Ilkka Mellin

Todennäköisyyslaskenta

Osa 2: Satunnaismuuttujat ja todennäköisyysjakaumat

Moniulotteiset satunnaismuuttujat ja jakaumat

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

>> **Kaksiulotteiset todennäköisyysjakaumat**

Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus

Kaksiulotteisten jakaumien odotusarvot ja varianssit

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 kokonais-kulutus (l *per capita* vuodessa) *riippuu* alkoholijuomien hinta-tasosta, käytettävissä olevista tuloista ja alkoholin saatavuudesta?
- Miten todennäköisyys sairastua keuhkosyöpään (p) *riippuu* tupakoinnin määrästä ja kestosta?
- Miten vehnän sato (t/ha) *riippuu* kesän keskilämpötilasta, sade-määrästä, maan muokkaus-tavoista, lannoituksesta ja tuholaisten torjunnasta?

Kaksiulotteiset satunnaismuuttujat

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

$$X : S \rightarrow \mathbb{R}$$

$$Y : R \rightarrow \mathbb{R}$$

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

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

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

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

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

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

jokin *tapahtuma*.

- Tällöin

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

Diskreetit kaksiulotteiset jakaumat: Symmetriset todennäköisyyskentät

- Olkoon (x_i, y_i), $i = 1, 2, \ldots , n$ erillisten tason pisteiden muodostama *diskreetti* pistejoukko.
- Määritellään *diskreetin kaksiulotteisen jakauman piste-todennäköisyysfunktio* kaavalla

$$f_{XY}(x_i, y_i) = \Pr(X = x_i \text{ ja } Y = y_i) = \frac{1}{n}, \; i = 1, 2, \ldots, n$$

- Tällöin satunnaismuuttuja (X, Y) määrittelee **symmetrisen todennäköisyyskentän**, jossa kaikki alkeistapahtumat

$$(x_i, y_i) \;, \; i = 1, 2, \ldots , n$$

ovat *yhtä todennäköisiä*.

Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 1/6

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

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta

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

  X: $\{1, 2, 3, 4, 5, 6\}$
  Y: $\{1, 2, 3, 4, 5, 6\}$

- Muodostetaan satunnaismuuttujien X ja Y *yhteisjakauma*.

Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 2/6

- Kahden nopanheiton tulosvaihtoehdot voidaan esittää seuraavana taulukkona:

Tulos 1. nopanheitosta	**Mahdolliset tulokset 2. nopanheitosta**					
1	1	2	3	4	5	6
2	1	2	3	4	5	6
3	1	2	3	4	5	6
4	1	2	3	4	5	6
5	1	2	3	4	5	6
6	1	2	3	4	5	6

- Siten satunnaismuuttujien X ja Y järjestetty pari (X, Y) määrittelee *diskreetin kaksiulotteisen satunnaismuuttujan*, jonka arvoina on 36 lukuparia

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

Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 3/6

- Koska noppa oletettiin *virheettömäksi* ja heittojen tulokset oletettiin *riippumattomiksi*, on luontevaa ajatella, että kahden nopanheiton tulosten muodostamat 36 tulosvaihtoehtoa

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

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

  $$f_{XY}(x,y) = \Pr(X = x, Y = y) = \Pr(X = x)\Pr(Y = y) = \frac{1}{6}\cdot\frac{1}{6} = \frac{1}{36}$$

  kun

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

Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 4/6

- Satunnaismuuttujien X ja Y *yhteisjakauman* pistetodennäköisyysfunktio

$$f_{XY}(x,y) = \Pr(X = x \text{ ja } Y = y) = \frac{1}{36}$$

$$x = 1,2,3,4,5,6\,;\; y = 1,2,3,4,5,6$$

voidaan esittää seuraavana taulukkona:

2. heiton silmäluku y						
6	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
1	1/36	1/36	1/36	1/36	1/36	1/36
	1	**2**	**3**	**4**	**5**	**6**

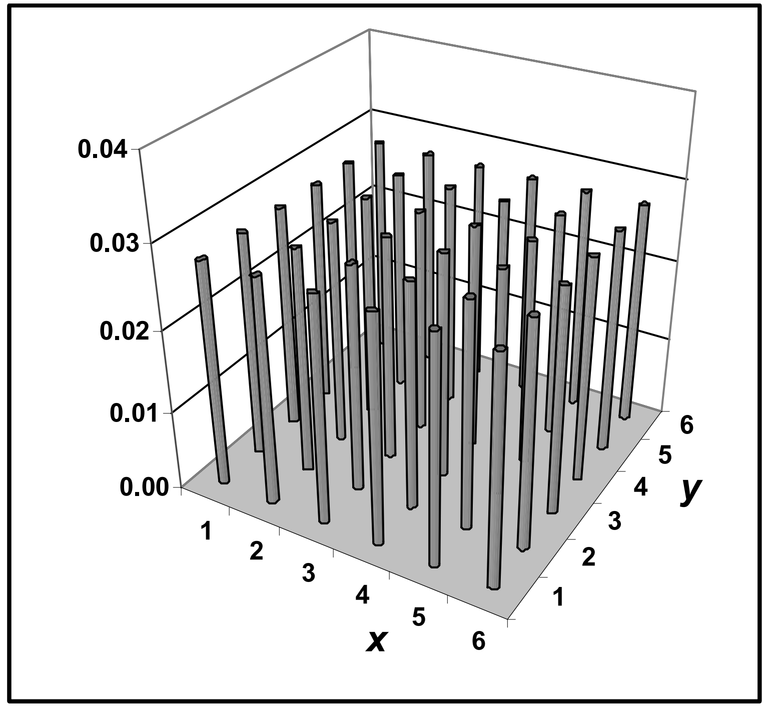
1. heiton silmäluku *x*

Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 5/6

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos 1. heitosta
  Y = tulos 2. heitosta

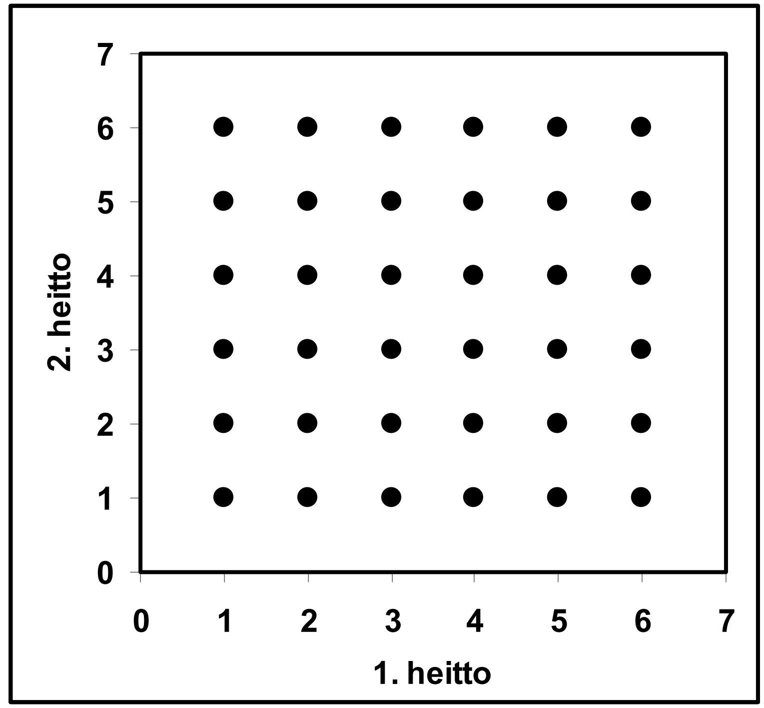
- Kuva oikealla esittää satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktiota*.
- Kuvassa on 36 pylvästä, joista jokaisen korkeus 1/36.



Diskreetit kaksiulotteiset jakaumat: 1. esimerkki nopanheitosta 6/6

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos 1. heitosta
  Y = tulos 2. heitosta
- Kuva oikealla havainnollistaa satunnaismuuttujien X ja Y *yhteisjakaumaa*.
- Kuvassa on 36 pistettä, joista jokaisen todennäköisyys on 1/36.



