Todennäköisyyslaskenta

Osa 2: Satunnaismuuttujat ja todennäköisyysjakaumat

- **Moniulotteiset satunnaismuuttujat ja jakaumat**

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

>> **Kaksiulotteiset todennäköisyysjakaumat**

Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus

Kaksiulotteisten jakaumien odotusarvot

Kovarianssi ja korrelaatio

Ehdolliset jakaumat ja odotusarvot

Satunnaistekijöiden väliset riippuvuudet

- ***Yhden* satunnaismuuttujan todennäköisyysjakaumat kuvaavat useimpia satunnaisilmiötä vain rajoitetusti**.
- Satunnaisilmiöihin liittyy tavallisesti **useita satunnaisia tekijöitä**, joiden väliset **riippuvuudet** ovat mielenkiinnon kohteina.
- Useiden satunnaisten tekijöiden välisten riippuvuuksien *mallintaminen* vaatii tekijöihin liittyvien satunnais-muuttujien **yhteisjakauman** tarkastelua.

Esimerkkejä riippuvuustarkasteluista

- Miten *työttömyysaste* Suomessa (% työvoimasta) **riippuu** BKT:n kasvuvauhdista, viennin volyymista ja BKT:n kasvuvauhdista muissa EU-maissa ja USA:ssa?
- Miten *alkoholin kokonaiskulutus* (l *per capita* vuodessa) **riippuu** alkoholijuomien hintatasosta, käytettävissä olevista tuloista ja alkoholin saatavuudesta?
- Miten *todennäköisyys sairastua keuhkosyöpään* (p) **riippuu** tupakoinnin määrästä ja kestosta?
- Miten *vehnän sato* (t/ha) **riippuu** kesän keskilämpötilasta, sademäärästä, maan muokkaustavoista, lannoituksesta ja tuholaisten torjunnasta?

Kaksiulotteiset satunnaismuuttujat

- Olkoot X ja Y satunnaismuuttujia, joiden otosavaruudet ovat R ja S.
- Tällöin

$$X : R \to \mathbb{R}$$

$$Y : S \to \mathbb{R}$$

- Olkoon $R \times S$ otosavaruuksien R ja S *karteesinen tulo*:

$$R \times S = \left\{ (r, s) \middle| r \in R, s \in S \right\}$$

- Satunnaismuuttujien X ja Y *järjestetty pari* (X, Y) määrittelee **kaksiulotteisen satunnaismuuttujan**:

$$(X, Y) : R \times S \to \mathbb{R}^2$$

Diskreetit kaksiulotteiset satunnaismuuttujat ja niiden jakaumat

- Olkoot X ja Y *diskreettejä* satunnaismuuttujia.

- Tällöin järjestetty pari (X, Y) määrittelee **diskreetin kaksiulotteisen satunnaismuuttujan**.

- Diskreetti kaksiulotteinen satunnaismuuttuja (X, Y) määrittelee **diskreetin kaksiulotteisen todennäköisyys-jakauman**, jota kutsutaan satunnaismuuttujien X ja Y **yhteisjakaumaksi**.

Diskreetit kaksiulotteiset jakaumat: Pistetodennäköisyysfunktio

- Reaaliarvoinen funktio

  $$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

  määrittelee *diskreettien satunnaismuuttujien X* ja *Y* **yhteisjakauman pistetodennäköisyysfunktion**, jos seuraavat ehdot pätevät:

(1) $f_{XY}(x, y) \geq 0$ kaikille x ja y

(2) $$\sum_x \sum_y f_{XY}(x, y) = 1$$

(3) $\Pr(X = x \text{ ja } Y = y) = f_{XY}(x, y)$

Diskreetit kaksiulotteiset jakaumat: Tapahtumien todennäköisyydet

- Olkoon

  $$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

  diskreettien satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio*.
- Olkoon

  $$A \subset \mathbb{R}^2$$
- Tällöin

$$\boxed{\Pr((X,Y) \in A) = \sum_{(x,y)} \sum_{\in A} f_{XY}(x,y)}$$

Diskreettien kaksiulotteisten jakaumien kertymäfunktiot

- Olkoon (X, Y) *diskreetti* kaksiulotteinen satunnaismuuttuja.
- Olkoon $\{x_1, x_2, x_3, \ldots\}$ satunnaismuuttujan X tulosvaihtoehtojen eli arvojen joukko.
- Olkoon $\{y_1, y_2, y_3, \ldots\}$ satunnaismuuttujan Y tulosvaihtoehtojen eli arvojen joukko.
- **Diskreetin kaksiulotteisen jakauman kertymäfunktio** on

$$\boxed{\begin{aligned} F_{XY}(x, y) &= \Pr(X \le x \text{ ja } Y \le y) \\ &= \sum_{x_i \le x} \sum_{y_i \le y} f_{XY}(x_i, y_i) \end{aligned}}$$

jossa f_{XY} satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio*.

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 1/9

- Heitetään *virheetöntä* noppaa kaksi kertaa.
- Määritellään *satunnaismuuttujat* X, Y ja Z:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta
  $Z = X + Y$ = silmälukujen summa

- Voimme olettaa, että 2. heiton tulos on *riippumaton* 1. heiton tuloksesta (ja kääntäen).
- Satunnaismuuttujien X, Y ja Z mahdolliset arvot:

  X: {1, 2, 3, 4, 5, 6}
  Y: {1, 2, 3, 4, 5, 6}
  Z: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

- Määrätään satunnaismuuttujien X ja Z *yhteisjakauma*.

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 2/9

- Muodostetaan ensin *summamuuttujan* $Z = X + Y$ *jakauma*.
- Muodostetaan sitä varten *aputaulukko*, joka esittää kaikkia mahdollisia tapoja, joilla nopanheittojen silmälukujen summa $Z = X + Y$ voi syntyä:

Silmälukujen summat $z = x + y$

2. heiton silmäluku y						
6	7	8	9	10	11	12
5	6	7	8	9	10	11
4	5	6	7	8	9	10
3	4	5	6	7	8	9
2	3	4	5	6	7	8
1	2	3	4	5	6	7
	1	**2**	**3**	**4**	**5**	**6**

1. heiton silmäluku x

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 3/9

- Aputaulukosta voidaan suoraan lukea 1. ja 2. nopanheiton silmälukujen *summan*

  $$Z = X + Y$$

  todennäköisyysjakauma:

Silmälukujen summat $z = x + y$ ja niiden todennäköisyydet

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

- Esimerkki:

  Summa 5 voi syntyä kahden nopanheiton tuloksena 4:llä eri tavalla:

  $$5 = 1 + 4 = 2 + 3 = 3 + 2 = 4 + 1$$

  joten todennäköisyys saada summaksi 5 on 4/36.

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 4/9

- Satunnaismuuttujien X ja Z = X + Y järjestetty pari (X, Z) määrittelee diskreetin kaksiulotteisen satunnaismuuttujan, jonka arvoina on 66 lukuparia

$$(x, z)\ ;\ x = 1, 2, 3, 4, 5, 6\ ;\ z = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$$

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 5/9

- Koska noppa oletettiin *virheettömäksi* ja heittojen tulokset oletettiin *riippumattomiksi*, 1. heiton tulos ja 1. ja 2. heiton tulosten ***mahdollisten*** summien muodostamat 36 tulosvaihtoehtoa

  $$(x, z)\ ;\ x = 1, 2, 3, 4, 5, 6\ ;\ z = x + y\ ;\ y = 1, 2, 3, 4, 5, 6$$

  ovat *yhtä todennäköisiä*.
- Tällöin satunnaismuuttujien X ja Y *yhteisjakauman piste-todennäköisyysfunktio* saa positiiviset arvot

  $$f_{XZ}(x,z) = \Pr(X = x \text{ ja } Z = z) = \frac{1}{36}$$

  kun

  $$x = 1, 2, 3, 4, 5, 6$$
  $$z = x + y$$
  $$y = 1, 2, 3, 4, 5, 6$$

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 6/9

- Satunnaismuuttujien X ja $Z = X + Y$ *yhteisjakauman piste-todennäköisyysfunktio* voidaan esittää seuraavana taulukkona:

Silmälukujen summa z						
12	0	0	0	0	0	1/36
11	0	0	0	0	1/36	1/36
10	0	0	0	1/36	1/36	1/36
9	0	0	1/36	1/36	1/36	1/36
8	0	1/36	1/36	1/36	1/36	1/36
7	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	0
5	1/36	1/36	1/36	1/36	0	0
4	1/36	1/36	1/36	0	0	0
3	1/36	1/36	0	0	0	0
2	1/36	0	0	0	0	0
	1	**2**	**3**	**4**	**5**	**6**

