TILASTOMATEMATIIKKA

Harjoitus 2 ratkaisut, kevät 2018

Harjoituksen teemoja ovat:

(i) Kertymäfunktio, pistetodennäköisyysfunktio ja tiheysfunktio

(ii) Diskreetti satunnaismuuttuja

(iii) Jatkuva satunnaismuuttuja

(iv) Tyypillisimmät jakaumat, jotka kantavat omaa nimeä

Tehtävien laskemista voi rytmittää esimerkiksi niin, että alkuviikosta käy läpi tehtäviä 1-5, joista tehtävä 1 liittyy satunnaismuuttujien luokitteluun ja loput tehtävät liittyvät diskreetteihin satunnaismuuttujiin. Tehtävät 6-9 liittyvät jatkuviin satunnaismuuttujiin. Toki myös muunlainen rytmittäminen on mahdollista. **Minimivaatimuksena laskuharjoituspisteille** ovat **tehtävät 2, 4 ja 5** alkuviikosta ja **tehtävät 6 ja 8** loppuviikosta.

1. Ilmoita kussakin seuraavista tapauksista, onko kyse jatkuvasta vai diskreetistä satunnaismuuttujasta. Määrää myös satunnaismuuttujan arvojoukko, mikäli se on mahdollista. Voidaanko mallintamisessa käyttää hyväksi jotain tunnettua jakaumaa?

   a) Vikojen lukumäärä neliömetrillä satunnaisesti valitussa paperirullassa.
   b) Kemikaalin konsentraatio liuoksessa.
   c) Liian pitkien pulttien osuus satunnaisesti valitussa pultteja sisältävässä laatikossa.
   d) Virheiden lukumäärä 1000 satunnaisesti valitussa rivissä ohjelmointikoodia.
   e) Satunnaisesti valitun metallilevyn murtolujuus.
   f) Elektronisen komponentin elinikä.

Ratkaisu:

a) Vikojen lukumäärä on ei-negativiinen kokonaisluku, sillä vikoja voi olla nolla, yksi, kaksi,... kappaletta, joten kyseessä on *diskreetti sm.*
Periaatteessa vikojen lukumäärälle X ei ole ylärajaa, joten arvojoukokoksi voidaan ottaa $S_X = \mathbb{N}_0 = \{0,1,2,\ldots\}$.
Vikojen lukumäärä jotakin yksikköä (tässä tapauksessa neliömetriä) kohden on tyypillinen Poisson-jakauman sovelluskohde, joten muuttujan X voidaan olettaa noudattavan Poisson-jakaumaa. On syytä korostaa, että Poisson-jakauman käyttäminen tarkoittaa matemaattisen mallin muodostamista ongelmalle. Todellinen vikojen lukumäärä voi periaatteessa olla mitä hyvänsä. Jos kuitenkin Poisson-jakauman käyttö antaa havaintojen kanssa sopusoinnussa olevia tuloksia, on sen käyttö perusteltua.

b) Vastaus riippuu konsentraation X määritelmästä. Jos konsentraatiolla tarkoitetaan liuenneen aineen massan suhdetta liuoksen tilavuuteen, voi konsentraatio saada periaatteessa minkä tahansa ei-negatiivisen reaalilukuarvon, jolloin arvojoukoksi voidaan valita $S_X = [0,\infty[$. Tokihan käytännössä konsentraatiolle täytyy olla jokin yläraja, mutta koska ylärajaa ei tiedetä, voidaan arvojoukoksi ottaa rajoittamaton väli $[0,\infty[$. Tällöin kyseessä on *jatkuva satunnaismuuttuja.*
Konsentraation voidaan olettaa noudattavan normaalijakaumaa, vaikka normaalijakauma saa kaikki reaalilukuarvot, kun taas konsentraatio ei voi olla negatiivinen. Jälleen normaalijakauman valinnalla muodostetaan matemaattinen malli, joka enemmän tai vähemmän kuvaa todellista tilannetta.

