MS-A0504 Todennäköisyyslaskennan ja tilastotieteen peruskurssi

2B Keskihajonta ja korrelaatio

Lasse Leskelä

Matematiikan ja systeemianalyysin laitos
Perustieteiden korkeakoulu
Aalto-yliopisto

Lukuvuosi 2018–2019
Periodi IV

Sisältö

Mitä odotusarvo kertoo jakaumasta?

Satunnaismuuttujan odotusarvo $\mathbb{E}(X)$:

- on X:n mahdollisten arvojen todennäköisyyksillä painotettu summa ($\sum_x x\, f(x)$ tai $\int x\, f(x) dx$)
- kertoo likiarvon keskiarvolle, joka saadaan suuresta määrästä riippumattomia X:n tavoin jakautuneita satunnaislukuja
- ei kerro mitään jakauman *leveydestä*

Esim

Diskreettejä satunnaismuuttujia, joiden odotusarvo on 1:

k	1
$\mathbb{P}(X = k)$	1

k	0	1	2
$\mathbb{P}(Z = k)$	$\frac{1}{2}$	0	$\frac{1}{2}$

k	0	1	2
$\mathbb{P}(Y = k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

k	0	1000000
$\mathbb{P}(W = k)$	0.999999	0.000001

Miten mitata satunnaismuuttujan poikkeamaa odotusarvosta?

Satunnaisluvun poikkeama odotusarvosta $\mu = \mathbb{E}(X)$ on satunnaisluku $|X - \mu|$. Sen odotusarvo $\mathbb{E}|X - \mu|$:

- kertoo likiarvon keskiarvolle $\frac{1}{n}\sum_{i=1}^{n}|X_i - \mu|$, joka saadaan suuresta määrästä riippumattomia X:n tavoin jakautuneita satunnaislukuja
- on optimoinnin kannalta hankala suure, koska funktio $x \mapsto |x|$ ei ole derivoituva nollassa.

Entä jos korvataan $|X - \mu|$ luvulla $(X - \mu)^2$?

Varianssi

[engl. *variance*]

Satunnaisluvun neliöpoikkeama odotusarvosta $\mu = \mathbb{E}(X)$ on satunnaisluku $(X - \mu)^2$. Sen odotusarvo eli satunnaisluvun X varianssi $\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2]$:

- kertoo likiarvon keskiarvolle $\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ suuresta määrästä riippumattomia X:n tavoin jakautuneita sat.lukuja
- on optimoinnin kannalta mukava suure, koska funktio $x \mapsto x^2$ on äärettömän monta kertaa derivoituva
- mittaa poikkeamaa neliöyksiköissä

	X	$\mathrm{Var}(X)$
Pituus	m	m^2
Aika	s	s^2
Tuotto	EUR	EUR^2

Tulos palautetaan alkuperäisiin mittayksiköihin ottamalla neliöjuuri.

Keskihajonta

[engl. *standard deviation*]

Satunnaisluvun keskihajonta $\mathrm{SD}(X) = \sqrt{\mathbb{E}[(X-\mu)^2]}$ on alkuperäisiin yksiköihin normitettu odotusarvoinen neliöpoikkeama odotusarvosta $\mu = \mathbb{E}(X)$.

$\mathrm{SD}(X)$ mittaa:

- X:n odotusarvoista normitettua poikkeamaa odotusarvosta
- X:n jakauman leveyttä

Diskreetti jakauma:

$$\mu = \sum_x x\, f(x)$$

$$\mathrm{SD}(X) = \sqrt{\sum_x (x-\mu)^2 f(x)}$$

Jatkuva jakauma:

$$\mu = \int x\, f(x)\, dx$$

$$\mathrm{SD}(X) = \sqrt{\int (x-\mu)^2\, f(x)\, dx}$$

Esimerkki: Odotusarvon 1 satunnaislukuja

Laske satunnaislukujen X, Y, Z keskihajonnat:

k	1
$\mathbb{P}(X = k)$	1

k	0	1	2
$\mathbb{P}(Y = k)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

k	0	2
$\mathbb{P}(Z = k)$	$\frac{1}{2}$	$\frac{1}{2}$

$$\mathrm{SD}(X) = \sqrt{\sum_k (k-\mu)^2 f_X(k)} = \sqrt{(1-1)^2 \times 1} = 0.$$

$$\mathrm{SD}(Y) = \sqrt{(0-1)^2 \times \frac{1}{3} + (1-1)^2 \times \frac{1}{3} + (2-1)^2 \times \frac{1}{3}} = \sqrt{\frac{2}{3}} \approx 0.82.$$

$$\mathrm{SD}(Z) = \sqrt{(0-1)^2 \times \frac{1}{2} + (1-1)^2 \times 0 + (2-1)^2 \times \frac{1}{2}} = 1.$$