Diskreetit kaksiulotteiset jakaumat: 2. 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: 2. 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: 2. 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: 2. esimerkki nopanheitosta 4/9

- 1. nopanheiton tulos ja 1. ja 2. nopanheiton tulosten kaikki mahdolliset summat voidaan esittää seuraavana taulukkona:

Tulos 1. nopanheitosta	**Mahdolliset summat 1. ja 2. nopanheiton tuloksista**					
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

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

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

Diskreetit kaksiulotteiset jakaumat: 2. esimerkki nopanheitosta 5/9

- Koska noppa oletettiin *virheettömäksi* ja heittojen tulokset oletettiin *riippumattomiksi*, on luontevaa ajatella, että 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 pistetodennäköisyysfunktion* saa positiiviset arvot

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

  kun

  $$\begin{aligned} &x = 1, 2, 3, 4, 5, 6 \\ &z = x + y \\ &y = 1, 2, 3, 4, 5, 6 \end{aligned}$$

Diskreetit kaksiulotteiset jakaumat: 2. esimerkki nopanheitosta 6/9

- Satunnaismuuttujien X ja $Z = X + Y$ *yhteisjakauman pistetodennäköisyysfunktio* 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: 2. esimerkki nopanheitosta 7/9

- Esimerkkejä:

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

$$\Pr(X = 1 \text{ ja } Z = 10) = 0$$

(ii) Oletetaan, että 1. nopalla on saatu 6. Tällöin silmälukujen summaksi *ei voi tulla* 3, joten

$$\Pr(X = 6 \text{ ja } Z = 3) = 0$$

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

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

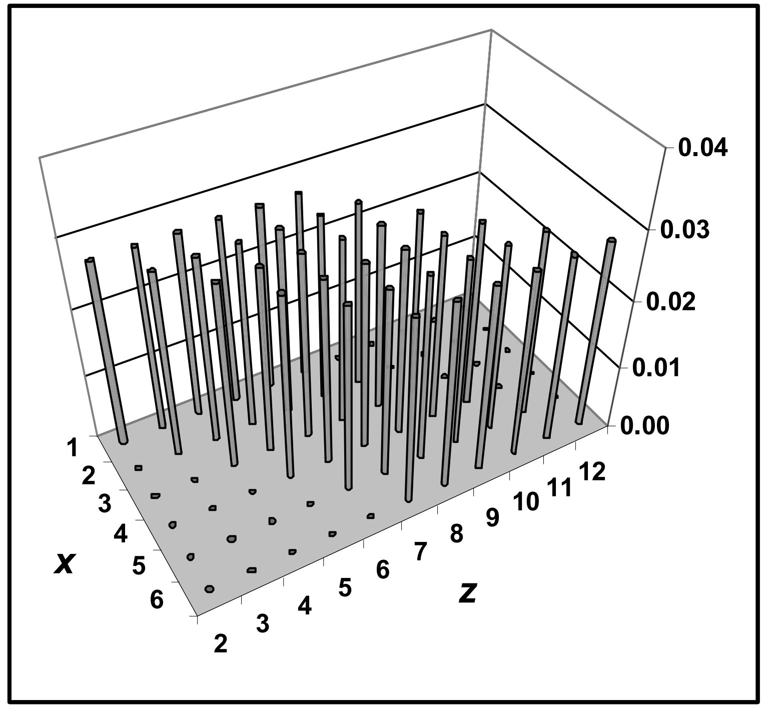
(iv) Oletetaan, että 1. nopalla on saatu 6. Tällöin silmälukujen summaksi *voi tulla* 7, 8, 9, 10, 11 tai 12, joten

$$\Pr(X = 6 \text{ ja } Z = z) = 1/36 \text{ ; } z = 7, 8, 9, 10, 11, 12$$

Diskreetit kaksiulotteiset jakaumat: 2. 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.

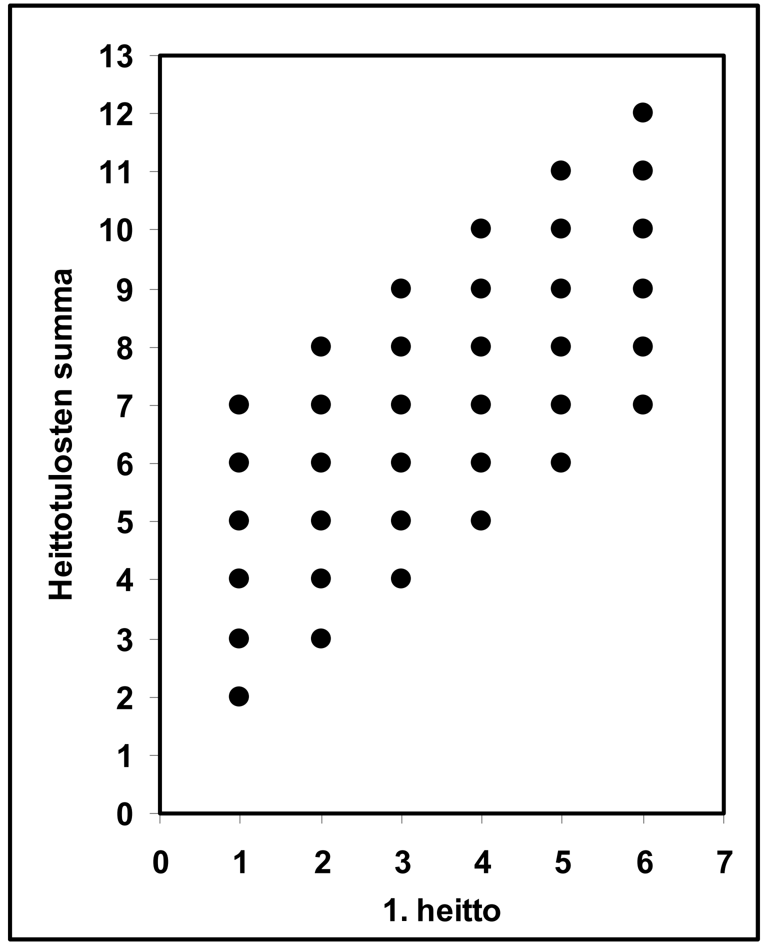


Diskreetit kaksiulotteiset jakaumat: 2. esimerkki nopanheitosta 9/9

- 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

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

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

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

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

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

Kaksinkertaiset integraalit: Integrointijärjestys

- Huomautus:

  Käytämme *kaksinkertaisten integraalien* yhteydessä seuraavaa sopimusta integrointijärjestyksestä:

$$\int_a^b \int_c^d f_{XY}(x,y)dydx = \int_a^b \left(\int_c^d f_{XY}(x,y)dy \right) dx$$

Jatkuvat kaksiulotteiset jakaumat: Tapahtumien todennäköisyydet

- Olkoon

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

jatkuvien satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio*.

- Olkoon

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

jokin *tapahtuma*.

- Tällöin

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

Kaksiulotteisten jakaumien kertymäfunktiot

- Olkoon (X, Y) satunnaismuuttujien X ja Y muodostama *järjestetty pari*.
- Satunnaismuuttujien X ja Y **yhteisjakauman kertymäfunktio** F_{XY} määritellään kaavalla:

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

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

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

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

Jatkuvien kaksiulotteisten jakaumien kertymäfunktiot

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

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

$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\,dv\,du$$

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 ja varianssit

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

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

- Satunnaismuuttujan Y **reunajakauman pistetodennäköisyysfunktio** on

$$f_Y(y) = \Pr(Y = y) = \sum_x f_{XY}(x, y)$$

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

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

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

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta

- Satunnaismuuttujien X ja Y mahdolliset arvot:

  X: $\{1, 2, 3, 4, 5, 6\}$
  Y: $\{1, 2, 3, 4, 5, 6\}$

- Määrätään satunnaismuuttujien X ja Y *reunajakaumat*.