c) Koska laatikon kokoa N ei tunneta, voi pulttien osuus saada periaatteessa minkä tahansa ei-negatiivisen rationaalilukuarvon, joten pulttien osuutta kuvaavan sm:n X arvojoukoksi voidaan valita

$$S_X = \left\{ \frac{m}{n} \,:\, 0 \leq m \in \mathbb{Z}, 0 < n \in \mathbb{Z} \right\}.$$

Tällöin kyseessä on *diskreetti sm.*
Jakaumasta emme voi periaatteessa sanoa mitään. Mutta jos laatikon koko tunnetaan, niin viallisten pulttien lukumäärä laatikossa noudattaa binomijakaumaa.

d) Koska virheiden lukumäärä voi olla periaatteessa mikä tahansa ei-negatiivinen kokonaisluku, voidaan tässäkin käyttää Poisson-jakaumaa, jolloin $S_X = \mathbb{N}_0$ ja kyseessä on *diskreetti sm.* Tarkkaan ottaen Poisson-jakauman käyttö ei ole oikein, sillä yhdellä rivillä ei voi olla loputtomasti merkkejä. Toisakseen, koska eri riveillä voi olla eri määrä merkkejä, emme voi käyttää binomijakaumaa. Toisaalta, koska rivejä on kohtuullisen paljon, kannattaisi käytännössä käyttää Poisson-jakaumaa, vaikka kullakin rivillä olisi sama määrä merkkejä ja siten binomijakauma olisi oikea jakaumamalli.

e) Satunnaisesti valitun metallilevyn murtolujuus voi olla periaatteessa mikä tahansa ei-negatiivinen reaaliluku, jolloin kyseessä olisi *jatkuva satunnaismuuttuja.* Kuten b)-kohdassa, jakaumamallina voisimme käyttää normaalijakaumaa, vaikka se ei sovikaan yhteen (teoreettisen) arvojoukon $[0,\infty[$ kanssa.

f) Elinikä X voi olla periaatteessa mikä tahansa ei-negatiivinen reaaliluku, joten arvojoukoksi voidaan valita $S_X = [0,\infty[$ ja siten kyseessä on *jatkuva sm.* Yksinkertaisin eliniän jakaumamalli on eksponenttijakauma.

2. Olkoot X ja Y riippumattomia satunnaismuuttujia, jotka noudattavat diskreettiä tasajakaumaa

Arvo	1	2	3	4
Todennäköisyys	0.25	0.25	0.25	0.25

a) Laske muuttujien X ja Y kertymäfunktio.

b) Esitä satunnaismuuttujien riippumattomuuden määritelmä (löytyy esimerkiksi luentomonisteesta). Laske todennäköisyys

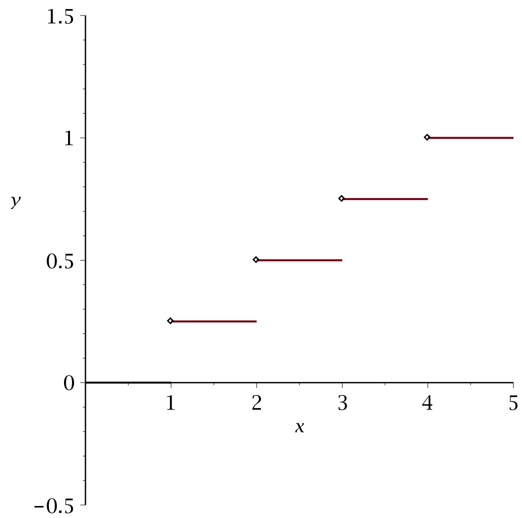
$$P(X \leq 2 \,\mathrm{ja}\, Y \leq 3) = P(\{X \leq 2\} \cap \{Y \leq 3\}).$$

c) Mitä arvoja saa muuttujien X ja Y aritmeettinen keskiarvo $Z = \frac{1}{2}(X+Y)$? Määrää muuttujan Z jakauma.

