

Insinöörimatematiikka 3

Mika Hirvensalo
`mikhirve@utu.fi`

Matematiikan ja tilastotieteen laitos
Turun yliopisto

2022

Määritelmä

Äärellisen (tai numeroituvan) joukon $X = \{x_1, \dots, x_n\} \subseteq R$ todennäköisyysjakauma on funktio joka toteuttaa ehdot $f(x) \geq 0$ ja $\sum_{x \in X} f(x) = 1$. Tällöin sanotaan, että f on *diskreetti todennäköisyysjakauma* ja että satunnaismuuttuja X saa arvon x_i todennäköisyydellä $f(x_i)$.

Määritelmä

Jos $f(x) \geq 0$ ja

$$\int_{-\infty}^{\infty} f(x) dx = 1,$$

sanotaan että f on jatkuva todennäköisyysjakauma.

Määritelmä

Jatkuvaan jakaumaan f liittyvä satunnaismuuttuja X saa arvon välillä $[a, b]$ todennäköisyydellä

$$\mathbb{P}(a \leq X \leq b) = \int_a^b f(x) dx$$

Määritelmä

Satunnaismuuttujan X keskiarvo (odotusarvo) on

$$\mu = \mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x) dx$$

Vertaa:

$$\mathbb{E}(X) = \frac{x_1 + x_2 + \dots + x_n}{n} = x_1 \cdot \frac{1}{n} + x_2 \cdot \frac{1}{n} + \dots + x_n \cdot \frac{1}{n}$$

(tasainen diskreetti jakauma)

$$\mathbb{E}(X) = x_1 \cdot f(x_1) + x_2 \cdot f(x_2) + \dots + x_n \cdot f(x_n)$$

(yleinen diskreetti jakauma)

Esimerkki

Painottamaton noppa: X saa arvot $\{1, 2, 3, 4, 5, 6\}$, jokaisen todennäköisyydellä $\frac{1}{6}$. Tällöin

$$\mathbb{E}(X) = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = 3,5.$$

Painotettu noppa: X saa arvon 6 todennäköisyydellä $\frac{1}{2}$ ja arvot $\{1, 2, 3, 4, 5\}$ todennäköisyydellä $\frac{1}{10}$. Tällöin

$$\mathbb{E}(X) = 1 \cdot \frac{1}{10} + 2 \cdot \frac{1}{10} + \dots + 5 \cdot \frac{1}{10} + 6 \cdot \frac{1}{2} = 4,5.$$

Esimerkki

Olkoon X silmälukujen summa kahden painottamattoman nopan heitossa. Tällöin

$$\mathbb{P}(X = 2) = \frac{1}{36}, \mathbb{P}(X = 3) = \frac{2}{36}, \dots, \mathbb{P}(X = 7) = \frac{6}{36},$$
$$\mathbb{P}(X = 8) = \frac{5}{36}, \dots, \mathbb{P}(X = 12) = \frac{1}{36}$$

Tällöin

$$\mathbb{E}(X) = 2 \cdot \frac{1}{36} + 3 \cdot \frac{2}{36} + \dots + 12 \cdot \frac{1}{36} = 7$$

Määritelmä

Satunnaismuuttujan X varianssi on (jatkuva jakauma)

$$\text{Var}(X) = \mathbb{E}((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

ja (diskreetti jakauma)

$$\text{Var}(X) = \mathbb{E}((X - \mu)^2) = \sum_{x \in X} (x - \mu)^2 f(x)$$

Varianssin neliöjuurta $\sigma = \sqrt{\text{Var}(X)}$ kutsutaan keskihajonnaksi.

