

TILASTOMATEMATIIKKA

Harjoitusviikko 3 ratkaisut, kevät 2024

Harjoituksen teemoja ovat

- (i) Diskreetin jakauman odotusarvo, varianssi ja keskihajonta
- (ii) Jatkuvan jakauman odotusarvo, varianssi ja keskihajonta
- (iii) Odotusarvon ja varianssin ominaisuudet
- (iv) Keskeinen raja-arvolause
- (v) Normaalijakauma-approksimaatio
- (vi) Jatkuvuuskorjaus

Tehtävät 1-5 liittyvät teemoihin (i)-(iii), joita käsitellään pääasiassa alkuviikon harjoituksissa. Teemoja (iv)-(vi) käsitellään tehtävissä 6-9, jotka on ensisijaisesti tarkoitettu loppuviikon tehtäviksi. Tehtäviä voi toki ratkoa haluamassaan järjestyksessä. **Minimivaatimuksena laskuharjoituspisteille ovat tehtävät 2 ja 3** alkuviikosta ja **tehtävät 6 ja 8** loppuviikosta.

1. Tarkastellaan diskreettiä satunnaismuuttujaa, jonka pistetodennäköisyydet ovat

x	0	1	2	3
$\mathbb{P}(X = x)$	1/8	3/8	3/8	1/8

- a) Määrää X :n odotusarvo ja varianssi.
- b) Olkoon $X \sim \text{Bin}(3, 1/2)$. Määrää X :n pistetodennäköisyydet. Mitä havaitset? Käytä kaavakokoelmaa X :n odotusarvon ja varianssin määrittämiseen.
- c) Simuloi Excelillä annettua jakaumaa *Data Analysis Toolpakin Random number generation*-toiminnolla generoimalla kokoa 10000 oleva otos jakaumasta. Poimi kunkin arvon esiintymiskerrat *COUNTIF*-komennolla. Mitä saat simuloituiksi pistetodennäköisyyksiksi? Voit katsoa mallia vaikkapa videosta

<https://www.youtube.com/watch?v=emtaBhH7kf0>.

Ratkaisu:

- a) Määritelmän mukaan odotusarvo on

$$\mu_X = \mathbb{E}(X) = \sum_{k=0}^3 k\mathbb{P}(X = k) = 0 \cdot \frac{1}{8} + 1 \cdot \frac{3}{8} + 2 \cdot \frac{3}{8} + 3 \cdot \frac{1}{8} = \underline{\underline{\frac{3}{2}}}$$

Varianssi voidaan laskea laskukaavalla

$$\text{Var}(X) = \sum_{k=0}^3 (k - \mu_X)^2 \mathbb{P}(X = k).$$

Menetellään tässä kuitenkin toisin, sillä usein varianssi on näppärämpi laskea kaavakokoelmastakin löytyvällä laskukaavalla

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Koska odotusarvo $\mathbb{E}(X)$ on jo laskettu, riittää laskea

$$\mathbb{E}(X^2) = \sum_{k=0}^3 k^2 \mathbb{P}(X = k) = 0^2 \cdot \frac{1}{8} + 1^2 \cdot \frac{3}{8} + 2^2 \cdot \frac{3}{8} + 3^2 \cdot \frac{1}{8} = 3,$$

josta saadaan

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \frac{3}{4}.$$

b) Binomijakauman pistetodennäköisyydet voidaan laskea laskukaavalla

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

Nyt $X \sim \text{Bin}(3, 1/2)$, joten $n = 3$ ja $p = \frac{1}{2}$. Sijoittamalla nämä vuorotellen yllä olevaan laskukaavaan saadaan täsmälleen sama jakauma kuin tehtävänannossa. Esimerkiksi

$$\mathbb{P}(X = 0) = \binom{3}{0} \left(\frac{1}{2}\right)^0 \left(1 - \frac{1}{2}\right)^{3-0} = \left(\frac{1}{2}\right)^3 = \frac{1}{8}.$$

Odotusarvon ja varianssin laskukaavojen mukaan $\mathbb{E}(X) = np = \frac{3}{2}$ ja $\text{Var}(X) = np(1-p) = \frac{3}{4}$, jotka myös ovat samat kuin a)-kohdassa lasketut.

c) Laaditaan ensin Exceliin tehtävänannon mukainen taulukko. Tehdään sitten simulaation numeroita $1, 2, \dots, 10000$ vastaava sarake Home-valikosta löytyvän Fill-toiminnon alla olevalla Series-toiminnolla. Simuloidaan simulaation järjestysnumeroiden viereen otos *Data Analysis Toolpakin Random Number Generation*-toiminnolla. Muuttujien lukumääräksi valitaan 1, sillä meillä on vain yksi muuttuja. Satunnaislukujen määräksi valitaan 10000. Jakaumaksi valitaan diskreetti. Value and Probability Input Range kohtaan maalataan alue, jossa arvot ja niitä vastaavat pistetodennäköisyydet ovat. Valitse tulostusalueeksi simulaation järjestysnumeron 1 viereinen solu. Sitten vaan kone laskemaan. Voit halutessasi ottaa mallia linkissä olevasta videosta.