Keskihajonta: Laskentakaava

Fakta

Satunnaisluku X odotusarvona $\mu = \mathbb{E}(X)$ toteuttaa

$$\mathsf{SD}(X) \;=\; \sqrt{\mathsf{Var}(X)} \;=\; \sqrt{\mathbb{E}(X^2) - \mu^2}.$$

Todistus.

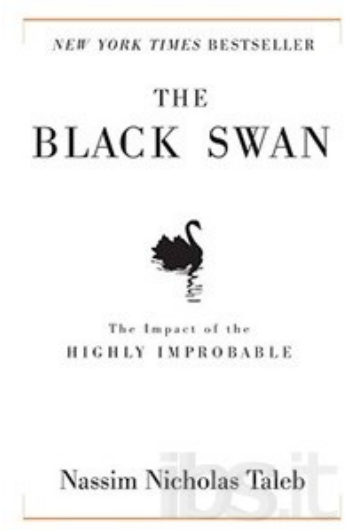
$$\begin{aligned}
\mathsf{Var}(X) \;=\; \mathbb{E}[(X-\mu)^2] \;&=\; \mathbb{E}[X^2 - 2\mu X + \mu^2] \\
&=\; \mathbb{E}[X^2] - \mathbb{E}[2\mu X] + \mathbb{E}[\mu^2] \\
&=\; \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\
&=\; \mathbb{E}[X^2] - \mu^2
\end{aligned}$$

$$\Longrightarrow \qquad \mathsf{SD}(X) \;=\; \sqrt{\mathsf{Var}(X)} \;=\; \sqrt{\mathbb{E}[X^2] - \mu^2}$$

Esimerkki: Musta joutsen

k	0	10^6
$\mathbb{P}(X = k)$	$1 - 10^{-6}$	10^{-6}

$$\mu = \mathbb{E}(X) = 1$$



Laske keskihajonta.

Tapa 1 (määritelmästä):

$$\begin{aligned} \mathrm{SD}(X) &= \sqrt{\sum_x (x-\mu)^2 f(x)} \\ &= \sqrt{(0-1)^2 \times (1-10^{-6}) + (10^6-1)^2 \times 10^{-6}} \approx 1000. \end{aligned}$$

Tapa 2 (laskentakaavan avulla):

$$\mathbb{E}(X^2) = \sum_x x^2 f(x) = 0^2 \times (1-10^{-6}) + (10^6)^2 \times 10^{-6} = 10^6.$$

$$\Longrightarrow \quad \mathrm{SD}(X) = \sqrt{\mathbb{E}(X^2) - \mu^2} = \sqrt{10^6 - 1^2} \approx 1000.$$

Esimerkki: Metro

Jos seuraavan metron saapumiseen kuluva aika X noudattaa välin $[0, 10]$ tasajakaumaa, niin saapumisajan odotusarvo $\mu = 5$. Laske keskihajonta.

Tapa 1 (määritelmästä):

$$\mathrm{SD}(X) = \sqrt{\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx} = \sqrt{\int_0^{10} (x-5)^2 \frac{1}{10}\,dx} = \cdots$$

Tapa 2 (laskentakaavan avulla):

$$\mathbb{E}(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^{10} x^2 \frac{1}{10}\,dx = \frac{1}{10}\Big|_0^{10} \frac{1}{3}x^3 \approx 33.33.$$

$$\Longrightarrow \quad \mathrm{SD}(X) = \sqrt{\mathbb{E}(X^2) - \mu^2} = \sqrt{33.33 - 5^2} \approx 2.89.$$

Siirretyn ja skaalatun satunnaisluvun keskihajonta

Fakta (Viime luento)

(i) $\mathbb{E}(a) = a$.