1. nopan silmäluku *x*

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 7/9

- Esimerkkejä edellisen kalvon taulukon muodostamisesta:

(i) Oletetaan, että 1. nopalla on saatu 2. Tällöin silmälukujen summaksi *ei voi tulla* 10, joten

$$\Pr(X = 2 \text{ ja } Z = 10) = 0$$

(ii) Oletetaan, että 1. nopalla on saatu 2. Tällöin silmälukujen summaksi *voi tulla* 3, 4, 5, 6, 7 tai 8, joten

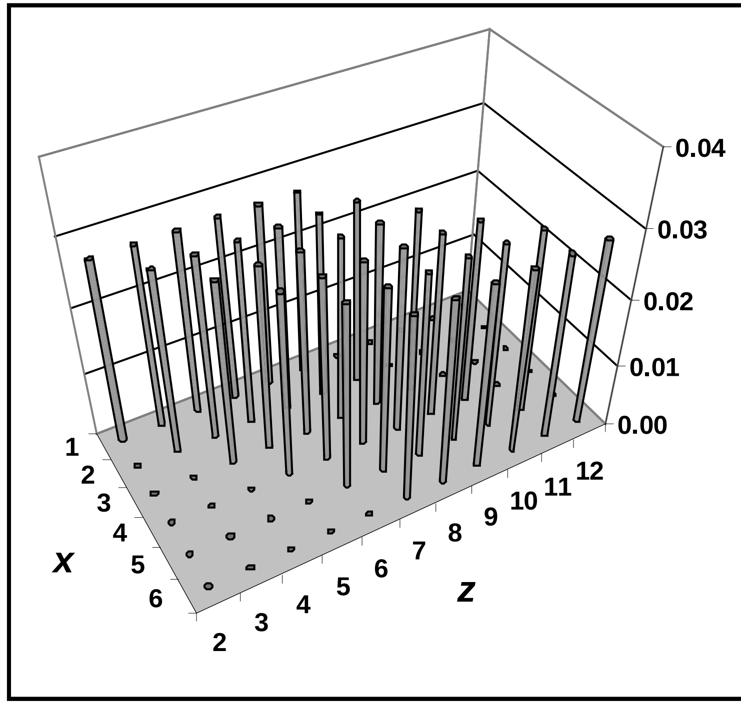
$$\Pr(X = 2 \text{ ja } Z = z) = 1/36 \,; \qquad z = 3, 4, 5, 6, 7, 8$$

Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 8/9

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos 1. heitosta
  Y = tulos 2. heitosta
  $Z = X + Y$

- Kuva oikealla esittää satunnaismuuttujien X ja Z *yhteisjakauman pistetodennäköisyysfunktiota*.
- Kuvassa on 36 pylvästä, joista jokaisen korkeus on 1/36.

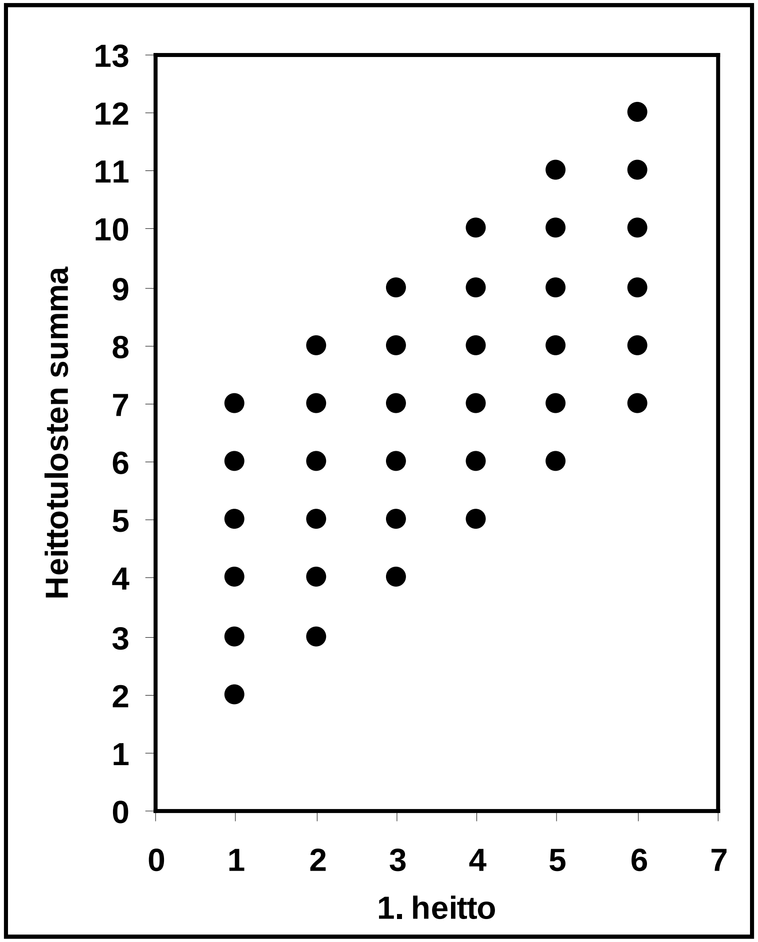


Diskreetit kaksiulotteiset jakaumat: esimerkki nopanheitosta 9/9

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos 1. heitosta
  Y = tulos 2. heitosta
  $Z = X + Y$

- Kuva oikealla havainnollistaa satunnaismuuttujien X ja Z *yhteisjakaumaa*.
- Kuvassa on 36 pistettä, joista jokaisen todennäköisyys on 1/36.



Jatkuvat kaksiulotteiset satunnaismuuttujat ja niiden jakaumat

- Olkoot X ja Y *jatkuvia* satunnaismuuttujia.

- Tällöin järjestetty pari (X, Y) määrittelee **jatkuvan kaksiulotteisen satunnaismuuttujan**.

- Kaksiulotteinen jatkuva satunnaismuuttuja (X, Y) määrittelee **jatkuvan kaksiulotteisen todennäköisyys-jakauman**, jota kutsutaan satunnaismuuttujien X ja Y **yhteisjakaumaksi**.

Jatkuvat kaksiulotteiset jakaumat: Tiheysfunktio

- Reaaliarvoinen jatkuva funktio

$$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

määrittelee *jatkuvien satunnaismuuttujien X* ja *Y* **yhteisjakauman tiheysfunktion**, jos seuraavat ehdot pätevät:

(1) $f_{XY}(x, y) \geq 0$ kaikille x ja y

(2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{XY}(x, y) dy dx = 1$

(3) $\Pr(a \leq X \leq b \text{ ja } c \leq Y \leq d) = \int_a^b \int_c^d f_{XY}(x, y) dy dx$

Jatkuvat kaksiulotteiset jakaumat: Tapahtumien todennäköisyydet

- Olkoon

$$f_{XY} : \mathbb{R}^2 \to \mathbb{R}$$

jatkuvien satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio*.

- Olkoon

$$A \subset \mathbb{R}^2$$

- Tällöin

$$\boxed{\Pr((X,Y) \in A) = \iint_A f_{XY}(x,y)dydx}$$

Jatkuvien kaksiulotteisten jakaumien kertymäfunktiot

- Olkoon (X, Y) *jatkuva* kaksiulotteinen satunnaismuuttuja.
- **Jatkuvan kaksiulotteisen jakauman kertymäfunktio** on

$$\boxed{\begin{aligned} F_{XY}(x,y) &= \Pr(X \le x \text{ ja } Y \le y) \\ &= \int_{-\infty}^{x}\int_{-\infty}^{y} f_{XY}(u,v)\,dv\,du \end{aligned}}$$

jossa f_{XY} satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio*.

Jatkuvan kaksiulotteisen jakauman tiheysfunktion ja kertymäfunktion yhteys

- Olkoon (X, Y) *jatkuva* kaksiulotteinen satunnaismuuttuja.
- Olkoon $F_{XY}(x, y)$ satunnaismuuttujien X ja Y yhteisjakauman *kertymäfunktio*.
- Jos *derivaatta*

$$\frac{\partial^2 F_{XY}(x,y)}{\partial x \partial y} = f_{XY}(x,y)$$

on olemassa ja on jatkuva, funktio $f_{XY}(x, y)$ on satunnaismuuttujien X ja Y yhteisjakauman *tiheysfunktio*.