Lasketaan lopuksi simulaation antamien arvojen esiintymiskertojen lukumäärä COUNTIF-toiminnolla. Oma simulaatio antoi seuraavat tulokset

Arvot	x	0	1	2	3
Pistetodennäköisyydet	Teoreettinen	0.125	0.375	0.375	0.125
$\mathbb{P}(X = x)$	Simuloitu	0.1288	0.3803	0.3669	0.124

2. Määrittää vakio c , kun tiedetään, että X on satunnaismuuttuja, jonka

a) pistetodennäköisyydet ovat

x	1	2	3	4	5
$\mathbb{P}(X = x)$	c	$2c$	$3c$	$2c$	c

Laske X :n odotusarvo ja keskihajonta. Mitä intuitiosi sanoo odotusarvosta?

b) tiheysfunktio on

$$f_X(x) = \begin{cases} c(3 - |x - 3|), & 0 \leq x \leq 6, \\ 0, & \text{muulloin.} \end{cases}$$

Laske X :n odotusarvo ja keskihajonta. Mitä intuitiosi sanoo odotusarvosta?

Ratkaisu: Tämä tehtävä itse asiassa ratkottiin c :n osalta jo edellisellä harjoitusviikolla, joskin hieman toisessa muodossa.

a) On hyvä tapa täydentää pistetodennäköisyyksien taulukkoa sarakkeella, johon tulee pistetodennäköisyyksien summa

x	1	2	3	4	5	Σ
$\mathbb{P}(X = x)$	c	$2c$	$3c$	$2c$	c	1

Koska pistetodennäköisyyksien summan tulee olla yksi, niin c voidaan ratkaista. On oltava

$$\sum_{k=1}^5 \mathbb{P}(X = k) = 1 \Leftrightarrow 9c = 1 \Leftrightarrow c = \frac{1}{9}.$$

Lasketaan odotusarvo ja varianssi. Hajonta saadaan varianssin neliöjuurena. Koska X on diskreetti sm, jonka arvojoukko on äärellinen, ovat sekä odotusarvo että varianssi äärellisenä olemassa. Odotusarvo $\mu_X = \mathbb{E}(X)$ lasketaan kaavalla

$$\mu_X = \sum_{k=1}^5 k \cdot \mathbb{P}(X = k) = 1 \cdot \frac{1}{9} + 2 \cdot \frac{2}{9} + 3 \cdot \frac{3}{9} + \dots + 5 \cdot \frac{1}{9} = \frac{1}{9}(1 + 4 + 9 + 8 + 5) = \frac{27}{9} = 3.$$

Varianssi $\sigma_X^2 = \text{Var}(X)$ lasketaan kaavalla

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2,$$

jossa odotusarvo $\mu_X = \frac{5}{3}$ on jo edellä laskettu ja $\mathbb{E}(X^2)$ lasketaan muunnoksen odotusarvona samaan tapaan kuin yllä ja saadaan

$$\mathbb{E}(X^2) = \sum_{k=1}^5 k^2 \mathbb{P}(X = k) = \frac{1}{9}(1^2 + 2^2 \cdot 2 + 3^2 \cdot 3 + 4^2 \cdot 2 + 5^2) = \frac{93}{9} = \frac{31}{3}.$$

Edellisen perusteella

$$\sigma_X^2 = \mathbb{E}(X^2) - \mu_X^2 = \frac{31}{3} - 3^2 = \frac{4}{3},$$

josta hajonnaksi tulee $\sigma_X = \frac{2}{\sqrt{3}} \approx 1.7$.

Toinen tapa varianssin laskemiseen on suora määritelmä

$$\sigma_X^2 = \mathbb{E}((X - \mu_X)^2) = \sum_{k=1}^5 (k - \mu_X)^2 \mathbb{P}(X = k) = \dots = \frac{4}{3},$$

jota kautta luonnollisesti päädytään samaan lopputulokseen kuin yllä.

Intuitiolla odotusarvon pitäisi olla 3, sillä se on "keskimäinen" arvo ja todennäköisyysjakauma on symmetrinen sen suhteen. Ehkäpä vielä taulukkoa konkreettisemmin tämä näkyy, jos piirtää todennäköisyyksistä histogrammin, mutta sehän me oltiin jo skaalausta vaille tehty harjoitusviikolla 2.

b) Koska tiheysfunktion integraali yli koko \mathbb{R} :n pitää olla yksi, saadaan integroimalla

$$\int_{-\infty}^{\infty} f_X(x) dx = \int_0^6 c(3 - |x - 3|) dx = c \int_0^3 x dx + c \int_3^6 (6 - x) dx = 9c = 1,$$

josta $c = \frac{1}{9}$. Sama voidaan perustella myös geometrisesti. Piirtämällä funktion kuvaaja nähdään, että muodostuu kolmio, jonka pinta-ala on 9, kun $c = 1$. Niinpä täytyy olla $c = \frac{1}{9}$, jotta ala olisi 1 kuten kuuluukin. Funktion kuvaaja tapauksessa $c = 1$ löytyy viime viikon harjoituksista.

