2A Jatkuvat satunnaismuuttujat

Tuntitehtävät

2A1 (Palkkajakauman malli.) Satunnaisesti valitun palkansaajan kuukausiansiota (eur) mallinnetaan satunnaismuuttujalla X, jolla on tiheysfunktio

$$f(x) = \begin{cases} \alpha c^\alpha x^{-\alpha-1}, & x > c, \\ 0, & \text{muuten}, \end{cases}$$

missä $\alpha = 1.6$ ja $c = 1500$.

(a) Laske X:n kertymäfunktio ja piirrä siitä kuva.

(b) Määritä X:n arvojoukko, eli miten pieniä ja miten suuria arvoja X voi saada?

(c) Laske todennäköisyys, että satunnaisesti valittu palkansaaja ansaitsee yli 15 000 eur/kk.

(d) Määritä sellainen ansiotaso z, että tasan 90% palkansaajista ansaitsee alle z euroa kuukaudessa.

Ratkaisu.

(a) Kertymäfunktion arvo pisteessä $x > c$ saadaan integraalina

$$F(x) = \int_{-\infty}^x f(t)\,dt = \int_c^x f(t)\,dt = \alpha c^\alpha \int_c^x t^{-\alpha-1}\,dt = \alpha c^\alpha \Big|_c^x (-\alpha^{-1} t^{-\alpha}) = 1 - c^\alpha x^{-\alpha}.$$

Pisteissä $x \le c$ kertymäfunktion arvo on $F(x) = 0$. Näin ollen

$$F(x) = \begin{cases} 1 - c^\alpha x^{-\alpha}, & x > c, \\ 0, & \text{muuten}. \end{cases}$$

(b) Koska $f(x) > 0$ kaikilla $x > c$ ja koska f on jatkuva välillä (c, ∞), todetaan että X saa mielivaltaisia arvoja välillä $(1500, \infty)$.

(c) Kysytty todennäköisyys on $1 - F(15000) = c^\alpha x^{-\alpha} = 1500^{1.6} \times 15000^{-1.6} \approx 2.51\%$.

(d) Merkitään $u = 0.90$ ja ratkaistaan yhtälö $F(z) = u$ eli $c^\alpha x^{-\alpha} = 1 - u$, josta saadaan

$$z = \frac{c}{(1-u)^{1/\alpha}} = \frac{1500}{0.10^{1/1.6}} \approx 6325.45.$$

Mallin mukaan siis 90% palkansaajista ansaitsee alle 6325.45 euroa ja 10% palkansaajista yli kyseisen tason. Lukua $z = 6325.45$ kutsutaan X:n jakauman 90% kvantiiliksi (tai fraktiiliksi).

Tässä tehtävässä tarkasteltu jakauma on nimeltään Pareto-jakauma muotoparametrina $\alpha = 1.6$ ja skaalaparametrina $c = 1500$.

2A2 (Molemmat myöhässä.) Tässä tehtävässä tutkimme yksinkertaista kahden muuttujan *jatkuvaa yhteisjakaumaa*. Monen muuttujan integraalilaskennasta tarvitsemme vain seuraavan yksinkertaisen huomion: Jos kahden muuttujan funktion arvo on jossain alueessa vakio, niin funktion integraali kyseisessä alueessa on alueen pinta-ala kertaa kyseinen vakio.

Ulla saapuu U minuuttia ja Venla V minuuttia myöhässä sopimastaan lounastapaamisesta, jonka piti alkaa klo 12:00. Kumpikin myöhästymisaika noudattaa jatkuvan välin $[0, 60]$ tasajakaumaa ja ne ovat toisistaan riippumattomat.

(a) Selvitä U:n jakauman tiheysfunktio f_U, V:n jakauman tiheysfunktio f_V sekä U:n ja V:n yhteisjakauman tiheysfunktio $f_{U,V}$.
Opastus: Lue luentomonisteen luvut 2.3–2.6. U ja V ovat jatkuvasti tasajakautuneita. Yhteisjakauman tiheyteen käytä lausetta 2.10. Ilmoita sekä funktion lauseke, että se millä muuttujien arvoilla se pätee. Muualla se on varmaan nolla?