Määritelmä

Jakaumafunktio eli kertymäfunktio on

$$F(x) = \mathbb{P}(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

(diskreetti jakauma) ja

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

(jatkuva jakauma)

Huomautus

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1,$$

$F(x)$ on kasvava funktio

Esimerkki

Kahden nopan heitossa

$$F(x) = 0, \text{ kun } x < 2,$$

$$F(x) = \frac{1}{16} \text{ kun } 2 \leq x < 3,$$

$$F(x) = \frac{1}{12} \text{ kun } 3 \leq x < 4,$$

$$F(x) = \frac{1}{6} \text{ kun } 4 \leq x < 5,$$

...

$$F(x) = \frac{35}{36} \text{ kun } 11 \leq x < 12,$$

$$F(x) = 1 \text{ kun } x \geq 12$$

Huomautus

Jos X on satunnaismuuttuja ja g koko reaaliakselilla määritelty funktio, niin $g(X)$ on myös satunnaismuuttuja ja

$$\mathbb{E}(g(X)) = \sum_{x_i} g(x_i) f(x_i)$$

(diskreetti jakauma) ja

$$\mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

(jatkuva jakauma)

Esimerkki

jos $a, b \in \mathbb{R}$ ja $g(X) = aX + b$

$$\begin{aligned}\mathbb{E}(aX + b) &= \int_{-\infty}^{\infty} (ax + b)f(x) dx \\ &= a \int_{-\infty}^{\infty} xf(x) dx + b \int_{-\infty}^{\infty} f(x) dx \\ &= a\mathbb{E}(X) + b\end{aligned}$$

Määritelmä

Satunnaismuuttujan X n :s origomomentti on (jatkuva jakauma)

$$\alpha_n(X) = \mathbb{E}(X^n) = \int_{-\infty}^{\infty} x^n f(x) dx$$

Määritelmä

Satunnaismuuttujan X n :s keskusmomentti on

$$\mu_n(X) = \mathbb{E}((X - \mu)^n) = \int_{-\infty}^{\infty} (x - \mu)^n f(x) dx$$

Huomautus

$\mu_0(X) = 1$, $\mu_1(X) = 0$ ja $\mu_2(X) = \text{Var}(X)$.

Keskusmomentti vs. origomomentti

$$\begin{aligned}\text{Var}(X) &= \mu_2 = \mathbb{E}((X - \mu)^2) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \\&= \int_{-\infty}^{\infty} (x^2 - 2\mu x + \mu^2) f(x) dx \\&= \int_{-\infty}^{\infty} x^2 f(x) dx - 2\mu \int_{-\infty}^{\infty} x f(x) dx + \mu^2 \int_{-\infty}^{\infty} f(x) dx \\&= \alpha_2 - 2\mu \cdot \mu + \mu^2 \cdot 1 = \alpha_2 - \mu^2 = \alpha_2 - \alpha_1^2 \\&= \mathbb{E}(X^2) - \mathbb{E}(X)^2\end{aligned}$$

Tasainen jakauma

$X = \{1, 2, \dots, n\}$, $\mathbb{P}(X = i) = \frac{1}{n}$ Tällöin

$$\mathbb{E}(X) = 1 \cdot \frac{1}{n} + 2 \cdot \frac{1}{n} + \dots + n \cdot \frac{1}{n} = \frac{1}{n}(1 + \dots + n) = \frac{1}{n} \cdot \frac{1}{2}n(n+1) = \frac{n+1}{2}$$

ja

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \sum_{k=1}^n k^2 \frac{1}{n} - \left(\frac{n+1}{2}\right)^2 = \frac{n^2 - 1}{12}$$

Binomijakauma

Toistetaan todennäköisyydellä p onnistuva koe (riippumattomasti) n kertaa. Satunnaismuuttujan X arvo on onnistuneiden kokeiden määrä ja tällöin

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

ja

$$\mathbb{E}(X) = np \text{ ja } \text{Var}(X) = np(1 - p)$$

Esimerkki

Lentokoneen moottori menee epäkuntoon todennäköisyydellä $p > 0$. Jos vähintään puolet moottoreista on kunnossa, kone selviytyy.

Todennäköisyys sille, että kaksimoottorinen kone ei selviydy, on $p \cdot p = p^2$.

Todennäköisyys sille, että nelimoottorinen kone ei selviydy, on

$$\binom{4}{3} p^3 (1-p) + \binom{4}{4} p^4 (1-p)^0 = 4p^3 - 3p^4.$$

$$4p^3 - 3p^4 < p^2 \Leftrightarrow p \in (0, \frac{1}{3}).$$

Binäärinen symmetrinen kanava

$0 \rightarrow 1$ ja $1 \rightarrow 0$ todennäköisyydellä p .