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

- Satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio*:

$$\Pr(X = x \text{ ja } Y = y) = f_{XY}(x, y) \; ; x = 1, 2, \ldots , 6 \; ; y = 1, 2, \ldots , 6$$

1. heiton silmäluku *x*

2. heiton silmäluku *y*	1	2	3	4	5	6
6	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
1	1/36	1/36	1/36	1/36	1/36	1/36

Diskreetin 2-ulotteisen jakauman reunajakaumat: 1. esimerkki nopanheitosta 3/5

- Satunnaismuuttujien X ja *Y reunajakaumien pistetodennäköisyys-funktiot* saadaan määräämällä yhteisjakauman todennäköisyydet antavassa taulukossa *rivi*- ja *sarakesummat*.
- Satunnaismuuttujan *X reunajakauman pistetodennäköisyysfunktio* on

$$f_X(x) = \Pr(X = x) = \sum_{y=1}^{6} f_{XY}(x, y) = 6 \times \frac{1}{36} = \frac{1}{6} \; ; x = 1, 2, 3, 4, 5, 6$$

koska

$$f_{XY}(x, y) = \frac{1}{36} \; ; x = 1, 2, 3, 4, 5, 6 \; ; y = 1, 2, 3, 4, 5, 6$$

- Satunnaismuuttujan *Y reunajakauman pistetodennäköisyysfunktio* on

$$f_Y(y) = \Pr(Y = y) = \sum_{x=1}^{6} f_{XY}(x, y) = 6 \times \frac{1}{36} = \frac{1}{6} \; ; y = 1, 2, 3, 4, 5, 6$$

koska

$$f_{XY}(x, y) = \frac{1}{36} \; ; x = 1, 2, 3, 4, 5, 6 \; ; y = 1, 2, 3, 4, 5, 6$$

Diskreetin 2-ulotteisen jakauman reunajakaumat: 1. esimerkki nopanheitosta 4/5

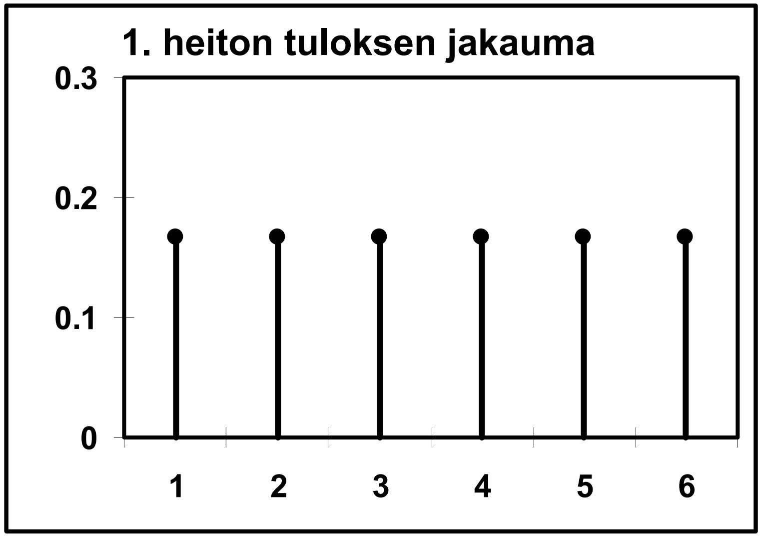
- Satunnaismuuttujien X ja Y *yhteisjakauman* ja *reunajakaumien pistetodennäköisyysfunktiot*:

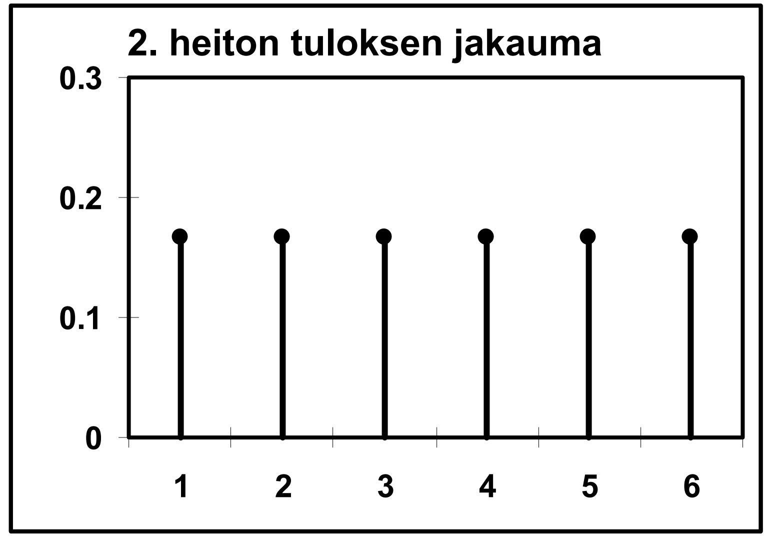
1. heiton silmäluku *x*

2. heiton silmäluku *y*

	1	2	3	4	5	6	Yht
6	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
5	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
4	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
3	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
2	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
1	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
Yht	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1**

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





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

  X = tulos 1. heitosta
  Y = tulos 2. heitosta

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

- Satunnaismuuttujien X ja Z *yhteisjakauman pistetodennäköisyysfunktio*:

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

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: 2. 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: 2. esimerkki nopanheitosta 4/5

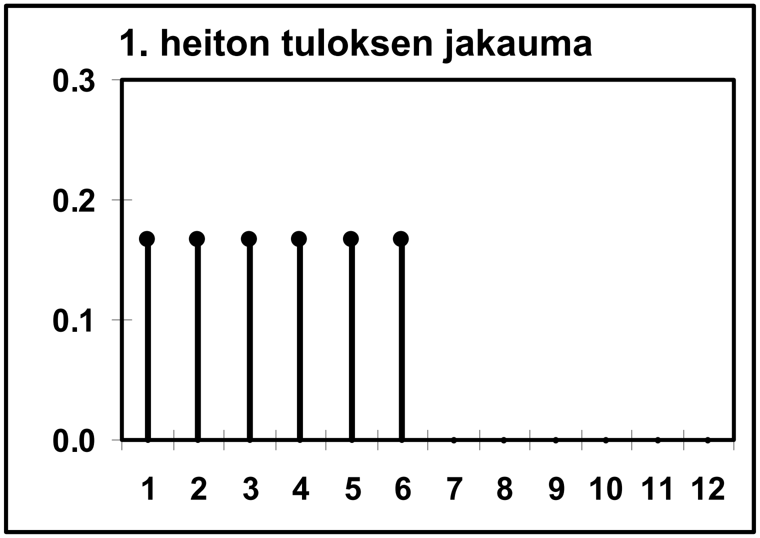
- Satunnaismuuttujien X ja $Z = X + Y$ *yhteisjakauma* ja *reunajakaumat*:

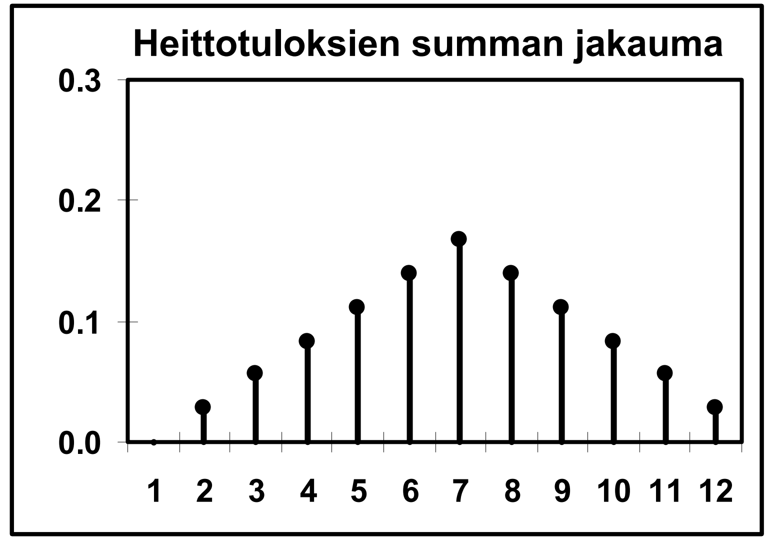
1. nopan silmäluku *x*

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**
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**

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





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

$$X = \text{tulos 1. heitosta}$$
$$Y = \text{tulos 2. heitosta}$$
$$Z = X + Y = \text{heittotulosten summa}$$

Jatkuvan kaksiulotteisen jakauman reunajakaumat

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

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

- Satunnaismuuttujan Y **reunajakauman tiheysfunktio** on

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

- 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 1:

  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 2:

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 1:

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 , $i = 1, 2, \ldots , p$ *yhteisjakauman* kertymäfunktio

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

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

- Määritelmä 2:

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 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 ja tapahtumien riippumattomuus: Perustelu

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

$$
\begin{aligned}
\Pr(a \le X \le b \text{ ja } c \le Y \le d) &= \int_a^b \int_c^d f_{XY}(x, y)dydx \\
&= \int_a^b \int_c^d f_X(x) f_Y(y)dydx \\
&= \int_a^b f_X(x) \int_c^d f_Y(y)dydx \\
&= \int_a^b f_X(x) \Pr(c \le Y \le d)dx \\
&= \Pr(c \le Y \le d) \int_a^b f_X(x)dx \\
&= \Pr(c \le Y \le d) \Pr(a \le X \le b)
\end{aligned}
$$

Satunnaismuuttujien riippumattomuus: 1. esimerkki nopanheitosta

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos (silmäluku) 1. heitosta

  Y = tulos (silmäluku) 2. heitosta

- Tarkastellaan satunnaismuuttujien X ja Y *yhteisjakaumaa*:

$$\begin{aligned}
\Pr(X = x \text{ ja } Y = y) &= \frac{1}{36} \\
&= \frac{1}{6} \cdot \frac{1}{6} \\
&= \Pr(X = x)\Pr(Y = y) \\
&\quad x = 1, 2, 3, 4, 5, 6 \,;\; y = 1, 2, 3, 4, 5, 6
\end{aligned}$$

- Siten satunnaismuuttujat X ja Y *ovat riippumattomia*.

Satunnaismuuttujien riippumattomuus: 2. 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*

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

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*

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

Reunajakaumien odotusarvot: Diskreetit jakaumat 1/2

- Olkoon *diskreettien* satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$.
- Satunnaismuuttujan X **odotusarvo** $\mathrm{E}(X) = \mu_X$ yhtyy satunnaismuuttujan X *reunajakauman odotusarvoon*:

$$\begin{aligned}\mathrm{E}(X) &= \sum_x \sum_y x f_{XY}(x,y) = \sum_x x \sum_y f_{XY}(x,y) \\ &= \sum_x x f_X(x)\end{aligned}$$

Reunajakaumien odotusarvot: Diskreetit jakaumat 2/2

- Olkoon *diskreettien* satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$.
- Satunnaismuuttujan Y **odotusarvo** $\mathrm{E}(Y) = \mu_Y$ yhtyy satunnaismuuttujan Y *reunajakauman odotusarvoon*:

$$\begin{aligned}\mathrm{E}(Y) &= \sum_x \sum_y y f_{XY}(x, y) = \sum_y y \sum_x f_{XY}(x, y) \\ &= \sum_y y f_Y(y)\end{aligned}$$

Reunajakaumien odotusarvot: Jatkuvat jakaumat 1/2

- Olkoon *jatkuvien* satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$.
- Satunnaismuuttujan X **odotusarvo** $\mathrm{E}(X) = \mu_X$ yhtyy satunnaismuuttujan X *reunajakauman odotusarvoon*:

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

$$= \int_{-\infty}^{+\infty} x f_X(x)\,dx$$

Reunajakaumien odotusarvot: Jatkuvat jakaumat 2/2

- Olkoon *jatkuvien* satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$.
- Satunnaismuuttujan Y **odotusarvo** $\mathrm{E}(Y) = \mu_Y$ yhtyy satunnaismuuttujan Y *reunajakauman odotusarvoon*:

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

$$= \int_{-\infty}^{+\infty} y f_Y(y)\,dy$$

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**.

Summan ja erotuksen odotusarvot

- Satunnaismuuttujien X ja Y **summan** $X + Y$ **odotusarvo**:

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

- Satunnaismuuttujien X ja Y **erotuksen** $X - Y$ **odotusarvo**:

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

- Huomautus:

  Satunnaismuuttujien summan ja erotuksen odotusarvojen kaavat on esitetty ilman perustelua luvussa **Jakaumien tunnusluvut**.

Summan ja erotuksen odotusarvot: Perustelu 1/2

- Esitetään satunnaismuuttujien summan odotusarvoa koskevan tuloksen perustelu kahden *jatkuvan* satunnaismuuttujan tapauksessa.
- Olkoot satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio*

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

ja X:n ja Y:n *reunajakaumien tiheysfunktiot* vastaavasti

$$f_X(x) \text{ ja } f_Y(y)$$

Summan ja erotuksen odotusarvot: Perustelu 2/2

- Tällöin
$$
\begin{aligned}
\mathrm{E}(X \pm Y) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x \pm y) f_{XY}(x,y)\,dydx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} [x f_{XY}(x,y) \pm y f_{XY}(x,y)]\,dydx \\
&= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x f_{XY}(x,y)\,dydx \pm \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y f_{XY}(x,y)\,dydx \\
&= \int_{-\infty}^{+\infty} x \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dydx \pm \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f_{XY}(x,y)\,dxdy \\
&= \int_{-\infty}^{+\infty} x f_X(x)\,dx \pm \int_{-\infty}^{+\infty} y f_Y(y)\,dy \\
&= \mathrm{E}(X) \pm \mathrm{E}(Y)
\end{aligned}
$$

Lineaarikombinaation odotusarvo

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

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

  ja olkoot a_i , i = 1, 2, … , k *vakioita*.
- Tällöin satunnaismuuttujien X_i , i = 1, 2, … , k **lineaarikombinaation**

  $$Y = a_1X_1 + a_2X_2 + \cdots + a_kX_k$$

  odotusarvo on

  $$\begin{aligned}\mathrm{E}(Y) &= \mathrm{E}(a_1X_1 + a_2X_2 + \cdots + a_kX_k) \\ &= a_1\,\mathrm{E}(X_1) + a_2\,\mathrm{E}(X_2) + \cdots + a_k\,\mathrm{E}(X_k) \\ &= a_1\mu_1 + a_2\mu_2 + \cdots + a_k\mu_k\end{aligned}$$

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**:

$$\begin{aligned}\mathrm{E}(XY) &= \mathrm{E}(X)\,\mathrm{E}(Y) \\ &= \mu_X \mu_Y\end{aligned}$$

- 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: Perustelu 1/2

- Esitetään riippumattomien satunnaismuuttujien tulon odotusarvoa koskevan tuloksen perustelu kahden *jatkuvan* satunnaismuuttujan tapauksessa.
- Olkoot satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio*

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

  ja X:n ja Y:n *reunajakaumien tiheysfunktiot* vastaavasti

  $$f_X(x) \text{ ja } f_Y(y)$$

- Olkoot satunnaismuuttujat X ja Y *riippumattomia*, jolloin

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

Riippumattomuus ja tulon odotusarvo: Perustelu 2/2

- Tällöin $$\begin{aligned}\mathrm{E}(XY) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_{XY}(x,y)\,dydx \\ &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf_X(x)f_Y(y)\,dydx \\ &= \int_{-\infty}^{+\infty} xf_X(x)\int_{-\infty}^{+\infty} yf_Y(y)\,dydx \\ &= \int_{-\infty}^{+\infty} xf_X(x)\,\mathrm{E}(Y)\,dx \\ &= \mathrm{E}(Y)\int_{-\infty}^{+\infty} xf_X(x)\,dx \\ &= \mathrm{E}(Y)\,\mathrm{E}(X)\end{aligned}$$

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.

Reunajakaumien varianssit

- Olkoot satunnaismuuttujien X ja Y *odotusarvot*

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

- Satunnaismuuttujien X ja Y **varianssit** yhtyvät vastaavien *reunajakaumien variansseihin*:

$$\begin{aligned}\mathrm{Var}(X) &= \mathrm{D}^2(X) = \sigma_X^2 \\ &= \mathrm{E}[(X-\mu_X)^2]\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(Y) &= \mathrm{D}^2(Y) = \sigma_Y^2 \\ &= \mathrm{E}[(Y-\mu_Y)^2]\end{aligned}$$

Varianssi vaihtelun mittana

- Satunnaismuuttujan X varianssi

  $$\mathrm{Var}(X)$$

  kuvaa satunnaismuuttujan X todennäköisyysjakauman *todennäköisyysmassan* **vaihtelua** *satunnaismuuttujan X oman odotusarvon* $\mathrm{E}(X)$ *ympärillä*.
- Satunnaismuuttujan Y varianssi

  $$\mathrm{Var}(Y)$$

  kuvaa satunnaismuuttujan Y todennäköisyysjakauman *todennäköisyysmassan* **vaihtelua** *satunnaismuuttujan Y oman odotusarvon* $\mathrm{E}(Y)$ *ympärillä*.

