14 Jatkuva jakauma

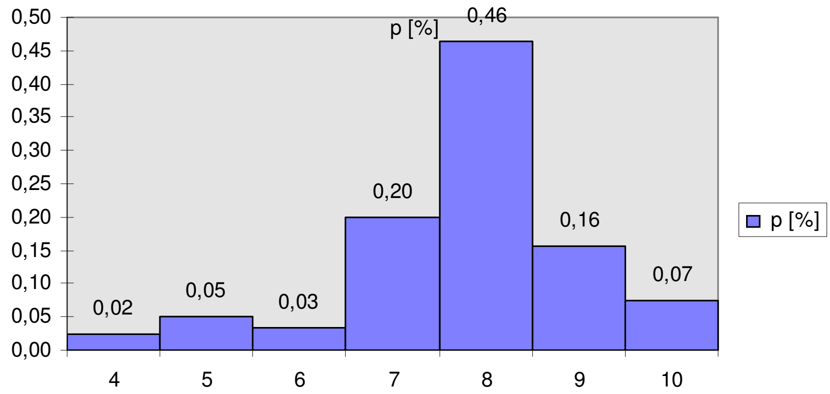
Edellä määriteltiin diskreetiksi satunnaismuuttujaksi sellainen, joka voi saada vain (hyppäyksittäin) erillisiä arvoja. Jatkuva satunnaismuuttuja voi saada mitä hyvänsä arvoja yleensä kuitenkin niin, että nämä ovat rajoitettuja. Tällöin sanotaan, että jatkuvan satunnaismuuttujan tapausavaruus on jokin reaaliakselin väli $a < \underline{x} < b$. Jatkuvan satunnaismuuttujan kentässä saatetaan tutkia apinapopulaation ikä-jakaumaa, alkuasukasheimon pituusjakaumaa, vehnäjauhopussin massajakaumaa, elektronisuihkun energiajakaumaa, jne.

Jatkuva jakauma ja sen todennäköisyysmassan vaihtelu hallitaan **tiheysfunktiolla.** Jos **diskreetin** satunnaismuuttujan todennäköisyysjakauman kuvaajana käytetään janadiagrammia, kyseessä on joukko x-akselilta alkavia, y-akselin suuntaisia janoja, joiden yhteispituus on 1. Jos taas käytetään kuvaajana histogrammia, toisissaan kiinni olevaa pylväsjoukkoa, muodostuu eräs tasoalue, jonka pinta-ala voidaan normittaa ykköseksi. Tästä on hyvin lyhyt ajatussiirtymä **jatkuvan satunnaismuuttujan tiheysfunktioon ja sen kuvaajaan.**

Käsitellään kuitenkin ennen täsmällisiä määritelmiä johdatteleva

Esim. 1 Suuressa lukiossa jakaantuivat erään kurssi äidinkielen arvosanat seuraavasti:

x	f	p [%]
4	3	0,024793
5	6	0,049587
6	4	0,033058
7	24	0,198347
8	56	0,46281
9	19	0,157025
10	9	0,07438
	121	1