Nyt, kun c on saatu, voidaan laskea odotusarvo. Odotusarvo on integraali

$$\mu_X = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^3 \frac{1}{9} x^2 dx + \int_3^6 \frac{1}{9} (6x - x^2) dx = \frac{1}{27} x^3 \Big|_{x=0}^3 + \frac{1}{9} (3x^2 - \frac{1}{3} x^3) \Big|_{x=3}^6 = 3.$$

Samaan lopputulokseen päädytään myös intuitiolla, sillä todennäköisyysjakauma on symmetrinen suoran $x = 3$ suhteen. Tämän näkee ehkä parhaiten viime viikon harjoitustehtävän 2 kuvasta.

Varianssi voidaan laskea kaavalla

$$\sigma_X^2 = \text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2.$$

Odotusarvo laskettiin jo edellä, joten riittää laskea odotusarvo $\mathbb{E}(X^2)$, joksi saadaan samalla tavalla kuin yllä integroimalla

$$\mathbb{E}(X^2) = \int_0^3 \frac{1}{9} x^3 dx + \int_3^6 \frac{1}{9} (6x^2 - x^3) dx = \dots = \frac{9}{4} + \frac{33}{4} = \frac{21}{2}.$$

Edellisen perusteella varianssi on

$$\text{Var}(X) = \sigma_X^2 = \mathbb{E}(X^2) - \mu_X^2 = \frac{21}{2} - 3^2 = \frac{3}{2},$$

josta hajonnaksi tulee $\sigma_X = \sqrt{\frac{3}{2}} \approx 1.2$.

3. Olkoot X_1, X_2 riippumattomia ja samalla tavalla jakautuneita satunnaismuuttujia. Laske muuttujan $X = \max\{X_1, X_2\}$
- odotusarvo ja varianssi, kun $X_i \sim \text{Bin}(1, p)$ (Bernoulli-jakautuneita).
 - odotusarvo ja varianssi, kun $X_i \sim \text{Tas}(0, 1)$.
 - odotusarvo (Excelillä) simuloimalla, kun $X_i \sim N(0, 1)$. Käytä käänteisfunktio menetelmää Excelin NORM.S.INV-toiminnolla. Simuloi esimerkiksi 10000 näytettä kummastakin muuttujasta X_i ja laske saamiesi maksimien keskiarvo. Vertaa harjoitusviikon 2, tehtävä 3. **Haaste:** laske odotusarvo $\mathbb{E}(X) = \frac{1}{\sqrt{\pi}}$ analyttisesti.

Ratkaisu:

Tämä on valitettavasti funktionaalista muunnosta $X = g(Y)$ hankalampi tehtävä, sillä jo tunnuslukujen laskeminen edellyttää X :n jakauman tuntemista ja tarkasteltavia muuttujia X_1, X_2 on kaksi. Toinen tapa muuttujien X_1 ja X_2 yhteisjakauman hyödyntäminen, mutta palataan siihen vasta kurssin loppupuolella. Onneksi a)- ja c)-kohdissa X :n jakauma (kertymäfunktio) selvitettiin luentoviikon 2 luento-esimerkissä 19.

- a) Viime viikon perusteella itse asiassa $X \sim \text{Bin}(1, 1 - (1 - p)^2)$ (on Bernoulli-jakautunut parametrilla $q = 1 - (1 - p)^2$), joten odotusarvo on

$$\mu_X = \mathbb{E}(X) = \sum_{k=0}^1 k \cdot \mathbb{P}(X = k) = 1 - (1 - p)^2.$$

Voidaan myös vedota kaavakokoelmasta löytyvään kaavaan $\mathbb{E}(X) = np$, kun $X \sim \text{Bin}(n, p)$. Varianssin laskemiseen voidaan käyttää kaavaa

$$\text{Var}(X) = \sigma_X^2 = \mathbb{E}(X^2) - \mu_X^2.$$

Itse asiassa $\mathbb{E}(X^2) = \mu_X$, sillä $1^2 = 1$ ja $0^2 = 0$. Siten alkuosan nojalla

$$\text{Var}(X) = \mu_X - \mu_X^2 = \mu_X(1 - \mu_X) = (1 - p)^2(1 - (1 - p)^2).$$

Toinen tapa on määritelmän

$$\text{Var}(X) = \mathbb{E}((X - \mu_X)^2) = \sum_{k=0}^1 (k - \mu_X)^2 \cdot \mathbb{P}(X = k)$$

käyttö. Kolmannessa tavassa voidaan hyödyntää kaavakokoelmasta löytyvää laskukaavaa $\text{Var}(X) = np(1 - p)$, kun $X \sim \text{Bin}(n, p)$.