Reunajakaumien varianssit: Diskreetit jakaumat

- Olkoon *diskreettien* satunnaismuuttujien X ja Y *yhteisjakauman pistetodennäköisyysfunktio* $f_{XY}(x, y)$ ja vastaavien reunajakaumien pistetodennäköisyyfunktiot $f_X(x)$ ja $f_Y(y)$.
- Tällöin satunnaismuuttujien X ja Y **varianssit** ovat *vakioita*

$$\mathrm{D}^2(X) = \sum_x (x - \mu_X)^2 f_X(x)$$

$$\mathrm{D}^2(Y) = \sum_y (y - \mu_Y)^2 f_Y(y)$$

jossa

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

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

Reunajakaumien varianssit: Jatkuvat jakaumat

- Olkoon *jatkuvien* satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktio* $f_{XY}(x, y)$ ja vastaavien reunajakaumien tiheysfunktiot $f_X(x)$ ja $f_Y(y)$.
- Tällöin satunnaismuuttujien X ja Y **varianssit** ovat *vakioita*

$$\mathrm{D}^2(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)dx$$

$$\mathrm{D}^2(Y) = \int_{-\infty}^{+\infty} (y - \mu_Y)^2 f_Y(y)dy$$

jossa

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

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

Varianssien vaihtoehtoiset laskukaavat

- Satunnaismuuttujien X ja Y *varianssien* kaavat voidaan kirjoittaa seuraaviin yhtäpitäviin muotoihin:

$$\begin{aligned}
\mathrm{D}^2(X) &= \mathrm{E}[(X-\mu_X)^2] \\
&= \mathrm{E}(X^2)-\mu_X^2 \\
&= \mathrm{E}(X^2)-[\mathrm{E}(X)]^2 \\
\mathrm{D}^2(Y) &= \mathrm{E}[(Y-\mu_Y)^2] \\
&= \mathrm{E}(Y^2)-\mu_Y^2 \\
&= \mathrm{E}(Y^2)-[\mathrm{E}(Y)]^2
\end{aligned}$$

Varianssien vaihtoehtoiset laskukaavat: Perustelu

- Varianssin vaihtoehtoisen laskukaavan perustelu satunnaismuuttujalle X:

$$
\begin{aligned}
\mathrm{D}^2(X) &= \mathrm{E}[(X-\mu_X)^2] \\
&= \mathrm{E}[X^2 - 2\mu_X X + \mu_X^2] \\
&= \mathrm{E}(X^2) - \mathrm{E}(2\mu_X X) + \mathrm{E}(\mu_X^2) \\
&= \mathrm{E}(X^2) - 2\mu_X \mathrm{E}(X) + \mu_X^2 \\
&= \mathrm{E}(X^2) - 2\mu_X \mu_X + \mu_X^2 \\
&= \mathrm{E}(X^2) - \mu_X^2
\end{aligned}
$$

Reunajakaumien standardipoikkeamat

- Olkoot satunnaismuuttujien X ja Y *odotusarvot*

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

- Satunnaismuuttujien X ja Y **standardipoikkeamat** yhtyvät vastaavien *reunajakaumien standardipoikkeamiin*:

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

Standardipoikkeama vaihtelun mittana

- Satunnaismuuttujan X standardipoikkeama

  $$\mathrm{D}(X)$$

  kuvaa satunnaismuuttujan X todennäköisyysjakauman *todennäköisyysmassan* **vaihtelua** *satunnaismuuttujan* X *oman odotusarvon* $\mathrm{E}(X)$ *ympärillä*.
- Satunnaismuuttujan Y standardipoikkeama

  $$\mathrm{D}(Y)$$

  kuvaa satunnaismuuttujan Y todennäköisyysjakauman *todennäköisyysmassan* **vaihtelua** *satunnaismuuttujan* Y *oman odotusarvon* $\mathrm{E}(Y)$ *ympärillä*.

Odotusarvo, varianssi ja standardipoikkeama: Suureiden dimensio eli laatu

- Satunnaismuuttujalla sekä sen *odotusarvolla* ja *standardipoikkeamalla on aina sama dimensio* eli *laatu*.

  Esimerkki:

  Jos satunnaismuuttujan laatuna on m (metri), niin myös niiden odotusarvon ja standardipoikkeaman laatuna on m.

- Satunnaismuuttujan ja sen *varianssilla ei ole sama dimensio* eli *laatu*.

  Esimerkki:

  Jos satunnaismuuttujan laatuna on m (metri), niin sen varianssin laatuna on m^2.

Odotusarvot ja standardipoikkeamat: 1. esimerkki nopanheitosta

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta

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

$$\mathrm{E}(X) = \sum_{x=1}^{6} x \Pr(X = x) = \frac{1}{6}\sum_{x=1}^{6} x = \frac{21}{6} = 3.5 = \mathrm{E}(Y)$$

$$\mathrm{E}(X^2) = \sum_{x=1}^{6} x^2 \Pr(X = x) = \frac{1}{6}\sum_{x=1}^{6} x^2 = \frac{91}{6} = \mathrm{E}(Y^2)$$

$$\mathrm{D}^2(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12} = 2.917 = \mathrm{D}^2(Y)$$

$$\mathrm{D}(X) = \sqrt{2.917} = 1.708 = \mathrm{D}(Y)$$

Odotusarvot ja standardipoikkeamat: 2. 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* ja 2. *momentit*:

$$\mathrm{E}(X) = \sum_{x=1}^{6} x \Pr(X = x) = \frac{21}{6} = 3.5 = \mathrm{E}(Y)$$

$$\mathrm{E}(Z) = \sum_{z=1}^{12} z \Pr(Z = z) = \frac{252}{36} = 7$$

$$\mathrm{E}(X^2) = \sum_{x=1}^{6} x^2 \Pr(X = x) = \frac{91}{6} = \mathrm{E}(X^2)$$

$$\mathrm{E}(Z^2) = \sum_{z=1}^{12} z^2 \Pr(Z = z) = \frac{1974}{36}$$

Odotusarvot ja standardipoikkeamat: 2. esimerkki nopanheitosta 2/2

- *Varianssit* ja *standardipoikkeamat* :

$$\mathrm{D}^2(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2 = \frac{35}{12} = 2.917 = \mathrm{D}^2(Y)$$

$$\mathrm{D}(X) = \sqrt{2.917} = 1.708 = \mathrm{D}(Y)$$

$$\mathrm{D}^2(Z) = \mathrm{E}(Z^2) - [\mathrm{E}(Z)]^2 = \frac{210}{36} = 5.833$$

$$\mathrm{D}(Z) = \sqrt{5.833} = 2.415$$

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

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*

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

- Satunnaismuuttujien X ja Y kovarianssi

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

kuvaa satunnaismuuttujien X ja Y *yhteisjakauman todennäköisyysmassan* **yhteisvaihtelua** *satunnaismuuttujien X ja Y odotusarvojen* $\mathrm{E}(X)$ *ja* $\mathrm{E}(Y)$ *määräämän pisteen* $(\mathrm{E}(X), \mathrm{E}(Y))$ *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:

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

- Huomautus:

  Vrt. kovarianssin vaihtoehtoista laskukaavaa varianssin vaihtoehtoisiin laskukaavoihin.

Kovarianssin vaihtoehtoinen laskukaava: Perustelu

- Satunnaismuuttujien kovarianssin vaihtoehtoisen laskukaavan perustelu:

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

Kovarianssi ja varianssit

- Satunnaismuuttujien kovarianssit itsensä kanssa yhtyvät satunnaismuuttujien varianseihin:

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

Odotusarvot, varianssit ja kovarianssi sekä lineaarimuunnokset 1/2

- Olkoot satunnaismuuttujien X ja Y odotusarvot, varianssit ja kovarianssi:

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

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

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

- Olkoot

$$W = a + bX$$

$$Z = c + dY$$

jossa a, b, c, $d \in \mathbb{R}$ ovat reaalisia vakioita.

