Todennäköisyysjakaumat

■ Sisältö
■ Hakemisto

ESITIEDOT: ■ todennäköisyyslaskenta, ■ määrätty integraali
KATSO MYÖS: ■ tilastomatematiikka

Diskreetit jakaumat

■ stokastinen muuttuja (diskreetti)

Olkoon X diskreetti stokastinen muuttuja, joka saa erilliset arvot $x_1, x_2, \dots, x_n$. (Näitä voi olla myös ääretön määrä: x_k, missä k saa arvoikseen kaikki luonnolliset luvut.)

■ todennäköisyys (funktio)
■ todennäköisyys (funktio)

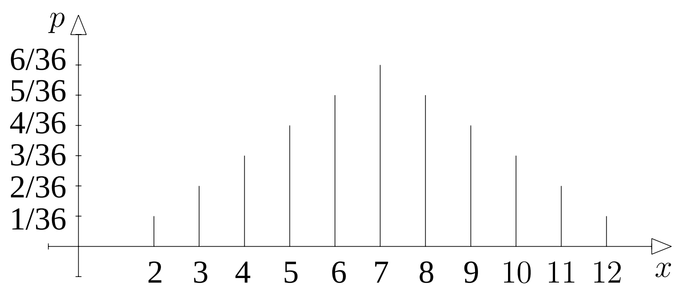
Diskreetin stokastisen muuttujan jakauma, *diskreetti jakauma*, so. muuttujan saamiin arvoihin liittyvät todennäköisyydet voidaan määritellä antamalla *pistetodennäköisyydet* p_k jokaiselle indeksille k:

$$P(X = x_k) = p_k.$$

■ summamerkintä

Nämä ovat ≥ 0 ja niiden summa on $\sum_k p_k = 1$.

Pistetodennäköisyydet voidaan graafisesti esittää pystysuorien janojen avulla. Esimerkiksi kahta noppaa heitettäessä pistelukujen summa on stokastinen muuttuja, joka saa arvot $2,\ 3,\ \dots,\ 12$. Näiden todennäköisyysjakauma ilmenee seuraavasta graafisesta esityksestä:



■ riippumattomuus (todennäköisyyslaskennassa)
■ komplementtitapahtuma

Usein esiintyvä diskreetti jakauma on *binomijakauma*. Tämä kuvaa saman kokeen toistamista, missä toistot ovat toisistaan riippumattomia ja kussakin toistossa tuloksena on joko A tai tämän komplementtitapahtuma $\bar{A}$. Jos näiden todennäköisyydet ovat $P(A) = p$ ja $P(\bar{A}) = 1 - p$, niin toistettaessa koe n kertaa saadaan tulos A täsmälleen k kertaa todennäköisyydellä

$$p_k = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n.$$

Diskreettejä jakaumia on paljon muitakin.

ESITIEDOT: ■ todennäköisyyslaskenta, ■ määrätty integraali
KATSO MYÖS: ■ tilastomatematiikka

■ Sisältö
■ Hakemisto

Jatkuvat jakaumat

Jatkuva stokastinen muuttuja X saa reaaliarvoja joltakin reaaliakselin väliltä $[a, b]$ (tai mahdollisesti kaikki reaaliarvot). Tällöin yksittäisen reaaliarvon todennäköisyys on $= 0$, vaikka arvo ei olekaan mahdoton. Sen sijaan todennäköisyys, että arvo osuu jollekin osavälille $[c, d]$, on yleensä positiivinen.

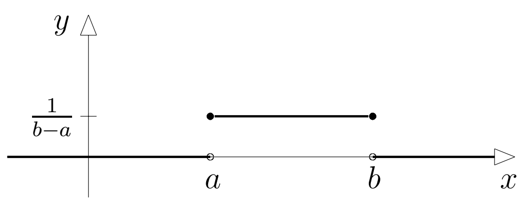
Jatkuvan muuttujan X jakauma, ns. *jatkuva jakauma* määritellään yleensä *tiheysfunktion* $f(x)$ avulla. Tämä on kaikkialla ≥ 0 ja todennäköisyys, että muuttujan X arvo on välillä $[c, d]$, saadaan integraalista:

$$P(c \leq X \leq d) = \int_c^d f(x)\,dx.$$

Jos integroidaan satunnaismuuttujan kaikkien mahdollisten arvojen muodostaman välin yli (voidaan ajatella koko reaaliakselin yli), on kyseessä varma tapahtuma ja siis $\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Yksinkertainen esimerkki jatkuvasta jakaumasta on *tasainen jakauma* välillä $[a, b]$. Tämän tiheysfunktio on

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & \text{kun } x \in [a, b], \\ 0 & \text{muulloin.} \end{cases}$$



Jatkuvia jakaumia on paljon muitakin; erityisen tärkeä on *normaalijakauma*, jota käsitellään edempänä.