Ratkaisu:

a) Muuttujat X ja Y ovat samalla tavalla jakautuneita satunnaismuuttujia, joiden kertymäfunktio on porrasfunktio

$$F(x) = \begin{cases} 0, & x < 1, \\ 0.25, & 1 \leq x < 2, \\ 0.5, & 2 \leq x < 3, \\ 0.75, & 3 \leq x < 4, \\ 1, & x \geq 4. \end{cases}$$



Kuva 1: Kertymäfunktion kuvaaja

b) Satunnaismuuttujien riippumattomuuden määritelmä löytyy luentomonisteen sivulta 28 (Määritelmä 5). Lyhykäisesti: satunnaismuuttujat X ja Y ovat riippumattomia, jos

$$P(\{X \leq x\} \cap \{Y \leq y\}) = P(X \leq x)P(Y \leq y)$$

kaikilla $x,y \in \mathbb{R}$.
Koska X ja Y ovat nyt riippumattomia, niin

$$P(X \leq 2 \,\mathrm{ja}\, Y \leq 3) = P(X \leq 2)P(Y \leq 3) = F(2)F(3) = 0.5 \cdot 0.75 = 0.375.$$

c) Selvitetään ensin aritmeettisen keskiarvon arvojoukko. Koska X ja Y saavat arvot 1,2,3,4, niin aritmeettisen keskiarvon arvojoukoksi saadaan

$$S_Z = \{1,1.5,2,2.5,3,3.5,4\}.$$

Mitä ovat näitä vastaavat pistetodennäköisyydet? Minimiarvoa $Z = 1$ vastaa ainoastaan yksi X:n ja Y:n arvon kombinaatio $X = 1$ ja $Y = 1$, joten

$$P(Z = 1) = P(X = 1\,\text{ja}\,Y = 1) \overset{\text{riipp.}}{=} P(X = 1)P(Y = 1) = 0.25 \cdot 0.25 = 0.0625.$$

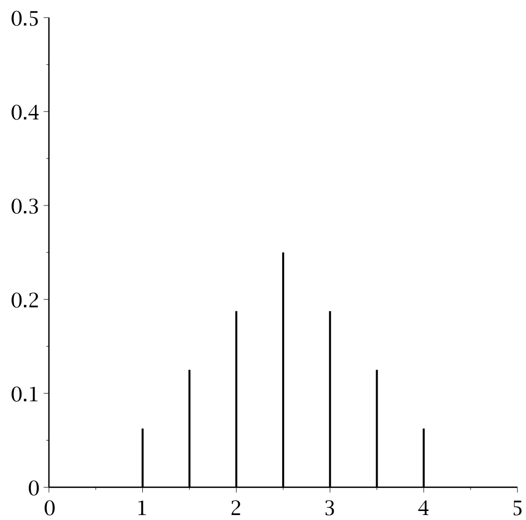
Arvolle $Z = 1.5$ taasen on 2 mahdollisuutta: joko $X = 1$ ja $Y = 2$ tai $X = 2$ ja $Y = 1$. Koska vaihtoehdot ovat toisensa poissulkevia, pistetodennäköisyys on näin ollen

$$\begin{aligned} P(Z = 1.5) &= P\left((\{X = 1\} \cap \{Y = 2\}) \cup (\{X = 2\} \cap \{Y = 1\})\right) \\ &\overset{\text{erillisiä}}{=} P(\{X = 1\} \cap \{Y = 2\}) + P(\{X = 2\} \cap \{Y = 1\}) \\ &\overset{\text{riipp.}}{=} P(X = 1)P(Y = 2) + P(X = 2)P(Y = 1) \\ &= 0.25 \cdot 0.25 + 0.25 \cdot 0.25 = 0.125. \end{aligned}$$

Vastaavalla tavalla voidaan laskea muut pistetodennäköisyydet. Vaihtoehtojen lukumäärä kasvaa arvoon $Z = 2.5$ saakka ja alkaa sitten taas vähentyä (symmetrisesti) arvoon $Z = 4$ saakka. Jakaumaksi (Z:n pistetodennäköisyydeksi) saadaan

z	1	1.5	2	2.5	3	3.5	4
$P(Z = z)$	0.0625	0.1250	0.1875	0.2500	0.1875	0.1250	0.0625

Jakauma ei siis enää ole tasainen. Alla on pistetodennäköisyyden kuvaaja, johon selvyyden vuoksi on merkitty pisteiden sijaan arvopisteiden kohdalle pystyviivat, joiden korkeudet vastaavat arvojen pistetodennäköisyyksiä.