(ii) $\mathbb{E}(bX) = b\mathbb{E}(X)$.

(iii) $\mathbb{E}(a + bX) = a + b\mathbb{E}(X)$.

Fakta

(i) $\mathrm{SD}(a) = 0$.

(ii) $\mathrm{SD}(bX) = |b|\,\mathrm{SD}(X)$.

(iii) $\mathrm{SD}(a + bX) = |b|\,\mathrm{SD}(X)$.

Todistus.

$$\begin{aligned}
\mathrm{Var}(a + bX) &= \mathbb{E}[\big(a + bX - \mathbb{E}[a + bX]\big)^2] \\
&= \mathbb{E}[\big(a + bX - a - b\mu\big)^2] \\
&= \mathbb{E}[\big(bX - b\mu\big)^2] \\
&= \mathbb{E}[b^2(X - \mu)^2] \;=\; b^2\mathbb{E}[(X - \mu)^2] \;=\; b^2\,\mathrm{Var}(X),
\end{aligned}$$

$$\mathrm{SD}(a + bX) \;=\; \sqrt{\mathrm{Var}(a + bX)} \;=\; \sqrt{b^2\,\mathrm{Var}(X)} \;=\; |b|\,\mathrm{SD}(X).$$

On siis todistettu (iii). Kohdat (i) ja (ii) ovat kohdan (iii) erikoistapauksia. □

Sisältö

Chebyshevin epäyhtälö: Poikkeamat odotusarvosta

Fakta (Chebyshevin epäyhtälö)

Jokaiselle satunnaisluvulle odotusarvona μ ja keskihajontana σ, tapahtuman
$\{X = \mu \pm 2\sigma\} = \{\mu - 2\sigma \leq X \leq \mu + 2\sigma\}$
todennäköisyys on vähintään

$$\mathbb{P}(X = \mu \pm 2\sigma) \geq \frac{3}{4}.$$

Pafnuty Chebyshev
1821–1894

Yleisemmin $\mathbb{P}(X = \mu \pm r\sigma) \geq 1 - \frac{1}{r^2}$ kaikilla $r \geq 1$.

- X:n arvo sijaitsee melko todennäköisesti (tn $\geq 75\%$) kahden keskihajonnan sisällä odotusarvostaan
- X:n arvo sijaitsee hyvin todennäköisesti (tn $\geq 99.9999\%$) tuhannen keskihajonnan sisällä odotusarvostaan

Esimerkki: Tiedostopalvelin

Palvelimelta ladatun satunnaisen tiedoston koko odotusarvoltaan 1000 kB ja keskihajonnaltaan 200 kB. Onko todennäköistä vai epätodennäköistä, että satunnaisen tiedoston koko on

(a) välillä 600–1400 kB?

(b) välillä 800–1200 kB?

Ratkaisu

(a) Chebyshevin epäyhtälöstä

$$\mathbb{P}(X \in [600, 1400]) \;=\; \mathbb{P}(X = \mu \pm 2\sigma) \;\geq\; 75\%,$$

joten tiedoston koko on melko todennäköisesti välillä 600–1400 kB.

(b) Chebyshev ei tällä tarkkuudella kerro mitään hyödyllistä, sillä

$$\mathbb{P}(X \in [800, 1200]) \;=\; \mathbb{P}(X = \mu \pm \sigma) \;\geq\; 1 - \frac{1}{1^2} \;=\; 0.$$

Ilman tarkempia tietoja tiedostojen kokojakaumasta ei tähän voida vastata mitään.

Esimerkki: Tiedostopalvelin ($\approx$ normaalijakauma)

Palvelimelta ladatun satunnaisen tiedoston koko odotusarvoltaan 1000 kB ja keskihajonnaltaan 200 kB. Onko todennäköistä vai epätodennäköistä, että satunnaisen tiedoston koko on

(a) välillä 600–1400 kB?

(b) välillä 800–1200 kB?

kun oletetaan, että tiedostojen kokojakauma $\approx$ normaalijakauma.

Ratkaisu

(a) Normaalijakauman taulukoista (tai R:llä `1-2*pnorm(-2)`)

$$\mathbb{P}(X \in [600, 1400]) = \mathbb{P}(X = \mu \pm 2\sigma) = \mathbb{P}(\frac{X-\mu}{\sigma} = 0 \pm 2) \approx 95\%,$$

joten tiedoston koko on todennäköisesti välillä 600–1400 kB.