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Kaksiulotteiset todennäköisyysjakaumat

>> Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus

Kaksiulotteisten jakaumien odotusarvot

Kovarianssi ja korrelaatio

Ehdolliset jakaumat ja odotusarvot

Diskreetin kaksiulotteisen jakauman reunajakaumat

- Olkoon $f_{XY}(x, y)$ *diskreetin* kaksiulotteisen jakauman pistetodennäköisyysfunktio.
- Satunnaismuuttujan X **reunajakauman pistetodennäköisyysfunktio** on

$$\boxed{f_X(x) = \Pr(X = x) = \sum_y f_{XY}(x, y)}$$

- Satunnaismuuttujien X ja Y reunajakaumat *yhtyvät* satunnaismuuttujien X ja Y todennäköisyysjakaumiin.

Diskreetin 2-ulotteisen jakauman reunajakaumat: esimerkki nopanheitosta 1/5

- Heitetään *virheetöntä* noppaa kaksi kertaa.
- Määritellään *satunnaismuuttujat* X, Y ja Z seuraavasti:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta
  $Z = X + Y$ = silmälukujen summa

- Satunnaismuuttujien X, Y ja Z mahdolliset arvot:

  X: {1, 2, 3, 4, 5, 6}
  Y: {1, 2, 3, 4, 5, 6}
  Z: {2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

- Määrätään satunnaismuuttujien X ja Z *reunajakaumat*.

Diskreetin 2-ulotteisen jakauman reunajakaumat: esimerkki nopanheitosta 2/5

- Satunnaismuuttujien X ja Z *yhteisjakauman pistetodennäköisyysfunktio*:

$$\Pr(X = x \text{ ja } Z = z) = f_{XZ}(x, z)\ ;\ x = 1, 2, \dots , 6\ ;\ z = 2, 3, \dots , 12$$

1. nopan silmäluku *x*

Silmälukujen summa *z*

	1	2	3	4	5	6
12	0	0	0	0	0	1/36
11	0	0	0	0	1/36	1/36
10	0	0	0	1/36	1/36	1/36
9	0	0	1/36	1/36	1/36	1/36
8	0	1/36	1/36	1/36	1/36	1/36
7	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	0
5	1/36	1/36	1/36	1/36	0	0
4	1/36	1/36	1/36	0	0	0
3	1/36	1/36	0	0	0	0
2	1/36	0	0	0	0	0

Diskreetin 2-ulotteisen jakauman reunajakaumat: esimerkki nopanheitosta 3/5

- Satunnaismuuttujien X ja Y *reunajakaumien pistetodennäköisyys-funktiot* saadaan määräämällä yhteisjakauman todennäköisyydet antavassa taulukossa *rivi*- ja *sarakesummat*.
- Esimerkkejä:

(i) Satunnaismuuttujan X *reunajakauman* pistetodennäköisyys, kun $X = 4$:

$$f_X(4) = \sum_{z=1}^{12} f_{XZ}(4,z)$$
$$= 0+0+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+\frac{1}{36}+0+0+0 = \frac{1}{6}$$

(ii) Satunnaismuuttujan Z *reunajakauman* pistetodennäköisyys, kun $Z = 10$:

$$f_Z(10) = \sum_{x=1}^{6} f_{XZ}(x,10) = 0+0+0+\frac{1}{36}+\frac{1}{36}+\frac{1}{36} = \frac{3}{36}$$

Diskreetin 2-ulotteisen jakauman reunajakaumat: esimerkki nopanheitosta 4/5

- Satunnaismuuttujien X ja $Z = X + Y$ *yhteisjakauma* ja *reunajakaumat*:

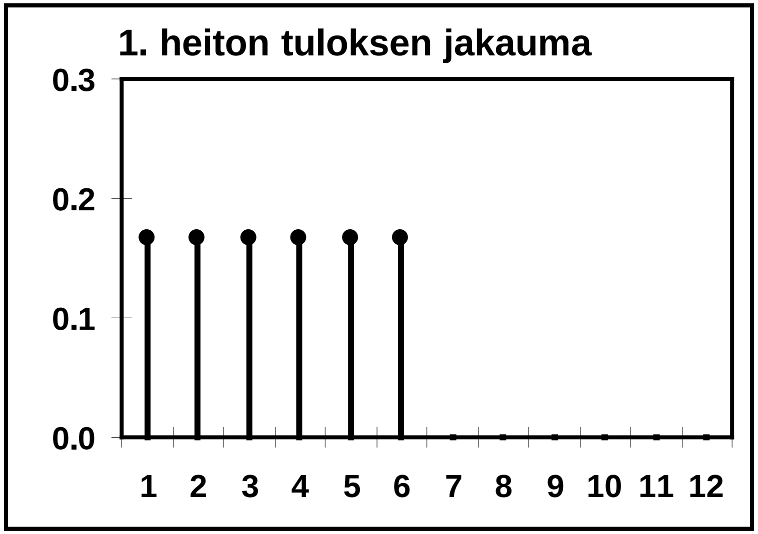
1. nopan silmäluku *x*

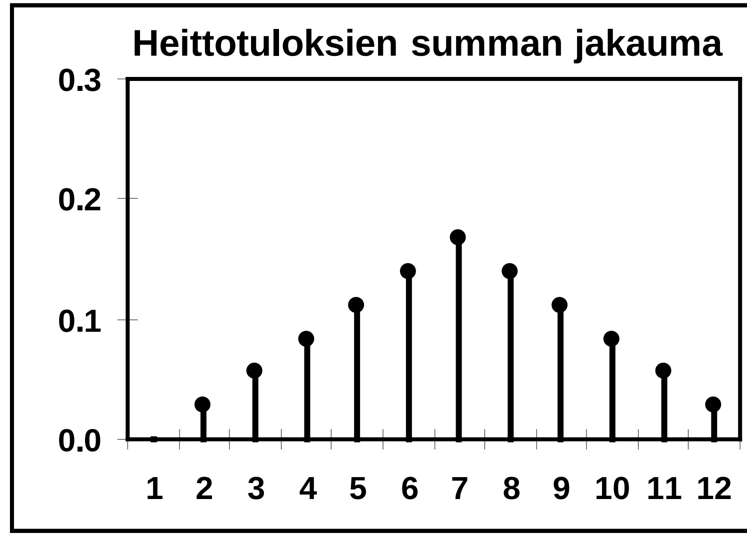
Silmälukujen summa *z*

	1	**2**	**3**	**4**	**5**	**6**	**Yht**	
12	0	0	0	0	0	1/36	**1/36**	
11	0	0	0	0	1/36	1/36	**2/36**	
10	0	0	0	1/36	1/36	1/36	**3/36**	$\leftarrow f_Z(10)$
9	0	0	1/36	1/36	1/36	1/36	**4/36**	
8	0	1/36	1/36	1/36	1/36	1/36	**5/36**	
7	1/36	1/36	1/36	1/36	1/36	1/36	**6/36**	
6	1/36	1/36	1/36	1/36	1/36	0	**5/36**	
5	1/36	1/36	1/36	1/36	0	0	**4/36**	
4	1/36	1/36	1/36	0	0	0	**3/36**	
3	1/36	1/36	0	0	0	0	**2/36**	
2	1/36	0	0	0	0	0	**1/36**	
Yht	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1**	

$\uparrow$
$f_X(4)$

Diskreetin 2-ulotteisen jakauman reunajakaumat: esimerkki nopanheitosta 5/5





- Kuvat yllä esittävät satunnaismuuttujien X ja Z *reunajakaumien pistetodennäköisyysfunktioita*:

  X = tulos 1. heitosta
  Y = tulos 2. heitosta
  $Z = X + Y$ = heittotulosten summa

Jatkuvan kaksiulotteisen jakauman reunajakaumat

- Olkoon $f_{XY}(x, y)$ *jatkuvan* kaksiulotteisen jakauman tiheysfunktio.
- Satunnaismuuttujan X **reunajakauman tiheysfunktio** on

$$\boxed{f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y)dy}$$

- Satunnaismuuttujien X ja Y reunajakaumat *yhtyvät* satunnaismuuttujien X ja Y todennäköisyysjakaumiin.

Satunnaismuuttujien riippumattomuus 1/2

- Oletukset:

(i) Olkoon satunnaismuuttujien X ja Y *yhteisjakauman* pistetodennäköisyys- tai tiheysfunktio $f_{XY}(x, y)$.

(ii) Olkoon satunnaismuuttujan X *reunajakauman* pistetodennäköisyys- tai tiheysfunktio $f_X(x)$.