- b) Vaikka a)-kohdassa jakaumamalli säilyi muunnoksessa, näin ei suinkaan ole yleisesti. Tässä on esimerkki sellaisesta. Käytetään samaa menetelyä kuin luentoviikon 2 esimerkissä 19, jonka mukaan

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\max\{X_1, X_2\} \leq x) = \mathbb{P}(\{X_1 \leq x\} \cap \{X_2 \leq x\}) = \mathbb{P}(X_1 \leq x)^2 = x^2$$

kaikilla $0 \leq x \leq 1$. Toisessa yhtäsuuruudessa käytettiin hyväksi tietoa, että maksimi on pienempää kuin x täsmälleen silloin, kun molemmat muuttujat ovat pienempää kuin x . Viimeistä edellisessä yhtäsuuruudessa käytettiin muuttujien riippumattomuutta ja sitä, että muuttujat ovat samalla tavalla jakautuneita. Viimeisessä yhtäsuuruudessa huomioitiin, että $X_1 \sim \text{Tas}(0, 1)$.

Tiheysfunktio saadaan derivoimalla

$$f_X(x) = \frac{d}{dx} F_X(x) = \begin{cases} 2x, & 0 < x < 1, \\ 0, & \text{muulloin.} \end{cases}$$

Odotusarvo saadaan nyt tavalliseen tapaan

$$\mu_X = \mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 2x^2 dx = \frac{2}{3} x^3 \Big|_{x=0}^1 = \frac{2}{3}.$$

Vastaavasti saadaan

$$\mathbb{E}(X^2) = \int_0^1 2x^3 dx = \frac{1}{2} x^4 \Big|_{x=0}^1 = \frac{1}{2}$$

varianssia varten, joten varianssi on

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2 = \frac{1}{2} - \left(\frac{2}{3}\right)^2 = \frac{1}{2} - \frac{4}{9} = \frac{1}{18}.$$

- c) Lasketaan Excelillä NORM.S.INV(RAND())-toiminnolla yhteen sarakkeeseen 10000 näyttöä muuttujasta X_1 . Tehdään sama toiseen sarakkeeseen muuttujalle X_2 . Poimitaan riittävästi näin saattujen lukujen maksimi ja lasketaan maksimille keskiarvo. Eräs simulaatio antoi keskiarvoksi noin 0.56. Huomaa, että tarkka arvo on $\mathbb{E}(X) = 1/\sqrt{\pi} \approx 0.56$ kahden desimaalin tarkkuudella.

Lasketaan lopuksi odotusarvo analyttisesti. Luentoviikolla 2 laskettiin maksimin kertymäfunktio. Tehdään se nopeasti uudelleen samaan tapaan kuin b)-kohdassa. Saadaan

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{X_1 \leq x\} \cap \{X_2 \leq x\}) = \mathbb{P}(X_1 \leq x)^2 = \Phi(x)^2,$$

josta derivoimalla saadaan tiheysfunktiksi

$$f_X(x) = \frac{d}{dx} F_X(x) = 2\varphi(x)\Phi(x).$$

Meillä pitäisi integroida funktio $x f_X(x) = 2x\varphi(x)\Phi(x)$. Tämä ei onnistu ihan suoraan, joten temppuillaan vähän. Onneksi meillä on käytössä osittaisintegrointi. Sitä varten palautetaan mieliin standardinormaalijakauman tiheysfunktion lauseke $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Valitaan osittaisintegrointikaavassa

$$\int_{-\infty}^{\infty} f'(x)g(x)dx = f(x)g(x) \Big|_{x=-\infty}^{\infty} - \int_{-\infty}^{\infty} f(x)g'(x)dx$$

funktiksi $f'(x) = x\varphi(x)$ ja funktiksi $g(x) = 2\Phi(x)$. Tällöin $f(x) = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = -\varphi(x)$ ja $g'(x) = 2\varphi(x)$. Silloin osittaisintegrointi antaa

$$\mu_X = \mathbb{E}(X) = \int_{-\infty}^{\infty} 2x\varphi(x)\Phi(x)dx = 2 \int_{-\infty}^{\infty} \varphi(x)^2 dx.$$

Sijoitustermi hävisi, sillä $\varphi(x) \rightarrow 0$ vikkelästi, kun $x \rightarrow \pm\infty$. Muokataan viimeistä integraalia seuraavasti

$$\varphi(x)^2 = \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\right)^2 = \left(\frac{1}{\sqrt{2\pi}}\right)^2 e^{-x^2} = \sigma \frac{1}{\sqrt{2\pi}} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}\right),$$

missä $\sigma = \frac{1}{\sqrt{2}}$. Koska sulussa on muuttujan $Y \sim N(0, \sigma^2)$ tiheysfunktio, on sen integraali yli koko \mathbb{R} :n yksi. Siten

$$\mu_X = \mathbb{E}(X) = 2\sigma \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \frac{1}{\sqrt{\pi}}.$$

4. Magneettisessa tiedon tallennuksessa oli havaittu virheiden määrän suurta talletusyksikköä kohti noudattavan Poisson-jakaumaa siten, että talletuksessa esiintyi keskimäärin yksi virheellinen bitti 100000 talletettua bittiä kohti. Tieto talletetaan sektoreina, missä sektori käsitti 4096 8-bittistä sanaa.
- a) Kuinka monta tallennusvirhettä esiintyi keskimäärin yhdessä sektorissa?
- b) Kuinka monta sektoria on keskimäärin käytävä läpi jotta ainakin yksi tallennusvirhe löytyy? Sektorit ovat toisistaan riippumattomat.