Odotusarvot, varianssit ja kovarianssi sekä lineaarimuunnokset 2/2

- Tällöin pätevät seuraavat kaavat:

$$\mathrm{E}(W) = a + b\,\mathrm{E}(X) = a + b\mu_X$$

$$\mathrm{Var}(W) = b^2\,\mathrm{Var}(X) = b^2\sigma_X^2$$

$$\mathrm{E}(Z) = c + d\,\mathrm{E}(Y) = c + d\mu_Y$$

$$\mathrm{Var}(Z) = d^2\,\mathrm{Var}(Y) = d^2\sigma_Y^2$$

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

Kovarianssi ja lineaarimuunnokset: Perustelu

- Jos satunnaismuuttujien X ja Y kovarianssi on Cov(X, Y), niin

$$\begin{aligned}
\mathrm{Cov}(W,Z) &= \mathrm{E}[(W-\mathrm{E}(W))(Z-\mathrm{E}(Z))] \\
&= \mathrm{E}[(a+bX-\mathrm{E}(a+bX))(c+dY-\mathrm{E}(c+dY))] \\
&= \mathrm{E}[(bX-\mathrm{E}(bX))(dY-\mathrm{E}(dY))] \\
&= bd\,\mathrm{E}[(X-\mathrm{E}(X))(Y-\mathrm{E}(Y))] \\
&= bd\,\mathrm{Cov}(X,Y)
\end{aligned}$$

Summan ja erotuksen varianssit

- Olkoot satunnaismuuttujien X ja Y *varianssit*

$$\mathrm{Var}(X) = \sigma_X^2 \qquad \mathrm{Var}(Y) = \sigma_Y^2$$

ja kovarianssi

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

- Tällöin

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

Summan ja erotuksen varianssit: Perustelu

- Summan ja erotuksen varianssia koskevat kaavat nähdään oikeaksi seuraavalla tavalla:

$$
\begin{aligned}
\mathrm{Var}(X \pm Y) &= \mathrm{E}[X \pm Y - \mathrm{E}(X \pm Y)]^2 \\
&= \mathrm{E}[X \pm Y - \mathrm{E}(X) \mp \mathrm{E}(Y)]^2 \\
&= \mathrm{E}[(X - \mathrm{E}(X)) \pm (Y - \mathrm{E}(Y))]^2 \\
&= \mathrm{E}[(X - \mathrm{E}(X))^2 + (Y - \mathrm{E}(Y))^2 \\
&\qquad \pm 2(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))] \\
&= \mathrm{E}[(X - \mathrm{E}(X))^2] + \mathrm{E}[(Y - \mathrm{E}(Y))^2] \\
&\qquad \pm 2\,\mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))] \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y) \pm 2\,\mathrm{Cov}(X, Y)
\end{aligned}
$$

Riippumattomuus ja kovarianssi

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

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

- Huomautus:

  Käänteinen *ei päde*: Siitä, että

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

  ei välttämättä seuraa, että satunnaismuuttujat X ja Y olisivat riippumattomia.

Riippumattomuus ja kovarianssi: Perustelu

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

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

- Siten

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

Riippumattomuus sekä summan ja erotuksen varianssit

- Olkoot satunnaismuuttujien X ja Y *varianssit*

  $$\mathrm{Var}(X) = \sigma_X^2 \qquad \mathrm{Var}(Y) = \sigma_Y^2$$

- Olkoot X ja Y ovat *riippumattomia*, jolloin

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

- Tällöin

  $$\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

- Huomautus:

  Riippumattomien satunnaismuuttujien summan ja erotuksen varianssien kaavat on esitetty ilman perustelua luvussa **Jakaumien tunnusluvut**.

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 **korrelaatiokerroin** on *vakio*

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

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.

Kovarianssi ja korrelaatio: Suureiden dimensio eli laatu

- Huomaa, että satunnaismuuttujien korrelaatio *on dimensioton* eli *laaduton suure* toisin kuin niiden kovarianssi.

Korrelaatiokerroin ja lineaarimuunnokset

- Olkoon satunnaismuuttujien X ja Y *korrelaatiokerroin* Cor(X, Y).
- Olkoot

$$W = a + bX$$
$$Z = c + dY$$

- Tällöin

$$\mathrm{Cor}(W, Z) = \mathrm{sgn}(bd)\,\mathrm{Cor}(X, Y)$$

jossa sgn on ns. *merkkifunktio*:

$$\mathrm{sgn}(x) = \begin{cases} +1, \text{ jos } x > 0 \\ \ \ 0, \text{ jos } x = 0 \\ -1, \text{ jos } x < 0 \end{cases}$$

Lineaarimuunnosten korrelaatiokerroin: Perustelu

- Olkoon satunnaismuuttujien X ja Y *korrelaatiokerroin* Cor(X, Y).
- Olkoot

$$W = a + bX$$
$$Z = c + dY$$

- Tällöin

$$\begin{aligned}
\mathrm{Cor}(W,Z) &= \frac{\mathrm{Cov}(W,Z)}{\sqrt{\mathrm{Var}(W)\,\mathrm{Var}(Z)}} \\
&= \frac{bd\,\mathrm{Cov}(X,Y)}{\sqrt{b^2\,\mathrm{Var}(X)\;d^2\,\mathrm{Var}(Y)}} \\
&= \mathrm{sgn}(bd)\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}} \\
&= \mathrm{sgn}(bd)\,\mathrm{Cor}(X,Y)
\end{aligned}$$

Korrelaatiokerroin: Ominaisuudet

- Olkoon satunnaismuuttujien X ja Y *korrelaatiokerroin* Cor(X, Y).
- Tällöin

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

(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: Ominaisuuden (i) perustelu 1/2

- **Väite**:

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

- **Perustelu**:

(1) Olkoot W ja Z satunnaismuuttujia ja olkoon $a \in \mathbb{R}$ *vakio*.

(2) $W^2 \geq 0,\ Z^2 \geq 0 \Rightarrow \mathrm{E}(W^2) \geq 0,\ \mathrm{E}(Z^2) \geq 0$

(3) $(aW - Z)^2 \geq 0 \Rightarrow \mathrm{E}(aW - Z)^2 \geq 0$

(4) Kohdasta (3) seuraa, että

$$\mathrm{E}(aW - Z)^2 = a^2\mathrm{E}(W^2) - 2a\mathrm{E}(WZ) + \mathrm{E}(Z^2) \geq 0$$

(5) Valitaan:

$$a = \frac{\mathrm{E}(WZ)}{\mathrm{E}(W^2)}$$

Korrelaatiokerroin: Ominaisuuden (i) perustelu 2/2

(6) Siten

$$\mathrm{E}(aW - Z)^2 = -\frac{[\mathrm{E}(WZ)]^2}{\mathrm{E}(W^2)} + \mathrm{E}(Z^2) \geq 0$$

josta seuraa, että

$$\frac{[\mathrm{E}(WZ)]^2}{\mathrm{E}(W^2)\,\mathrm{E}(Z^2)} \leq 1 \qquad (*)$$

(7) Väite seuraa epäyhtälöstä ($*$), kun valitaan

$W = X - \mathrm{E}(X)$ ja $Z = Y - \mathrm{E}(Y)$

ja otetaan saadusta epäyhtälöstä neliöjuuri.

- **Huomautus**:

  Epäyhtälöstä ($*$) seuraa eräs ns. *Schwarzin epäyhtälön* monista muodoista:

  $$[\mathrm{E}(WZ)]^2 \leq \mathrm{E}(W^2)\,\mathrm{E}(Z^2)$$

Korrelaatiokerroin: Ominaisuuden (ii) perustelu

- **Väite**:

  Jos X ja Y ovat *riippumattomia*, niin

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

- **Perustelu**:

  Väite seuraa siitä, että jos X ja Y ovat riippumattomia, niin

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

- **Huomautus**:

  Käänteinen *ei päde*: Siitä, että

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

  ei välttämättä seuraa, että satunnaismuuttujat X ja Y olisivat riippumattomia.