Kuva 2: Muuttujan Z pistetodennäköisyysfunktio

3. Jokainen muropaketti sisältää lapsia ilahduttavan muovisen eläinhahmon. Oletetaan, että hahmoja on 6 ja että kukin hahmo on yhtä todennäköinen. Olkoon yksi eläimistä tiikeri. Iiro Ikiteekkari ostaa lapselleen 10 muropakettia. Millä todennäköisyydellä

   a) lapsi saa vähintään yhden tiikerin?

   b) lapsi saa ensimmäisen tiikerin viimeisestä paketista?

Ratkaisu:

a) Merkitään A_i = "lapsi saa tiikerin i:nnestä paketista", jolloin $P(A_i) = \frac{1}{6}$ ja $P(\overline{A_i}) = \frac{5}{6}$.

$$\begin{aligned}&P(\text{“lapsi saa vähintään yhden tiikerin”}) = 1 - P(\text{“lapsi ei saa yhtään tiikeriä”})\\ &= 1 - P(\overline{A_1})P(\overline{A_2})\cdots P(\overline{A_{10}})\\ &= 1 - \left(\frac{5}{6}\right)^{10} \simeq \underline{0{,}838}\end{aligned}$$

Koska kyseessä on toistokoe, voidaan myös ratkaista tehtävä tarkastelemalla binomijakautunutta satunnaismuuttujaa X, joka ilmoittaa löytyneiden tiikereiden lukumäärän kymmenessä riippumattomassa toistossa. $X \sim \mathrm{Bin}(n{,}p)$, $n = 10$, $p = \frac{1}{6}$. Tällöin

$$\begin{aligned}&P(\text{“lapsi saa vähintään yhden tiikerin”}) = P(X \geq 1)\\ &= 1 - P(X = 0)\\ &= 1 - \binom{n}{0}p^0(1-p)^{10} = 1 - (1-p)^{10} \simeq \underline{0{,}838}\end{aligned}$$

b) Todennäköisyys, että lapsi saa tiikerin vasta viimeisestä paketista on

$$\begin{aligned}P(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_9} \cup A_{10}) &= P(\overline{A_i})^9 P(A_{10})\\ &= \left(\frac{5}{6}\right)^9 \frac{1}{6}\\ &\simeq \underline{0{,}0323}\end{aligned}$$

Jos tässä merkitään satunnaismuuttujalla Y sitä, monennellako kerralla saadaan ensimmäinen tiikeri, niin on Y geometrisesti jakautunut, $Y \sim \mathrm{Geo}(p)$. Tällöin

$$P(Y = k) = (1-p)^{k-1} \cdot p, k = 1{,}2{,}...$$

ja kysytty todennäköisyys on $P(Y = 10) = (1-p)^9 p \simeq \underline{0{,}0323}$.

4. Erittäin laaja havupuualue oli neulaskadon takia jäämässä tuotantotavoitteestaan. Tuottoennusteen tekemiseksi tietty osa metsää jaettiin hehtaarin suuruisiin ruutuihin ja laskettiin neulaskadon vaivaamien puiden lukumäärä ruutua kohti. Laskennan tuloksena havaittiin neulaskadon vaivaamien puiden lukumäärä/ruutu olevan Poisson-jakautuneen ja sellaisten ruutujen, joissa neulaskadon vaivaamia puita ei ollut lainkaan, osuuden olevan 7 %.

   a) Mikä oli oli keskimääräinen neulaskadon vaivaamien puiden lukumäärä ruutua kohti?

   b) Millä todennäköisyydellä ruudussa esiintyi ainakin 2 neulaskadon vaivaamaa puuta?