(b) Normaalijakauman taulukoista (tai R:llä `1-2*pnorm(-1)`)

$$\mathbb{P}(X \in [800, 1200]) = \mathbb{P}(X = \mu \pm \sigma) = \mathbb{P}\left(\frac{X-\mu}{\sigma} = 0 \pm 1\right) \approx 68\%,$$

joten tiedoston koko on melko todennäköisesti välillä 800–1200 kB.

Esimerkki: Tiedostopalvelin (diskreetti jakauma)

Palvelimelta ladatun satunnaisen tiedoston koko on odotusarvoltaan 1000 kB ja keskihajonnaltaan 200 kB. Onko todennäköistä vai epätodennäköistä, että satunnaisen tiedoston koko on

(a) välillä 600–1400 kB?

(b) välillä 800–1200 kB?

kun tiedostojen kokojakauma on

k	750 kB	1000 kB	1250 kB
$\mathbb{P}(X = k)$	32%	36%	32%

Ratkaisu

Suoraan jakauman taulukosta nähdään, että tiedoston koko on varmuudella (tn $= 100\%$) välillä $\mu \pm 2\sigma = [600, 1400]$, mutta melko epätodennäköisesti (tn $= 36\%$) välillä $\mu \pm \sigma = [800, 1200]$.

Chebyshevin epäyhtälön todistus

Todistus.

Jatkuvalle satunnaisluvulle X, jonka tiheysfunktio on $f(x)$, pätee

$$\begin{aligned}
\mathbb{P}\Big(|X-\mu| \geq r\sigma\Big) &= \int_{\{x:|x-\mu|\geq r\sigma\}} f(x)\,dx \\
&\leq \int_{\{x:|x-\mu|\geq r\sigma\}} \underbrace{\frac{(x-\mu)^2}{(r\sigma)^2}}_{\geq 1} \; f(x)\;dx \\
&\leq \int_{-\infty}^{\infty} \frac{(x-\mu)^2}{(r\sigma)^2}\, f(x)\,dx \;=\; \frac{\mathrm{Var}(X)}{r^2\sigma^2} \;=\; \frac{1}{r^2}.
\end{aligned}$$

Diskreetin satunnaisluvun tapaus seuraa samaan tapaan. □

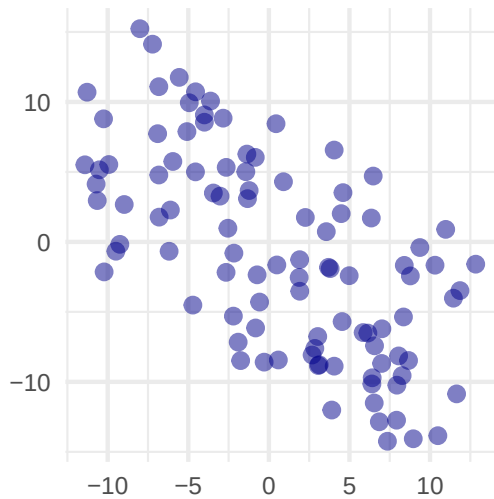
Huom:
Chebyshevin epäyhtälön avulla voidaan todistaa suurten lukujen laki.
https://en.wikipedia.org/wiki/Law_of_large_numbers