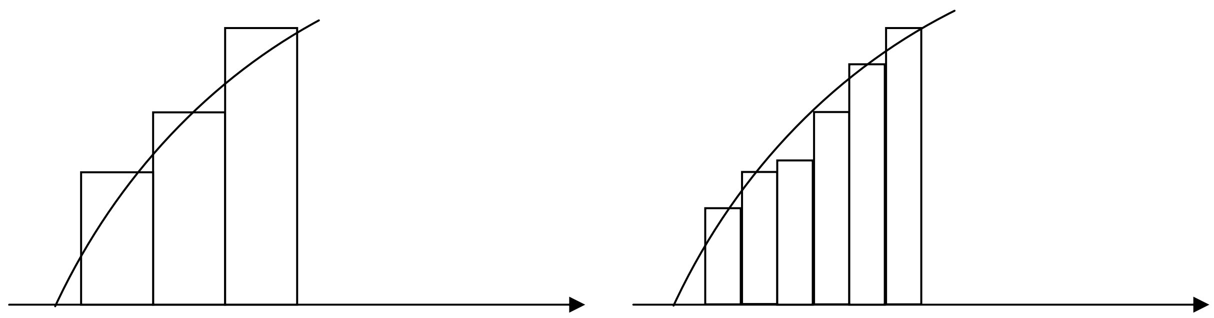


Esim. 1 Äidinkielen kurssiarvosanojen jakauma

Kuvion piirtäminen on normitettu siten, että pylväiden yhteenlaskettu pinta-ala on = 1. Tämä käy valitsemalla jokaisen pylvään leveydeksi = 1, sillä onhan suhteellisten frekvenssien summa = 1. Jos joku olisi kiinnostunut sellaisesta seikasta, millä todennäköisyydellä yksittäinen kurssille osallistunut oppilas on saanut arvosanan seitsemän, eikä tunneta frekvenssijakauman taulukkoa, ja vaikka tunnettaisiinkin, vastaus on seiskapylvään pinta-ala. Ellei kuvion kokonaispinta-alaa ole normitettu ykköseksi, vastaus olisi seiskapylvään alan suhde pylväiden yhteenlaskettuun pinta-alaan.

Jos oppilasmäärä olisi hyvin paljon suurempi ja jos oppilaat arvosteltaisiin kokonaisten numeroitten sijasta esimerkiksi kymmenesosanumeron tarkkuudella, oheinen pylväskuvio (histogrammi) muistuttaisi paljon enemmän ns. Gaussin kellokäyrää, johon palataan normaalijakauman yhteydessä.

Kouluarvostelussa ei hevin ryhdytä arvostelua suorittamaan sadasosanumeron tarkkuudella, mutta esimerkiksi fysikaalisissa mittauksissa tällainen olisi aivan luontevaa. Kun tutkitaan suurta joukkoa (eli n on hyvin suuri) ja kun käytetään hyvin tiheää luokkaväliä, niin suhteellisia frekvenssejä kuvaavan histogrammin pylväät ovat varsin kapeita. Jos tällöin pylväistä piirrettäisiin näkyviin vain niiden vaakasuorat yläreunat, niin todennäköisyysmassan jakautumista kuvaava sahalaitaviiva alkaisi yhä enemmän muistuttaa jatkuvan funktion kuvaajaa. Tällä on yleensä yksi maksimiarvo ja kaikki havaintoarvot ovat keskittyneet enemmän tai vähemmän tiiviisti tämän maksimiarvon läheisyyteen. Jos jakauman keskihajonta on pieni, havaintoarvot sijaitsevat varsin kapealla alueella, kuten diskreetin jakauman keskihajonnan merkityksen selvittelyjen yhteydessä todettiin.



Kuva. Jako tihenee kaksinkertaiseksi

Tiheysfunktion ominaisuuksiin viitattiin jo edellä. Että tämä ei koskaan saa negatiivista arvoa, käy ymmärrettäväksi, jos kuvittelee sen muodostuvan tiheäjakoisen histogrammin pylväiden yläreunoista. Nehän vastaavat (esim. oppilasjoukon tapauksessa) tietyn arvosanan saaneiden lukumäärää, joka ei voi olla negatiivinen.

****

MÄÄRITELMÄ 15 Funktio f on satunnaismuuttujan $\underline{x}$ tiheysfunktio, jos se täyttää kaksi ehtoa:

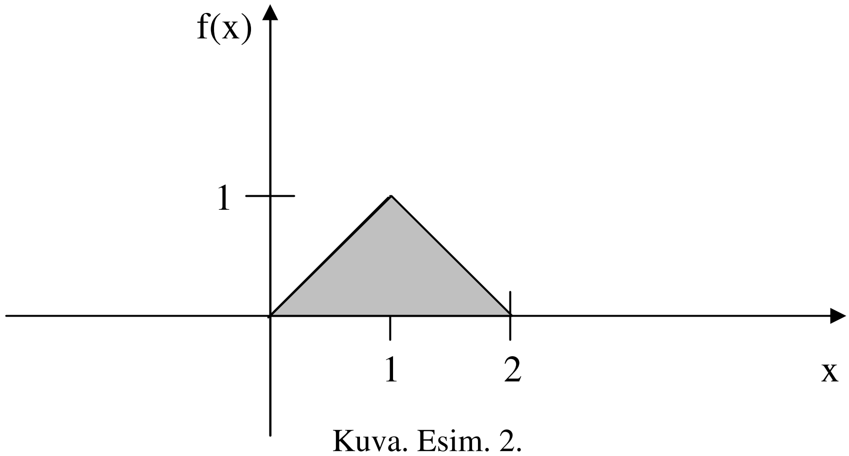
- $f(x) \geq 0$ kaikilla x:n arvoilla
- funktion f kuvaaja rajoittaa yhdessä x-akselin kanssa tasopinnan, jonka ala on tasan yksi.