(b) Laske todennäköisyys, että Ulla saapuu ennen klo 12:20.
Opastus: Käytä pelkästään U:n jakaumaa.

(c) Laske geometrian avulla todennäköisyys, että Ulla saapuu ennen klo 12:15 ja Venla saapuu klo 12.30 ja 12:45 välillä.
Opastus: Väritä $[0, 60]\times[0, 60]$ -neliöön yllämainittua tapahtumaa vastaava suorakulmio ja selvitä sen pinta-ala. Käytä sitten tehtävänannossa annettua vihjettä kahden muuttujan integraalista.

(d) Laske (c)-kohdan todennäköisyys ilman geometriaa, hyödyntämällä U:n ja V:n riippumattomuutta.
Opastus: Osaat laskea U:ta koskevien tapahtumien todennäköisyyksiä, ja V:tä koskevien tapahtumien todennäköisyyksiä. Käytä kaavaa (2.17).

(e) Laske todennäköisyys, että Ulla saapuu yli 30 minuuttia Venlan jälkeen. Käytä samaa piirtämistekniikkaa kuin (c)-kohdassa.
Opastus: Alueen päättely vaatii hiukan miettimistä. Kokeile kiinnittää V:lle jokin arvo ja tutki, millainen U on silloin oltava. Kun alue on selvillä ja piirretty, selvitä sen pinta-ala. Muistele, miten kolmion pinta-ala lasketaan.

Ratkaisu.

(a) U ja V noudattavat molemmat välin $[0, 60]$ tasajakaumaa, jolla on tiheysfunktio

$$f_U(x) = f_V(x) = \begin{cases} \frac{1}{60}, & 0 < x < 60, \\ 0, & \text{muuten.} \end{cases}$$

Koska U ja V ovat riippumattomat, voidaan yhteisjakauman tiheysfunktio kirjoittaa tulomuodossa $f_{U,V}(x,y) = f_U(x)f_V(y)$. Näin ollen yhteisjakaumalla on tiheysfunktio

$$f_{U,V}(x,y) \;=\; \begin{cases} \frac{1}{3600}, & 0 < x < 60,\ 0 < y < 60, \\ 0, & \text{muuten.} \end{cases}$$

Tämä yhteisjakauma on neliön $[0,60] \times [0,60]$ tasajakauma. Kannattaa huomata, että tiheys on neliön sisällä kaikkialla sama vakio, eli tasainen (josta jakauma saa nimensä).

(b) Kysytty todennäköixsyys on

$$\int_0^{20} f_U(x)\,dx \;=\; 20 \times \frac{1}{60} \;=\; \frac{1}{3}.$$

(c) Edellä laskettiin, että yhteisjakauman tiheys on kaikkialla neliössä $[0,60] \times [0,60]$ sama vakio 1/3600. Nyt tarkasteltavaa tapahtumaa vastaa neliö $A = [0,15] \times [30,45]$, sillä jos lukupari (U,V) on tässä neliössä, niin Ulla on saapunut viimeistään 15 min myöhässä ja Venla 30–45 min myöhässä. Neliön A pinta-ala on 225. Todennäköisyys, että (U,V) osuu tähän neliöön, on tiheysfunktion 1/3600 integraali neliön yli, eli (vihjeen mukaisesti)

$$225 \times (1/3600) = 225/3600 = 1/16 = 6.25\%.$$

Todennäköisyys on itse asiassa neliön A pinta-alan osuus koko neliön $[0,60] \times [0,60]$ pinta-alasta. Mieti miksi. (Vihje: Tiheysfunktio on vakio.)

