MTTTP5, luento 13.11.2018

Kertausta

Jos $X \sim N(\mu, \sigma^2)$, niin $Z = (X - \mu)/\sigma \sim N(0, 1)$.

- $P(X \leq a) \quad = P((X - \mu)/\sigma \leq (a - \mu)/\sigma)$

  $= \Phi((a - \mu)/\sigma)$

- $P(X \geq a) \quad = 1 – P(X \leq a)$

  $= 1 - P((X - \mu)/\sigma \leq (a - \mu)/\sigma)$

  $= 1 - \Phi((a - \mu)/\sigma)$

- $P(a \leq X \leq b) \quad = P(X \leq b) - P(X \leq a)$

  $= \Phi((b - \mu)/\sigma) - \Phi((a - \mu)/\sigma)$

3.5.5 Normaalijakauma (jatkuu)

Normaalijakaumaan liittyviä keskeisiä tuloksia

- $Jos\ X \sim N(\mu, \sigma^2),\ niin\ aX + b \sim N(a\mu + b, a^2\sigma^2)$

- $Jos\ X_1, X_2, \dots, X_n\ ovat\ riippumattomia\ ja\ X_i \sim N(\mu_i,\ \sigma_i^2), niin$

$$X_1 + X_2 + \cdots + X_n \sim N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

Esim. Lentomatkustajien tavaroiden painon oletetaan vaihtelevan siten, että ne painavat keskimäärin 20 kg keskihajonnan ollessa 5 kg. Oletetaan lisäksi painon vaihtelevan normaalijakauman mukaisesti. Eräs lentokonetyyppi kuljettaa 100 matkustajaa. Millä todennäköisyydellä matkatavaroiden yhteispaino ylittää 2150 kg?

Yhteispaino $Y = X_1 + X_2 + \ldots + X_{100}$, missä $X_i \sim N(20,25)$.

$E(Y) = E(X_1) + E(X_2) + \ldots + E(X_{100}) = 100 \cdot 20 = 2000$

$Var(Y) = Var(X_1) + Var(X_2) + \ldots + Var(X_{100}) = 100 \cdot 25$

$= 2500$

$Y \sim N(2000, 2500)$

$P(Y > 2150) = 1 - P(Y \leq 2150)$

$= 1 - \Phi((2150-2000)/50)$

$= 1 - \Phi(3) = 1 - 0{,}9987 = 0{,}0013.$

- $Jos\ X_1, X_2, \ldots, X_n\ ovat\ riippumattomia\ \ ja$

$$E(X_i) = \mu_i$$

$$Var(X_i) = \sigma_i^2,$$

$niin$

$$X_1 + X_2 + \cdots + X_n \overset{likimain}{\sim} N(\mu_1 + \mu_2 + \cdots + \mu_n, \sigma_1^2 + \sigma_2^2 + \cdots + \sigma_n^2)$$

Esim. Lentoyhtiötä pyydetään kuljettamaan 100 lammasta. Yhtiöllä on käytössä kone, joka voi ottaa kuljetettavakseen 5000 kg. Aiemmin on punnittu 1000 vastaavanlaista lammasta ja saatu keskiarvoksi 45 kg, hajonnaksi 3 kg ja painot ovat vaihdelleet välillä 37 kg – 56 kg. Voiko yhtiö ottaa pyydetyn 100 lampaan lastin kuljetettavakseen?

Yhteispaino $Y = X_1 + X_2 + ... + X_{100}$, missä $E(X_i) = 45$, $Var(X_i) = 9$

$E(Y) = E(X_1) + E(X_2) + ... + E(X_{100}) = 100 \cdot 45 = 4500$

$Var(Y) = Var(X_1) + Var(X_2) + ... + Var(X_{100}) = 100 \cdot 9$
$= 900$

$Y \sim N(4500, 900)$, likimain

$P(Y > 5000) = 1 - P(Y \leq 5000)$
$\approx 1 - \Phi((5000-4500)/30) = 1 - \Phi(16{,}67) \approx 0$

On siis lähes mahdotonta, että raja ylittyisi. Lampaat voi hyvin ottaa kuljetettavaksi. Liian varovainen arvio olisi $100 \cdot 56 = 5600$.