Olkoon satunnaismuuttujan Y *reunajakauman* pistetodennäköisyys- tai tiheysfunktio $f_Y(y)$.

- Määritelmä:

Satunnaismuuttujat X ja Y ovat **riippumattomia**, jos ja vain, jos

$$f_{XY}(x, y) = f_X(x)f_Y(y)$$

Satunnaismuuttujien riippumattomuus 2/2

- Oletukset:

(i) Olkoon satunnaismuuttujien X ja Y *yhteisjakauman* kertymäfunktio $F_{XY}(x, y)$.

(ii) Olkoon satunnaismuuttujan X *reunajakauman* kertymäfunktio $F_X(x)$.

Olkoon satunnaismuuttujan Y *reunajakauman* kertymäfunktio $F_Y(y)$.

- Määritelmä:

Satunnaismuuttujat X ja Y ovat **riippumattomia**, jos ja vain, jos

$$F_{XY}(x, y) = F_X(x)F_Y(y)$$

Satunnaismuuttujien riippumattomuus: Yleistys 1/2

- Oletukset:

(i) Olkoon satunnaismuuttujien X_i , $i = 1, 2, \ldots , p$ *yhteisjakauman* pistetodennäköisyys- tai tiheysfunktio

$$f(x_1, x_2, \ldots , x_p)$$

(ii) Olkoot satunnaismuuttujien X_i , $i = 1, 2, \ldots , p$ *reunajakaumien* pistetodennäköisyys- tai tiheysfunktiot $f(x_i)$, $i = 1, 2, \ldots , p$.

- Määritelmä:

Satunnaismuuttujat X_i , $i = 1, 2, \ldots , p$ ovat **riippumattomia**, jos ja vain, jos

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

Satunnaismuuttujien riippumattomuus: Yleistys 2/2

- Oletukset:

(i) Olkoon satunnaismuuttujien X_i , $i = 1, 2, \dots , p$ *yhteisjakauman* kertymäfunktio

$$F(x_1, x_2, \dots , x_p)$$

(ii) Olkoot satunnaismuuttujien X_i , $i = 1, 2, \dots , p$ *reunajakaumien* kertymäfunktiot $F(x_i)$, $i = 1, 2, \dots , p$.

- Määritelmä:

Satunnaismuuttujat X_i , $i = 1, 2, \dots , p$ ovat **riippumattomia**, jos ja vain, jos

$$F(x_1, x_2, \dots , x_p) = F(x_1)F(x_2) \cdots F(x_p)$$

Satunnaismuuttujien riippumattomuus ja tapahtumien riippumattomuus

- Olkoot satunnaismuuttujat X ja Y *riippumattomia*.
- Tällöin

$$\Pr(a \le X \le b \text{ ja } c \le Y \le d) = \Pr(a \le X \le b)\Pr(c \le Y \le d)$$

- Huomautus:

Vrt. *riippumattomien tapahtumien tulosääntö*:

$$\Pr(A \cap B) = \Pr(A)\Pr(B)$$

Satunnaismuuttujien riippumattomuus: esimerkki nopanheitosta

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta
  $Z = X + Y$ = silmälukujen summa

- Tarkastellaan satunnaismuuttujien X ja Z *yhteisjakaumaa*.
- Esimerkiksi:

$$\Pr(X = 1 \text{ ja } Z = 8) = 0 \neq \frac{1}{6} \cdot \frac{5}{36} = \Pr(X = 1)\Pr(Z = 8)$$

- Siten satunnaismuuttujat X ja Z *eivät ole riippumattomia*.

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Diskreetin kaksiulotteisen satunnaismuuttujan funktion yleinen odotusarvo

- Olkoon satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$.
- Olkoon

$$g : \mathbb{R}^2 \to \mathbb{R}$$

jatkuva funktio.
- Tällöin satunnaismuuttujan $g(X, Y)$ **odotusarvo** on *vakio*

$$\boxed{\mathrm{E}(g(X,Y)) = \sum_x \sum_y g(x,y) f_{XY}(x,y)}$$

Jatkuvan kaksiulotteisen satunnaismuuttujan funktion yleinen odotusarvo

- Olkoon satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$.
- Olkoon

$$g : \mathbb{R}^2 \to \mathbb{R}$$

  jatkuva funktio.
- Tällöin satunnaismuuttujan $g(X, Y)$ **odotusarvo** on *vakio*

$$\boxed{\mathrm{E}(g(X,Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) f_{XY}(x,y) dydx}$$

Reunajakaumien odotusarvot: Diskreetit jakaumat

- Olkoon *diskreettien* satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$.

- Voimme helposti laskea satunnaismuuttujan X **odotusarvo** $\mathrm{E}(X)$ jos tiedämme $f_{XY}(x, y)$:

$$\mathrm{E}(X) = \sum_x \sum_y x f_{XY}(x, y)$$

Reunajakaumien odotusarvot: Jatkuvat jakaumat

- Olkoon *jatkuvien* satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$.

- Voimme helposti laskea satunnaismuuttujan X **odotusarvo** $\mathrm{E}(X)$ jos tiedämme $f_{XY}(x, y)$:

$$\mathrm{E}(X) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{XY}(x, y)\,dy\,dx$$

Odotusarvot ja todennäköisyysmassan painopiste

- Olkoot satunnaismuuttujien X ja Y *odotusarvot*

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

- Tällöin kaksiulotteisen satunnaismuuttujan (X, Y) *odotusarvo* on *järjestetty pari*

$$(\mathrm{E}(X), \mathrm{E}(Y)) = (\mu_X, \mu_Y)$$

- Satunnaismuuttujien X ja Y odotusarvojen muodostama järjestetty pari (μ_X, μ_Y) määrää satunnaismuuttujien X ja Y *yhteisjakauman todennäköisyysmassan* **painopisteen**.

Riippumattomuus ja tulon odotusarvo

- Olkoot satunnaismuuttujien X ja Y *odotusarvot*

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

- Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, niin **tulon** XY **odotusarvo on odotusarvojen tulo**:

$$\mathrm{E}(XY) = \mathrm{E}(X)\,\mathrm{E}(Y)$$
$$= \mu_X \mu_Y$$

- Huomautus:

  Käänteinen *ei päde*: Siitä, että

  $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$

  ei seuraa, että satunnaismuuttujat X ja Y ovat riippumattomia.

Riippumattomuus ja tulon odotusarvo: Yleistys

- Olkoot satunnaismuuttujien X_i , i = 1, 2, … , k *odotusarvot*

$$\mathrm{E}(X_i) = \mu_i \ , \ i = 1, 2, \ldots, k$$

- Jos satunnaismuuttujat X_i , i = 1, 2, … , k ovat *riippumattomia*, niin

$$\mathrm{E}(X_1 X_2 \cdots X_k) = \mathrm{E}(X_1)\,\mathrm{E}(X_2)\cdots\mathrm{E}(X_k)$$
$$= \mu_1 \mu_2 \cdots \mu_k$$

- Huomautus:

Käänteinen *ei päde*: Siitä, että

$$\mathrm{E}(X_1 X_2 \cdots X_k) = \mathrm{E}(X_1)\,\mathrm{E}(X_2)\cdots\mathrm{E}(X_k)$$

ei seuraa, että satunnaismuuttujat X_i , i = 1, 2, … , k ovat riippumattomia.

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Kaksiulotteiset todennäköisyysjakaumat

Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus

Kaksiulotteisten jakaumien odotusarvot

>> Kovarianssi ja korrelaatio

Ehdolliset jakaumat ja odotusarvot

Kovarianssi

- Satunnaismuuttujien riippuvuus voi voimakkuudeltaan vaihdella täydellisestä riippumattomuudesta aina täydelliseen funktionaaliseen riippuvuuteen.

Kovarianssi

- Olkoot satunnaismuuttujien X ja Y *odotusarvot*

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{E}(Y) = \mu_Y$$

- Satunnaismuuttujien X ja Y **kovarianssi** on *vakio*

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

- Satunnaismuuttujien X ja Y kovarianssi Cov(X, Y)

  kuvaa satunnaismuuttujien X ja Y *yhteisjakauman toden-näköisyysmassan* **yhteisvaihtelua** *satunnaismuuttujien X ja Y odotusarvojen* E(X) *ja* E(Y) *määräämän pisteen* (E(X), E(Y)) *ympärillä*.

