Tilastollinen päättely

1. Otosjakaumat

1.1. Otos, otostunnusluvut ja niiden otosjakaumat

Arvonta, Havainto, Havaintoarvo, Otos, Otosjakauma, Otostunnusluku, Riippumattomuus, Satunnaismuuttuja, Satunnaisotos, Tilastollinen malli, Todennäköisyysjakauma, Tunnusluku, Yhteisjakauma

1.2. Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Odotusarvot ja varianssit

Aritmeettinen keskiarvo, Havainto, Havaintoarvo, Keskivirhe, Odotusarvo, Otos, Otosjakauma, Otoskeskihajonta, Otostunnusluku, Otosvarianssi, Riippumattomuus, Satunnaismuuttuja, Satunnaisotos, Standardointi, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.3. Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Otos normaalijakaumasta

Aritmeettinen keskiarvo, Havainto, Havaintoarvo, χ^2-jakauma, Momenttiemäfunktio, Normaalijakauma, Odotusarvo, Otos, Otosjakauma, Otostunnusluku, Otosvarianssi, Riippumattomuus, Satunnaismuuttuja, Satunnaismuuttujien muunnokset, Satunnaisotos, Standardointi, *t*-jakauma, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.4. Aritmeettisen keskiarvon otosjakauma: Suurten otosten tuloksia

Approksimatiivinen jakauma, Aritmeettinen keskiarvo, Asymptoottinen jakauma, Havainto, Havaintoarvo, Jakaumakonvergenssi, Keskeinen raja-arvolause, Momenttiemäfunktio, Normaalijakauma, Odotusarvo, Otos, Otosjakauma, Otostunnusluku, Otosvarianssi, Riippumattomuus, Satunnaismuuttuja, Satunnaismuuttujien summa, Satunnaisotos, Standardointi, Stokastinen konvergenssi, Suurten lukujen laki, Taylorin sarja, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.5. Suhteellisen frekvenssin otosjakauma

Approksimatiivinen jakauma, Aritmeettinen keskiarvo, Asymptoottinen jakauma, Bernoulli-jakauma, Binomijakauma, Havainto, Havaintoarvo, Jakaumakonvergenssi, Keskeinen raja-arvolause, Normaalijakauma, Odotusarvo, Otos, Otosjakauma, Otostunnusluku, Riippumattomuus, Satunnaismuuttuja, Satunnaisotos, Standardointi, Stokastinen konvergenssi, Suurten lukujen laki, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.6. Järjestystunnusluvut

Binomijakauma, Diskreetti jakauma, Havainto, Havaintoarvo, Jatkuva jakauma, Järjestetty otos, Järjestystunnusluku, Kertymäfunktio, Otos, Otosjakauma, Otostunnusluku, Riippumattomuus, Satunnaismuuttuja, Satunnaisotos, Tiheysfunktio, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.7. Delta-menetelmä

Approksimatiivinen jakauma, Asymptoottinen jakauma, Delta-menetelmä, Havainto, Havaintoarvo, Jakaumakonvergenssi, Keskeinen raja-arvolause, Muunnos, Normaalijakauma, Odotusarvo, Otos, Otosjakauma, Otostunnusluku, Riippumattomuus, Satunnaismuuttuja, Satunnaisotos, Slutskyn lause, Stokastinen konvergenssi, Suurten lukujen laki, Taylorin sarja, Todennäköisyysjakauma, Tunnusluku, Varianssi

1.1. Otos, otostunnusluvut ja niiden otosjakaumat

Satunnaisotos

Olkoot

$$X_1, X_2, \ldots, X_n$$

riippumattomia, *identtisesti jakautuneita* satunnaismuuttujia, joilla on siis *sama* piste-todennäköisyys- tai tiheysfunktio

$$f(x)$$

Tällöin sanomme, että satunnaismuuttujat

$$X_1, X_2, \ldots, X_n$$

muodostavat (yksinkertaisen) **satunnaisotoksen** jakaumasta $f(x)$ ja kutsumme satunnaismuuttujia $X_1, X_2, \ldots, X_n$ **havainnoiksi**.

Oletetaan, että satunnaismuuttujat

$$X_1, X_2, \ldots, X_n$$

saavat *otannan tuloksena* havaituiksi arvoikseen **havaintoarvot**

$$x_1, x_2, \ldots, x_n$$

Merkitään tätä seuraavalla tavalla:

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

Huomautuksia:

(i) Jos satunnaismuuttujat

$$X_1, X_2, \ldots, X_n$$

muodostavat *satunnaisotoksen* jakaumasta $f(x)$, niin niiden havaitut arvot

$$x_1, x_2, \ldots, x_n$$

on saatu *toistamalla toisistaan riippumattomia arvontoja n kertaa samoin*, *jakaumasta* $f(x)$ *saatavin todennäköisyyksin*.

(ii) Satunnaismuuttujien $X_1, X_2, \ldots, X_n$ havaitut arvot $x_1, x_2, \ldots, x_n$ ovat *kiinteitä* eli *ei-satunnaisia*, mutta ne *vaihtelevat toisistaan riippumatta* ja *satunnaisesti otoksesta toiseen*.

(iii) Satunnaisuus otoksessa *liittyy siis siihen*, *miten havaintoarvot vaihtelevat otoksesta toiseen*.

(iv) Satunnaisuus *ei siis liity otannan tuloksena saatuihin havaintoarvoihin* (jotka ovat kiinteitä lukuja!), *vaan otoksen poimintatapaan*.

Yksinkertaisen satunnaisotoksen tilastollinen malli

Olkoon

$$X_1, X_2, \ldots, X_n$$

satunnaisotos satunnaismuuttujan X jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio on $f(x)$. Tällöin satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *yhteisjakauma* muodostaa **tilastollisen mallin** *havaintoarvojen satunnaiselle vaihtelulle otoksesta toiseen*.

Koska satunnaismuuttujat

$$X_1, X_2, \ldots, X_n$$

on tässä oletettu riippumattomiksi, niin satunnaismuuttujien $X_1, X_2, \ldots, X_n$ *yhteisjakauma* saadaan satunnaismuuttujien $X_1, X_2, \ldots, X_n$ reunajakaumien tulona:

$$f(x_1, x_2, \ldots, x_n) = f(x_1)f(x_2)\cdots f(x_n)$$

jossa

$$X_i \sim f(x_i),\ i = 1, 2, \ldots, n$$

Otostunnusluvut

Olkoon

$$X_1, X_2, \ldots, X_n$$

satunnaisotos satunnaismuuttujan X jakaumasta, jonka pistetodennäköisyys- tai tiheysfunktio on $f(x)$. Tällöin *havainnot* $X_1, X_2, \ldots, X_n$ ovat *riippumattomia*, *identtisesti jakautuneita satunnaismuuttujia*, joilla on *sama pistetodennäköisyys-* tai *tiheysfunktio* $f(x)$:

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim f(x),\ i = 1, 2, \ldots, n$$

Olkoon

$$T = g(X_1, X_2, \ldots, X_n)$$

jokin *satunnaismuuttujien* $X_1, X_2, \ldots, X_n$ (mitallinen) *funktio*. Tällöin satunnaismuuttujaa T kutsutaan (**otos**-) **tunnusluvuksi**.

Oletetaan, että satunnaismuuttujat $X_1, X_2, \ldots, X_n$ saavat otannan tuloksena havaituiksi arvoikseen *havaintoarvot* $x_1, x_2, \ldots, x_n$:

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$$

Tällöin tunnusluku

$$T = g(X_1, X_2, \ldots, X_n)$$

saa havaituksi arvokseen t funktion g arvon pisteessä $(x_1, x_2, \ldots, x_n)$:

$$t = g(x_1, x_2, \ldots, x_n)$$

Tunnusluku T on satunnaismuuttujien $X_1, X_2, \ldots, X_n$ funktiona satunnaismuuttuja. Tunnusluvun T jakaumaa kutsutaan *tunnusluvun T* **otosjakaumaksi**. Tunnusluvun T otosjakauma muodostaa **tilastollisen mallin** *tunnusluvun T havaittujen arvojen satunnaiselle vaihtelulle otoksesta toiseen*.

1.2. Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Odotusarvot ja varianssit

Satunnaisotos

Olkoon

$$X_1, X_2, \ldots , X_n$$

satunnaisotos jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2. Tällöin *havainnot* $X_1, X_2, \ldots ,$ X_n ovat *riippumattomia* satunnaismuuttujia, joiden odotusarvo on μ ja varianssi on σ^2:

$$X_1, X_2, \ldots, X_n \perp$$
$$\mathrm{E}(X_i) = \mu \text{ , } i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1, 2, \ldots, n$$

Aritmeettinen keskiarvo

Määritellään havaintojen $X_1, X_2, \ldots , X_n$ **aritmeettinen keskiarvo** kaavalla

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

Aritmeettinen keskiarvo kuvaa *havaintoarvojen keskimääräistä arvoa*. Aritmeettinen keskiarvo on satunnaismuuttujien $X_1, X_2, \ldots , X_n$ funktiona *satunnaismuuttuja*, jonka arvo vaihtelee satunnaisesti otoksesta toiseen.

Otosvarianssi

Määritellään havaintojen $X_1, X_2, \ldots , X_n$ **otosvarianssi** kaavalla

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$$

jossa

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

on havaintojen $X_1, X_2, \ldots , X_n$ *aritmeettinen keskiarvo*.