(d) Riipumattomuuden perusteella

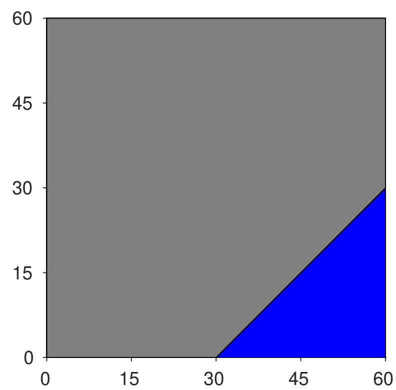
$$\mathrm{P}(U < 15,\ 30 < V < 45) \;=\; \mathrm{P}(U < 15)\ \mathrm{P}(30 < V < 45) \;=\; \frac{1}{4} \times \frac{1}{4} \;=\; \frac{1}{16}.$$

Tämä tulos on sama kuin (c)-kohdassa, niin kuin pitääkin.

(e) Tarkasteltava tapahtuma voidaan kirjoittaa muodossa $\{(U,V) \in A\}$, missä

$$A = \{(x,y) : x \geq y + 30\}.$$

Aluetta voi päätellä esim. seuraavasti. Korkeudella $y = 0$ alueeseen kuuluvat pisteet, joissa $x \geq 30$. Korkeudella $y = 10$ ne pisteet, joissa $x \geq 40$, ja niin edelleen. Itse asiassa aluetta A rajoittaa suora $x = y + 30$ (ja lisäksi (U,V):n arvojoukkoa kuvaavan neliön rajat) ja alue on suorasta oikealle. Kuvasta nähdään, että alue on muodoltaan kolmio.



Kolmion A pinta-ala on $\frac{1}{2} \times 30^2 = 450$. Näin ollen kysytty todennäköisyys on

$$\mathrm{P}(U \geq V + 30) \quad = \quad \frac{450}{3600} = 1/8 = 12.5\%.$$

Kotitehtävät

2A3 (Satelliitin toiminta) Maata kiertävässä satelliitissa on anturi, jonka toiminta-aika X (yksikkönä vuosi) noudattaa eksponenttijakaumaa *taajuusparametrilla* $\lambda = 0.5$ (yksikkönä 1/vuosi).

(a) Etsi eksponenttijakauman tiheys- ja kertymäfunktio luentomonisteesta tai luentokalvoilta (1B). Tarkista kertymäfunktiota derivoimalla, että siten saadaan tiheysfunktio.

(b) Käyttäen kertymäfunktiota, laske todennäköisyydet tapahtumille $A = \{X \leq 1\}$, $B = \{X > 4\}$, ja $C = \{4 < X \leq 5\}$. Ilmoita tulokset ainakin kuudella desimaalilla. Selitä sanoin, mitä nämä kolme tapahtumaa tarkoittavat.

(c) Käyttäen b-kohdan tuloksia ja ehdollisen todennäköisyyden määritelmää, laske todennäköisyys $\mathrm{P}(C \mid B)$. Selitä sanoin mitä tässä laskettiin. Vertaa tapahtuman A todennäköisyyteen.

(d) Tarkastellaan hyvin lyhyttä aikaväliä $h = 0.01$ (vuotta). Jos anturi on kestänyt johonkin ajanhetkeen asti, mikä on todennäköisyys, että se hajoaa seuraavan h vuoden aikana? Vertaa tätä todennäköisyyttä lukuun λh ja selitä, miksi λ:aa kutsutaan eksponenttijakauman taajuusparametriksi (engl. rate parameter).

Arviointiohje. 0.5 pistettä per kohta. Tehtävän kokonaispistemäärä pyöristetään ylöspäin lähimpään kokonaislukuun.

Ratkaisu.

(a) Kertymäfunktio on pisteissä $t > 0$

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

ja nolla muualla (eli pisteissä $t \leq 0$). Derivoimalla tätä t:n suhteen saadaan funktio

$$F'(t) = \lambda e^{-\lambda t},$$

kun $t > 0$, ja nolla muualla, eli se tiheysfunktio joka pitikin saada.