Kovarianssi: Diskreetit jakaumat

- Olkoon *diskreettien* satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$.
- Tällöin satunnaismuuttujien X ja Y **kovarianssi** on *vakio*

$$\mathrm{Cov}(X,Y) = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f_{XY}(x,y)$$

jossa

$$\mu_X = \mathrm{E}(X)$$

$$\mu_Y = \mathrm{E}(Y)$$

Kovarianssi:
Jatkuvat jakaumat

- Olkoon *jatkuvien* satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$.
- Tällöin satunnaismuuttujien X ja Y **kovarianssi** on *vakio*

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x-\mu_X)(y-\mu_Y)\, f_{XY}(x,y)dxdy$$

jossa

$$\mu_X = \mathrm{E}(X)$$

$$\mu_Y = \mathrm{E}(Y)$$

Kovarianssi: Vaihtoehtoinen laskukaava

- Satunnaismuuttujien X ja Y *kovarianssin* kaava voidaan kirjoittaa seuraaviin yhtäpitäviin muotoihin:

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

- Huomautus:

  Vrt. kovarianssin vaihtoehtoista laskukaavaa varianssin vaihtoehtoisiin laskukaavoihin.

Kovarianssi ja varianssit

- Satunnaismuuttujien kovarianssit itsensä kanssa yhtyvät satunnaismuuttujien varriansseihin:

$$
\begin{aligned}
\mathrm{Cov}(X,X) &= \mathrm{E}[(X-\mu_X)(X-\mu_X)] \\
&= \mathrm{Var}(X) \\
&= \sigma_X^2 \\
\mathrm{Cov}(Y,Y) &= \mathrm{E}[(Y-\mu_Y)(Y-\mu_Y)] \\
&= \mathrm{Var}(Y) \\
&= \sigma_Y^2
\end{aligned}
$$

Kovarianssiin liittyviä laskukaavoja

- Olkoot $W = a + bX$ jossa $a, b, c, d \in \mathbb{R}$.

  $Z = c + dY$

  Tällöin

  $$\operatorname{Cov}(W, Z) = bd \operatorname{Cov}(X, Y) = bd\sigma_{XY}$$

- Summan ja erotuksen varianssi yleisessä tapauksessa

$$\boxed{\begin{aligned}\operatorname{Var}(X \pm Y) &= \operatorname{Var}(X) + \operatorname{Var}(Y) \pm 2\operatorname{Cov}(X, Y) \\ &= \sigma_X^2 + \sigma_Y^2 \pm 2\sigma_{XY}\end{aligned}}$$

Riippumattomuus ja kovarianssi

- Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, niin

$$\mathrm{Cov}(X,Y)=0$$

mutta käänteinen *ei aina päde*.

- Perustelu: Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, niin

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$$

Siten

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)\\ &= \mathrm{E}(X)\,\mathrm{E}(Y)-\mathrm{E}(X)\,\mathrm{E}(Y)\\ &= 0\end{aligned}$$

Korrelaatiokerroin 1/2

- Oletetaan, että satunnaismuuttujilla X ja Y on seuraavat odotusarvot, varianssit ja kovarianssi:

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{Var}(X) = \mathrm{D}^2(X) = \sigma_X^2$$

$$\mathrm{E}(Y) = \mu_Y \qquad \mathrm{Var}(Y) = \mathrm{D}^2(Y) = \sigma_Y^2$$

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

Korrelaatiokerroin 2/2

- Satunnaismuuttujien *X* ja *Y* (Pearsonin tulomomentti-) **korrelaatiokerroin** on *vakio*

$$
\begin{aligned}
\mathrm{Cor}(X,Y) &= \rho_{XY} \\
&= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \\
&= \frac{\sigma_{XY}}{\sigma_X \sigma_Y}
\end{aligned}
$$

Korrelaatiokerroin lineaarisen riippuvuuden mittana

- Satunnaismuuttujien X ja Y korrelaatiokerroin

  $\mathrm{Cor}(X, Y)$

  mittaa satunnaismuuttujien X ja Y **lineaarisen riippuvuuden** *voimakkuutta*:

  (i) Mitä *suurempi* on

  $|\mathrm{Cor}(X, Y)|$

  sitä *voimakkaampaa* on satunnaismuuttujien X ja Y välinen lineaarinen riippuvuus.

  (ii) Mitä *pienempi* on

  $|\mathrm{Cor}(X, Y)|$

  sitä *heikompaa* on satunnaismuuttujien X ja Y välinen lineaarinen riippuvuus.

Korrelaatiokerroin: Ominaisuudet

- Olkoon satunnaismuuttujien X ja Y *korrelaatiokerroin* $\mathrm{Cor}(X, Y)$.
- Tällöin

(i) $-1 \le \mathrm{Cor}(X,Y) \le +1$

(ii) Jos X ja Y ovat *riippumattomia*, niin $\mathrm{Cor}(X,Y) = 0$

(iii) $\mathrm{Cor}(X,Y) = \pm 1$, jos ja vain, jos

$$Y = \alpha + \beta X$$

jossa α ja β ovat reaalisia *vakiota*, $\beta \neq 0$ ja lisäksi

$$\mathrm{sgn}(\mathrm{Cor}(X,Y)) = \mathrm{sgn}(\beta)$$

Korrelaatiokerroin: esimerkki nopanheitosta 1/2

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta
  $Z = X + Y$ = silmälukujen summa

- Tarkastellaan satunnaismuuttujien X ja Z *yhteisjakaumaa*.
- *Odotusarvot*, *varianssit* ja *standardipoikkeamat*:

$\mathrm{E}(X) = 21/6 = 3.5$ $\quad$ $\mathrm{E}(Z) = 252/36 = 7$

$\mathrm{D}^2(X) = 35/12 = 2.917$ $\quad$ $\mathrm{D}^2(Z) = 210/36 = 5.833$

$\mathrm{D}(X) = 1.708$ $\quad$ $\mathrm{D}(Z) = 2.415$

Korrelaatiokerroin: 2. esimerkki nopanheitosta 2/2

- Satunnaismuuttujat X ja Z *eivät ole riippumattomia*.
- Lasketaan ensin *kovarianssi*:

$$\mathrm{E}(XZ) = \sum_{x=1}^{6}\sum_{z=2}^{12} xz\,\Pr(X = x, Z = z) = \frac{987}{36}$$

$$\mathrm{Cov}(X, Z) = \mathrm{E}(XZ) - \mathrm{E}(X)\,\mathrm{E}(Z) = \frac{987}{36} - \frac{21}{6}\cdot\frac{42}{6} = \frac{105}{36} = 2.917$$

- *Korrelaatiokertoimen* arvo on:

$$\mathrm{Cor}(X, Z) = \frac{\mathrm{Cov}(X, Z)}{\mathrm{D}(X)\,\mathrm{D}(Z)} = \frac{1}{\sqrt{2}} = 0.707$$

Korreloimattomuus

- Jos

  $$\mathrm{Cov}(X,Y) = 0$$

  tai yhtäpitävästi

  $$\mathrm{Cor}(X,Y) = 0$$

  niin sanomme, että satunnaismuuttujat X ja Y ovat **korreloimattomia**.
- Huomautus:

  Satunnaismuuttujien korreloimattomuudesta *ei välttämättä seuraa* niiden riippumattomuus. Korrelaatio mittaa satunnaismuuttujien *lineaarisen* riippuvuuden voimakkuutta. Siten korreloimattomien satunnaismuuttujien välillä voi olla jopa eksakti epälineaarinen riippuvuus.

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Kaksiulotteiset todennäköisyysjakaumat
Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus
Kaksiulotteisten jakaumien odotusarvot
Kovarianssi ja korrelaatio
\>> Ehdolliset jakaumat ja odotusarvot

Ehdolliset jakaumat

- Kahden satunnaismuuttujan X ja Y yhteisjakauma kuvailee muuttujien yhteistä käyttäytymistä.

- Kun halutaan kuvailla, miten tieto toisen muuttujan saamasta arvosta vaikuttaa toiseen muuttujaan liittyviin todennäköisyyksiin, tarkastellaan ehdollisia jakaumia.

Ehdolliset jakaumat

- Olkoon satunnaismuuttujien X ja Y yhteisjakauman *pistetodennäköisyys*- tai *tiheysfunktio* $f_{XY}(x, y)$.
- Olkoot satunnaismuuttujien X ja Y *reunajakaumien* pistetodennäköisyys- tai tiheysfunktiot $f_X(x)$ ja $f_Y(y)$.