Ratkaisu:

Meitä kiinnostava satunnaismuuttuja on $X =$ ”virheiden lukumäärä sektoria kohden”, joka on tehtävänannon mukaan Poisson-jakautunut, $X \sim \text{Poi}(a)$ parametrilla

$$a = \frac{4096 \cdot 8}{100000} = \frac{32768}{100000} \approx 0.328,$$

sillä yksi sektori pitää sisällään 32768 bittiä.

- a) Koska Poisson-jakauman parametri a ilmoittaa keskimääräisen arvon, eli $E(X) = a$, niin yhdessä sektorissa esiintyy keskimäärin

$$E(X) = a \approx 0.328$$

tallennusvirhettä.

- b) Olkoon Y vähintään yhden tallennusvirheen sisältävien sektoreiden lukumäärä, jolloin $Y \sim \text{Bin}(n, p)$, missä n on tuntematon ja p on todennäköisyys sille, että sektorissa on vähintään yksi tallennusvirhe. Tehtävänannon mukaan

$$p = P(X \geq 1) = 1 - P(X = 0) = 1 - \frac{a^0}{0!} e^{-a} = 1 - e^{-\frac{32768}{100000}} \approx 0.279.$$

Tällöin ainakin yhden tallennusvirheen sisältäviä sektoreita on keskimäärin

$$E(Y) = np$$

kappaletta. Koska tallennusvirheitä pitää olla keskimäärin ainakin yksi, saadaan

$$E(Y) = np \geq 1 \Rightarrow n \geq \frac{1}{p} \approx 3.58,$$

eli on käytävä läpi keskimäärin 4 sektoria vähintään yhden virheen löytämiseksi. Tätä voidaan ajatella myös geometrisen jakauman, joka löytyy luentomonisteen sivulta 30. Olkoon Y sen sektorin järjestysluku, jolla esiintyy ensimmäinen tallennusvirhe. Tällöin Y noudattaa geometrista jakaumaa parametrilla p ja merkitään $Y \sim \text{Geo}(p)$, missä p on todennäköisyys sille, että sektorissa on vähintään yksi virhe, eli $p = P(X \geq 1)$. Luentomonisteen sivun 44 mukaan

$$E(Y) = \frac{1}{p} \approx 3.58.$$

5. Erään kemikaalin pitoisuus [g/cm^3] tietyssä ainesosuksessa on normaalijakautunut satunnaismuuttuja, jolla on odotusarvo μ ja hajonta $\sigma = 0.004 g/cm^3$. Tehtävänä on arvioida odotusarvoa μ otoksen avulla. Kuinka monta toisistaan riippumatonta näytettä on otettava, jotta näytteestä mitattujen kemikaalipitoisuuksien keskiarvo poikkeaisi μ :stä suuntaan tai toiseen korkeintaan $0.002 g/cm^3$ todennäköisyydellä 0,9?

Ratkaisu:

Olkoon $X_i =$ ”kemikaalin pitoisuus näytteessä i ”. Tehtävänannon mukaan $X \sim N(\mu, \sigma^2)$. Kemikaalipitoisuuden keskiarvoa kuvaava satunnaismuuttuja on aritmeettinen keskiarvo

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

Luentokalvojen Esimerkin 23 ja sivun 41 mukaan \bar{X} on normaalijakautunut parametreilla

$$E(\bar{X}) = \mu \quad \text{ja} \quad \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}.$$

Tähän palataan useasti myöhemminkin.

Yllä näytteiden lukumäärä n on tuntematon suure, joka ratkaistaan ehdosta

$$P(\mu - 0.002 \leq \bar{X} \leq \mu + 0.002) \geq 0.9.$$

Standardisoidaan \bar{X} , jolloin

$$Z = \frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

ja päästään taulukkoarvoihin käsiksi. Standardisoimalla saadaan

$$\begin{aligned} P(\mu - 0.002 \leq \bar{X} \leq \mu + 0.002) &= P(\mu + 0.002 - \mu \leq \bar{X} - \mu \leq \mu + 0.002 - \mu) \\ &= P\left(\frac{-0.002}{\sigma/\sqrt{n}} \leq Z \leq \frac{0.002}{\sigma/\sqrt{n}}\right) \\ &= \Phi\left(\frac{0.002}{\sigma/\sqrt{n}}\right) - \Phi\left(-\frac{0.002}{\sigma/\sqrt{n}}\right) \quad (Z \text{ on jatk. sm.}) \\ &= 2\Phi\left(\frac{0.002}{\sigma/\sqrt{n}}\right) - 1 \quad (\text{symmetriaominaisuus}) \\ &\geq 0.9, \end{aligned}$$