Edellisten tulosten perusteella saadaan otoskeskiarvoon liittyvät tulokset

- $Jos\ X_i \sim N(\mu,\ \sigma^2)\ ja\ X_i{:}t\ riippumattomia, niin$

$$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)\ \ ja$$

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)\ \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Esim. 3.5.14 GMAT-testiä käytetään useiden yliopistojen pääsykokeena. Kokeen tuloksen on todettu noudattavan normaalijakaumaa odotusarvona 525 ja keskihajontana 100. Sadan pyrkijän ryhmä osallistui ennen pääsykoetta valmennuskurssille. Pääsykokeessa heidän GMAT-testin keskiarvo oli 541,4. Menestyivätkö he pääsykokeessa muita paremmin?

$$\overline{X} \sim N\left(525, \frac{100^2}{100}\right)$$

$$P\left(\overline{X} \geq 541{,}4\right) = 1 - P\left(\overline{X} < 541{,}4\right)$$

$$= 1 - \Phi\left(\frac{541{,}4 - 525}{10}\right) = 1 - \Phi(1{,}64) = 0{,}0505$$

Eivät menestyneet paremmin kuin muut, koska ei ole harvinaista saada otoskeskiarvoa, joka suurempi kuin 541,4 silloin, kun menestyminen tavanomaista.

Esim. Auton sytytystulppien valmistaja väittää, että tulpat kestävät keskimäärin 60 000 km keskihajonnan ollessa 6 000 km sekä vaihtelu luonnehdittavissa normaalijakaumalla. Tutkit väitettä ja valitset satunnaisesti 4 tulppaa, joiden keskimääräiseksi kestoksi saat 55 500 km. Voitko uskoa valmistajan väitteen?

Jos valmistajan väite tosi, niin

$$\overline{X} \sim N\left(60000, \frac{6000^2}{4}\right).$$

$$P\left(\overline{X} \leq 55500\right) = \Phi\left(\frac{55500 - 60000}{\frac{6000}{2}}\right) = \Phi(-1{,}5)$$

$$= 1 - \Phi(1{,}5) = 1 - 0{,}9332 = 0{,}0668$$

Uskotaan valmistajan väite, koska väitteen ollessa tosi ei ole harvinaista saada otosta, jonka keskiarvo alle 55500.

- $Jos\ X_1, X_2, \dots X_n\ ovat\ riippumattomia, E(X_i) = \mu, Var(X_i) = \sigma^2,$

$niin$

$$X_1 + X_2 + \cdots + X_n \overset{likimain}{\sim} N(n\mu, n\sigma^2)$$

ja

$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n) \overset{likimain}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

Binomijakaumaa voidaan approksimoida normaalijakaumalla

- $Jos\ X \sim Bin(n,p), niin\ X \overset{likimain}{\sim} N\big(np, np(1-p)\big), kun\ n\ suuri.$

Esim. Tutkittiin uuden menetelmän käyttökelpoisuutta ihosairauden hoidossa. Vanhan menetelmän avulla 60 % potilasta parani. Uudella menetelmällä 72 potilasta sadasta parani. Onko uusi menetelmä vanhaa parempi?

Olkoon X = parantuneiden lukumäärä.

Jos uusi menetelmä toimii vanhan tavoin, niin

$X \sim \mathrm{Bin}(100, 0{,}6)$, $E(X) = 60$, $\mathrm{Var}(X) = 24$, joten

$X \sim N(60, 24)$ likimain.

Tällöin $P(X \geq 72) = 1 - P(X \leq 71)$

$\approx 1 - \Phi((71-60)/\sqrt{24})$

$= 1 - \Phi(2{,}26) = 0{,}0119.$

Binomijakaumasta laskettuna $P(X \geq 72) = 0{,}00843.$

Jos toimisi vanhan tavoin, niin olisi harvinaista saada parantuneita enemmän kuin 71. Päätellään uuden olevan parempi.

Esim. 3.5.12 Tentissä on 100 väittämää, jotka ovat tosia tai epätosia. Vastataan kaikkiin kysymyksiin arvaamalla.

Olkoon X = oikeiden vastausten lukumäärä.

X ~ Bin(100,1/2), joten

$$P(X = k) = \binom{100}{k}\left(\frac{1}{2}\right)^{100}, \qquad k = 0, 1, 2, \ldots 100$$

$$P(X \leq 60) = \sum_{k=0}^{60} \binom{100}{k}\left(\frac{1}{2}\right)^{100} = 0{,}9824$$

http://homepage.stat.uiowa.edu/~mbognar/applets/bin.html

$E(X) = 100/2 = 50$, $Var(X) = 100/4 = 25$, joten

$$X \overset{likimain}{\sim} N(50, 25)$$

$$P(X \leq 60) \approx \Phi\left(\frac{60-50}{5}\right) = \Phi(2) = 0{,}9772$$