Otosvarianssi kuvaa *havaintoarvojen vaihtelua* niiden aritmeettisen keskiarvon ympärillä. Otosvarianssi on satunnaismuuttujien $X_1, X_2, \ldots , X_n$ funktiona *satunnaismuuttuja*, jonka arvo vaihtelee satunnaisesti otoksesta toiseen.

Otosvarianssin kaava voidaan kirjoittaa myös muotoihin

$$s^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right)$$

Perustelu:

$$
\begin{aligned}
s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2-2\bar{X}X_i+\bar{X}^2) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-2\bar{X}\sum_{i=1}^{n}X_i+\sum_{i=1}^{n}\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-2n\bar{X}^2+n\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right) \\
&= \frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^2-\frac{1}{n}\left(\sum_{i=1}^{n}X_i\right)^2\right)
\end{aligned}
$$

■

Otoskeskihajonta

Määritellään havaintojen $X_1, X_2, \ldots, X_n$ (**otos-**) **keskihajonta** kaavalla

$$
s=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2}
$$

Otoskeskihajonta kuvaa *havaintoarvojen vaihtelua* niiden aritmeettisen keskiarvon ympärillä. Otoskeskihajonta on satunnaismuuttujien $X_1, X_2, \ldots, X_n$ funktiona *satunnaismuuttuja*, jonka arvo vaihtelee satunnaisesti otoksesta toiseen.

Huomaa, että havaintoarvojen otoskeskihajonta on mitattu samoissa yksiköissä kuin niiden aritmeettinen keskiarvo toisin kuin havaintoarvojen otosvarianssi, jonka mittayksikkönä on havaintoarvojen mittayksikön neliö.

Aritmeettisen keskiarvon odotusarvo ja varianssi

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2. Tällöin havaintojen $X_1, X_2, \ldots, X_n$ *aritmeettisen keskiarvon*

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

odotusarvo on

$$
\mathrm{E}(\bar{X})=\mu
$$

ja *varianssi* on

$$
\mathrm{Var}(\bar{X})=\mathrm{D}^2(\bar{X})=\frac{\sigma^2}{n}
$$

Perustelu:

Olkoot $X_1, X_2, \dots, X_n$ *riippumattomia* satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu \text{ , } i = 1, 2, \ldots, n$$

$$\mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1, 2, \ldots, n$$

Tällöin

$$\begin{aligned}
\mathrm{E}(\bar{X}) &= \mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n} \mu \\
&= \frac{1}{n} n\mu = \mu
\end{aligned}$$

ja

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}(X_i) && \mid \text{Havaintojen riippumattomuus} \\
&= \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2 \\
&= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}
\end{aligned}$$

■

Huomautuksia:

(i) *Yhtälö*

$$\mathrm{E}(\bar{X}) = \mu$$

pätee, *vaikka havainnot* $X_1, X_2, \dots, X_n$ *eivät olisi riippumattomia*, koska summan odotusarvo on *aina* summan tekijöiden odotusarvojen summa.

(ii) Sen sijaan yhtälön

$$\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

todistamisessa joudutaan vetoamaan siihen, että *riippumattomille satunnaismuuttujille pätee se*, *että summan varianssi on summan tekijöiden varianssien summa*.

Aritmeettisen keskiarvon $\bar{X}$ *standardipoikkeamaa*

$$\mathrm{D}(\bar{X}) = \frac{\sigma}{\sqrt{n}}$$

kutsutaan usein **keskiarvon keskivirheeksi** ja se kuvaa aritmeettisen keskiarvon otosvaihtelua oman odotusarvonsa μ ympärillä.

Standardoidun aritmeettisen keskiarvon odotusarvo ja varianssi

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2, jolloin havaintojen $X_1, X_2, \ldots, X_n$ aritmeettisen keskiarvon

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

odotusarvo on

$$\mathrm{E}(\bar{X}) = \mu$$

ja varianssi on

$$\mathrm{Var}(\bar{X}) = \mathrm{D}^2(\bar{X}) = \frac{\sigma^2}{n}$$

Tällöin *standardoidun satunnaismuuttujan*

$$Z = \frac{\bar{X} - \mathrm{E}(\bar{X})}{\mathrm{D}(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

odotusarvo ja *varianssi* ovat

$$\mathrm{E}(Z) = 0$$
$$\mathrm{Var}(Z) = 1$$

Otosvarianssin odotusarvo ja varianssi

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ, varianssi on σ^2 ja 4. keskusmomentti on μ_4. Tällöin havaintojen $X_1, X_2, \ldots, X_n$ *otosvarianssin*

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

odotusarvo on

$$\mathrm{E}(s^2) = \sigma^2$$

ja *varianssi* on

$$\mathrm{Var}(s^2) = \frac{1}{n-1}\left[\left(1-\frac{1}{n}\right)\mu_4 - \left(1-\frac{3}{n}\right)\sigma^4\right]$$

Suurille n pätee approksimatiivisesti

$$\mathrm{Var}(s^2) =_a \frac{1}{n-1}(\mu_4 - \sigma^4)$$

Jos lisäksi voidaan olettaa, että havainnot $X_1, X_2, \ldots, X_n$ ovat *normaalijakautuneita*, niin

$$\mathrm{Var}(s^2) = \frac{2\sigma^4}{n-1}$$

Perustelu:

Todistamme tässä vain otosvarianssin odotusarvoa koskeva tuloksen.

Olkoot $X_1, X_2, \ldots, X_n$ riippumattomia satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu \,,\ i = 1, 2, \ldots, n$$

$$\mathrm{Var}(X_i) = \sigma^2 \,,\ i = 1, 2, \ldots, n$$

Todetaan ensin, että

$$\begin{aligned}
\sum_{i=1}^n (X_i - \mu)^2 &= \sum_{i=1}^n [(X_i - \bar{X}) + (\bar{X} - \mu)]^2 \\
&= \sum_{i=1}^n (X_i - \bar{X})^2 + 2\sum_{i=1}^n (X_i - \bar{X})(\bar{X} - \mu) + \sum_{i=1}^n (\bar{X} - \mu)^2 \\
&= (n-1)s^2 + n(\bar{X} - \mu)^2
\end{aligned}$$

koska

$$s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$$

ja

$$\sum_{i=1}^n (X_i - \bar{X}) = \sum_{i=1}^n X_i - n\bar{X} = 0$$

Siten

$$\begin{aligned}
\mathrm{E}(s^2) &= \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}\mathrm{E}\left[\sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^n \mathrm{E}(X_i - \mu)^2 - n\,\mathrm{E}(\bar{X} - \mu)^2\right] \\
&= \frac{1}{n-1}\left[\sum_{i=1}^n \sigma^2 - n\frac{\sigma^2}{n}\right] \\
&= \frac{1}{n-1}(n\sigma^2 - \sigma^2) \\
&= \frac{n-1}{n-1}\sigma^2 = \sigma^2
\end{aligned}$$

koska

$$\mathrm{E}(X_i - \mu)^2 = \mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1, 2, \ldots, n$$

ja

$$\mathrm{E}(\bar{X} - \mu)^2 = \mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

■

1.3. Aritmeettisen keskiarvon ja otosvarianssin otosjakaumat: Otos normaalijakaumasta

Aritmeettisen keskiarvon otosjakauma, kun otos on normaalijakautunut

Olkoon $X_1, X_2, \ldots , X_n$ *satunnaisotos* normaalijakaumasta $\mathrm{N}(\mu,\sigma^2)$. Tällöin havaintojen $X_1, X_2, \ldots , X_n$ aritmeettinen keskiarvo

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

noudattaa **normaalijakaumaa** $\mathrm{N}(\mu,\sigma^2/n)$:

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Perustelu:

Olkoot $X_1, X_2, \ldots , X_n$ riippumattomia ja samaa normaalijakaumaa noudattavia satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu \text{ , } i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1, 2, \ldots, n$$

Edellä on todettu, että *ilman normaalisuusoletustakin pätee*

$$\mathrm{E}(\bar{X}) = \mu$$

$$\mathrm{Var}(\bar{X}) = \mathrm{D}^2(\bar{X}) = \frac{\sigma^2}{n}$$

Lauseen todistamiseksi pitää siis näyttää, että satunnaismuuttujien $X_1, X_2, \ldots , X_n$ aritmeettisen keskiarvon $\bar{X}$ jakauma on normaalijakauma. Käytämme todistamisessa hyväksi normaalijakauman *momenttiemäfunktiota*.

Jos

$$X \sim \mathrm{N}(\mu,\sigma^2)$$

niin satunnaismuuttujan X momenttiemäfunktio on

$$m_X(t) = \mathrm{E}(e^{tX}) = \exp(\mu t + \tfrac{1}{2}\sigma^2 t^2)$$

Momenttiemäfunktion yleisistä ominaisuuksista seuraa, että riippumattomien ja samoin jakautuneiden satunnaismuuttujien $X_1, X_2, \ldots, X_n$ aritmeettisen keskiarvon $\bar{X}$ momenttiemäfunktio on muotoa

$$m_{\bar{X}}(t) = [m(t/n)]^n$$

jossa $m(t)$ on satunnaismuuttujien $X_1, X_2, \ldots, X_n$ momenttiemäfunktio.