josta saadaan ehto

$$\Phi\left(\frac{0.002}{\sigma/\sqrt{n}}\right) \geq 0.95.$$

Normaalijakauman taulukosta saadaan

$$\frac{0.002}{\sigma/\sqrt{n}} \geq 1.645 \Rightarrow \sqrt{n} \geq \frac{1.645 \cdot 0.004}{0.002} \Rightarrow n \geq 11,$$

eli on otettava vähintään $n = 11$ näytettä halutun tarkkuuden saavuttamiseksi.

6. Eräällä tehtaalla valmistetuista mikrosiruista 2 % on viallisia. Poimitaan satunnaisotannalla 100 mikrosirua tarkastukseen. Laske normaalijakauma-approksimaatiolla todennäköisyys, että tarkastuksessa löydetään 5 viallista mikrosirua. Vertaa harjoitusviikon 2 tehtävään 3.

Ratkaisu: Olkoon

$X =$ ”viallisten mikrosirujen lukumäärä tarkistuksessa”,

jolloin $X \sim \text{Bin}(100, 0.02)$, kun oletetaan, että mikrosirujen viallisuus on toisistaan riippumaton. Binomijakaumamallin mukaan tunnusluvuiksi saadaan

$$\mathbb{E}(X) = np = 100 \cdot 0.02 = 2 \quad \text{ja} \quad \text{Var}(X) = \sigma_X^2 = np(1-p) = 100 \cdot 0.02 \cdot 0.98 = 1.96.$$

Standardoidaan X :

$$Z = \frac{X - \mathbb{E}(X)}{\sigma_X} = \frac{X - 2}{\sqrt{1.96}}.$$

Jatkuvuuskorjauksella ja standardoimalla muuttuja X saadaan kysytyksi todennäköisyydeksi

$$\begin{aligned} \mathbb{P}(X = 5) &= \mathbb{P}(5 \leq X \leq 5) = \mathbb{P}(4.5 < X \leq 5.5) \quad (\text{jatkuvuuskorjaus}) \\ &= \mathbb{P}\left(\frac{4.5 - 2}{\sqrt{1.96}} < Z \leq \frac{5.5 - 2}{\sqrt{1.96}}\right) \quad (\text{standardointi}) \\ &= \mathbb{P}\left(Z \leq \frac{3.5}{\sqrt{1.96}}\right) - \mathbb{P}\left(Z \leq \frac{2.5}{\sqrt{1.96}}\right) \\ &\approx \Phi(2.5) - \Phi(1.79) \quad (\text{normaalijakauma-approksimaatio}) \\ &= 0.9938 - 0.9633 = 0.0305. \quad (\text{taulukosta}) \end{aligned}$$

Excel antaa 0.0309 neljän desimaalin tarkkuudella.

Verrataan tulosta harjoitusviikon 2 tehtävässä 3 laskettuun tarkkaan todennäköisyyteen 0.0353. Suhteellinen virhe on $\frac{0.0353 - 0.0305}{0.0353} \approx 14\%$. Virhe ei ole valtavan suuri, kun otetaan huomioon, että laskettiin approksimaatio jatkuvalla jakaumalla iskreetin jakauman yksittäisen pisteen todennäköisyydelle, joka oli vieläpä aika pieni. Tarkkuus olisi luonnollisesti parantunut, jos olisi kysytty jonkun välin todennäköisyyttä.

7. Erään metallikomponentin väsymisrajaa X [MPa] mallinnetaan log-normaalijakaumalla, eli $\ln X \sim N(\mu, \sigma^2)$. On havaittu, että väsymisrajan mediaani Md on 400 [Mpa] ja että $\sigma^2 = 0.15$.

- a) Määrittää parametri μ . Käytä hyväksi tietoa

$$P(X \leq \text{Md}) = 0.5.$$

- b) Mikä on väsymisrajan odotusarvo?

- c) Mitataan 30 komponentin väsymisraja. Laske normaalijakauma-approksimaatiolla todennäköisyys, että väsymisrajan keskiarvo on vähintään 350 [MPa]? Oletetaan, että mittaukset ovat riippumattomia ja samalla tavalla jakautuneita.

Vihje: Log-normaalijakauman hajonnan kaavan voi kaivaa vaikka Wikipediasta.

Ratkaisu:

- a) Koska $\ln X \sim N(\mu, \sigma^2)$, niin annetun tiedon perusteella

$$P(X \leq \text{Md}) = P(\ln X \leq \ln \text{Md}) \stackrel{\text{stand.}}{=} P\left(\frac{\ln X - \mu}{\sigma} \leq \frac{\ln \text{Md} - \mu}{\sigma}\right) = 0.5.$$

Koska

$$Z = \frac{\ln X - \mu}{\sigma} \sim N(0, 1),$$

niin standardinormaalijakauman symmetrian perusteella

$$\frac{\ln \text{Md} - \mu}{\sigma} = 0 \Leftrightarrow \ln \text{Md} = \mu.$$

Parametri μ on siten $\ln \text{Md} = \ln 400$.