****

Esim. 2 Osoita, että funktio f: $f(x) = \begin{cases} x, \text{ kun } 0 \leq x \leq 1 \\ 2 - x, \text{ kun } 1 < x \leq 2 \\ 0, \text{ muulloin} \end{cases}$ on tiheysfunktio.

Piirrä sen kuvaaja.

Välillä $0 < x \leq 1$ kuvaajana on nouseva suora $y = x$, jonka saamat arvot tunnetusti ovat tällöin positiivisia, ja $f(0) = 0$. Välillä $1 < x < 2$ kuvaaja on laskeva suora, joka tällä välillä saa niinikään vain positiivisia arvoja, sillä $2 - x > 0 \Leftrightarrow 2 > x \Leftrightarrow x < 2$ ja $f(2) = 0$. Kun kaikilla muilla x:n arvoilla toteutuu ehto $f(x) = 0$, niin funktion f ei-negatiivisuus on mitä ilmeisin tosiasia.



Kuva. Esim. 2.

Tiheysfunktio rajoittaa x-akselin kanssa tasakylkisen kolmion, jonka kanta on 2 ja korkeus 1. Sen pinta-ala on todellakin 1 ja funktio f täyttää molemmat tiheysfunktiolle asetetut vaatimukset.

Mikäli edellinen esimerkki kuvaisi jonkin todellisen satunnaismuuttujan $\underline{x}$ todennäköisyysmassan jakautumista, tiedettäisiin arvojen keskittyneen pisteen $x = 1$ ympäristöön. Tiheysfunktion avulla voitaisiin laskea sellaisia todennäköisyyksiä, kuin $P(\underline{x} < 1.2)$, $P(0.8 \leq \underline{x} < 1.76)$ taikka $P(\underline{x} > 0.3)$.

$P(\underline{x} < 1.2)$ = pinta-ala, jota rajoittavat x-akseli, funktion f kuvaaja ja suora $x = 1.2$, siitä vasempaan.
$P(0.8 \leq \underline{x} < 1.76)$ = pinta-ala, jota rajoittavat x-akseli, funktion f kuvaaja sekä suorat $x = 0.8$ ja $x = 1.76$
$P(\underline{x} > 0.3) = 1 - P(\underline{x} \leq 0.3) = 1$ – pinta-ala, jota rajoittavat x-akseli, funktion f kuvaaja ja suora $x = 0.3$ oikean käden suuntaan.
Kuvattujen pinta-alojen laskeminen onnistuu, mikäli asetetuin ehdoin syntyvät geometriset kuviot ovat niin yksinkertaisia, että niiden pinta-alan lasku hallitaan. Tiheysfunktio ei kuitenkaan ole aina ensiasteinen polynomi, ja tällöin pinta-alan laskeminen melkein aina vaatii integraalilaskentaa tai kertymäfunktion yhtälön tuntemista.