- Satunnaismuuttujan X **ehdollinen jakauma** satunnaismuuttujan Y suhteen (ehdolla $Y = y$) on

$$\boxed{f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)} \text{ , jos } f_Y(y) > 0}$$

Riippumattomuus ja ehdolliset jakaumat

- Olkoot satunnaismuuttujat X ja Y *riippumattomia*.

- Tällöin

  $$f_{XY}(x, y) = f_X(x)\, f_Y(y)$$

  ja satunnaismuuttujan X *ehdollinen jakauma* satunnaismuuttujan Y suhteen yhtyy satunnaismuuttujan X *reunajakaumaan*:

  $$f_{X|Y}(x|y) = f_X(x)\ ,\ \text{jos}\ f_Y(y) > 0$$

- Tällöin satunnaismuuttujan X ehdollinen jakauma satunnaismuuttujan Y suhteen *ei riipu* ehtomuuttujan Y arvoista.

Ehdolliset odotusarvot: Diskreetit jakaumat

- Olkoot satunnaismuuttujat X ja Y *diskreettejä*.

- Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

$$\boxed{\mathrm{E}(X \mid Y = y) = \sum_x x f_{X|Y}(x \mid y)}$$

Ehdolliset odotusarvot: Jatkuvat jakaumat

- Olkoot satunnaismuuttujat X ja Y *jatkuvia*.

- Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

$$\boxed{\mathrm{E}(X \mid Y = y) = \int_{-\infty}^{+\infty} x f_{X|Y}(x \mid y)\,dx}$$

Riippumattomuus ja ehdolliset odotusarvot

- Jos satunnaismuuttujat X ja Y ovat *riippumattomia*, ehdolliset odotusarvot *yhtyvät* niiden *reunajakaumien odotusarvoihin*.
- Jos siis X ja Y ovat *riippumattomia*, seuraava pätee:

$$\mathrm{E}(X \mid Y) = \mathrm{E}(X)$$

$$\mathrm{E}(Y \mid X) = \mathrm{E}(Y)$$

- Tällöin satunnaismuuttujan X ehdollinen odotusarvo satunnaismuuttujan Y suhteen *ei riipu* ehtomuuttujan Y arvoista.

Regressiofunktiot ja -käyrät

- Tarkastellaan satunnaismuuttujan X *ehdollista odotusarvoa*

$$\mathrm{E}(X \mid Y = y)$$

ehtomuuttujan Y arvojen y *funktiona*.

- Tätä funktiota kutsutaan satunnaismuuttujan X **regressiofunktioksi** satunnaismuuttujan Y suhteen.
- Satunnaismuuttujan X regressiofunktio muuttujan Y suhteen määrittelee **regressiokäyrän**

$$\begin{aligned} x &= g_y(y) \\ &= \mathrm{E}(X \mid Y = y) \end{aligned}$$

Regressiokäyrät ja ennustaminen 1/4

- Oletetaan, että satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyys*- tai *tiheysfunktio* f_{XY} tunnetaan.

- Haluamme **ennustaa** satunnaismuuttujan X (tai Y) arvon satunnaismuuttujan Y (tai X) saaman arvon perusteella.

Regressiokäyrät ja ennustaminen 2/4

- Tehtävä 1:
  (i) Haluamme *ennustaa* satunnaismuuttujan X arvon satunnaismuuttujan Y saaman arvon perusteella.
  (ii) Olkoon ennustettu arvo $d(X \mid Y)$.
  (iii) Miten ennuste $d(X \mid Y)$ valitaan *optimaalisella tavalla*?

- Tehtävä 2:
  (i) Haluamme *ennustaa* satunnaismuuttujan Y arvon satunnaismuuttujan X saaman arvon perusteella.
  (ii) Olkoon ennustettu arvo $d(Y \mid X)$.
  (iii) Miten ennuste $d(Y \mid X)$ valitaan *optimaalisella tavalla*?

Regressiokäyrät ja ennustaminen 3/4

- Tehtävän 1 ratkaisu:

  Valitaan $d(X \mid Y)$ siten, että *ennusteen* **keskineliövirhe**

  $$\mathrm{E}[X - d(X \mid Y)]^2$$

  minimoituu.
- Voidaan osoittaa, että keskineliövirhe minimoituu valinnalla

  $$d(X \mid Y) = \mathrm{E}(X \mid Y)$$
- Siten *ehdollinen odotusarvo* $\mathrm{E}(X \mid Y)$ on *keskineliövirheen mielessä optimaalinen ennuste* satunnaismuuttujan X saamille arvoille.

Regressiokäyrät ja ennustaminen 4/4

- Tehtävän 2 ratkaisu:

  Valitaan $d(Y \mid X)$ siten, että *ennusteen* **keskineliövirhe**

  $$\mathrm{E}[Y - d(Y \mid X)]^2$$

  minimoituu.
- Voidaan osoittaa, että keskineliövirhe minimoituu valinnalla

  $$d(Y \mid X) = \mathrm{E}(Y \mid X)$$
- Siten *ehdollinen odotusarvo* $\mathrm{E}(Y \mid X)$ on *keskineliövirheen mielessä optimaalinen ennuste* satunnaismuuttujan Y saamille arvoille.

Todennäköisyyslaskenta

Osa 3: Todennäköisyysjakaumia

- **Kaksiulotteinen normaalijakauma**

Kaksiulotteinen normaalijakauma

- *Kaksiulotteinen normaalijakauma* on *normaalijakauman* (ks. lukua **Jatkuvia jakaumia**) *kaksiulotteinen yleistys*.
- Huomautus:

  Normaalijakauman yleistystä *p-ulotteiseen avaruuteen* ($p > 1$) kutsutaan **multinormaalijakaumaksi** tai ***p*-ulotteiseksi normaalijakaumaksi**.

Kaksiulotteinen normaalijakauma ja sen tiheysfunktio 1/2

- Satunnaismuuttujat X ja Y noudattavat **kaksiulotteista normaalijakaumaa**, jos niiden *yhteisjakauman* **tiheysfunktio** on muotoa

$$f_{XY}(x,y)=\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho_{XY}^2}}\exp\left\{-\frac{1}{2(1-\rho_{XY}^2)}Q(x,y)\right\}$$

jossa

$$Q(x,y)=\left(\frac{x-\mu_X}{\sigma_X}\right)^2+\left(\frac{y-\mu_Y}{\sigma_Y}\right)^2-2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

- Merkintä:

$$(X, Y) \sim \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

Kaksiulotteinen normaalijakauma ja sen tiheysfunktio 2/2

- Kaksiulotteisen normaalijakauman

  $N_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$

  parametrien on toteutettava seuraavat ehdot:

  $-\infty < \mu_X < +\infty \qquad \sigma_X > 0$

  $-\infty < \mu_Y < +\infty \qquad \sigma_Y > 0$

  $-1 < \rho_{XY} < +1$

Kaksiulotteinen normaalijakauman parametrit

- Olkoon

$$(X, Y) \sim \mathrm{N}_2(\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho_{XY})$$

- Kaksiulotteisen normaalijakauman **parametreina**, jotka täysin määräävät jakauman, ovat satunnaismuuttujien X ja Y **odotusarvot** ja **varianssit** sekä niiden **korrelaatio**:

$$\mathrm{E}(X) = \mu_X \qquad \mathrm{Var}(X) = \sigma_X^2$$

$$\mathrm{E}(Y) = \mu_Y \qquad \mathrm{Var}(Y) = \sigma_Y^2$$

$$\mathrm{Cor}(X, Y) = \rho_{XY}$$

- Lisäksi

$$\mathrm{Cov}(X, Y) = \sigma_{XY} = \rho_{XY}\sigma_X\sigma_Y$$

Tiheysfunktion ominaisuudet

- Kaksiulotteisen normaalijakauman tiheysfunktio määrittelee *pinnan*

$$z = f_{XY}(x, y)$$

kolmiulotteisessa avaruudessa.

- Pinnalla on *maksimi* satunnaismuuttujien X ja Y odotusarvojen μ_X ja μ_Y määräämässä jakauman todennäköisyysmassan *painopisteessä* (μ_X, μ_Y).
- Pinnan *muodon* määräävät **tasa-arvoellipsit**

$$Q(x,y) = \left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2 - 2\rho_{XY}\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right)$$

$$= c \text{ (vakio)}$$

Tasa-arvoellipsien ominaisuudet 1/3

- Kaksiulotteisen normaalijakauman tiheysfunktion muodostaman pinnan muodon määräävillä *tasa-arvoellipseillä* on seuraavat *ominaisuudet*:

  (i) Ellipsien *keskipisteenä* on jakauman todennäköisyys-massan *painopiste*

  $$(\mu_X, \mu_Y)$$

  (ii) Ellipsien *eksentrisyys* on sekä korrelaatiokertoimen ρ_{XY} että standardipoikkeamien σ_X ja σ_Y funktio.