Siten riippumattomien samaa normaalijakaumaa $\mathrm{N}(\mu,\sigma^2)$ noudattavien satunnaismuuttujien momenttiemäfunktioksi saadaan

$$\begin{aligned} m_{\bar{X}}(t) &= [m(t/n)]^n \\ &= [\exp(\mu(t/n) + \tfrac{1}{2}\sigma^2(t/n)^2)]^n \\ &= [\exp(n(\mu t/n + \tfrac{1}{2}\sigma^2(t/n)^2))] \\ &= [\exp(\mu t + \tfrac{1}{2}(\sigma^2/n)t^2)] \end{aligned}$$

mikä on normaalijakauman $\mathrm{N}(\mu,\sigma^2/n)$ momenttiemäfunktio. Siten

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

■

Standardoidun aritmeettisen keskiarvon otosjakauma, kun otos on normaalijakautunut

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* normaalijakaumasta $\mathrm{N}(\mu,\sigma^2)$, jolloin havaintojen $X_1, X_2, \ldots, X_n$ aritmeettinen keskiarvo

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

noudattaa normaalijakaumaa $\mathrm{N}(\mu,\sigma^2/n)$:

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Tällöin *standardoitu satunnaismuuttuja*

$$Z = \frac{\bar{X} - \mathrm{E}(\bar{X})}{\mathrm{D}(\bar{X})} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

noudattaa *standardoitua normaalijakaumaa* N(0,1):

$$Z \sim \mathrm{N}(0,1)$$

Otosvarianssin otosjakauma, kun otos on normaalijakautunut

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* normaalijakaumasta $\mathrm{N}(\mu,\sigma^2)$. Tällöin satunnaismuuttuja

$$Y = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2$$

noudattaa **χ^2-jakaumaa** vapausastein n:

$$Y \sim \chi^2(n)$$

Perustelu:

Olkoot $X_1, X_2, \ldots , X_n$ riippumattomia ja samaa normaalijakaumaa noudattavia satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu \,,\; i = 1, 2, \ldots, n$$

$$\mathrm{Var}(X_i) = \sigma^2 \,,\; i = 1, 2, \ldots, n$$

Määritellään satunnaismuuttuja Y kaavalla

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

Koska havainnot $X_1, X_2, \ldots , X_n$ ovat *riippumattomia* ja noudattavat *samaa* normaalijakaumaa $\mathrm{N}(\mu,\sigma^2)$, niin *standardoidut* satunnaismuuttujat

$$Y_i = \frac{X_i - \mu}{\sigma} \,,\; i = 1, 2, \ldots, n$$

ovat riippumattomia ja *noudattavat standardoitua normaalijakaumaa* N(0,1):

$$Y_i \sim \mathrm{N}(0,1) \,,\; i = 0, 1, 2, \ldots, n$$

Siten satunnaismuuttuja Y on *riippumattomien*, *standardoitua normaalijakaumaa* N(0,1) *noudattavien satunnaismuuttujien* Y_i , $i = 1, 2, \ldots , n$ *neliösumma*:

$$Y = \sum_{i=1}^{n} Y_i^2$$

Suoraan χ^2-jakauman määritelmästä seuraa, että satunnaismuuttuja Y noudattaa χ^2*-jakaumaa* vapausastein n:

$$Y \sim \chi^2(n)$$

■

Olkoon $X_1, X_2, \ldots , X_n$ *satunnaisotos* normaalijakaumasta $\mathrm{N}(\mu,\sigma^2)$. Tällöin satunnaismuuttuja

$$V = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2$$

noudattaa **χ^2-jakaumaa** vapausastein $(n - 1)$:

$$V \sim \chi^2(n-1)$$

Perustelu:

Olkoot $X_1, X_2, \ldots , X_n$ riippumattomia ja samaa normaalijakaumaa noudattavia satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu \,,\; i = 1, 2, \ldots, n$$

$$\mathrm{Var}(X_i) = \sigma^2 \,,\; i = 1, 2, \ldots, n$$

Määritellään satunnaismuuttuja V kaavalla

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{(n-1)s^2}{\sigma^2}$$

Satunnaismuuttujat

$$U_i = \frac{X_i - \bar{X}}{\sigma}, \ i = 1, 2, \ldots, n$$

eivät ole riippumattomia, koska

$$\sum_{i=1}^{n} U_i = 0$$

Voidaan kuitenkin osoittaa, että satunnaismuuttuja V voidaan esittää *riippumattomien, standardoitua normaalijakaumaa* N(0,1) *noudattavien satunnaismuuttujien*

V_i , $i = 1, 2, \ldots , n - 1$

neliösummana (ks. alla esitettyä perustelua normaalijakautuneen otoksen aritmeettisen keskiarvon ja otosvarianssin riippumattomuudelle):

$$V = \sum_{i=1}^{n-1} V_i^2$$

Suoraan χ^2-jakauman määritelmästä seuraa, että satunnaismuuttuja V noudattaa χ^2-*jakaumaa* vapausastein $(n - 1)$:

$$Y \sim \chi^2(n-1)$$

■

Huomautuksia:

(i) Satunnaismuuttuja

$$Y = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2$$

noudattaa χ^2-jakaumaa, jonka vapausasteiden lukumäärä on sama kuin havaintojen lukumäärä n.

(ii) Kun satunnaismuuttujasta Y siirrytään satunnaismuuttujaan

$$V = \sum_{i=1}^{n} \left(\frac{X_i - \bar{X}}{\sigma} \right)^2 = \frac{(n-1)s^2}{\sigma^2}$$

menetetään yksi vapausaste.

(iii) Yhden vapausasteen menetys on seurausta siitä, että kun *riippumattomista* satunnaismuuttujista

$$Y_i = \frac{X_i - \mu}{\sigma}, \ i = 1, 2, \ldots, n$$

siirrytään satunnaismuuttujiin

$$U_i = \frac{X_i - \bar{X}}{\sigma}\,,\; i = 1, 2, \ldots, n$$

korvaamalla odotusarvoparametri μ *estimaattorillaan* $\bar{X}$, syntyy yksi (lineaarinen) side-ehto

$$\sum_{i=1}^{n} U_i = 0$$

(iv) Oletukset havaintojen *riippumattomuudesta* ja *samasta jakaumasta* ovat välttämättömiä otosvarianssin eksaktia eli tarkkaa otosjakaumaa koskevalle tulokselle.

Aritmeettisen keskiarvon ja otosvarianssin riippumattomuus

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* normaalijakaumasta $\mathrm{N}(\mu,\sigma^2)$ ja olkoon havaintojen $X_1, X_2, \ldots,$ X_n aritmeettinen keskiarvo

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

ja otosvarianssi

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Tällöin

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$V = \frac{(n-1)s^2}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{\sigma}\right)^2 \sim \chi^2(n-1)$$

Lisäksi $\bar{X}$ ja s^2 ovat *riippumattomia* ja

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} \sim t(n-1)$$

Perustelu:

Olkoot $X_1, X_2, \ldots, X_n$ riippumattomia ja samaa normaalijakaumaa noudattavia satunnaismuuttujia, joille

$$\mathrm{E}(X_i) = \mu\,,\; i = 1, 2, \ldots, n$$

$$\mathrm{Var}(X_i) = \sigma^2\,,\; i = 1, 2, \ldots, n$$

Olkoot havaintojen $X_1, X_2, \ldots, X_n$ aritmeettinen keskiarvo

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

ja otosvarianssi

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Otoksen $X_1, X_2, \ldots, X_n$ *yhteisjakauman tiheysfunktio* voidaan kirjoittaa havaintojen riippumattomuuden ja normaalisuuden takia seuraavaan muotoon:

$$f(x_1, x_2, \ldots, x_n) = (2\pi)^{-\frac{1}{2}n}\sigma^{-n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

Määritellään *lineaarimuunnos*

$$\begin{cases} Y_1 = \frac{1}{\sqrt{n}}X_1 + \frac{1}{\sqrt{n}}X_2 + \frac{1}{\sqrt{n}}X_3 + \cdots + \frac{1}{\sqrt{n}}X_n \\ Y_2 = \frac{1}{\sqrt{2}}X_1 - \frac{1}{\sqrt{2}}X_2 \\ Y_3 = \frac{1}{\sqrt{6}}X_1 + \frac{1}{\sqrt{6}}X_2 - \frac{2}{\sqrt{6}}X_3 \\ \quad\vdots \\ Y_n = \frac{1}{\sqrt{n(n-1)}}X_1 + \frac{1}{\sqrt{n(n-1)}}X_2 + \frac{1}{\sqrt{n(n-1)}}X_3 + \cdots - \frac{n-1}{\sqrt{n(n-1)}}X_n \end{cases}$$

Muunnos voidaan esittää *matriisein* muodossa

$$\mathbf{Y} = \mathbf{BX}$$

jossa

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$$
$$\mathbf{X} = (X_1, X_2, \ldots, X_n)$$

ja $n\times n$-matriisi

$$\mathbf{B} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \cdots & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \frac{1}{\sqrt{n(n-1)}} & \cdots & -\frac{n-1}{\sqrt{n(n-1)}} \end{bmatrix}$$

on ortogonaalinen:

$$\mathbf{B}'\mathbf{B} = \mathbf{BB}' = \mathbf{I}$$

Matriisi **B** nähdään helposti ortogonaaliseksi seuraavalla päättelyllä:

Määritellään $n\times n$-matriisi

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & -1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & -2 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & -3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & 1 & 1 & \cdots & -(n-2) & 0 \\ 1 & 1 & 1 & 1 & \cdots & 1 & -(n-1) \end{bmatrix}$$

On helppo nähdä, että matriisin **C** rivit ovat selvästi kohtisuorassa toisiaan vastaan. Matriisi **B** saadaan matriisista **C** normeeraamalla sen rivit niin, että niiden pituudeksi tulee 1.