Ratkaisu: Olkoon X = ”neulaskadon vaivaamien puiden lukumäärä/ha”. Tehtävänannon mukaan $X \sim \mathrm{Poi}(a)$, missä parametria a ei vielä tiedetä.

a) Poisson-jakauman parametri a ilmoittaa satunnaismuuttujan keskimääräisen arvon. Tehtävänannon mukaan

$$P(X = 0) = \mathrm{e}^{-a} = 0.07 \Rightarrow a = -\ln 0.07 \approx 2.659,$$

neulaskadon vaivaamia puita on keskimäärin 2.7 kappaletta hehtaaria kohti.

b) Kysytään todennäköisyyttä $P(X \geq 2)$, joka ainakin funktiolaskimella on helpoin laskea komplementin kautta

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - \mathrm{e}^{\ln 0.07}(1 - \ln 0.07) \approx 74\%.$$

5. Valmistajan ilmoituksen mukaan halvoista elektronisista komponenteista AA5 keskimäärin 5 % ei täytä spesifikaatioita. Ostaja osti AA5:ttä monen tuhannen kappaleen erissä ja teki seuraavan päätöksentekomallin erän hyväksymiseksi vastaanottotarkastuksen yhteydessä: Erästä valittiin satunnaisesti 10 komponenttia, jotka testattiin. Jos kaikki testatut komponentit toteuttivat spesifikaatiot, niin erä hyväksyttiin. Jos ainakin 2 testatuista komponenteista ei toteuttanut spesifikaatioita, erä hylättiin. Jos testatuista täsmälleen yksi ei toteuttanut spesifikaatioita, otettiin uusi 10 yksilön satunnainen näyte, ja erä hyväksyttiin, jos uudessa näytteessä kaikki komponentit täyttävät spesifikaatiot. Jos valmistajan väite pitää paikkansa, niin millä todennäköisyydellä

   a) erä hyväksytään ensimmäisen näytteen perusteella?

   b) lopullinen päätös tehdään eli erä hylätään tai hyväksytään ensimmäisen näytteen perusteella?

   c) erä hyväksytään?

Ratkaisu: Olkoon

$$X_i = \text{"spesifikaation täyttävien komponenttien lukumäärä näytteessä } i\text{"}, \quad i = 1{,}2.$$

Koska komponentit toimitetaan monen tuhannen kappaleen erissä, voidaan olettaa, että $X_i \sim \mathrm{Bin}(10{,}0.95)$, $i = 1{,}2$, ja että X_1 ja X_2 ovat riippumattomia.

a) Erä hyväksytään ensimmäisen näytteen perusteella, kun $X_1 = 10$, joten kysytty todennäköisyys on

$$P(X_1 = 10) = \binom{10}{10} 0.95^{10} \cdot 0.05^{10-10} \approx 60\%.$$

b) Lopullinen päätös tehdään ensimmäisen näytteen perusteella täsmälleen silloin, kun $X_1 \neq 9$, joten kysytty todennäköisyys on

$$P(X_1 \neq 9) = 1 - P(X = 9) = 1 - \binom{10}{9} 0.95^{9} \cdot 0.05^{10-9} \approx 68\%.$$

c) Erä hyväksytään joko ensimmäisen tai toisen näytteen perusteella. Kysytään tapahtuman $\{X_1 = 10\} \cup (\{X_1 = 9\} \cap \{X_2 = 10\})$ todennäköisyyttä. Koska yhdisteen tapahtumat ovat erillisiä sekä X_1 ja X_2 ovat riippumattomia, on kysytty todennäköisyys

$$p = \binom{10}{10} 0.95^{10} \cdot 0.05^{0} + \binom{10}{9} 0.95^{9} \cdot 0.05^{1} \cdot \binom{10}{10} 0.95^{10} \cdot 0.05^{0} \approx 79\%.$$