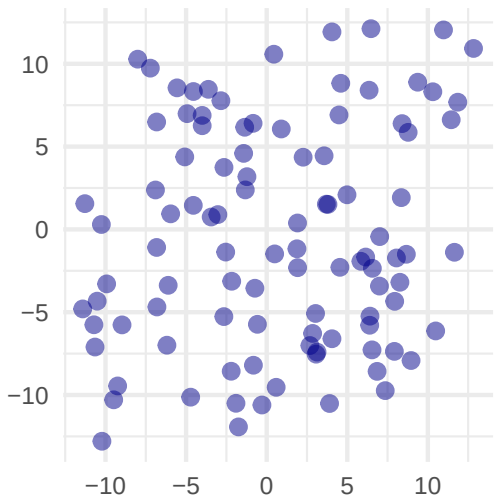
Sisältö

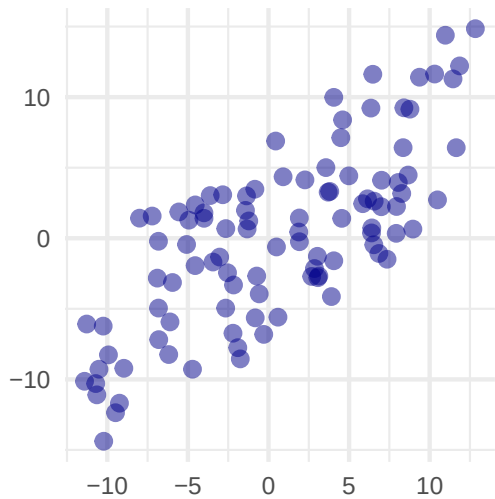
Yhteisvaihtelu

Keskihajonta mittaa yhden satunnaismuuttujan vaihtelua odotusarvonsa ympärillä.

Miten mitataan kahden satunnaismuuttujan X ja Y yhteisvaihtelua (suunta ja voimakkuus)?







Kovarianssi

[engl. *covariance*]

$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$, mittaa satunnaismuuttujien X:n ja Y:n yhteisvaihtelun suuntaa ja voimakkuutta.

Diskreetti yhteisjakauma:

$$\sum_x \sum_y (x - \mu_X)(y - \mu_Y) f(x, y)$$

Jatkuva yhteisjakauma:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y) f(x, y) dx dy.$$

Kovarianssi

- on > 0, kun $X - \mu_X$ ja $Y - \mu_Y$ ovat usein samanmerkkiset
- on < 0, kun $X - \mu_X$ ja $Y - \mu_Y$ ovat usein erimerkkiset
- mittaa yhteisvaihtelua neliöyksiköissä (m^2, s^2, EUR^2, ...)

Kovarianssia ei normiteta ottamalla neliöjuurta (miksi)? (Voi olla neg.)

Kovarianssi: Laskentakaava

Fakta

$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Todistus.

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathbb{E}[XY - \mu_X Y - \mu_Y X + \mu_X \mu_Y] \\
&= \mathbb{E}[XY] - \mu_X \mathbb{E}[Y] - \mu_Y \mathbb{E}[X] + \mathbb{E}[\mu_X \mu_Y] \\
&= \mathbb{E}[XY] - \mu_X \mu_Y - \mu_Y \mu_X + \mu_X \mu_Y \\
&= \mathbb{E}[XY] - \mu_X \mu_Y.
\end{aligned}
$$

Kovarianssin lineaarisuus

Fakta

Kovarianssioperaattori $(X, Y) \mapsto \mathrm{Cov}(X, Y)$ *on symmetrinen ja molempien argumenttiensa suhteen lineaarinen:*

$\mathrm{Cov}(Y, X) = \mathrm{Cov}(X, Y)$

$\mathrm{Cov}(X_1 + X_2, Y) = \mathrm{Cov}(X_1, Y) + \mathrm{Cov}(X_2, Y).$

$\mathrm{Cov}(X, Y_1 + Y_2) = \mathrm{Cov}(X, Y_1) + \mathrm{Cov}(X, Y_2).$

$\mathrm{Cov}(aX, Y) = a\,\mathrm{Cov}(X, Y)$

Yleisesti:

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$$

Kovarianssin lineaarisuus: Todistus

Merkitään $Y = \sum_{j=1}^{n} b_j Y_j$. (i) Kovarianssin laskentakaavasta ja odotusarvon lineaarisuudesta

$$\begin{aligned}
\operatorname{Cov}\left(\sum_i a_i X_i, Y\right) &= \mathbb{E}[(\sum_i a_i X_i) Y] - \mathbb{E}[(\sum_i a_i X_i)]\mathbb{E}[Y] \\
&= \sum_i a_i \mathbb{E}[X_i Y] - \left(\sum_i a_i \mathbb{E}[X_i]\right) \mathbb{E}[Y] \\
&= \sum_i a_i \mathbb{E}[X_i Y] - \sum_i a_i \mathbb{E}[X_i]\mathbb{E}[Y] \\
&= \sum_i a_i \left(\mathbb{E}[X_i Y] - \mathbb{E}[X_i]\mathbb{E}[Y]\right) = \sum_i a_i \operatorname{Cov}(X_i, Y).
\end{aligned}$$