Koska muunnos

$$\mathbf{Y} = \mathbf{BX}$$

on ortogonaalinen, niin muunnosta vastaavan *Jacobin determinantin* itseisarvo = 1.

Koska

$$Y_1 = \tfrac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n) = \sqrt{n}\bar{X}$$

ja

$$\begin{aligned} Y_1^2 + Y_2^2 + \cdots + Y_n^2 &= \mathbf{Y'Y} = \mathbf{X'B'BX} = \mathbf{X'X} = X_1^2 + X_2^2 + \cdots + X_n^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n\bar{X}^2 \end{aligned}$$

niin

$$Y_2^2 + \cdots + Y_n^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 = (n-1)s^2$$

Koska

$$\begin{aligned} \sum_{i=1}^{n}(X_i - \mu)^2 &= \sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n}(\bar{X} - \mu)^2 \\ &= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2 \\ &= Y_2^2 + \cdots + Y_n^2 + (Y_1 - \sqrt{n}\mu)^2 \end{aligned}$$

niin satunnaismuuttujien $Y_1, Y_2, \ldots , Y_n$ yhteisjakauman tiheysfunktioksi saadaan

$$\begin{aligned} f(y_1, y_2, \ldots, y_n) &= \frac{1}{(2\pi)^{\frac{1}{2}n}\sigma^n} e^{-\frac{1}{2\sigma^2}\left[(Y_1 - \sqrt{n}\mu)^2 + Y_2^2 + \cdots + Y_n^2\right]} \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(Y_1 - \sqrt{n}\mu)^2} \times \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}Y_2^2} \times \cdots \times \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}Y_n^2} \end{aligned}$$

Edellä esitetystä seuraa, että satunnaismuuttujat

$$Y_1, Y_2, \ldots , Y_n$$

ovat riippumattomia ja normaalijakautuneita:

$$\begin{aligned} &Y_1, Y_2, \ldots, Y_n \perp \\ &Y_1 = \sqrt{n}\bar{X} \sim \mathrm{N}(\sqrt{n}\mu, \sigma^2) \\ &Y_i \sim \mathrm{N}(0, \sigma^2)\ ,\ i = 2, \ldots, n \end{aligned}$$

Lisäksi

$$s^2 = \frac{1}{n-1}(Y_2^2 + \cdots + Y_n^2) = \frac{\sigma^2}{n-1}\left[\left(\frac{Y_2}{\sigma}\right)^2 + \cdots + \left(\frac{Y_n}{\sigma}\right)^2\right]$$

jossa

$$\frac{Y_i}{\sigma} \sim \mathrm{N}(0,1)\,,\ i = 2,\ldots,n$$

Siten olemme todistaneet, että

$$\bar{X} \perp s^2$$

$$\bar{X} \sim \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

Todistetaan vielä, että

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

Edellä todistetusta aritmeettista keskiarvoa $\bar{X}$ koskevasta jakaumatuloksesta seuraa, että

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

Lisäksi

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

ja

$$\bar{X} \perp s^2$$

Suoraan *Studentin t-jakauman* määritelmästä seuraa, että

$$t = \frac{\bar{X}-\mu}{s/\sqrt{n}} = \frac{\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{1}{n-1}\left(\dfrac{(n-1)s^2}{\sigma^2}\right)}} \sim t(n-1)$$

■

1.4. Aritmeettisen keskiarvon otosjakauma: Suurten otosten tuloksia

Suurten lukujen laki

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2. Määritellään havaintojen $X_1, X_2, \ldots, X_n$ *aritmeettinen keskiarvo* kaavalla

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

Heikon suurten lukujen lain mukaan satunnaismuuttujien $X_1, X_2, \ldots, X_n$ aritmeettisten keskiarvojen $\bar{X}_n$ muodostama jono *konvergoi* havaintojen lukumäärän n kasvaessa rajatta *stokastisesti* kohti satunnaismuuttujien yhteistä odotusarvoa μ:

$$\bar{X}_n \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

Perustelu:

Olkoon X_i , $i = 1, 2, 3, \ldots$ jono *riippumattomia* satunnaismuuttujia, joilla on *sama* odotusarvo μ ja varianssi on σ^2:

$$X_1, X_2, X_3, \ldots \perp$$
$$\mathrm{E}(X_i) = \mu \text{ , } i = 1, 2, 3, \ldots$$
$$\mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1, 2, 3, \ldots$$

Olkoon

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

satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ *aritmeettinen keskiarvo*.

Tshebyshevin epäyhtälön mukaan

$$\Pr\left(\left|\bar{X}_n - \mu\right| > \varepsilon\right) \le \frac{\sigma^2}{n\varepsilon^2}$$

Koska epäyhtälön oikea puoli $\to 0$, kun $n \to \infty$, niin

$$\bar{X}_n \xrightarrow[n\to\infty]{\mathrm{P}} \mu$$

■

Huomautuksia:

(i) *Heikko suurten lukujen laki* voidaan ilmaista sanoin seuraavasti:

Samoin jakautuneiden satunnaismuuttujien *aritmeettinen keskiarvo lähestyy muuttujien lukumäärän kasvaessa muuttujien yhteistä odotusarvoa* sellaisella tavalla, että (**suurten**) **poikkeamien todennäköisyys satunnaismuuttujien yhteisestä odotusarvosta tulee yhä pienemmäksi eli** (**suuret**) **poikkeamat tulevat yhä *harvinaisemmiksi***.

(ii) Suurten lukujen lakeja voidaan pitää matemaattisena formulointina empiirisen todennäköisyyden käsitteeseen liittyvälle **tilastollisen stabiliteetin** käsitteelle.

(iii) Suurten lukujen laeista on olemassa *yleisempiä muotoja*, joissa voidaan lieventää *samoinjakautuneisuus*- ja *riippumattomuusoletuksia*.

Keskeinen raja-arvolause

Olkoon $X_1, X_2, \ldots , X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2. Olkoon

$$Y_n = \sum_{i=1}^{n} X_i$$

havaintojen $X_1, X_2, \ldots , X_n$ *summa*. Summan Y_n odotusarvo ja varianssi ovat

$$\mathrm{E}(Y_n) = n\mu$$
$$\mathrm{D}^2(Y_n) = n\sigma^2$$

Standardoidaan summa Y_n :

$$Z_n = \frac{Y_n - \mathrm{E}(Y_n)}{\mathrm{D}(Y_n)} = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

Annetaan havaintojen lukumäärän n kasvaa rajatta. Tällöin *satunnaismuuttujan* Z_n *jakauma konvergoi* (jakaumakonvergenssin mielessä) *kohti standardoitua normaalijakaumaa* N(0,1):

$$\lim_{n\to\infty} \Pr\left(\frac{Y_n - n\mu}{\sigma\sqrt{n}} \le z\right) = \Phi(z)$$

jossa Φ on standardoidun normaalijakauman N(0,1) kertymäfunktio.

Merkitsemme tätä

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} \sim_a \mathrm{N}(0,1)$$

ja sanomme: Standardoitu satunnaismuuttuja Z_n noudattaa **asymptoottisesti standardoitua normaalijakaumaa N(0,1)**.

Perustelu:

Olkoon X_1, X_2, X_3, … jono *riippumattomia*, *samoin jakautuneita* satunnaismuuttujia, joiden odotusarvo on μ ja varianssi on σ^2:

$$X_1, X_2, X_3, \ldots \perp$$
$$\mathrm{E}(X_i) = \mu \text{ , } i = 1,2,3,\ldots$$
$$\mathrm{Var}(X_i) = \sigma^2 \text{ , } i = 1,2,3,\ldots$$

Oletetaan, että satunnaismuuttujien X_i , i = 1, 2, 3, … *momenttiemäfunktio*

$$m(t) = \mathrm{E}(e^{tX_i}) \text{ , } i = 1,2,3,\ldots$$

on olemassa jossakin origon ympäristössä.