  (iii) Ellipsi on *sitä eksentrisempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

  $$|\rho_{XY}|$$

Tasa-arvoellipsien ominaisuudet 2/3

(iv) Jos

$$\rho_{XY} = 0$$

ellipsien *pääakselit* ovat koordinaattiakseleiden suuntaiset.

(v) Jos

$$\rho_{XY} = 0$$

ja lisäksi

$$\sigma_X = \sigma_Y$$

niin ellipsit ovat *ympyröitä*.

(vi) Jos

$$\rho_{XY} = \pm 1$$

niin ellipsit surkastuvat *janoiksi*.

Kaksiulotteinen normaalijakauma

Tasa-arvoellipsien ominaisuudet 3/3

- *Tasa-arvoellipsien pääakselit* ovat satunnaismuuttujien X ja Y **kovarianssimatriisin**

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_X^2 & \sigma_{XY} \\ \sigma_{XY} & \sigma_Y^2 \end{bmatrix}$$

ominaisvektoreiden suuntaiset ja niiden pituudet suhtautuvat toisiinsa kuten matriisin $\mathbf{\Sigma}$ **ominaisarvojen** *neliöjuuret*.

Esimerkki:
Jakauman määrittely

- Olkoon

  $(X, Y) \sim N_2(4, 3, 2, 1, 0.7)$

- Jakauman *parametrit* ovat

  $$\mathrm{E}(X) = \mu_X = 4 \qquad \mathrm{Var}(X) = \sigma_X^2 = 2$$

  $$\mathrm{E}(Y) = \mu_Y = 3 \qquad \mathrm{Var}(Y) = \sigma_Y^2 = 1$$

  $$\mathrm{Cor}(X, Y) = \rho_{XY} = 0.7$$

- Siten

  $$\mathrm{Cov}(X, Y) = \rho_{XY}\sigma_X\sigma_Y = 0.7 \times \sqrt{2} \times 1 = 0.9899$$

Esimerkki: Tiheysfunktion kuvaaja

- Olkoon

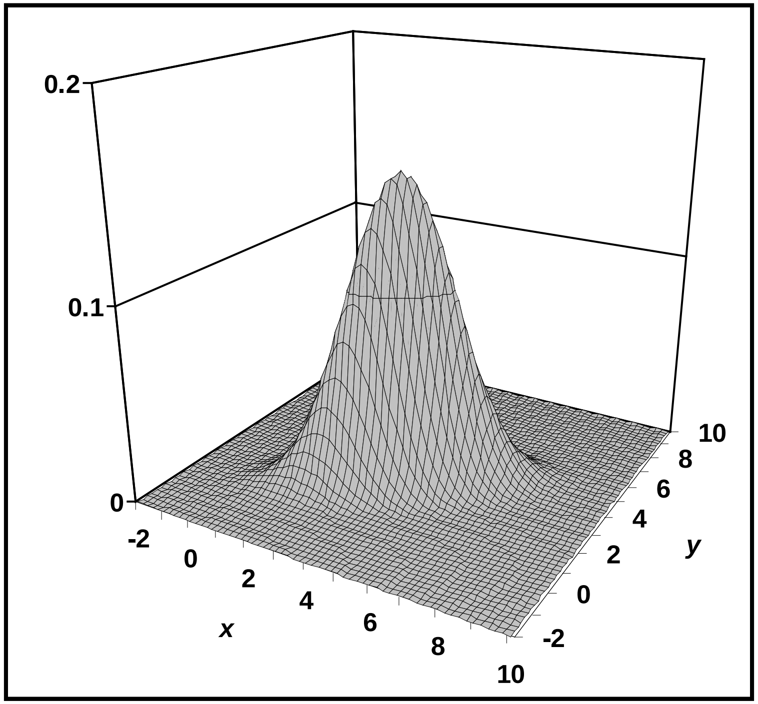
  $$(X, Y) \sim \mathrm{N}_2(4, 3, 2, 1, 0.7)$$

  jolloin

  $$\mu_X = 4 \qquad \sigma_X^2 = 2$$
  $$\mu_Y = 3 \qquad \sigma_Y^2 = 1$$
  $$\rho_{XY} = 0.7$$

- Kuva oikealla esittää jakauman *tiheysfunktiota*

  $$f_{XY}(x, y)$$



Kaksiulotteinen normaalijakauma

Esimerkki: Tasa-arvoellipsit

- Olkoon

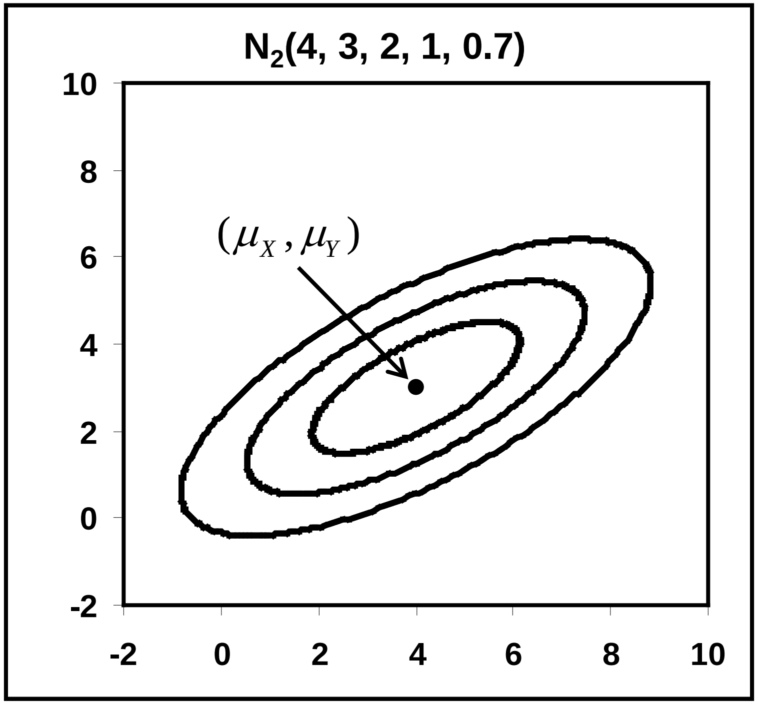
  $(X, Y) \sim \mathrm{N}_2(4, 3, 2, 1, 0.7)$

  jolloin

  $$\mu_X = 4 \qquad \sigma_X^2 = 2$$
  $$\mu_Y = 3 \qquad \sigma_Y^2 = 1$$
  $$\rho_{XY} = 0.7$$

- Kuva oikealla esittää jakauman tiheysfunktion kuvaajan *tasa-arvoellipsejä*, jotka vastaavat (likimäärin) *todennäköisyyksiä* 68 %, 95 % ja 99.7 %.

  Esimerkiksi *uloimman* ellipsin sisään jää n. 99.7 % jakauman todennäköisyysmassasta.



Reunajakaumat

- Voidaan osoittaa, että kaksiulotteisen normaalijakauman **reunajakaumat** ovat *normaalisia*:

  $X \sim \mathrm{N}(\mu_X,\ \sigma_X^2)$

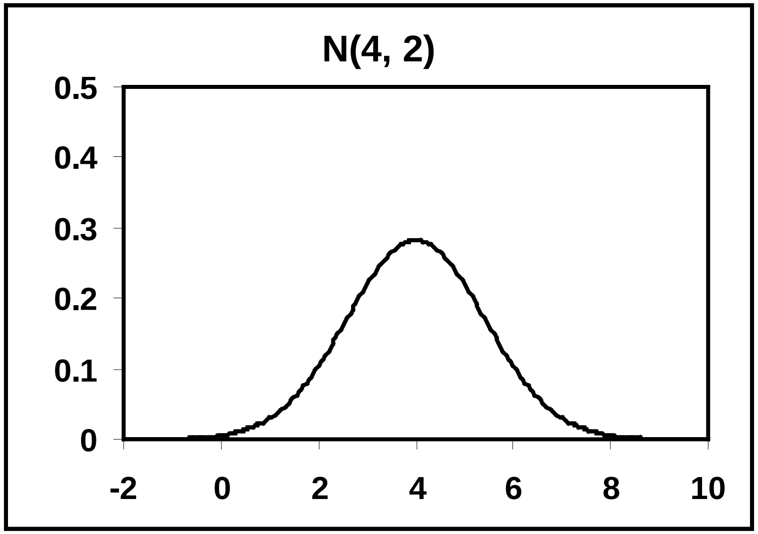
  $Y \sim \mathrm{N}(\mu_Y,\ \sigma_Y^2)$