Symmetrian ja kohdan (i) avulla

$$\begin{aligned}
\sum_i a_i \operatorname{Cov}(X_i, Y) &= \sum_i a_i \operatorname{Cov}(Y, X_i) \\
&= \sum_i a_i \operatorname{Cov}(\sum_j b_j Y_j, X_i) \\
&= \sum_i a_i \sum_j b_j \operatorname{Cov}(Y_j, X_i) \\
&= \sum_i \sum_j a_i b_j \operatorname{Cov}(X_i, Y_j).
\end{aligned}$$

Kovarianssi: Yhteenveto

Satunnaislukujen X ja Y kovarianssi on

$$\mathrm{Cov}(X,Y) \;=\; \mathbb{E}[(X-\mu_X)(Y-\mu_Y)] \;=\; \mathbb{E}(XY)-\mathbb{E}(X)\mathbb{E}(Y)$$

missä $\mu_X = \mathbb{E}(X)$ ja $\mu_Y = \mathbb{E}(Y)$.

Diskreetti yhteisjakauma:

$$\sum_x \sum_y (x-\mu_X)(y-\mu_Y)\, f(x,y)$$

Jatkuva yhteisjakauma:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\mu_X)(y-\mu_Y)\, f(x,y)\,dx\,dy.$$

Kovarianssi on symmetrinen ja lineaarinen:

$$\mathrm{Cov}(Y,X) \;=\; \mathrm{Cov}(X,Y)$$

$$\mathrm{Cov}\left(\sum_{i=1}^{m} a_i X_i, \sum_{j=1}^{n} b_j Y_j\right) \;=\; \sum_{i=1}^{m}\sum_{j=1}^{n} a_i b_j \,\mathrm{Cov}(X_i, Y_j)$$

Korrelaatio

[engl. *correlation*]

Kovarianssia ei normiteta ottamalla neliöjuurta (miksi)?
(Varianssi on aina ≥ 0, mutta kovarianssi voi olla neg.)

Korrelaatio

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)}$$

mittaa satunnaislukujen X ja Y yhteisvaihtelun suuntaa ja voimakkuutta normitetuissa yksiköissä.

Riippumattomien satunnaislukujen korrelaatio

Fakta

Jos X ja Y ovat stokastisesti riippumattomat, niin $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ *ja* $\mathrm{Cor}(X, Y) = 0$.

Todistus.

Diskreetti.

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_x \sum_y xy f_{X,Y}(x, y) \\
&= \sum_x \sum_y xy f_X(x) f_Y(y) \\
&= \left(\sum_x x f_X(x) \right) \left(\sum_y y f_Y(y) \right) = \mathbb{E}(X)\mathbb{E}(Y).
\end{aligned}$$

Kovarianssin laskukaavasta

$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = \mathbb{E}(X)\mathbb{E}(Y) - \mathbb{E}(X)\mathbb{E}(Y) = 0.$

Siis myös $\mathrm{Cor}(X, Y) = 0$. □

Esimerkki: Kaksi binaarista satunnaismuuttujaa

X ja Y ovat joukossa $\{-1, +1\}$ tasajakautuneita.

Lisäksi $c = \mathbb{P}(X = +1, Y = +1)$.

Määritä X:n ja Y:n yhteisjakauma ja korrelaatio.

	Y		
X	-1	$+1$	Yht
-1	c	$\frac{1}{2} - c$	$\frac{1}{2}$
$+1$	$\frac{1}{2} - c$	c	$\frac{1}{2}$
Yht	$\frac{1}{2}$	$\frac{1}{2}$	

$\mathbb{E}(X) = 0$
$\mathbb{E}(X^2) = (-1)^2 \times \frac{1}{2} + (+1)^2 \times \frac{1}{2} = 1$
$\mathrm{SD}(X) = \sqrt{\mathbb{E}(X^2) - (\mathbb{E}(X))^2} = \sqrt{1 - 0^2} = 1$

$\mathbb{E}(Y) = \mathbb{E}(X) = 0$, $\mathrm{SD}(Y) = \mathrm{SD}(X) = 1$.