6. Oletetaan, että Suomen korkeakouluopiskelijoiden älykkyysosamäärä ÄO on jakautunut normaalijakauman mukaan parametreilla $\mu = 115$ ja $\sigma = 10$.

a) Poimitaan umpimähkään tästä joukosta yksi opiskelija. Määrää todennäköisyydet seuraaville tapahtumille

i) ÄO ≤ 100,	iii) $105 <$ ÄO ≤ 135,
ii) ÄO > 150,	iv) ÄO $= 120$.

b) Määrää raja a, jonka yläpuolella korkeakouluopiskelijoiden ÄO on 5 %:n todennäköisyydellä.

Ratkaisu: Merkitään X = "Suomen korkeakouluopiskelijoiden ÄO".
$X \sim \mathrm{N}(\mu,\sigma^2)$, missä $\mu = 115$ ja $\sigma = 10$. Olkoon $Z = \frac{X-\mu}{\sigma}$, jolloin $Z \sim \mathrm{N}(0,1)$.

a) Umpimähkään valitulle opiskelijalle

i) Ehtoa ÄO ≤ 100 vastaa nyt $X \leq 100$, joten todennäköisyys on:

$$\begin{aligned} P(X \leq 100) &= P(\frac{X-115}{10} < \frac{100-115}{10}) \\ &= P(Z < -1{,}5) \\ &= \Phi(-1{,}5) = 1 - \Phi(1{,}5) \\ &\simeq 1 - 0{,}9332 \quad (\Phi\text{:n likiarvot saadaan taulukosta}) \\ &= \underline{0{,}0668}. \end{aligned}$$

ii) $P(X > 150) = P(Z > 3{,}5) = \Phi(3{,}5) = 1 - \Phi(3{,}5) \simeq 1 - 0.9998 = \underline{0{,}0002}$.

iii) $P(105 < X \leq 135) = P(-1 < Z \leq 2) = P(Z \leq 2) - P(Z < -1)$
$= \Phi(2) - \Phi(-1) = \Phi(2) - (1 - \Phi(1)) \simeq 0{,}9772 - (1 - 0{,}8413) = \underline{0{,}8185}$

iv) $P(X = 120) = \underline{0}$

b) Määrätään $b = \frac{a-\mu}{\sigma}$ siten ,että $P(Z > b) = 0{,}05$. Ratkaistaan siitä haluttu a:n arvo.

$$\begin{aligned} P(Z > b) &= 1 - P(Z < b) = 0{,}05 \\ P(Z < b) &= \Phi(b) = 0{,}95 \\ b &= 1{,}645 \quad \text{luetaan taulukosta} \\ \frac{a-115}{10} &= 1{,}645 \\ a &= 115 + 16{,}45 = \underline{131{,}45} \end{aligned}$$

7. Eloonjäämisfunktio S (*survival function*) ilmoittaa todennäköisyyden, että henkilö on elossa tietyn ajan kuluttua. Oletetaan, että syöpäpotilaan eloonjäämisfunktio on muotoa $S(t) = 1 - F_X(t)$, missä t on syöpädiagnoosista kulunut aika vuosina ja X on elinajan ilmoittava satunnaismuuttuja, jonka oletetaan noudattavan Weibull-jakaumaa parametreilla $\alpha = 0.98$ ja $\beta = 0.30$.

   a) Esitä Weibull-jakauman tiheysfunktion lauseke (löytyy esimerkiksi luentomonisteesta).

   b) Laske todennäköisyys, että henkilö on elossa 5 vuotta syöpädiagnoosin jälkeen.

   c) Mikäli henkilö on elossa 5 vuotta syöpädiagnoosin jälkeen, niin millä todennäköisyydellä hän elää vielä 5 vuotta?