Olkoon

$$Y_n = \sum_{i=1}^{n} X_i$$

satunnaismuuttujien X_i , i = 1, 2, … , *n summa*. Summamuuttujan Y_n odotusarvo ja varianssi ovat

$$\mathrm{E}(Y_n) = n\mu$$
$$\mathrm{D}^2(Y_n) = n\sigma^2$$

Standardoidaan summa Y_n :

$$Z_n = \frac{Y_n - \mathrm{E}(Y_n)}{\mathrm{D}(Y_n)} = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

Standardoidun satunnaismuuttujan Z_n odotusarvo ja varianssi ovat

$$\mathrm{E}(Z_n) = 0$$
$$\mathrm{Var}(Z_n) = 1$$

Siirrytään tarkastelemaan *keskistettyjä satunnaismuuttujia*

$$T_i = X_i - \mu \text{ , } i = 1, 2, 3\ldots$$

Satunnaismuuttujien T_i odotusarvo ja varianssi ovat

$$\mathrm{E}(T_i) = 0 \text{ , } i = 1, 2, 3, \ldots$$

$$\mathrm{D}^2(T_i) = \sigma^2 \text{ , } i = 1, 2, 3, \ldots$$

Keskistettyjen muuttujien T_i avulla standardoitu muuttuja Z_n voidaan kirjoittaa muotoon

$$\begin{aligned} Z_n &= \frac{Y_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \\ &= \frac{1}{\sigma\sqrt{n}}(T_1 + T_2 + \cdots + T_n) \end{aligned}$$

Koska satunnaismuuttujien X_i , $i = 1, 2, 3, \ldots$ momenttiemäfunktio on olemassa jossakin origon ympäristössä myös keskitettyjen muuttujien

$$T_i = X_i - \mu \text{ , } i = 1, 2, 3\ldots$$

momenttiemäfunktio on olemassa jossakin origon ympäristössä.

Olkoon

$$m(t) = \mathrm{E}(e^{tT_i})$$

satunnaismuuttujien T_i , $i = 1, 2, 3, \ldots$ momenttiemäfunktio.

Koska *riippumattomien satunnaismuuttujien summan momenttiemäfunktio on summan tekijöiden momenttiemäfunktioiden tulo*, niin satunnaismuuttujan

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}(T_1 + T_2 + \cdots + T_n)$$

momenttiemäfunktio $m_n(t)$ voidaan esittää muodossa

$$m_n(t) = \left[m\left(\frac{t}{\sigma\sqrt{n}} \right) \right]^n$$

jossa $m(t)$ on satunnaismuuttujien T_i , $i = 1, 2, 3, \ldots$ momenttiemäfunktio.

Satunnaismuuttujien T_i , $i = 1, 2, 3, \ldots$ momenttiemäfunktio $m(t)$ voidaan kehittää jossakin pisteen $t = 0$ ympäristössä *Taylorin sarjaksi*

$$m(t) = 1 + \alpha_1 t + \frac{\alpha_2}{2}t^2 + t^2\eta(t)$$

jossa

$$\alpha_k = \mathrm{E}(T_i^k) \text{ , } k = 1, 2, 3, \ldots$$

on satunnaismuuttujien T_i , $i = 1, 2, 3, \ldots$ k. origomomentti ja $\eta(t) \to 0$, kun $t \to 0$.

Koska

$$\alpha_1 = \mathrm{E}(T_i) = 0 \text{ , } i = 1,2,3,\ldots$$
$$\alpha_2 = \mathrm{E}(T_i^2) = \mathrm{D}^2(T_i) = \sigma^2 \text{ , } i = 1,2,3,\ldots$$

niin

$$m(t) = 1 + \alpha_1 t + \frac{\alpha_2}{2}t^2 + t^2\eta(t) = 1 + \frac{\sigma^2}{2}t^2 + t^2\eta(t)$$

Sijoitetaan satunnaismuuttujien T_i , i = 1, 2, 3, … momenttiemäfunktion $m(t)$ sarjakehitelmä

$$m(t) = 1 + \frac{\sigma^2}{2}t^2 + t^2\eta(t)$$

satunnaismuuttujan

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}} = \frac{1}{\sigma\sqrt{n}}(T_1 + T_2 + \cdots + T_n)$$

momenttiemäfunktion lausekkeeseen

$$m_n(t) = \left[m\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n$$

Saamme sijoituksen tuloksena lausekkeen

$$\begin{aligned} m_n(t) &= \left[1 + \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 \eta\left(\frac{t}{\sigma\sqrt{n}}\right) \right]^n \\ &= \left[1 + \frac{1}{n}\left(\frac{t^2}{2} + \frac{t^2}{\sigma^2}\eta\left(\frac{t}{\sigma\sqrt{n}}\right) \right) \right]^n \end{aligned}$$

jossa

$$\lim_{n\to\infty} \eta\left(\frac{t}{\sigma\sqrt{n}}\right) = 0$$

jokaiselle kiinteälle t.

Eksponenttifunktion ominaisuuksien perusteella

$$m_n(t) = \left[1 + \frac{1}{n}\left(\frac{t^2}{2} + \frac{t^2}{\sigma^2}\eta\left(\frac{t}{\sigma\sqrt{n}}\right) \right) \right]^n \xrightarrow[n\to\infty]{} e^{t^2/2}$$

Koska

$$e^{t^2/2}$$

on *standardoidun normaalijakauman* N(0,1) *momenttiemäfunktio*, näemme, että satunnaismuuttujien

$$Z_n = \frac{Y_n - n\mu}{\sigma\sqrt{n}}$$

muodostama jono *konvergoi jakaumaltaan* eli *heikosti* kohti standardoitua normaalijakaumaa N(0,1):

$$Z_n \xrightarrow[n\to\infty]{\mathrm{L}} Z \sim \mathrm{N}(0,1)$$

■

Huomautuksia:

(i) Keskeisen raja-arvolauseen mukaan usean samoin jakautuneen satunnaismuuttujan *summa on* (*tietyin ehdoin*) *approksimatiivisesti normaalinen* (*lähes*) *riippumatta yhteenlaskettavien jakaumasta*. Yhteenlaskettavien ei tarvitse olla edes *jatkuvia*, vaan ne voivat olla jopa *diskreettejä*.

(ii) Approksimaation *hyvyys riippuu* yhteenlaskettavien satunnaismuuttujien lukumäärästä, niiden jakaumasta ja erityisesti niiden jakauman *vinoudesta*:

– Approksimaation *hyvyys paranee*, kun yhteenlaskettavien satunnaismuuttujien *lukumäärä kasvaa*.

– Jos yhteenlaskettavien satunnaismuuttujien jakauma on *symmetrinen* (tai lähellä symmetristä) approksimaatio on suhteellisen hyvä jo melko pienillä yhteenlaskettavien lukumäärillä.

– Jos yhteenlaskettavien satunnaismuuttujien jakauma on *epäsymmetrinen*, hyvä approksimaatio vaatii selvästi enemmän yhteenlaskettavia.

(iii) Keskeisestä raja-arvolauseesta on olemassa *yleisempiä muotoja*, joissa voidaan lieventää *samoinjakautuneisuus*- ja *riippumattomuusoletuksia*.

Standardoidun aritmeettisen keskiarvon asymptoottinen jakauma

Olkoon $X_1, X_2, \ldots, X_n$ *satunnaisotos* jakaumasta, jonka odotusarvo on μ ja varianssi on σ^2. Olkoon

$$Y_n = \sum_{i=1}^{n} X_i$$

havaintojen $X_1, X_2, \ldots, X_n$ *summa*, jolloin havaintojen $X_1, X_2, \ldots, X_n$ *aritmeettinen keskiarvo* saadaan kaavalla

$$\bar{X}_n = \frac{1}{n} Y_n$$

Standardoidaan satunnaismuuttuja $\bar{X}_n$:

$$Z_n = \frac{\bar{X} - \mathrm{E}(\bar{X})}{\mathrm{D}(\bar{X})} = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}}$$

Koska

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma / \sqrt{n}} = \frac{Y_n - n\mu}{\sigma \sqrt{n}}$$

keskeisen raja-arvolauseen tulos ilmaistaan usein seuraavassa muodossa: *Standardoidun aritmeettisen keskiarvon* $\bar{X}_n$ *jakauma konvergoi havaintojen lukumäärän n kasvaessa rajatta kohti standardoitua normaalijakaumaa* N(0,1).

Aritmeettisen keskiarvon approksimatiivinen jakauma

Keskeisen raja-arvolauseesta seuraa, että riippumattomien, samoin jakautuneiden satunnaismuuttujien X_i , $i = 1, 2, \ldots , n$ aritmeettinen keskiarvo $\bar{X}_n$ on suurille (mutta äärellisille) havaintojen lukumäärille n *approksimatiivisesti normaalinen parametrein* μ *ja* σ^2/n:

$$\bar{X}_n \sim_a \mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

Huomaa, että tarkasti ottaen *emme saisi sanoa* (kuten usein tehdään), *että aritmeettinen keskiarvo* $\bar{X}_n$ *on asymptoottisesti normaalinen parametrein* μ *ja* σ^2/n, koska aritmeettisen keskiarvon $\bar{X}_n$ otosjakauma

$$\mathrm{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

lähestyy havaintojen lukumäärän n kasvaessa rajatta yhden pisteen jakaumaa

$$\Pr(X = \mu) = 1$$

1.5. Suhteellisen frekvenssin otosjakauma

Frekvenssi ja suhteellinen frekvenssi

Olkoon P jokin otosavaruuden S alkioiden *ominaisuus*. Jos otosavaruuden S alkiolla x on ominaisuus P, merkitään

$$P(x)$$

Olkoon

$$A = \{x \in S \mid P(x)\}$$

niiden otosavaruuden S alkioiden osajoukko, joilla on ominaisuus P. Oletetaan, että tapahtuman A todennäköisyys on

$$\Pr(A) = p$$

Poimitaan otosavaruudesta S *satunnaisotos*, jonka *koko* on n ja olkoon

$$f$$

niiden havaintoyksiköiden **frekvenssi**, joilla on ominaisuus P. Tällöin

$$\hat{p} = \frac{f}{n}$$

on vastaava **suhteellinen frekvenssi**.