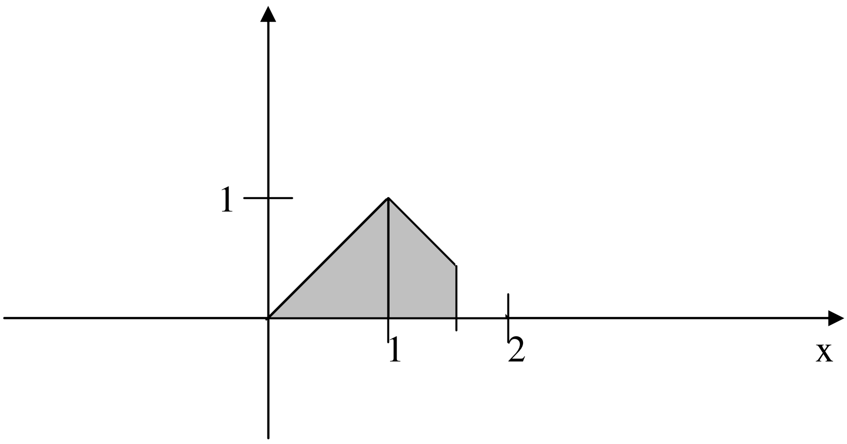
Esim. 3 Määritä edellisesimerkin jakauman kertymäfunktio. Piirrä sen kuvaaja pääpiirtein.

Kertymäfunktio F vastaa kysymykseen, millä todennäköisyydellä satunnaismuuttuja $\underline{x}$ on pienempi kuin x, ts. $F(x) = P(\underline{x} \leq x)$. Kyseessä on ala, jota rajoittavat x-akseli, tiheysfunktion kuvaaja sekä y-akselin suuntainen suora x = x̲.

Olkoon aluksi x $\le$ 0. Tällöin ei erotu mitään pinta-alaltaan nollasta eroavaa geometrista kuviota. Siten F(x) = 0 kaikilla x $\le$ 0.

Olkoon 0 < x < 1. Tällöin tiheysfunktio rajoittaa x-akselin kanssa suorakulmaisen kolmion, jonka kanta on x ja korkeus y = h = x ja pinta-ala = $x^2/2 = F(x)$.

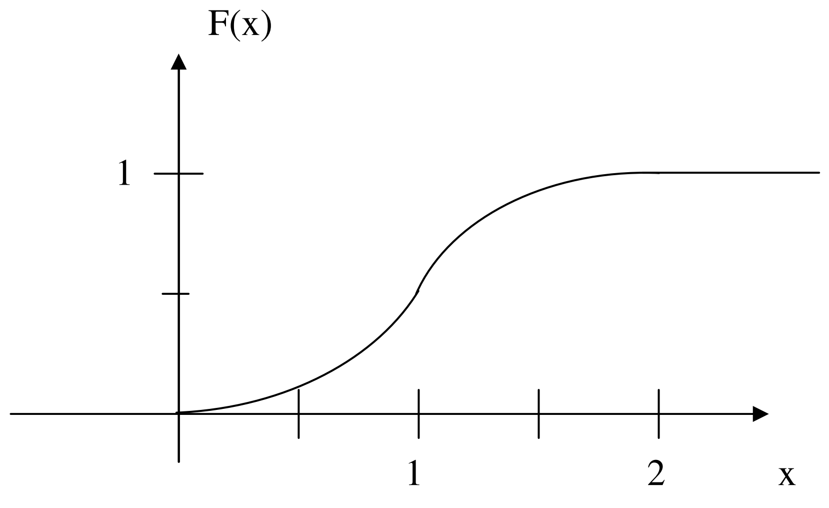
Olkoon 1 $\le$ x < 2. Tällöin tiheysfunktio rajoittaa x-akselin kanssa kuvion, jonka voidaan katsoa muodostuvan suorakulmaisesta kolmiosta (kateetit 1 ja 1) sekä puolisuunnikkaasta, jonka korkeus h = x – 1 ja kannat 1 ja 2 – x. Rajoittuvan alueen pinta-ala on



$$\frac{1}{2} + (x-1)\cdot\frac{1+2-x}{2} = \frac{1}{2} + \frac{(x-1)(3-x)}{2} = \frac{-2+4x-x^2}{2} = F(x)$$

Olkoon lopuksi x $\ge$ 2. Kaikki todennäköisyysmassa sijaitsee minkä tahansa x:n arvon vasemmalla puolella, jolle x $\ge$ 2. Siten F(x) = 1.

$$F(x) = \begin{cases} 0, \text{kun } 0 \le 0 \\ \dfrac{x^2}{2}, \text{kun } 0 < x \le 1 \\ \dfrac{-2+4x-x^2}{2}, \text{kun } x < 1 < 2 \\ 1, \text{kun} x \ge 2 \end{cases}$$