- b) Luento-esimerkin 22 mukaan

$$E(X) = \exp\left(\mu + \frac{1}{2}\sigma^2\right) = \exp(\ln 400) \exp(\sigma^2/2) = 400e^{0.075}.$$

- c) Olkoon $X_i =$ "väsymisraja i :nnessä mittauksessa". Tarkastellaan keskiarvoa

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{30} \sum_{i=1}^{30} X_i.$$

Kohdasta b) saadaan

$$E(X_i) = E(X) = 400e^{0.075} \approx 431$$

ja Wikipediasta

$$\begin{aligned} \sigma_{X_i} = \sigma_X &= \sqrt{\text{Var}(X_i)} = \sqrt{(\exp(\sigma^2) - 1) \exp(2\mu + \sigma^2)} \\ &= \sqrt{(\exp(0.15) - 1) \exp(0.15) 400}. \end{aligned}$$

Yllä termi $\exp(2\mu)$ sievennettiin seuraavasti

$$\exp(2\mu) = \exp(2 \ln 400) = \exp(\ln 400^2) = 400^2.$$

Koska $E(\bar{X}) = E(X)$ ja $\sigma_{\bar{X}} = \sigma_X / \sqrt{n}$, niin normaaliapproksimaatiolla kysytyksi todennäköisyydeksi saadaan

$$\begin{aligned} P(\bar{X} > 350) &= P\left(\frac{\bar{X} - E(\bar{X})}{\sigma_{\bar{X}}} > \frac{(350 - 400e^{0.075}) \sqrt{30}}{\sqrt{(\exp(0.15) - 1) \exp(0.15) 400}}\right) \\ &\approx P(Z > -2.56) = \Phi(2.56) \approx 0.995. \end{aligned}$$

8. Oletetaan, että satunnaisesti valittu nuorukainen ostaa suosittua energiajuomaa yhden pullon todennäköisyydellä 0.9 ja kaksi pulloa todennäköisyydellä 0.1.
- Laske normaalijakauma-approksimaatiolla todennäköisyys, että 100 nuorukaista ostaa vähintään 115 pulloa.
 - Laske normaalijakauma-approksimaatiolla, kuinka monta pulloa kauppiaan on varattava myyntiin, jotta 100 nuorukaiselle riittää juomia vähintään 95 % todennäköisyydellä.
 - Laske edelliset kohdat tarkasti tilanteeseen sopivalla jakaumamallilla. Mitä oletuksia täytyy tehdä, jotta homma onnistuu?

Ratkaisu:

- a) Olkoon

$$X_i = \text{”nuoren } i \text{ ostamien pullojen lukumäärä”}, \quad i = 1, 2, \dots, 100,$$

jolloin muuttuja $S_{100} = \sum_{i=1}^{100} X_i$ ilmoittaa 100 nuoren ostamien pullojen lukumäärän. Käytetään normaalijakauma-approksimaatiota. Sitä varten täytyy laskea muuttujien X_i odotusarvo

$$\mu_{X_i} = \sum_{k=1}^2 k \cdot \mathbb{P}(X_i = k) = 0.9 \cdot 1 + 0.1 \cdot 2 = 1.1$$

ja varianssi

$$\text{Var}(X_i) = \sigma_{X_i}^2 = \mathbb{E}(X_i^2) - \mu_{X_i}^2 = 0.9 \cdot 1^2 + 0.1 \cdot 2^2 = 0.09,$$

josta saadaan hajonnaksi $\sigma_{X_i} = 0.3$. Summamuuttujan tunnusluvut vastaavasti ovat

$$\mu_{S_{100}} = \mathbb{E}(S_{100}) = 100 \cdot 1.1 = 110 \quad \text{ja} \quad \sigma_{S_{100}} = \sqrt{\text{Var}(S_{100})} = 10 \cdot 0.3 = 3.$$

Kysytään todennäköisyyttä $\mathbb{P}(S_{100} \geq 115)$, joksi saadaan

$$\begin{aligned} \mathbb{P}(S_{100} \geq 115) &= \mathbb{P}(S_{100} \geq 114.5) && \text{(jatkuvuuskorjaus)} \\ &= \mathbb{P}\left(\frac{S_{100} - \mathbb{E}(S_{100})}{\sigma_{S_{100}}} \geq \frac{114.5 - 110}{3}\right) && \text{(standardointi)} \\ &= 1 - \mathbb{P}\left(\frac{S_{100} - \mathbb{E}(S_{100})}{\sigma_{S_{100}}} < \frac{114.5 - 110}{3}\right) \\ &\approx 1 - \Phi(1.5) && \text{(norm. approksimaatio)} \\ &= 1 - 0.9332 \approx 6.7\%. && \text{(taulukosta)} \end{aligned}$$

- b) Tässä kysytään pienintä pullojen lukumäärää n , jolle $\mathbb{P}(S_{100} \leq n) \geq 0.95$. Standardoimalla saadaan

$$\mathbb{P}(S_{100} \leq n) = \mathbb{P}\left(\frac{S_{100} - \mathbb{E}(S_{100})}{\sigma_{S_{100}}} \leq \frac{n - 110}{3}\right) \geq 0.95.$$

Taulukosta tai numeerisesti saadaan (normaalijakauman 95 % kvantiili) $\frac{n-110}{3} \approx 1.645$, josta löydetään pienimmäksi n :n kokonaislukuarvoksi $n = 115$.

