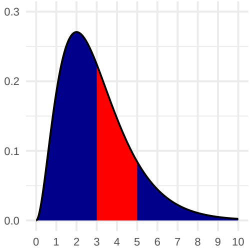
X on diskreetti, jos sen jakauma voidaan esittää funktion $f_X(x) \geq 0$ avulla muodossa

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



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

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



Diskreetin jakauman tiheysfunktio

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

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

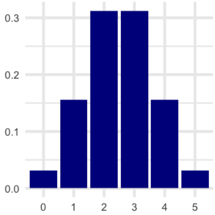
Vastaavasti mikä tahansa yo. ehdot toteuttava funktio on jonkin diskreetin jakauman tiheysfunktio.

Diskreetin jakauman tiheysfunktio

Kun satunnaismuuttujan arvojoukko on pieni, kannattaa jakauma esittää taulukkona.

Esim (Kruunien lukumäärä 5:llä kolikonheitolla)

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



Suuren arvojoukon tapauksessa jakauma kannattaa esittää tiheysfunktion avulla.

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 parametreina $n = 5\;000\;000$ 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 $\mathrm{P}(X = x) \;=\; 0$.

Tiheysfunktio on todennäköisyys suhteessa reaalilukujen esitystarkkuuteen; sen jatkuvuuspisteissä pienillä $h > 0$ arvoilla

$$f_X(x) \;\approx\; \frac{\mathrm{P}(X = x \pm h/2)}{h}$$

Kertymäfunktio ja tiheysfunktio

Jatkuvan satunnaisluvun

- kertymäfunktio saadaan tiheysfunktion integraalina

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

- tiheysfunktio saadaan kertymäfunktion derivaattana

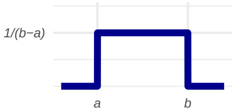
$$f_X(x) = F_X'(x)$$

kertymäfunktion derivoituvuuspisteissä.

Esim. Jatkuva tasajakauma

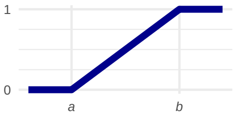
Funktio

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



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

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

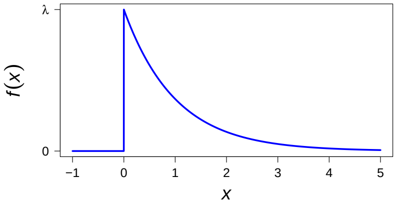


Sijoittamalla $a = 0$ ja $b = 10$ saadaan metron odotusajan jakauma.

Eksponenttijakauma

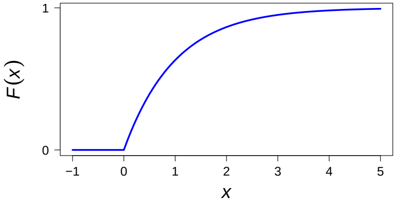
Eksponenttijakauman parametrina $\lambda > 0$ tiheysfunktio on

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



Integroimalla tiheysfunktiota $\Longrightarrow$ kertymäfunktio on

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



Eksponenttijakauman muistittomuus

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

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

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

Tulkinta

Huolimatta siitä, kuinka kauan bussia on jo odotettu, on tn että pitää odottaa vielä t min lisää sama kuin tn, että juuri pysäkille saapunut henkilö odottaa yli t min.

Satunnaisluvut — yhteenveto

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

Sisältö

Satunnaismuuttujien yhteisjakauma

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

Esim (Kaksi noppaa)

Noppien silmälukujen X_1 ja X_2 yhteisjakauma on

	X_2					
X_1	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

eli tulojoukon $\{1, \ldots, 6\} \times \{1, \ldots, 6\}$ tasajakauma.

Ensimmäisen nopan ja noppien maksimin yhteisjakauma

$\mathrm{P}(X_1 = 1, M = 1) = \mathrm{P}(X_1 = 1, X_2 = 1) = \frac{1}{36}$

Yleisesti:

- $k < i \implies \mathrm{P}(X_1 = i, M = k) = 0$
- $k = i \implies \mathrm{P}(X_1 = i, M = k) = \mathrm{P}(X_1 = i, X_2 \leq i) = \frac{k}{36}$
- $k > i \implies \mathrm{P}(X_1 = i, M = k) = \mathrm{P}(X_1 = i, X_2 = k) = \frac{1}{36}$

Satunnaismuuttujien X_1 ja M yhteisjakauma:

	M					
X_1	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	0	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	0	0	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	0	0	0	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	0	0	0	0	$\frac{5}{36}$	$\frac{1}{36}$
6	0	0	0	0	0	$\frac{6}{36}$

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

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

ja

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

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

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

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

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

Yhteisjakauman reunajakaumat

Rivi- ja sarakesummia kutsutaan yhteisjakauman reunajakaumiksi.

	M						
X_1	1	2	3	4	5	6	Yht
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	0	$\frac{2}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	0	0	$\frac{3}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	0	0	0	$\frac{4}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	0	0	0	0	$\frac{5}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	0	0	0	0	0	$\frac{6}{36}$	$\frac{1}{6}$
Yht	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$	

Rivisummista saadaan X_1:n jakauma
Sarakesummista saadaan M:n jakauma

Silmälukujen yhteisjakauman reunajakaumat

	X_2						
X_1	1	2	3	4	5	6	Yht
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
Yht	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	

Rivisummista saadaan X_1:n jakauma
Sarakesummista saadaan X_2: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

Ehdolliset jakauma

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:

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

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

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

Satunnaismuuttujien riippuvuus ja riippumattomuus

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

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

Yhtäpitävästi:

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

tai

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

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

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?
$\mathrm{P}(X_1 = 1, X_2 = 1) \ = ?$

Satunnaisotanta palauttaen ja palauttamatta

Palauttaen

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

Palauttamatta

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

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

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

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

Esim. Satunnaisotanta palauttamatta

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

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

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

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

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

Sisältö

Ääretön diskreetti arvojoukko

Diskreetin satunnaismuuttujan mahdollisten arvojen joukko voi olla ääretön.

Esim (Kimblen alkuvaihe)

$N =$ nopanheittojen lukumäärä, kunnes saadaan kuutonen.

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

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 $t \in [0, 9]$,

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

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

Onko Y:n jakauma diskreetti vai jatkuva?

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