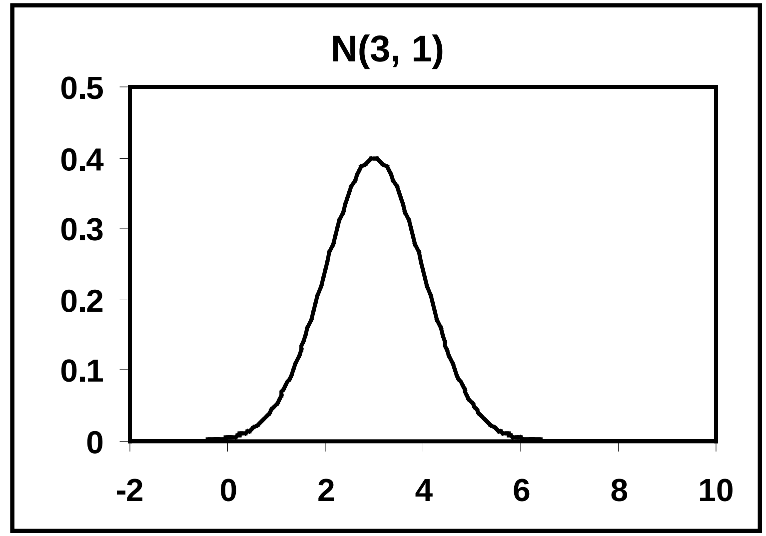
  ja niiden *tiheysfunktiot* ovat

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_X} \exp\left\{ -\frac{1}{2} \left(\frac{x-\mu_X}{\sigma_X} \right)^2 \right\}$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_Y} \exp\left\{ -\frac{1}{2} \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right\}$$

Esimerkki: Reunajakaumat





- Olkoon

  $(X, Y) \sim \mathrm{N}_2(4, 3, 2, 1, 0.7)$

- Kuvat yllä esittävät satunnaismuuttujien X ja Y *reunajakaumia*:

  $X \sim \mathrm{N}(4, 2)$

  $Y \sim \mathrm{N}(3, 1)$

Reunajakaumat

- Huomautus:

  Jos yhteisjakauman reunajakaumat ovat normaalijakaumia, ei yhteisjakauman tarvitse olla kaksiulotteinen normaalijakauma.

Korreloimattomuus *vs* riippumattomuus

- **Kaksiulotteisen normaalijakauman tapauksessa satunnaismuuttujien *X* ja *Y korreloimattomuus on yhtäpitävää niiden riippumattomuuden kanssa*.**
- Huomautuksia:
  - Satunnaismuuttujien riippumattomuudesta *seuraa aina* niiden korreloimattomuus.
  - Satunnaismuuttujien korreloimattomuudesta *ei yleisesti seuraa* niiden riippumattomuus.

Ehdolliset jakaumat

- Kaksiulotteisen normaalijakauman **ehdolliset jakaumat** ovat *normaalisia*:

$$(X \mid Y = y) \sim \mathrm{N}\left(\mu_{X|Y}, \sigma^2_{X|Y}\right)$$

jossa

$$\mu_{X|Y} = \mathrm{E}(X \mid Y = y) = \mu_X + \rho_{XY}\frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

$$\sigma^2_{X|Y} = \mathrm{Var}(X \mid Y = y) = (1 - \rho^2_{XY})\sigma^2_X$$

Ehdolliset odotusarvot

- Satunnaismuuttujan X **ehdollinen odotusarvo** eli **regressiofunktio** satunnaismuuttujan Y suhteen

$$\mathrm{E}(X \mid Y = y) = \mu_X + \rho_{XY} \frac{\sigma_X}{\sigma_Y}(y - \mu_Y)$$

  on *lineaarinen* satunnaismuuttujan Y arvojen y suhteen.

- Satunnaismuuttujan Y **ehdollinen odotusarvo** eli **regressiofunktio** satunnaismuuttujan X suhteen

$$\mathrm{E}(Y \mid X = x) = \mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

  on *lineaarinen* satunnaismuuttujan X arvojen x suhteen.

Regressiosuorat

- Kaksiulotteisen normaalijakauman *regressiokäyrät* ovat **suoria**, joiden yhtälöt voidaan kirjoittaa satunnais-muuttujan *X* saamien arvojen *x* funktioina seuraaviin muotoihin:

  (i) *x*:n regressiosuora *y*:n suhteen:

  $$y = \mu_Y + \frac{1}{\rho_{XY}} \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

  (ii) *y*:n regressiosuora *x*:n suhteen:

  $$y = \mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

Regressiosuorien ominaisuudet 1/4

- Regressiosuorilla on seuraavat ominaisuudet:

(i) Molemmat regressiosuorat kulkevat jakauman todennäköisyysmassan painopisteen (μ_X, μ_Y) kautta.

(ii) Molempien regressiosuorien kulmakertoimilla ja satunnaismuuttujien X ja Y korrelaatiokertoimella ρ_{XY} on *aina sama merkki*:

– Suorat ovat *nousevia*, jos $\rho_{XY} > 0$.

– Suorat ovat *laskevia*, jos $\rho_{XY} < 0$.

(iii) x:n regressiosuora y:n suhteen on *aina jyrkempi* kuin y:n regressiosuora x:n suhteen, koska

$$\rho_{XY}^2 \leq 1$$

Regressiosuorien ominaisuudet 2/4

(iv) x:n regressiosuora y:n suhteen on *sitä loivempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

$$|\rho_{XY}|$$

(v) y:n regressiosuora x:n suhteen on *sitä jyrkempi mitä voimakkaammin satunnaismuuttujat X ja Y ovat korreloituneita* eli mitä suurempi on

$$|\rho_{XY}|$$

Kaksiulotteinen normaalijakauma

Regressiosuorien ominaisuudet 3/4

(vii) Molemmat regressiosuorat ovat *sitä jyrkempiä mitä pienempi on satunnaismuuttujan X varianssi*

$$\sigma_X^2$$

(vi) Molemmat regressiosuorat ovat *sitä jyrkempiä mitä suurempi on satunnaismuuttujan Y varianssi*

$$\sigma_Y^2$$

(viii) Regressiosuorat *yhtyvät* täsmälleen silloin, kun

$$\rho = \pm 1$$

Regressiosuorien ominaisuudet 4/4

(ix) Jos $\rho = 0$, niin regressiosuorat ovat *kohtisuorassa toisiaan vastaan* ja x:n regressiosuora y:n suhteen on

$$x = \mu_X$$

ja y:n regressiosuora x:n suhteen on

$$y = \mu_Y$$

jolloin x:n saamat arvot *eivät riipu* y:n saamista arvoista ja y:n saamat arvot *eivät riipu* x:n saamista arvoista.

Kaksiulotteinen normaalijakauma

Esimerkki:
Regressiosuorat 1/2

- Olkoon

$$(X, Y) \sim N_2(4, 3, 2, 1, 0.7)$$

- x:n regressiosuora muuttujan y suhteen on

$$y = \mu_Y + \frac{1}{\rho_{XY}} \times \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$= 3 + \frac{1}{0.7} \times \frac{1}{\sqrt{2}}(x - 4) = -1.0406 + 1.0101x$$

- y:n regressiosuora muuttujan x suhteen on

$$y = \mu_Y + \rho_{XY} \frac{\sigma_Y}{\sigma_X}(x - \mu_X)$$

$$= 3 + 0.7 \frac{1}{\sqrt{2}}(x - 4) = 1.0201 + 0.4950x$$

Esimerkki: Regressiosuorat 2/2

- Olkoon

$$(X, Y) \sim \mathrm{N}_2(4, 3, 2, 1, 0.7)$$

- Kuva oikealla esittää jakauman tiheysfunktion kuvaajan *tasa-arvoellipsejä*, jotka vastaavat (likimäärin) *todennäköisyyksiä* 68 %, 95 % ja 99.7 %.
- Kuvan suorista *jyrkempi*

$$y = -1.0406 + 1.0101 \times x$$

on *x:n regressiosuora y:n suhteen* ja suorista *loivempi*

$$y = 1.0201 + 0.4950 \times x$$

on *y:n regressiosuora x:n suhteen*.