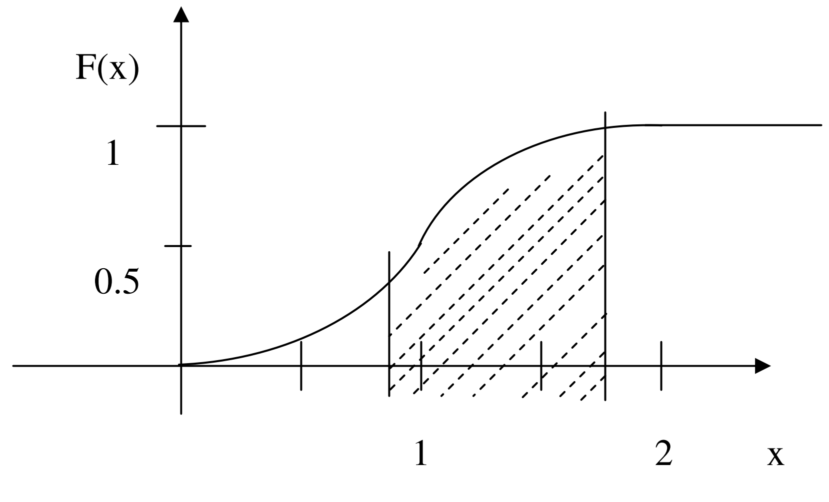


Esim. 4 Määritä kertymäfunktion avulla esimerkkeihin 6.54 – 6.55 liittyvät todennäköisyydet a) P($\underline{x}$ < 1.2), b) P(0.8 ≤ $\underline{x}$ < 1.76) ja c) P($\underline{x}$ > 0.3).

a) $P(\underline{x} < 1.2) = F(1.2) = \dfrac{-2 + 4 \cdot 1.2 - 1.2^2}{2} = 0.68.$

b) $P(0.8 \leq \underline{x} < 1.76) = F(1.76) - F(0.8) = \dfrac{-2 + 4 \cdot 1.76 - 1.76^2}{2} - \dfrac{0.8^2}{2} =$
$= 0.9712 - 0.32 = 0.6512$ (kuvio alla)

c) $P(\underline{x} > 0.3) = 1 - P(x < 0.3) = 1 - \dfrac{0.3^2}{2} = 1 - \dfrac{0{,}09}{2} = 0.1545.$



Kuva. Esim. 4. b).

Jos olet perehtynyt integraalilaskentaan, niin tiheysfunktiosta f lähtien voit määrittää kertymäfunktion pinta-alatulkintaan nojautuen määrätyn integraalin avulla:

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

mistä välittömästi voit päätellä erinomaisen hyödyllisen keinon tarkistaa, onko kertymäfunktion määritys mennyt oikein:

Tiheysfunktio = kertymäfunktion derivaatta

Kun kertymäfunktio tunnetaan, satunnaismuuttujaan liittyviä todennäköisyyksiä on helppo laskea. Niinpä

$$P(a \leq \underline{x} < b) = F(b) - F(a) = \int_{a}^{b} f(x)dx.$$

Huomasit ehkä, että esimerkeissä on käsitelty puoliavoimia välejä. Sillä, onko tapahtuma A puoliavoin, avoin vaiko suljettu väli, ei ole käytännössä mitään merkitystä. Usein on mukava tarkastella puoliavoimia välejä välin alkupiste mukana, jolloin toinen reunapiste ei kuulu samanaikaisesti kahteen eri tapaukseen. Määritelmästä seuraa sekin, että yksityisen satunnaismuuttujan todennäköisyys on nolla, vaikka $\underline{x}$ sijaitsisi tiheysfunktion nollasta poikkeavalla alueella. Tämä taas perusteellisemmin selvittäen johtuu siitä, että sellaisella suorakulmiolla, jonka kanta on nolla, on pinta-alakin nolla.

Kun kysytään, millä todennäköisyydellä umpimähkään valitun japanilaisen pituus on 165 cm, tarkoitetaan käytännössä, millä todennäköisyydellä henkilön pituus kuuluu välille 164.5 cm $\leq$ pituus < 165.5 cm eikä tarkoiteta todennäköisyyttä P(pituus = 165.0000000000000000000.......... cm), missä desimaalipisteen jälkeen seuraavia nollia on äärettömän monta.