- c) Oletetaan, että *nuorten ostoskäyttäytyminen on toisistaan riippumatonta*. Tällöin tehtävän ratkaisemiseen voidaan käyttää binomijakaumamallia. Nimittäin, 100 nuorta ostaa pulloja vähintään 100 ja korkeintaan 200. Jokainen 100:n ylittävä pullo ostetaan todennäköisyydellä 0.1, joten $S_{100} = 100 + Y$, missä $Y \sim \text{Bin}(100, 0.1)$. Kohdan a) todennäköisyydeksi saadaan

$$\mathbb{P}(S_{100} \geq 115) = \mathbb{P}(Y \geq 15) = 1 - \mathbb{P}(Y \leq 14) \approx 7.3\%.$$

Todennäköisyys on laskettu Excelin BINOM.DIST-komennolla. Prosenttiyksikön tarkkuudella normaalijakauma-approksimaatio antoi saman tuloksen. Kohdassa b) kysytään pienintä n :ää, jolle $\mathbb{P}(S_{100} \leq n) = \mathbb{P}(Y \leq n - 100) \geq 0.95$. Excelin BINOM.INV-komennolla binomijakauman $Y \sim \text{Bin}(100, 0.1)$ 95 % kvantiiliksi (tai persentiiliksi) saadaan 15, joten $n = 115$. Tämä on sama kuin normaaliapproksimaatiolla saatu tulos. Alussa tekemämme oletus voidaan luonnollisesti asettaa kyseenalaiseksi. Silloin emme itse asiassa olisi voineet laskea b)-kohtaakaan samalla tavalla kuin laskimme sen, sillä summan varianssin laskenta vaikeutuu.

9. Fyysikot käyttävät satunnaiskävelyä diffuusion tai yleisesti partikkelien satunnaisliikkeen mallintamiseen. Partikkeli suorittaa siirtymiä ajanhetkillä $t = 1, 2, \dots$ ja sen paikka S_n ajanhetkellä n voidaan ajatella olevan siirtymien X_1, \dots, X_n summa. Oletetaan, että siirtymät ovat riippumattomia ja samalla tavalla jakautuneita niin, että partikkeli voi samalla todennäköisyydellä joko pysyä paikallaan tai siirtyä yhden yksikön verran (+1) oikealle tai vasemmalle (-1). Määrää normaalijakauma-approksimaatiolla todennäköisyys, että 10000 siirtymän jälkeen partikkeli on vähintään 100 yksikköä lähtöpisteestä oikealle.

Ratkaisu:

Olkoon X_i partikkelin siirtymä hetkellä i .

Tällöin $X_i \in \{-1, 0, 1\}$ ja $P(X_i = -1) = P(X_i = 0) = P(X_i = 1) = \frac{1}{3}$.

$$E(X_i) = (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0$$

$$E(X_i^2) = (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} = \frac{2}{3}$$

$$\text{Var}(X_i) = E(X_i^2) - E(X_i)^2 = \frac{2}{3} - 0^2 = \frac{2}{3}$$

Kappaleen paikka hetkellä k on $S_k = \sum_{i=1}^k X_i$ ja voidaan laskea sille

$$E(S_k) = E\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k E(X_i) = 0$$

$$\text{Var}(S_k) = \text{Var}\left(\sum_{i=1}^k X_i\right) = \sum_{i=1}^k \text{Var}(X_i) = k \cdot \frac{2}{3}$$

Hetkellä $k = 10000$ on $S_{10000} = S \stackrel{\text{likim.}}{\sim} N(0, 10000 \cdot \frac{2}{3})$ ja todennäköisyys, että kappale on vähintään 100 yksikköä lähtöpisteestä oikealla käyttäen jatkuvuuskorjausta on

$$P(S \geq 100 - \frac{1}{2}) = 1 - P(S < 99,5)$$

$$= 1 - P\left(\frac{S - 0}{100\sqrt{\frac{2}{3}}} < \frac{99,5 - 0}{100\sqrt{\frac{2}{3}}}\right)$$

$$= 1 - \Phi\left(0,995 \cdot \sqrt{\frac{3}{2}}\right)$$

$$= 1 - \Phi(1,21862) \simeq 1 - 0,888506$$

$$= \underline{\underline{0,111}}$$