(b)

$$
\begin{aligned}
\mathrm{P}(A) &= \mathrm{P}(X \le 1) = F(1) = 1 - e^{-0.5\cdot 1} \approx 0.393470 \\
\mathrm{P}(B) &= \mathrm{P}(X > 4) = 1 - F(4) = e^{-0.5\cdot 4} \approx 0.135335 \\
\mathrm{P}(C) &= \mathrm{P}(4 < X \le 5) = F(5) - F(4) = (1 - e^{-0.5\cdot 5}) - (1 - e^{-0.5\cdot 4}) \approx 0.053250
\end{aligned}
$$

Selitys:

- A = anturi hajoaa ensimmäisen vuoden aikana
- B = anturi toimii ainakin neljä vuotta
- C = anturi toimii jonkin ajan välillä 4–5 vuotta; toisin sanoen, kun alkuhetkestä on kulunut neljä vuotta, anturi toimii vielä, mutta hajoaa sitten seuraavan vuoden aikana.

(c) Huomataan, että $C \subseteq B$, joten $\mathrm{P}(C \cap B) = \mathrm{P}(C)$. Ehdollisen todennäköisyyden määritelmästä

$$\mathrm{P}(C \mid B) = \frac{\mathrm{P}(C \cap B)}{\mathrm{P}(B)} \approx \frac{0.053250}{0.135335} \approx 0.393470.$$

Riippuen välitulosten tarkkuudesta viimeiset desimaalisi voivat vähän poiketa. Tulos 0.393470 laskettiin täydellä konetarkkuudella.

Tässä laskettiin todennäköisyys, että anturi hajoaa viidennen vuoden aikana, **jos** se on toiminnassa vielä neljän vuoden jälkeen eli viidennen vuoden alkaessa.

Havaitaan (numeerisesti), että $\mathrm{P}(C|B) = \mathrm{P}(A)$. Toisin sanoen, anturilla oli noin 40% todennäköisyys hajota *ensimmäisen* vuoden aikana; mutta **jos** se on kestänyt neljä vuotta, sillä on *siinä tapauksessa* tämä sama 40% todennäköisyys hajota *seuraavan* vuoden aikana.

Tätä ominaisuutta kutsutaan eksponenttijakauman *muistittomuudeksi*. Toisin sanoen, riippumatta siitä, kuinka kauan anturi on kestänyt, sillä aina sama todennäköisyys hajota *seuraavan* vuoden (tms. ajanjakson) aikana. Tämä on *eksponenttijakauman erikoisominaisuus.*

Voidaan hyvin kuvitella tilanteita, joissa jonkin laitteen elinikä *ei noudata* eksponenttijakaumaa, jos esimerkiksi vanhalla laitteella onkin suurempi (tai pienempi) todennäköisyys hajota kuin uudella.

(d) Muistittomuusominaisuuden perusteella sillä ei ole merkitystä, kauanko anturi on kestänyt. Sen todennäköisyys hajota seuraavan h vuoden aikana on sama kuin oli sen todennäköisyys ensimmäisten $h = 0.01$ vuoden aikana, eli

$$\mathrm{P}(X \le h) = 1 - e^{-\lambda h} = 1 - e^{-0.005} \approx 0.004988.$$

Tämä on numeerisesti lähellä lukua $\lambda h = 0.5 \cdot 0.01 = 0.005$.

Selitys: Kullakin lyhyellä aikavälillä (jos anturi vielä toimii aikavälin alkaessa) hajoamistapahtumia on odotettavissa taajuudella 0.5 hajoamista vuodessa.

2A4 (Sumea logiikka.) Sumeassa logiikassa lauseilla on 0-1 totuusarvon sijasta reaaliarvoinen totuusarvo välillä $[0,1]$. Tutkijat mallintavat erään lauseen totuusarvoa satunnaismuuttujalla X, jolla on tiheysfunktio

$$f(x) = \begin{cases} cx(1-x), & 0 \le x \le 1, \\ 0, & \text{muuten.} \end{cases}$$

(a) Etsi vakion c arvo.