Frekvenssi f kuvaa A-tyyppisten alkioiden *lukumäärää* otoksessa ja vastaava suhteellinen frekvenssi $\hat{p} = f/n$ kuvaa A-tyyppisten alkioiden *suhteellista osuutta* otoksessa.

Sekä frekvenssi f että vastaava suhteellinen frekvenssi $\hat{p} = f/n$ ovat *satunnaismuuttujia*, joiden saamat *arvot vaihtelevat satunnaisesti otoksesta toiseen*.

Frekvenssin jakauma

Olkoon A jokin otosavaruuden S *tapahtuma*:

$$A \subset S$$

ja olkoon

$$\Pr(A) = p$$

Poimitaan otosavaruudesta S *satunnaisotos*, jonka *koko* on n ja olkoon

$$f$$

A-tyyppisten alkioiden *lukumäärä* eli *frekvenssi* otoksessa. Tällöin tapahtuman A frekvenssi f noudattaa **binomijakaumaa** parametrein n ja p:

$$f \sim \text{Bin}(n, p)$$

Binomijakauman ominaisuuksien perusteella satunnaismuuttujan f **odotusarvo** ja **varianssi** ovat

$$\text{E}(f) = np$$

$$\text{Var}(f) = \text{D}^2(f) = npq$$

jossa $q = 1 - p$.

Perustelu:

Olkoon A jokin otosavaruuden S *tapahtuma* ja olkoon

$$\Pr(A) = p$$

Tällöin tapahtuman A *komplementtitapahtuman* A^c (tapahtuma A *ei satu*) *todennäköisyys* on

$$\Pr(A^c) = 1 - \Pr(A) = 1 - p = q$$

Määritellään *diskreetti satunnaismuuttuja* X:

$$X = \begin{cases} 1, & \text{jos tapahtuma } A \text{ sattuu} \\ 0, & \text{jos tapahtuma } A \text{ ei satu} \end{cases}$$

Tällöin satunnaismuuttujan X *jakauma* on

$$\Pr(X = 1) = p$$

$$\Pr(X = 0) = 1 - p = q$$

Määritelmän mukaan satunnaismuuttuja X noudattaa **Bernoulli-jakaumaa** parametrinaan p:

$$X \sim \text{Bernoulli}(p)$$

On helppo nähdä, että

$$\text{E}(X) = p$$

$$\text{Var}(X) = \text{D}^2(X) = pq$$

Toistetaan *samaa* Bernoulli-koetta n kertaa, jossa n on *kiinteä*, *etukäteen päätetty* luku. Oletetaan, että koetoistot ovat *riippumattomia* ja tarkkaillaan tapahtuman A sattumista koetoistojen aikana.

Oletetaan, että

$$\Pr(A) = p$$

$$\Pr(A^c) = 1 - p = q$$

Määritellään *diskreetit satunnaismuuttujat* X_i , $i = 1, 2, \ldots , n$:

$$X_i = \begin{cases} 1, & \text{jos tapahtuma } A \text{ sattuu kokeessa } i \\ 0, & \text{jos tapahtuma } A \text{ ei satu kokeessa } i \end{cases}$$

Tällöin

$$X_i \sim \text{Bernoulli}(p) \text{ , } i = 1, 2, \ldots, n$$

Lisäksi satunnaismuuttujat X_i , $i = 1, 2, \ldots , n$ ovat *riippumattomia*:

$$X_1, X_2, \ldots, X_n \perp$$

Määritellään *diskreetti satunnaismuuttuja* f :

f = Tapahtuman A *frekvenssi* n-kertaisessa Bernoulli-kokeessa

Suoraan binomijakauman määritelmän mukaan frekvenssi f noudattaa **binomijakaumaa** parametrein n ja p:

$$f \sim \text{Bin}(n, p)$$

Selvästi

$$f = \sum_{i=1}^{n} X_i$$

koska luku 1 esiintyy summassa $f = \sum X_i$ *täsmälleen* yhtä monta kertaa kuin tapahtuma A sattuu koetoistojen aikana.

Tämä merkitsee sitä, että *binomijakautunut satunnaismuuttuja* voidaan aina esittää *riippumattomien Bernoulli-jakautuneiden satunnaismuuttujien summana*.

Koska *satunnaismuuttujien summan odotusarvo on aina satunnaismuuttujien odotusarvojen summa*, niin frekvenssin $f = \sum X_i$ *odotusarvo* on

$$\text{E}(f) = \text{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{E}(X_i) = \sum_{i=1}^{n} p = np$$

Koska *riippumattomien satunnaismuuttujien summan varianssi on satunnaismuuttujien varianssien summa*, niin frekvenssin $f = \sum X_i$ *varianssi* on

$$\text{D}^2(X) = \text{D}^2\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{D}^2(X_i) = \sum_{i=1}^{n} pq = npq$$

■

Suhteellisen frekvenssin odotusarvo ja varianssi

Olkoon A jokin otosavaruuden S *tapahtuma*:

$$A \subset S$$

ja olkoon

$$\Pr(A) = p$$

Poimitaan otosavaruudesta S *satunnaisotos*, jonka *koko* on n ja olkoon

$$f$$

A-tyyppisten alkioiden *lukumäärä* eli *frekvenssi* otoksessa. Tällöin tapahtuman A suhteellisen frekvenssin $\hat{p} = f/n$ **odotusarvo** ja **varianssi** ovat

$$\mathrm{E}(\hat{p}) = p$$

$$\mathrm{Var}(\hat{p}) = \mathrm{D}^2(\hat{p}) = \frac{pq}{n}$$

jossa $q = 1 - p$.

Perustelu:

Suhteellisen frekvenssin $\hat{p} = f/n$ odotusarvo ja varianssi saadaan suoraan frekvenssin f odotusarvoa ja varianssia koskevista tuloksista satunnaismuuttujan odotusarvon ja varianssin yleisten ominaisuuksien perusteella:

$$\mathrm{E}(\hat{p}) = \mathrm{E}\left(\frac{f}{n}\right) = \frac{1}{n}\mathrm{E}(f) = \frac{1}{n}np = p$$

$$\mathrm{Var}(\hat{p}) = \mathrm{Var}\left(\frac{f}{n}\right) = \frac{1}{n^2}\mathrm{Var}(f) = \frac{1}{n^2}npq = \frac{pq}{n}$$

■

Suhteellisen frekvenssin $\hat{p} = f/n$ *standardipoikkeamaa*

$$\mathrm{D}(\hat{p}) = \sqrt{\frac{pq}{n}}$$

kutsutaan tavallisesti **suhteellisen frekvenssin keskivirheeksi** ja se kuvaa suhteellisen frekvenssin otosvaihtelua oman odotusarvonsa p ympärillä.

Koska suhteellisen frekvenssin $\hat{p} = f/n$ odotusarvo

$$\mathrm{E}(\hat{p}) = p$$

ja varianssi on

$$\mathrm{Var}(\hat{p}) = \frac{pq}{n}$$

niin suhteellisen frekvenssin otosjakauma *keskittyy yhä voimakkaammin tapahtuman A todennäköisyyden p ympärille*, *kun otoskoko n kasvaa*.

Suhteellinen frekvenssi ja suurten lukujen laki

Olkoon A jokin otosavaruuden S *tapahtuma*:

$$A \subset S$$

ja olkoon

$$\Pr(A) = p$$

Poimitaan otosavaruudesta *S satunnaisotos*, jonka *koko* on *n* ja olkoon

$$\hat{p}_n = f / n$$

A-tyyppisten alkioiden *suhteellinen lukumäärä* eli *suhteellinen frekvenssi* otoksessa.

Heikon suurten lukujen lain mukaan *suhteellisten frekvenssien* $\hat{p}_n = f / n$ muodostama jono *konvergoi* havaintojen lukumäärän *n* kasvaessa rajatta *stokastisesti* kohti tapahtuman *A* todennäköisyyttä *p*:

$$\hat{p}_n \xrightarrow[n\to\infty]{\mathrm{P}} p$$

Perustelu:

Suhteellisen frekvenssin asymptoottista raja-arvoa koskeva tulos seuraa aritmeettisen keskiarvon asymptoottista käyttäytymistä koskevasta *heikosta suurten lukujen laista*, kun huomataan, että

$$\hat{p}_n = \frac{f}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

jossa

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \mathrm{Bernoulli}(p)$$

jolloin

$$\mathrm{E}(X_i) = p \text{ , } i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \mathrm{D}^2(X_i) = pq \text{ , } i = 1, 2, \ldots, n$$

■

Suhteellisen frekvenssin asymptoottinen jakauma

Olkoon *A* jokin otosavaruuden *S tapahtuma*:

$$A \subset S$$

ja olkoon

$$\Pr(A) = p$$

Poimitaan otosavaruudesta *S satunnaisotos*, jonka *koko* on *n* ja olkoon

$$\hat{p}_n = f / n$$

A-tyyppisten alkioiden *suhteellinen lukumäärä* eli *suhteellinen frekvenssi* otoksessa.