Ratkaisu:

a) Weibull-jakauman tiheysfunktion lauseke löytyy esimerkiksi luentomonisteen sivulta 41 tai vaikkapa Wikipedia-artikkelista

https://en.wikipedia.org/wiki/Weibull_distribution

Tiheysfunktio on

$$f(t) = \begin{cases} \alpha\beta t^{\beta-1}e^{-\alpha t^{\beta}}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Huomaa, että Wikipediassa ja luentomonisteessa esitetty tiheysfunktion määritelmä poikkevat toisistaan parametrien osalta. Wikipediassa esitetyn tiheysfunktion parametrit ovat λ ja k, joiden yhteys yllä esitettyihin parametreihin α ja β on

$$\beta = k \quad \text{ja} \quad \alpha = \frac{1}{\lambda^{\beta}}.$$

Vaikka tiheysfunktion lauseke saattaa näyttää aluksi kummalliselta, huomioi, että eksponenttifunktion edessä on merkkiä vaille eksponentin derivaatta, joten tiheysfunktio on helppo integroida ja kertymäfunktioksi saadaan suoraan

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t = \int_{0}^{x} f(t)\mathrm{d}t = \begin{cases} 1 - e^{-\alpha x^{\beta}}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Weibull-jakauma on eksponenttijakauman yleistys, joka ottaa huomioon, että komponentin (tässä tapauksessa ihmisen) vioittumistodennäköisyys voi muuttua iän myötä.

b) Kysytään todennäköisyyttä $P(X > 5)$, joka voidaan laskea kertymäfunktion avulla seuraavasti

$$P(X > 5) = 1 - P(X \leq 5) = 1 - F(5) = S(5) = e^{-0.98 \cdot 5^{0.30}} \approx 0.20 = 20\%.$$

c) Nyt kysytään ehdollista todennäköisyyttä

$$P(X > 10 | X > 5) = \frac{P(X > 10 \,\text{ja}\, X > 5)}{P(X > 5)} = \frac{P(X > 10)}{P(X > 5)} = \frac{S(10)}{S(5)} \approx 0.69 = 69\%.$$

8. Tietokanta oli vuorokaudessa käytettävissä 24 h. Kehnosta tietoliikenneprotokollasta johtuen tietokantaan yhteyden saamisen keskimääräinen odotusaika oli 5 sekuntia. Pitkäaikaisen kokemuksen perusteella odotusajan oli todettu noudattavan eksponenttijakaumaa.

   a) Esitä eksponenttijakauman tiheysfunktion lauseke (löytyy esimerkiksi luentomonisteesta).

   b) Millä todennäköisyydellä yhteyden saantiin kuluu ainakin 20 sekuntia?

   c) Jos yhteyden saamista on odotettu jo minuutti, niin millä todennäköisyydellä yhteys saadaan seuraavan 20 sekunnin aikana?

   d) Vuorokauden aikana tietokantaan otettiin 10 yhteyttä toisistaan riippumatta. Millä todennäköisyydellä jokaisen yhteyden odottaminen kesti korkeintaan 5 sekuntia?

Ratkaisu: Olkoon X = "yhteyden saantiin kuluva aika". Tehtävänannon mukaan $X \sim \mathrm{Exp}(a)$. Koska eksponenttijakaumassa parametrin a käänteisluku ilmoittaa muuttujan keskimääräisen arvon, niin $\frac{1}{a} = 5 \Leftrightarrow a = \frac{1}{5}$.

a) Tiheysfunktio on

$$f_X(x) = \begin{cases} a\mathrm{e}^{-ax}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

b) Kysytty todennäköisyys on

$$P(X \geq 20) = 1 - P(X < 20) \overset{X \text{ jatk. sm.}}{=} 1 - P(X \leq 20) = 1 - F_X(20) = \mathrm{e}^{-4} \approx 1.8\%.$$

Kertymäfunktion käytön sijaan voidaan myös integroida

$$P(X \geq 20) = \int_{20}^{\infty} f_X(x)\mathrm{d}x,$$

joka luonnollisesti antaa saman lopputuloksen (edellyttäen, että integroinnin on tehnyt oikein).