Siirrettäessä 2-pituisia sanoja 00, 01, 10, ja 11 virhetodennäköisyys on

$$\binom{2}{1} p(1-p) + \binom{2}{2} p^2 = 2p - p^2$$

Koodaus

00 $\rightarrow$ 00000, 01 $\rightarrow$ 00111, 10 $\rightarrow$ 11100 ja 11 $\rightarrow$ 11011 nostaa lähetyssaikaa 2,5-kertaiseksi. Merkitään

$C = \{00000, 00111, 11100, 11011\}$ ja dekodausfunktio δ määritellään Hamming-etäisyyden perusteella, siis esim.

$$\delta(10000) = \delta(01000) = \delta(00100) = \delta(00010) = \delta(00001) = 00000.$$

Jos Hamming-etäisyys d_H kahteen eri sanaan on yhtä suuri, määritellään δ mielivaltaisesti. Esim. $\delta(10101) = 11100$ mutta voisi olla myös 00111.

Huomautus

Todennäköisyys sille, että lähetettäessä sana $\mathbf{c}$ saadaankin sana $\mathbf{x}$ on

$$\mathbb{P}(\mathbf{x} \mid \mathbf{c}) = p^{d_H(\mathbf{x}, \mathbf{c})} (1 - p)^{5 - d_H(\mathbf{x}, \mathbf{c})}$$

Esimerkiksi

$$\mathbb{P}(00001 \mid 00000) = (1 - p)(1 - p)(1 - p)(1 - p)p = p^1(1 - p)^4$$

ja

$$\mathbb{P}(01001 \mid 00000) = (1 - p)p(1 - p)(1 - p)p = p^2(1 - p)^3.$$

Määritelmä

Jos $\mathbf{c} \in C$, merkitään

$$E_{\mathbf{c}} = \{\mathbf{x} \in \mathbb{F}_2^5 \mid \delta(\mathbf{x}) \neq \mathbf{c}\}$$

ja

$$D_{\mathbf{c}} = \{\mathbf{x} \in \mathbb{F}_2^5 \mid \delta(\mathbf{x}) = \mathbf{c}\} = \mathbb{F}_2^5 \setminus E_{\mathbf{c}}.$$

$E_{\mathbf{c}}$ on siis niiden sanojen joukko, jotka eivät dekodaudu koodisanaksi $\mathbf{c}$ ja $D_{\mathbf{c}}$ sen komplementti.

Määritelmä

$$\mathbb{P}_{er}(\mathbf{c}) = \sum_{\mathbf{x} \in E_{\mathbf{c}}} \mathbb{P}(\mathbf{x} \mid \mathbf{c}) = 1 - \sum_{\mathbf{x} \in D_{\mathbf{c}}} \mathbb{P}(\mathbf{x} \mid \mathbf{c})$$

kertoo dekodausvirheen todennäköisyyden kun syötteenä on koodisana $\mathbf{c} \in C$.

Keskiarvo

$$\mathbb{P}_{er}(C) = \frac{1}{|C|} \sum_{\mathbf{c} \in C} \mathbb{P}_{er}(\mathbf{c})$$

Ilmoittaa virhetodennäköisyyden kun käytetään koodia C , olettaen että koodisanat jakautuvat yhtä todennäköisesti.

Dekoodautuvat sanat

$$D_{00000} = \{00000, 00001, 00010, 00100, 01000, 01001, 01010, 10000, 10001, 10010\}$$

$$D_{00111} = \{00011, 00101, 00110, 00111, 01101, 01110, 01111, 10101, 10110, 10111\}$$

$$D_{11100} = \{01100, 10100, 11000, 11100, 11101, 11110\}$$

$$D_{11011} = \{01011, 10011, 11001, 11010, 11011, 11111\}$$

Virhetodennäköisyydet

$$\mathbb{P}_{er}(00000) = 1 - (1 - p)^5 - 5(1 - p)^4 p - 4(1 - p)^3 p^2$$

Virhetodennäköisyydet

$$\mathbb{P}_{er}(00000) = 1 - (1 - p)^5 - 5(1 - p)^4 p - 4(1 - p)^3 p^2$$

$$\mathbb{P}_{er}(00111) = 1 - (1 - p)^5 - 5(1 - p)^5 p - 4(1 - p)^3 p^2$$

$$\mathbb{P}_{er}(11100) = 1 - (1 - p)^5 - 5(1 - p)^4 p$$

$$\mathbb{P}_{er}(00111) = 1 - (1 - p)^5 - 5(1 - p)^4 p$$

Keskiarvo

Koodin C virhetodennäköisyys on keskiarvo edellisistä:

$$\mathbb{P}_{er}(C) = 8p^2 - 14p^3 + 9p^4 - 2p^5$$

Koodattu vs. koodaamaton

$$\begin{aligned}\frac{\mathbb{P}_{er}(C)}{2p - p^2} &= \frac{p^2(8 - 14p + 9p^2 - 2p^3)}{2p - 2p^2} \\ &= p(4 - 5p + 2p^2) \\ &\approx 4p\end{aligned}$$

Jos esim. $p = \frac{1}{10^6}$, on $4p = \frac{1}{250000}$

Määritelmä

Jos $X = \mathbb{N} \cup \{0\}$, on X :llä Poisson-jakauma, jos

$$\mathbb{P}(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

Huomautus

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

Huomautus

Poisson-jakauma on hyvä approksimaatio Binomijakaumalle, mikäli p on pieni ja n suuri. Tarkemmin: Jos merkitään $p = \lambda/n$, on

$$\binom{n}{k} p^k (1-p)^{n-k} \xrightarrow{n \rightarrow \infty} e^{-\lambda} \frac{\lambda^k}{k!}$$

Approksimaatio binomijakaumalle

Jos merkitään $p = \frac{\lambda}{n}$ ($\lambda = np = \mathbb{E}(X)$)

$$\begin{aligned} & \binom{n}{k} p^k (1-p)^{n-k} \\ = & \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ = & \frac{\lambda^k}{k!} \frac{n(n-1)\dots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\ \xrightarrow{n \rightarrow \infty} & e^{-\lambda} \frac{\lambda^k}{k!} \end{aligned}$$

Lause

Poisson-jakaumalle

$$\mathbb{E}(X) = \lambda$$

ja

$$\text{Var}(X) = \lambda$$

Lause

Jos $f(k, t) = \mathbb{P}(k \text{ tapahtumaa aikavälillä } [0, t])$ ja

- tapahtumat ovat riippumattomat
- $\mathbb{P}(\text{yksi tapahtuma } h\text{-pituisella aikavälillä}) = (v + g(h)) \cdot h$,
missä $g(h) \xrightarrow{h \rightarrow 0} 0$.
- $\mathbb{P}(k > 1 \text{ tapahtumaa } h\text{-pituisella aikavälillä}) \xrightarrow{h \rightarrow 0} 0$.

niin $f(k, t) = \frac{e^{-vt}(vt)^k}{k!} = e^{-\lambda} \frac{\lambda^k}{k!}$, missä $\lambda = vt$.

Esimerkki

Sadasta signaalista keskimäärin yksi muuttuu kanavassa. Kun lähetetään 200 toisistaan riippumatonta signaalia, mikä on todennäköisyys sille että ainakin kolme muuttuu?

Binomijakaumaa käyttäen:

$$\begin{aligned}\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X < 3) \\ &= 1 - \binom{200}{0} 0,01^0 \cdot 0,99^{200} - \binom{200}{1} 0,01^1 \cdot 0,99^{199} \\ &\quad - \binom{200}{2} 0,01^2 \cdot 0,98^{198} \\ &= 0,323321 \dots\end{aligned}$$

Esimerkki

Poisson-jakaumaa käyttäen: Tässä $p = 0,01$ ja $n = 200$, joten $\lambda = \mathbb{E}(X) = 0,01 \cdot 200 = 2$ ja

$$\begin{aligned}\mathbb{P}(X \geq 3) &= 1 - \mathbb{P}(X < 3) \\ &= 1 - e^{-2} \left(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!} \right) \\ &= 1 - e^{-2}(1 + 2 + 2) \\ &= 0,323324 \dots\end{aligned}$$