(b) Laske X:n kertymäfunktio ja piirrä siitä kuva.

(c) Laske todennäköisyys, että totuusarvo on vähintään 0.75.

(d) Määritä jakauman moodi eli piste, jossa tiheysfunktio $f(x)$ saavuttaa suurimman arvonsa.

Arviointiohje. 0.5 pistettä per kohta. Jos ei ole osannut ratkaista vakion c arvoa, mutta kohdat (b)-(d) ovat kuitenkin oikein (vakion arvoa vaille), saa kustakin silti 0.5 pistettä. (b)-kohdassa ei tarvitse erikseen merkata, millä arvoilla kertymäfunktio saa arvon 0 tai 1. Tehtävän kokonaispistemäärä pyöristetään ylöspäin lähimpään kokonaislukuun.

Ratkaisu.

(a) Tiheysfunktion tulee integroitua ykköseksi yli määrittelyalueensa, eli nyt tulee siis olla,

$$1 = \int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 cx(1-x)\,dx.$$

Koska,

$$\int_0^1 x(1-x)\,dx = \int_0^1 x - x^2\,dx = \Big|_0^1 (\frac{1}{2}x^2 - \frac{1}{3}x^3) = \frac{1}{2} - \frac{1}{3} = \frac{1}{6},$$

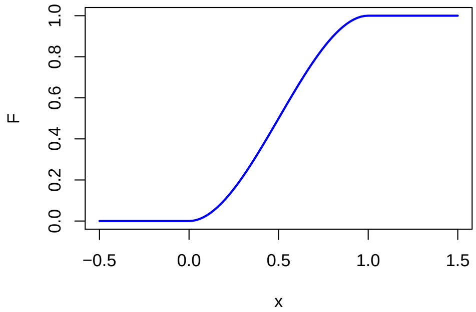
tulee siis olla $c = 6$.

(b) Koska kaikki jakauman todennäköisyysmassa on välillä $[0,1]$, tulee olla $F(x) = 0$, kun $x < 0$, ja $F(x) = 1$, kun $x > 1$. Kertymäfunktion arvo pisteessä $0 \le x \le 1$ saadaan integraalina

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_0^x f(t)\,dt = 6\int_0^x t(1-t)\,dt = 6\Big|_0^x (\frac{1}{2}t^2 - \frac{1}{3}t^3) = x^2(3-2x).$$

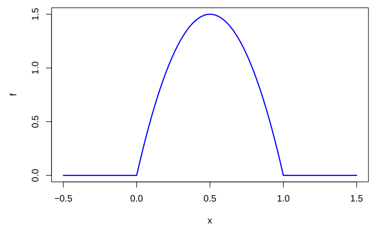
Näin ollen kertymäfunktio ja sen kuvaaja ovat

$$F(x) = \begin{cases} 0, & x < 0, \\ x^2(3-2x), & 0 \le x \le 1, \\ 1, & x > 1. \end{cases}$$



(c) Kysytty todennäköisyys on $P(X \geq 0.75) = 1 - P(X < 0.75) = 1 - F(0.75) = 1 - (0.75)^2 \cdot (3 - 2 \cdot 0.75) \approx 15.6\%$.

(d) Välin $[0, 1]$ ulkopuolella tiheysfunktio on nolla, joten maksimikohdan tulee löytyä tämän välin sisältä. Välillä $[0, 1]$ tiheysfunktio on alaspäin aukeava paraabeli $f(x) = -6x^2 + 6x$, joka saa välin päätepisteissä arvon nolla, joten maksimikohta löytyy siis paraabelin derivaatan nollakohdasta.



Välillä $[0, 1]$,

$$f'(x) = -12x + 6,$$

joka menee nollaan kun $x = 1/2$, joten maksimi saavutetaan tässä pisteessä. Löydetty piste on satunnaismuuttujan X *moodi*, eli tietyssä mielessä kaikkein todennäköisin satunnaismuutujan arvo (kaikista pisteistä sen ympäristössä on eniten todennäköisyysmassaa).