Korrelaatiokerroin: Ominaisuuden (iii) perustelu 1/3

- **Väite**:

$$\mathrm{Cor}(X,Y) = \pm 1 \quad \Rightarrow \quad Y = \alpha + \beta X \ , \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \beta \neq 0$$

- **Perustelu**:

(1) Ominaisuuden (i) perustelussa kohdan (4) epäyhtälön vasen puoli on 2. asteen polynomi muuttujan a suhteen:

$$h(a) = \mathrm{E}(aW - Z)^2 = a^2\mathrm{E}(W^2) - 2a\mathrm{E}(WZ) + \mathrm{E}(Z^2)$$

(2) Ominaisuuden (i) perustelun kohdan (3) mukaan $h(a) \geq 0 \ \forall a$.

(3) $h(a) > 0$ kaikille a, jos ja vain jos 2. asteen yhtälön $h(a) = 0$ *diskriminantti*

$$D = 4[\mathrm{E}(WZ)]^2 - 4\,\mathrm{E}(W^2)\,\mathrm{E}(Z^2) < 0$$

- **Huomautus**:

Korrelaatiokertoimen ominaisuus (i) seuraa myös kohdan (3) epäyhtälöstä.

Korrelaatiokerroin: Ominaisuuden (iii) perustelu 2/3

(4) $h(a) = 0$ täsmälleen silloin, kun 2. asteen yhtälön $h(a) = 0$ *diskriminantti*

$$D = 4[\mathrm{E}(WZ)]^2 - 4\,\mathrm{E}(W^2)\,\mathrm{E}(Z^2) = 0$$

(5) Siten $\mathrm{Cor}(X, Y) = \pm 1$ täsmälleen silloin, kun

$$\mathrm{E}(aW - Z)^2 = 0$$

mikä on yhtäpitävää sen kanssa, että

$$\Pr(aW - Z = 0) = 1$$

(6) Siten

$$Z = aW$$

todennäköisyydellä 1.

(7) Väite seuraa, kun valitaan $W = X - \mathrm{E}(X)$ ja $Z = Y - \mathrm{E}(Y)$ ja merkitään

$$\beta = a \text{ ja } \alpha = \mathrm{E}(Y) - a\,\mathrm{E}(X)$$

Korrelaatiokerroin: Ominaisuuden (iii) perustelu 3/3

- **Väite**:

$$\mathrm{Cor}(X,Y) = \pm 1 \quad \Leftarrow \quad Y = \alpha + \beta X \text{ , } \alpha \in \mathbb{R}, \beta \in \mathbb{R}, \beta \neq 0$$

- **Perustelu**:

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

$$\begin{aligned} \mathrm{Cor}(X,Y) &= \mathrm{Cor}(X, \alpha + \beta X) \\ &= \mathrm{sgn}(\beta)\mathrm{Cor}(X,X) \\ &= \mathrm{sgn}(\beta) \\ &= \pm 1 \end{aligned}$$

jossa sgn(·) on ns. *merkkifunktio*:

$$\mathrm{sgn}(x) = \begin{cases} +1, \text{ jos } x > 0 \\ 0, \text{ jos } x = 0 \\ -1, \text{ jos } x < 0 \end{cases}$$

Korrelaatiokerroin: 1. esimerkki nopanheitosta

- Heitetään *virheetöntä* noppaa kaksi kertaa:

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta

- Tarkastellaan satunnaismuuttujien X ja Y *yhteisjakaumaa*.
- *Odotusarvot*, *varianssit* ja *standardipoikkeamat* (laskettu aikaisemmin):

  $$\mathrm{E}(X) = \mathrm{E}(Y) = 21/6 = 3.5$$
  $$\mathrm{D}^2(X) = \mathrm{D}^2(Y) = 35/12 = 2.917$$
  $$\mathrm{D}(X) = \mathrm{D}(Y) = 1.708$$

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

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

Korrelaatiokerroin: 2. 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* (laskettu aikaisemmin):

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

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

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

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.

Moniulotteiset satunnaismuuttujat ja todennäköisyysjakaumat

Kaksiulotteiset todennäköisyysjakaumat

Kaksiulotteisten jakaumien reunajakaumat ja riippumattomuus

Kaksiulotteisten jakaumien odotusarvot ja varianssit

Kovarianssi ja korrelaatio

>> Ehdolliset jakaumat ja odotusarvot

Ehdollinen todennäköisyys

- Olkoot A ja B tapahtumia ja $\Pr(B) \neq 0$.
- Tapahtuman *A ehdollinen todennäköisyys* tapahtuman B suhteen on

$$\Pr(A \mid B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

- *Ehdollisen jakauman määritelmä* mukailee ehdollisen todennäköisyyden määritelmää.

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

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

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

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

Ehdolliset jakaumat: Kommentteja

- Satunnaismuuttujan X ehdollinen jakauma

  $$f_{X|Y}(x|y)$$

  satunnaismuuttujan Y suhteen (ehdolla $Y = y$) *riippuu yleisessä tapauksessa ehtomuuttujan Y arvosta y.*
- Satunnaismuuttujan Y ehdollinen jakauma

  $$f_{Y|X}(y|x)$$

  satunnaismuuttujan X suhteen (ehdolla $X = x$) *riippuu yleisessä tapauksessa ehtomuuttujan X arvosta x.*

Riippumattomuus ja ehdolliset jakaumat 1/3

- 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 $f_Y(y)$.
- Olkoot satunnaismuuttujat X ja Y *riippumattomia*.
- Tällöin

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

Riippumattomuus ja ehdolliset jakaumat 2/3

- Jos satunnaismuuttujat X ja Y ovat riippumattomia, satunnaismuuttujan X *ehdollinen jakauma* satunnaismuuttujan Y suhteen yhtyy satunnaismuuttujan X *reunajakaumaan*:

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

- Jos satunnaismuuttujat X ja Y ovat riippumattomia, satunnaismuuttujan Y *ehdollinen jakauma* satunnaismuuttujan X suhteen yhtyy satunnaismuuttujan Y *reunajakaumaan*:

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

Riippumattomuus ja ehdolliset jakaumat 3/3

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

  (i) Satunnaismuuttujan X ehdollinen jakauma satunnaismuuttujan Y suhteen *ei riipu* ehtomuuttujan Y arvoista.

  (ii) Satunnaismuuttujan Y ehdollinen jakauma satunnaismuuttujan X suhteen *ei riipu* ehtomuuttujan X arvoista.

Ehdolliset odotusarvot: Diskreetit jakaumat

- Olkoot satunnaismuuttujat X ja Y *diskreettejä*.
- Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

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

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

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

Ehdolliset odotusarvot: Jatkuvat jakaumat

- Olkoot satunnaismuuttujat X ja Y *jatkuvia*.
- Satunnaismuuttujan X **ehdollinen odotusarvo** satunnaismuuttujan Y suhteen on satunnaismuuttujan X ehdollisen jakauman odotusarvo:

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

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

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

Ehdolliset odotusarvot: Kommentteja

- Satunnaismuuttujan X ehdollinen odotusarvo

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

  satunnaismuuttujan Y suhteen (ehdolla $Y = y$) *riippuu yleisessä tapauksessa ehtomuuttujan Y arvosta y.*
- Satunnaismuuttujan Y ehdollinen odotusarvo

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

  satunnaismuuttujan X suhteen (ehdolla $X = x$) *riippuu yleisessä tapauksessa ehtomuuttujan X arvosta x.*

Riippumattomuus ja ehdolliset odotusarvot 1/2

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

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

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

Riippumattomuus ja ehdolliset odotusarvot 2/2

- Olkoot satunnaismuuttujat X ja Y *riippumattomia*.
- Tällöin:
  - (i) Satunnaismuuttujan X ehdollinen odotusarvo satunnaismuuttujan Y suhteen *ei riipu* ehtomuuttujan Y arvoista.
  - (ii) Satunnaismuuttujan Y ehdollinen odotusarvo satunnaismuuttujan X suhteen *ei riipu* ehtomuuttujan X arvoista.