Suhteellinen frekvenssi $\hat{p} = f / n$ *noudattaa suurissa otoksissa approksimatiivisesti normaalijakaumaa*:

$$\hat{p} = \frac{f}{n} \sim_a \mathrm{N}\left(p, \frac{pq}{n}\right)$$

Siten *standardoitu satunnaismuuttuja*

$$Z = \frac{\hat{p} - p}{\sqrt{pq/n}}$$

noudattaa suurissa otoksissa approksimatiivisesti standardoitua normaalijakaumaa:

$$Z \sim_a \mathrm{N}(0,1)$$

Perustelu:

Suhteellisen frekvenssin otosjakaumaa koskeva asymptoottinen tulos seuraa *keskeisestä raja-arvolauseesta*, kun huomataan, että

$$\hat{p} = \frac{f}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

jossa

$$X_1, X_2, \ldots, X_n \perp$$
$$X_i \sim \text{Bernoulli}(p)$$

jolloin

$$\mathrm{E}(X_i) = p \text{ , } i = 1, 2, \ldots, n$$
$$\mathrm{Var}(X_i) = \mathrm{D}^2(X_i) = pq \text{ , } i = 1, 2, \ldots, n$$

■

1.6. Järjestystunnusluvut

Järjestystunnusluvut: määritelmä

Olkoon

$$X_1, X_2, \ldots, X_n$$

satunnaisotos. Tällöin havainnot $X_1, X_2, \ldots, X_n$ ovat *riippumattomia*, *samaa jakaumaa noudattavia satunnaismuuttujia*.

Jos satunnaismuuttujien $X_1, X_2, \ldots, X_n$ havaitut arvot järjestetään *suuruusjärjestykseen pienimmästä suurimpaan*, saamme *järjestystunnusluvut*

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

ja sanomme, että

$$X_{(j)} \text{ , } j = 1, 2, \ldots, n$$

on ***j*. järjetystunnusluku**. Järjestystunnusluvut $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ toteuttavat ehdot

$$X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$$

Erityisesti

$$X_{(1)} = \textbf{minimi}$$

$$X_{(n)} = \textbf{maksimi}$$

$(X_{(1)}, X_{(n)})$ = **vaihteluväli**

$X_{(n)} - X_{(1)}$ = **vaihteluvälin pituus**

Järjestystunnuslukujen jakaumat: Otos diskreetistä jakaumasta

Olkoon

$$X_1, X_2, \ldots, X_n$$

satunnaisotos diskreetistä jakaumasta, jonka *pistetodennäköisyysfunktio* on

$$f_X(x_i) = p_i$$

jossa

$$x_1 \le x_2 \le x_3 \le \cdots$$

ovat satunnaismuuttujan X mahdolliset arvot kasvavassa järjestyksessä.

Olkoot

$$P_0 = 0$$
$$P_1 = p_1$$
$$P_2 = p_1 + p_2$$
$$\ldots$$
$$P_i = p_1 + p_2 + \cdots + p_i$$
$$\ldots$$

Olkoot

$$X_{(1)}, X_{(2)}, \ldots, X_{(n)}$$

otosta $X_1, X_2, \ldots, X_n$ vastaavat järjestystunnusluvut. Tällöin

$$\Pr(X_{(j)} \le x_i) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1-P_i)^{n-k}$$

ja

$$\Pr(X_{(j)} = x_i) = \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1-P_i)^{n-k} - P_{i-1}^k (1-P_{i-1})^{n-k} \right]$$

Perustelu:

Kiinnitetään i ja määritellään satunnaismuuttuja Y seuraavalla tavalla:

Y = niiden havaintojen $X_1, X_2, \ldots, X_n$ *lukumäärä*, jotka ovat pienempiä kuin x_i

Kutsutaan jokaiselle $X_1, X_2, \ldots, X_n$ tapahtumaa

$$\{X_j \le x_i\}\ , j = 1, 2, \ldots, n$$

onnistumiseksi ja tapahtumaa

$$\{X_j > x_i\} = \{X_j \le x_i\}^c\ , j = 1, 2, \ldots, n$$

epäonnistumiseksi. Siten satunnaismuuttuja Y kuvaa onnistumisten lukumäärää otoksessa $X_1, X_2, \ldots, X_n$.

Onnistumisen todennäköisyys

$$P_i = \Pr(X_j \le x_i)$$

on *sama* kaikille j, koska satunnaismuuttujat $X_1, X_2, \ldots, X_n$ ovat samoin jakautuneita. Lisäksi onnistumiset (ja epäonnistumiset) ovat tapahtumina *riippumattomia* jokaiselle j, koska satunnaismuuttujat $X_1, X_2, \ldots, X_n$ ovat riippumattomia.

Siten

$$Y \sim \text{Bin}(n, P_i)$$

jolloin

$$\Pr(Y = j) = \binom{n}{j} P_i^j (1-P_i)^{n-j}$$

Tapahtumat

$$\{X_{(j)} \le x_i\}$$

ja

$$\{Y \ge j\}$$

ovat *ekvivalentteja*, koska tällöin *ainakin* j havainnoista $X_1, X_2, \ldots, X_n$ on pienempiä tai yhtä suuria kuin x_i .

Siten satunnaismuuttujan $X_{(j)}$ *kertymäfunktio* on muotoa

$$\Pr(X_{(j)} \le x_i) = \Pr(Y \ge j) = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1-P_i)^{n-k}$$

ja satunnaismuuttujan $X_{(j)}$ *pistetodennäköisyysfunktio* on muotoa

$$\begin{aligned}\Pr(X_{(j)} = x_i) &= \Pr(X_{(j)} = x_i) - \Pr(X_{(j)} = x_{i-1}) \\ &= \sum_{k=j}^{n} \binom{n}{k} \left[P_i^k (1-P_i)^{n-k} - P_{i-1}^k (1-P_{i-1})^{n-k} \right]\end{aligned}$$

mikä pätee myös, kun $i = 1$, koska $P_0 = 0$.

■

Järjestystunnuslukujen jakaumat: Otos jatkuvasta jakaumasta

Olkoon

$$X_1, X_2, \ldots, X_n$$

satunnaisotos jatkuvasta jakaumasta, jonka *kertymäfunktio* on $F_X(x)$ ja tiheysfunktio on $f_X(x)$.

Tällöin *j. järjestystunnusluvun* $X_{(j)}$ *tiheysfunktio* on

$$f_{X_{(j)}}(x) = \frac{n!}{(j-1)!(n-j)!}[F_X(x)]^{j-1} f_X(x)[1-F_X(x)]^{n-j}$$

Perustelu:

Olkoon $x \in \mathbb{R}$. Määritellään satunnaismuuttuja Y seuraavalla tavalla:

Y = niiden havaintojen $X_1, X_2, \ldots, X_n$ *lukumäärä*, jotka ovat pienempiä kuin x

Kutsutaan jokaiselle $X_1, X_2, \ldots, X_n$ tapahtumaa

$\{X_j \leq x\}$, $j = 1, 2, \ldots, n$

onnistumiseksi ja tapahtumaa

$\{X_j > x\} = \{X_j \leq x\}^c$, $j = 1, 2, \ldots, n$

epäonnistumiseksi. Siten satunnaismuuttuja Y kuvaa onnistumisten lukumäärää otoksessa $X_1, X_2, \ldots, X_n$.

Onnistumisen todennäköisyys

$\Pr(X_j \leq x) = F_X(x)$

on *sama* kaikille j, koska satunnaismuuttujat $X_1, X_2, \ldots, X_n$ ovat samoin jakautuneita. Lisäksi onnistumiset (ja epäonnistumiset) ovat tapahtumina *riippumattomia* jokaiselle j, koska satunnaismuuttujat $X_1, X_2, \ldots, X_n$ ovat riippumattomia.

Siten

$$Y \sim \text{Bin}(n, F_X(x))$$

jolloin

$$\Pr(Y = j) = \binom{n}{j}[F_X(x)]^j[1-F_X(x)]^{n-j}$$

Tapahtumat

$\{X_{(j)} \leq x\}$

ja

$\{Y \geq j\}$

ovat *ekvivalentteja*, koska tällöin *ainakin* j havainnoista $X_1, X_2, \ldots, X_n$ on pienempiä tai yhtä suuria kuin x. Siten satunnaismuuttujan $X_{(j)}$ *kertymäfunktio* on muotoa

$$F_{X_{(j)}}(x) = \Pr(X_{(j)} \leq x) = \Pr(Y \geq j) = \sum_{k=j}^{n}\binom{n}{k}[F_X(x)]^k[1-F_X(x)]^{n-k}$$

Satunnaismuuttujan $X_{(j)}$ *tiheysfunktioksi* saadaan

$$
\begin{aligned}
f_{X_{(j)}}(x) &= \frac{d}{dx} F_{X_{(j)}}(x) \\
&= \sum_{k=j}^{n} \binom{n}{k} \Big\{ k[F_X(x)]^{k-1} f_X(x)[1-F_X(x)]^{n-k} \\
&\qquad -(n-k)[F_X(x)]^k f_X(x)][1-F_X(x)]^{n-k-1} \Big\} \\
&= \binom{n}{j} j[F_X(x)]^{j-1} f_X(x)[1-F_X(x)]^{n-j} \\
&\qquad + \sum_{k=j+1}^{n} \binom{n}{k} k[F_X(x)]^{k-1} f_X(x)[1-F_X(x)]^{n-k} \\
&\qquad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F_X(x)]^k f_X(x)][1-F_X(x)]^{n-k-1} \\
&= \frac{n!}{(j-1)!(n-j)!}[F_X(x)]^{j-1} f_X(x)[1-F_X(x)]^{n-j} \\
&\qquad + \sum_{k=j}^{n-1} \binom{n}{k+1} (k+1)[F_X(x)]^k f_X(x)[1-F_X(x)]^{n-k-1} \\
&\qquad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F_X(x)]^k f_X(x)][1-F_X(x)]^{n-k-1} \\
&= \frac{n!}{(j-1)!(n-j)!}[F_X(x)]^{j-1} f_X(x)[1-F_X(x)]^{n-j}
\end{aligned}
$$

Viimeinen yhtälö perustuu siihen, että

$$
\binom{n}{k+1}(k+1) = \frac{n!}{k!(n-k-1)!} = \binom{n}{k}(n-k)
$$

■

1.7. Delta-menetelmä

Taylorin sarjakehitelmä

Oletetaan, että funktiolla $g(x)$ on *r. derivaatta*

$$
g^{(r)}(x) = \frac{d^r}{dx^r} g(x)
$$

pisteessä $x = a$. Tällöin

$$
T_r(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i \ , \ a \in \mathbb{R}
$$

on funktion $g(x)$ **Taylorin polynomi** astetta r.