c) Kysytään todennäköisyyttä

$$\begin{aligned} P(X \leq 80 | X > 60) &= \frac{P(\{X \leq 80\} \cap \{X > 60\})}{P(X > 60)} = \frac{P(60 < X \leq 80)}{P(X > 60)} \\ &= \frac{F_X(80) - F_X(60)}{1 - F_X(60)} \\ &= \frac{\mathrm{e}^{-12} - \mathrm{e}^{-16}}{\mathrm{e}^{-12}} \approx 98.2\% \end{aligned}$$

Huomaa, että saatu tulos on sama kuin todennäköisyys $P(X \leq 20)$ eksponenttijakauman *muistinmenetysominaisuuden* mukaisesti.

d) Olkoon X_i = "i:nnen yhteyden saantiin kuluva aika", jolloin muuttujat $X_i \sim \mathrm{Exp}(\frac{1}{5})$ ovat riippumattomia. Lasketaan

$$P(X \leq 5) = F_X(5) = 1 - \mathrm{e}^{-1}.$$

Koska X_i:t ovat riippumattomia ja samalla tavalla jakautuneita kuin X, on kysytty todennäköisyys

$$P\left(\bigcap_{i=1}^{10} \{X_i \leq 5\}\right) = P(X \leq 5)^{10} = (1 - \mathrm{e}^{-1})^{10} \approx 1\%.$$

9. Olkoon $X \sim N(0,1)$. Laske kertymäfunktio ja tiheysfunktio muuttujille

a) X^2 (χ^2-jakauma [khii toiseen jakauma] vapausasteilla 1),

b) e^X (log-normaalijakauma).

Ratkaisu:

a) Merkitään $Y = X^2$, $X \sim \mathrm{N}(0,1)$.
Jos $y < 0$, niin $P(Y < y) = 0$. Tällöin $F(y) = 0$, joten myös $f(y) = \frac{d}{dy}F(y) = 0$.
Jos $y \geq 0$, niin kertymäfunktioksi saadaan

$$\begin{aligned}
F(y) &= P(Y < y) = P(X^2 < y) = P(|X| < \sqrt{y}) \\
&= P(-\sqrt{y} < X < \sqrt{y}) = \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \\
&= 2\Phi(\sqrt{y}) - 1 \\
&= 2\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, dt - 1 \\
&= \underline{\sqrt{\frac{2}{\pi}} \int_{-\infty}^{\sqrt{y}} e^{-\frac{t^2}{2}}\, dt - 1}
\end{aligned}$$

Tiheysfunktio saadaan derivoimalla tästä (huomaa, että $\frac{d}{dy}\Phi(y) = \varphi(y)$)

$$\begin{aligned}
f(y) &= \frac{d}{dy}F(y) \\
&= \frac{d}{dy}\left(2\Phi(\sqrt{y}) - 1\right) \\
&= 2\varphi(\sqrt{y})\frac{d}{dy}\sqrt{y} \\
&= \frac{\varphi(\sqrt{y})}{\sqrt{y}} \\
&= \underline{\frac{e^{-\frac{y}{2}}}{\sqrt{2\pi y}}}
\end{aligned}$$

b) Merkitään $Z = e^X$, jolloin $\ln(Z) = X \sim \mathrm{N}(0,1)$. Lasketaan kertymäfunktio

$$\begin{aligned}
F(z) &= P(Z < z) = P(e^X < z) = P(X < \ln z) = \Phi(\ln z) \\
&= \underline{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln z} e^{-\frac{t^2}{2}}\, dt}
\end{aligned}$$

Tiheysfunktio saadaan derivoimalla

$$\begin{aligned}
f(z) &= \frac{d}{dz}\Phi(\ln z) = \varphi(\ln z) \cdot \frac{1}{z} = \frac{\varphi(\ln z)}{z} \\
&= \underline{\frac{e^{-\frac{(\ln z)^2}{2}}}{\sqrt{2\pi}z}}
\end{aligned}$$