■ stokastinen muuttuja (jatkuva)
■ väli (reaaliakselin)
■ todennäköisyys (funktio)
■ todennäköisyys (funktio)
■ määrätty integraali
■ tapahtuma (todennäköisyyslaskennassa)

■ Sisältö
■ Hakemisto

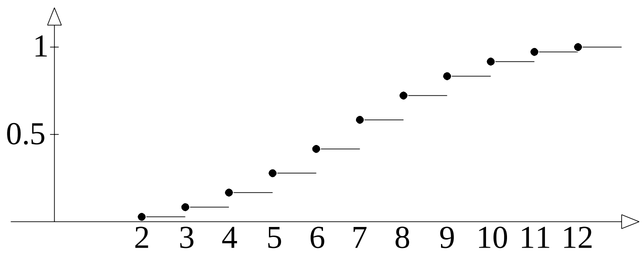
ESITIEDOT: ■ todennäköisyyslaskenta, ■ määrätty integraali
KATSO MYÖS: ■ tilastomatematiikka

Kertymäfunktio

Diskreetin tai jatkuvan stokastisen muuttujan X *kertymäfunktioksi* kutsutaan funktiota $F(x) = P(X \leq x)$. Koska kyseessä on todennäköisyys, on $0 \leq F(x) \leq 1$ kaikilla x. Funktio on kaikkialla kasvava (so. jos $x_1 < x_2$, niin $F(x_1) \leq F(x_2)$; funktio voi siis olla paloittain vakio).

■ stokastinen muuttuja (diskreetti)
■ stokastinen muuttuja (jatkuva)
■ todennäköisyys (funktio)
■ todennäköisyys (funktio)
■ kasvava (funktio)
■ kasvava (funktio)

Diskreetin satunnaismuuttujan kertymäfunktio on porrasfunktio; esimerkkinä olkoon kahta noppaa heitettäessä saatavan pistesumman kertymäfunktio:



Jatkuvan jakauman kertymäfunktiolle $F(x)$ ja tiheysfunktiolle $f(x)$ pätee

■ määrätty integraali

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx.$$

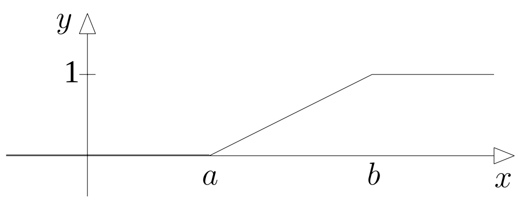
Jos tiheysfunktio on jatkuva, pätee myös

■ jatkuvuus
■ derivoituvuus
■ integraalifunktio

$$F'(x) = f(x).$$

Esimerkkinä jatkuvan satunnaismuuttujan kertymäfunktiosta olkoon väliä $[a, b]$ vastaavan tasaisen jakauman kertymäfunktio:

$$F(x) = \begin{cases} 0, & \text{kun } x \leq a, \\ \dfrac{x-a}{b-a}, & \text{kun } a \leq x \leq b, \\ 1, & \text{kun } x \geq b. \end{cases}$$



■ Sisältö
■ Hakemisto

Jakauman tunnusluvut

Erilaisia jakaumia pyritään luonnehtimaan sopivilla *tunnusluvuilla*.

Tärkein näistä on *odotusarvo*, jota voidaan luonnehtia stokastisen muuttujan saamien arvojen painotetuksi keskiarvoksi, jossa painoina ovat todennäköisyydet.

Jos diskreetti satunnaismuuttuja X saa arvot x_k todennäköisyyksillä p_k, odotusarvo on

$$E(X) = \sum_k x_k p_k.$$

Jos jatkuvan satunnaismuuttujan tiheysfunktio on $f(x)$, on odotusarvo

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

Tämäkin on todennäköisyyksillä painotettu keskiarvo, kuten nähdään tarkastelemalla integraalia vastaavaa Riemannin summaa.