Lauseketta

$$g(x) - T_r(x) = R_r(x)$$

sanotaan Taylorin polynomin *jäännöstermiksi*. Voidaan osoittaa, että jäännöstermillä $R_r(x)$ on seuraava ominaisuus: Jos x lähestyy rajatta pistettä a, niin jäännöstermi $R_r(x)$ lähestyy rajatta nollaa *nopeammin* kuin Taylorin polynomin korkeimman asteen termi:

$$\lim_{x \to a} \frac{g(x) - T_r(x)}{(x-a)^r} = 0$$

Jos funktion $g(x)$ $(r+1)$. *derivaatta* $g^{(r+1)}(x)$ *on olemassa* pisteessä a, niin jäännöstermi on muotoa

$$R_r(x) = g(x) - T_r(x) = \int_a^x \frac{g^{(r+1)}(x)}{r!}(x-t)^r \, dt$$

Siten saamme funktion $g(x)$ **Taylorin sarjakehitelmäksi** pisteen a ympäristössä

$$g(x) = \sum_{i=0}^{r} \frac{g^{(i)}(a)}{i!}(x-a)^i + R$$

jossa $R \to 0$, kun $x \to a$.

Tarvitsemme jatkossa lähinnä vain 1. tai 2. asteen Taylorin polynomeja.

Tunnusluvun funktion odotusarvo ja varianssi

Olkoot

$$T_i\,,\ i = 1, 2, \ldots, k$$

satunnaismuuttujia, joiden *odotusarvot* ovat

$$\mathrm{E}(T_i) = \theta_i\,,\ i = 1, 2, \ldots, k$$

Määritellään vektorit

$$\mathbf{T} = (T_1, T_2, \ldots, T_k)$$

ja

$$\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$$

Olkoon $g(\mathbf{T})$ satunnaismuuttujan $\mathbf{T}$ derivoituva funktio. Haluamme määrätä *approksimatiivisen varianssin* satunnaismuuttujalle $g(\mathbf{T})$.

Olkoon

$$g_i'(\boldsymbol{\theta}) = \frac{\partial}{\partial t_i} g(\mathbf{t}) \Big|_{t_1=\theta_1, \ldots, t_k=\theta_k}$$

funktion $g(\mathbf{T})$ 1. derivaatta muuttujan t_i suhteen pisteessä $t_1 = \theta_1, \ldots, t_k = \theta_k$.

Funktion g *Taylorin sarjakehitelmä* pisteen $\boldsymbol{\theta}$ ympäristössä on muotoa

$$g(\mathbf{t}) = g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i) + R$$

jossa $R \to 0$, kun $t_1 \to \theta_1, \ldots, t_k \to \theta_k$.

Siten (approksimatiivisesti)

$$g(\mathbf{t}) \approx g(\boldsymbol{\theta}) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(t_i - \theta_i)$$

Ottamalla tästä *odotusarvo* saadaan

$$\mathrm{E}(g(\mathbf{T})) \approx \mathrm{E}(g(\boldsymbol{\theta})) + \sum_{i=1}^{k} g_i'(\boldsymbol{\theta})\,\mathrm{E}(T_i - \theta_i) = g(\boldsymbol{\theta})$$

koska

$$\mathrm{E}(T_i) = \theta_i \,,\ i = 1, 2, \ldots , k$$

Voimme siten approksimoida funktion $g(\mathbf{T})$ *varianssia* lausekkeella

$$\begin{aligned}
\mathrm{Var}(g(\mathbf{T})) &\approx \mathrm{E}\left\{[g(\mathbf{T}) - g(\boldsymbol{\theta})]^2\right\} \\
&\approx \mathrm{E}\left\{\left[\sum_{i=1}^{k} g_i'(\boldsymbol{\theta})(T_i - \theta_i)\right]^2\right\} \\
&= \sum_{i=1}^{k} [g_i'(\boldsymbol{\theta})]^2\,\mathrm{Var}(T_i) + 2\sum_{i>j} g_i'(\boldsymbol{\theta})g_j'(\boldsymbol{\theta})\,\mathrm{Cov}(T_i, T_j)
\end{aligned}$$

Erityisesti, jos satunnaismuuttujat T_i , $i = 1, 2, \ldots , k$ ovat *korreloimattomia*, niin

$$\mathrm{Var}(g(\mathbf{T})) \approx \sum_{i=1}^{k} [g_i'(\boldsymbol{\theta})]^2\,\mathrm{Var}(T_i)$$

Delta-menetelmä

Käyttämällä edellä johdettuja approksimaatioita funktion $g(\mathbf{T})$ odotusarvolla ja varianssille, saamme seuraavan *yleistyksen keskeiselle raja-arvolauseelle*:

Olkoon

$$Y_1, Y_2, Y_3, \ldots$$

satunnaismuuttujien jono, jolle

$$\sqrt{n}(Y_n - \theta) \to \mathrm{N}(0, \sigma^2)$$

jakaumakonvergenssin mielessä. Olkoon g funktio, jolle

$$g'(\theta) \neq 0$$

jokaiselle kiinteälle θ. Tällöin

$$\sqrt{n}(g(Y_n) - g(\theta)) \to \mathrm{N}(0, \sigma^2[g'(\theta)]^2)$$

jakaumakonvergenssin mielessä.

Perustelu:

Funktion $g(Y_n)$ *Taylorin sarjakehitelmä* pisteen $Y_n = \theta$ ympäristössä on muotoa

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + R$$

jossa jäännöstermi $R \to 0$, kun $Y_n \to \theta$.

Voidaan osoittaa, että $Y_n \to \theta$ *stokastisesti*, koska $Y_n \to \theta$ *jakaumakonvergenssin mielessä*. Siten $R \to 0$ *stokastisesti*. Soveltamalla ns. *Slutskyn lausetta* (ks. alla) todetaan, että

$$\sqrt{n}(g(Y_n) - g(\theta)) = g'(\theta)\sqrt{n}(Y_n - \theta) \to \mathrm{N}(0, \sigma^2[g'(\theta)]^2)$$

■

Toisen kertaluvun delta-menetelmä

Jos *delta-menetelmässä*

$$g'(\theta) = 0$$

niin ottamalla funktion g *Taylorin sarjakehitelmään* mukaan myös 2. asteen termi, saadaan kehitelmä

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + \frac{g''(\theta)}{2}(Y_n - \theta)^2 + R'$$

jossa jäännöstermi $R' \to 0$, kun $Y_n \to \theta$. Siten

$$g(Y_n) - g(\theta) = \frac{g''(\theta)}{2}(Y_n - \theta)^2 + R'$$

koska oletuksen mukaan $g'(\theta) = 0$.

Koska standardoitua normaalijakaumaa N(0,1) noudattavan satunnaismuuttujan neliö noudattaa $\chi^2(1)$-*jakaumaa* χ^2-jakauman määritelmän mukaan, niin

$$\frac{n(Y_n - \theta)^2}{\sigma^2} \to \chi^2(1)$$

jakaumakonvergenssin mielessä.

Soveltamalla ns. *Slutskyn lausetta* (ks. alla) saadaan tästä seuraava tulos:

Olkoon

$$Y_1, Y_2, Y_3, \ldots$$

satunnaismuuttujien jono, jolle

$$\sqrt{n}(Y_n - \theta) \to \mathrm{N}(0, \sigma^2)$$

jakaumakonvergenssin mielessä. Olkoon g funktio, jolle

$$g'(\theta) = 0$$

mutta

$$g''(\theta) \neq 0$$

jokaiselle kiinteälle θ. Tällöin

$$n(g(Y_n) - g(\theta)) \to \sigma^2 \frac{g''(\theta)}{2}\chi^2(1)$$

jakaumakonvergenssin mielessä.

Slutskyn lause

Oletetaan, että

$$X_n \to X$$

jakaumakonvergenssin mielessä ja

$$Y_n \to a$$

stokastisesti. Tällöin

(i) $$Y_n X_n \to aX$$

(ii) $$Y_n + X_n \to a + X$$

jakaumakonvergenssin mielessä.