$\mathbb{E}(XY) = (-1)^2 \times c + 2 \times (-1)(+1) \times (\frac{1}{2} - c) + (+1)^2 c = 4c - 1$
$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 4c - 1$

$$\mathrm{Cor}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)} = 4c - 1$$

Esimerkki: Lineaarinen deterministinen riippuvuus

$Y = a + bX$, missä X noudattaa jotain (tunnettua tai tuntematonta) jakaumaa odotusarvona $\mathbb{E}(X) = \mu$ ja keskihajontana $\mathrm{SD}(X) = \sigma$.
Laske X:n ja Y:n korrelaatio.

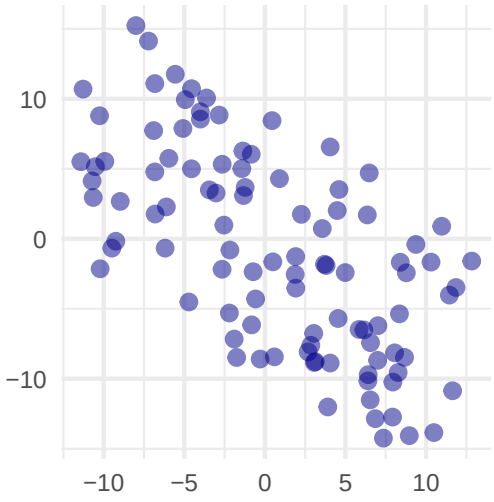
$$\mathrm{Cov}(X, Y) \;=\; \mathrm{Cov}(X, a + bX) \;=\; \mathrm{Cov}(X, a) + \mathrm{Cov}(X, bX) \;=\; b\,\mathrm{Var}(X).$$

$$\mathrm{SD}(Y) \;=\; \mathrm{SD}(a + bX) \;=\; |b|\,\mathrm{SD}(X)$$

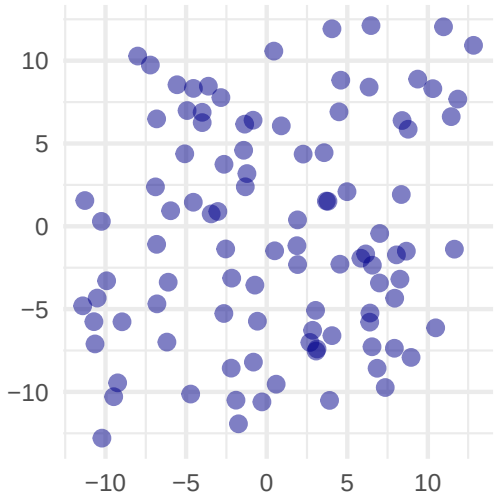
$$\mathrm{Cor}(X, Y) \;=\; \frac{\mathrm{Cov}(X, Y)}{\mathrm{SD}(X)\,\mathrm{SD}(Y)} \;=\; \frac{b\,\mathrm{Var}(X)}{|b|\,\mathrm{SD}(X)^2} \;=\; \frac{b}{|b|}.$$

$$\mathrm{Cor}(X, Y) \;=\; \begin{cases} +1, & b > 0, \\ 0, & b = 0, \\ -1, & b < 0. \end{cases}$$

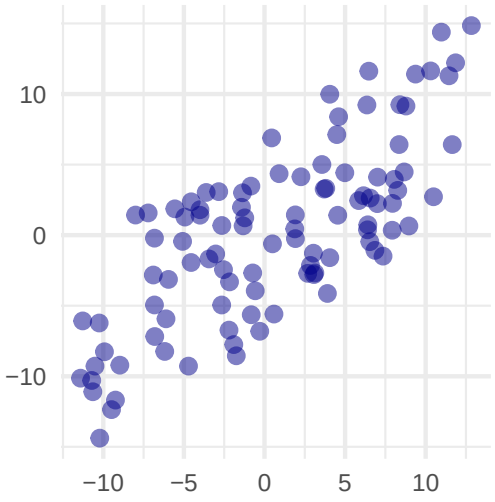
Yhteisjakaumasta simuloituja lukupareja



$\rho = -0.60$



$\rho = 0.28$



$\rho = 0.80$

Seuraavalla kerralla puhutaan satunnaismuuttujien summista ja normaaliapproksimaatiosta...