Esimerkiksi yhden nopan heitossa odotusarvo on

$$1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = 3.5\,;$$

kyseessä on silmälukujen suora keskiarvo, koska kaikki pistetodennäköisyydet ovat samoja.

Binomijakauman ja väliä $[a, b]$ vastaavan tasaisen jakauman odotusarvoiksi saadaan

$$\sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np \qquad \text{ja} \qquad \int_a^b \frac{x}{b-a}\, dx = \frac{a+b}{2}.$$

Jakauman laajuutta — *hajontaa* — odotusarvon $E(X) = \mu$ ympärillä voidaan mitata tarkastelemalla etäisyyttä $|X - \mu|$ (joka itse on stokastinen muuttuja). Tämän neliön odotusarvo on jakauman *varianssi*:

$$\sigma^2 = E((X-\mu)^2).$$

Varianssin neliöjuuri $\sigma = \sqrt{E((X-\mu)^2)}$ on jakauman *keskihajonta*.

Diskreetin ja jatkuvan jakauman tapauksessa varianssille voidaan johtaa lausekkeet

$$\sigma^2 = \sum_k (x_k - \mu)^2 p_k \qquad \text{ja} \qquad \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx.$$

Näiden avulla saadaan esimerkiksi binomijakauman varianssiksi $np(1-p)$ ja tasaisen jakauman varianssiksi $(b-a)^2/12$.

■ stokastinen muuttuja (diskreetti)
■ stokastinen muuttuja (jatkuva)
■ keskiarvo (painotettu)
■ summamerkintä
■ määrätty integraali
■ Riemannin summa
■ keskihajonta

■ Sisältö
■ Hakemisto

Normaalijakauma

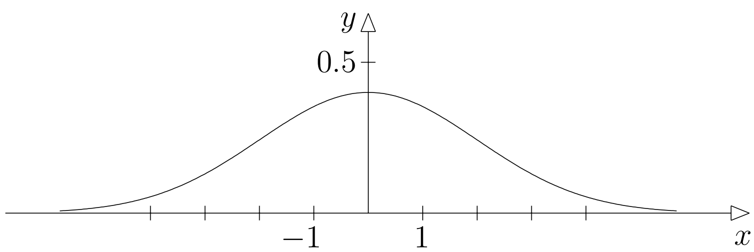
Jatkuvista jakaumista tärkein on *normaalijakauma*. Sitä tarvitaan sekä monissa sovelluksissa että teoreettisissa tarkasteluissa kuvaamaan ilmiöitä, joissa keskialueen arvot ovat tietyllä tavalla todennäköisempiä kuin kummankin ääripään arvot.

Standardinormaalijakauman tiheysfunktio on

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \quad x \in \mathbb{R}.$$

Tämän kuvaajaa kutsutaan Gaussin *kellokäyräksi*.

■ Gauss

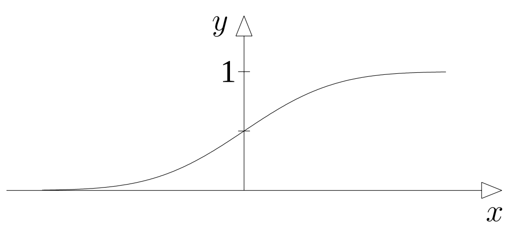


Standardinormaalijakauman kertymäfunktio on

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt,$$

missä integraalia ei voida laskea tavallisten alkeisfunktioiden avulla. Useat tietokoneohjelmat kuitenkin kykenevät laskemaan funktiolle $\Phi(x)$ numeerisia arvoja. Näitä on myös taulukoituina.

■ integraalifunktio
■ alkeisfunktio



Standardinormaalijakauman odotusarvo on 0 ja keskihajonta 1.

Normaalijakaumasta käytetään usein myös yleistä muotoa, jossa odotusarvo on $= \mu$ ja keskihajonta $= \sigma$. Tiheysfunktio ja kertymäfunktio ovat tällöin

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(x-\mu)^2/(2\sigma^2)} = \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma}\right)$$
$$F(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{-\infty}^{x} e^{-(t-\mu)^2/(2\sigma^2)}\, dt = \Phi\left(\frac{x-\mu}{\sigma}\right).$$