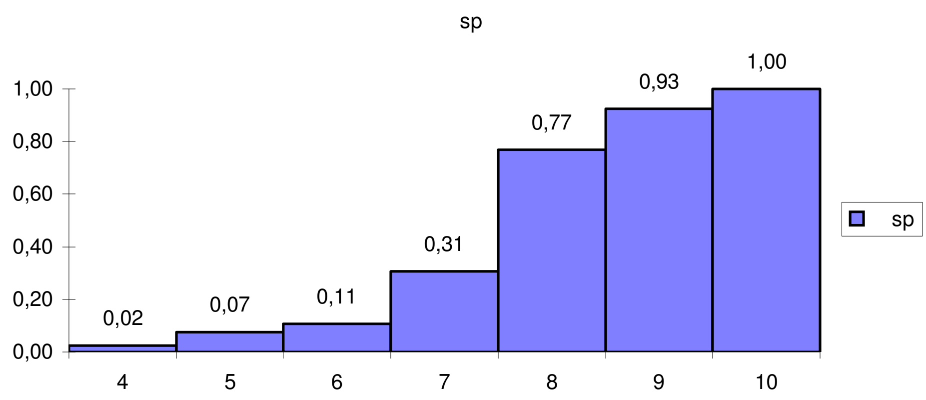
Esim. 5 Jos satunnaismuuttuja x kuvaa umpimähkään valitun japanilaisen pituutta, niin kertymäfunktion arvo F(173) antaa todennäköisyyden sille, että kyseinen pituus alle 173 cm. Edelleen F(169) – F(163) antaa todennäköisyyden sille, että umpimähkään valitun japanilaisen pituus on vähemmän kuin 169 cm, mutta enemmän kuin 163 cm.

Tässä japanilaisten pituutta kuvaavaa kertymäfunktiota sen paremmin kuin tiheysfunktiotakaan ei tunneta.

Esim. 6 Palataan uudelleen esimerkkiin 6.53, jossa tarkasteltiin erään lukion erään äidinkielen kurssin arvosanoja. Lasketaanpa arvosanoille vielä ns. summafrekvenssit ja suhteelliset summafrekvenssit:

x	f	s	p	sp
4	3	3	0,02	0,02
5	6	9	0,05	0,07
6	4	13	0,03	0,11
7	24	37	0,20	0,31
8	56	93	0,46	0,77
9	19	112	0,16	0,93
10	9	121	0,07	1,00
	121			

Piirretään suhteellisten frekvenssien summakäyrä ja katsellaan miten se on yhteydessä kertymäfunktioon aivan kuten suhteellisten frekvenssien kuvaaja on yhteydessä tiheysfunktioon.



Kun tulkitaan histogrammin pylväät todennäköisyysmassaksi, niin nähdään, että kukaan ei saa alle nelosen olevaa arvosanaa eikä toisaalta ole ketään, jonka kurssiarvosana olisi enemmän kuin 10. Edelleen nähdään helposti esim. todennäköisyys sille, että umpimähkään valitun oppilaan arvosana on enintään 8 eli $P(\underline{x} \le 8) = 0.77$. Tämä siis saadaan muiden paitsi 9 – ja 10 – pylväiden yhteenlaskettuna pinta-alana.

Jatkuvan jakauman odotusarvo ja keskihajonta määritellään määrätyn integraalin avulla. Kun integraalilaskentaa ei ole opiskeltu, tyydytään vain määritelmän esittelyyn:

****

MÄÄRITELMÄ 16

Olkoon $\underline{x}$ satunnaismuuttuja, jonka tiheysfunktio f on annettu. Tällöin

$$\boxed{\begin{aligned} E\underline{x} &= \int_{-\infty}^{\infty} f(x) \cdot x \, dx \\ D\underline{x} &= \sqrt{\int_{-\infty}^{\infty} f(x) \cdot (x - E\underline{x})^2 \, dx} \end{aligned}}$$

****