Ehdolliset odotusarvot satunnaismuuttujina: Iteroidun odotusarvon laki

- Ehdolliset odotusarvot voidaan tulkita *satunnais-muuttujiksi* ehtomuuttujan suhteen.
- Siten satunnaismuuttujan Y *ehdollisen odotusarvon odotusarvoksi* (satunnaismuuttujan X suhteen) saadaan

$$\mathrm{E}_X\left[\mathrm{E}(Y|X)\right] = \mathrm{E}(Y)$$

- Siten satunnaismuuttujan X *ehdollisen odotusarvon odotusarvoksi* (satunnaismuuttujan Y suhteen) saadaan

$$\mathrm{E}_Y\left[\mathrm{E}(X|Y)\right] = \mathrm{E}(X)$$

- Kaavat tunnetaan nimellä **iteroidun odotusarvon laki**.

Iteroidun odotusarvon laki: Perustelu

- Iteroidun odotusarvon laki nähdään oikeaksi *jatkuvan* jakauman tapauksessa seuraavalla tavalla:

$$
\begin{aligned}
\mathrm{E}_X\left[\mathrm{E}(Y|X)\right] &= \int_{-\infty}^{+\infty} \mathrm{E}(Y|X) f_X(x)dx = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} y f_{Y|X}(y|x)dy\right] f_X(x)dx \\
&= \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} y f_{Y|X}(y|x) f_X(x)dy\right]dx = \int_{-\infty}^{+\infty}\left[\int_{-\infty}^{+\infty} y f_{XY}(x,y)dy\right]dx \\
&= \int_{-\infty}^{+\infty} y \int_{-\infty}^{+\infty} f_{XY}(x,y)dxdy = \int_{-\infty}^{+\infty} y f_Y(y)dy \\
&= \mathrm{E}(Y)
\end{aligned}
$$

Regressiofunktiot ja -käyrät 1/2

- 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**

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

Regressiofunktiot ja -käyrät 2/2

- Tarkastellaan satunnaismuuttujan Y *ehdollista odotusarvoa*

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

ehtomuuttujan X arvojen x *funktiona*.

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

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

Regressiofunktiot ja -käyrät: Kommentteja

- Olkoon

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

satunnaismuuttujan X *regressiokäyrä* satunnaismuuttujan Y saamien arvojen y suhteen.

- Olkoon

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

satunnaismuuttujan Y *regressiokäyrä* satunnaismuuttujan X saamien arvojen x suhteen.

- Vaikka funktiot g_y ja g_x *ovat funktioina täysin määrättyjä*, *sattuma määrää*, *mikä funktioiden arvoista realisoituu*.

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 $d(Y|X)$ siten, että *ennusteen* **keskineliövirhe**

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

  minimoituu.
- Voidaan osoittaa, että keskineliövirhe minimoituu valinnalla

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

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 1/6

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

  X = tulos (silmäluku) 1. heitosta
  Y = tulos (silmäluku) 2. heitosta
- Satunnaismuuttujien X ja Y mahdolliset arvot:

  X: $\{1, 2, 3, 4, 5, 6\}$
  Y: $\{1, 2, 3, 4, 5, 6\}$
- Määrätään satunnaismuuttujien X ja Y *ehdolliset jakaumat*.

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 2/6

- Satunnaismuuttujien X ja Y *yhteisjakauma* ja *reunajakaumat*:

Todennäköisyydet Pr(*X* = *x*, *Y* = *y*)

2. heiton silmäluku *y* \ 1. heiton silmäluku *x*	1	2	3	4	5	6	Yht
6	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
5	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
4	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
3	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
2	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
1	1/36	1/36	1/36	1/36	1/36	1/36	**1/6**
Yht	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1/6**	**1**

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 3/6

- Kalvon 2/6 satunnaismuuttujan *X ehdolliset jakaumat* saadaan jakamalla yhteisjakauman todennäköisyydet antavassa taulukossa kunkin *rivin todennäköisyydet vastaavilla rivisummilla* eli *vastaavilla reunatodennäköisyyksillä*.
- Esimerkki:

  Satunnaismuuttujan *X ehdollinen jakauma satunnaismuuttujan Y suhteen*, kun $Y = 3$:

$$f_{X|Y}(x \mid Y=3) = \frac{f_{XY}(x,3)}{f_Y(3)} = \frac{1/36}{1/6} = \frac{1}{6} \ ; \ x = 1,2,3,4,5,6$$

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 4/6

• Satunnaismuuttujan X *ehdolliset jakaumat* ovat taulukon *riveinä*:

Todennäköisyydet Pr($X = x \mid Y = y$)

2. heiton silmäluku y \ 1. heiton silmäluku x	1	2	3	4	5	6	Yht
6	1/6	1/6	1/6	1/6	1/6	1/6	1
5	1/6	1/6	1/6	1/6	1/6	1/6	1
4	1/6	1/6	1/6	1/6	1/6	1/6	1
3	1/6	1/6	1/6	1/6	1/6	1/6	1
2	1/6	1/6	1/6	1/6	1/6	1/6	1
1	1/6	1/6	1/6	1/6	1/6	1/6	1
Yht	1	1	1	1	1	1	

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 5/6

- Kalvon 2/6 satunnaismuuttujan *Y ehdolliset jakaumat* saadaan jakamalla yhteisjakauman todennäköisyydet antavassa taulukossa kunkin *sarakkeen todennäköisyydet vastaavilla sarakesummilla* eli *vastaavilla reunatodennäköisyyksillä*.
- Esimerkki:

Satunnaismuuttujan *Y ehdollinen jakauma satunnaismuuttujan X suhteen*, kun $X = 4$:

$$f_{Y|X}(y|X=4) = \frac{f_{XY}(4,y)}{f_X(4)} = \frac{1/36}{1/6} = \frac{1}{6}\ ;\ y = 1,2,3,4,5,6$$

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 1. esimerkki nopanheitosta 6/6

- Satunnaismuuttujan *Y ehdolliset jakaumat* ovat taulukon *sarakkeina*:

Todennäköisyydet Pr(*Y* = *y* | *X* = *x*)

2. heiton silmäluku *y* / 1. heiton silmäluku *x*

	1	2	3	4	5	6	Yht
6	1/6	1/6	1/6	1/6	1/6	1/6	1
5	1/6	1/6	1/6	1/6	1/6	1/6	1
4	1/6	1/6	1/6	1/6	1/6	1/6	1
3	1/6	1/6	1/6	1/6	1/6	1/6	1
2	1/6	1/6	1/6	1/6	1/6	1/6	1
1	1/6	1/6	1/6	1/6	1/6	1/6	1
Yht	1	1	1	1	1	1	

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 2. esimerkki nopanheitosta 1/4

- 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

- 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 satunnaismuuttujan Z *ehdolliset jakaumat* satunnaismuuttujan X suhteen.

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 2. esimerkki nopanheitosta 2/4

- 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**
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**

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 2. esimerkki nopanheitosta 3/4

- Kalvon 2/4 satunnaismuuttujan Z *ehdolliset jakaumat* saadaan jakamalla yhteisjakauman todennäköisyydet antavassa taulukossa kunkin *sarakkeen todennäköisyydet vastaavilla sarakesummilla* eli *vastaavilla reunatodennäköisyyksillä*.
- Esimerkkejä:

(i) Tapahtuman $Z = 8$ *ehdollinen todennäköisyys*, kun ehtotapahtumana on $X = 3$:

$$\Pr(Z=8 \mid X=3) = f_{Z|X}(8 \mid X=3) = \frac{f_{ZX}(8,3)}{f_X(3)} = \frac{1/36}{1/6} = \frac{1}{6}$$

(ii) Tapahtuman $Z = 10$ *ehdollinen todennäköisyys*, kun ehtotapahtumana on $X = 3$:

$$\Pr(Z=10 \mid X=3) = f_{Z|X}(10 \mid X=3) = \frac{f_{ZX}(10,3)}{f_X(3)} = \frac{0}{1/6} = 0$$

Diskreetin 2-ulotteisen jakauman ehdolliset jakaumat: 2. esimerkki nopanheitosta 4/4

- Satunnaismuuttujan Z *ehdolliset jakaumat* ovat taulukon *sarakkeina*:

1. nopan silmäluku *x*

Silmälukujen summa *z*

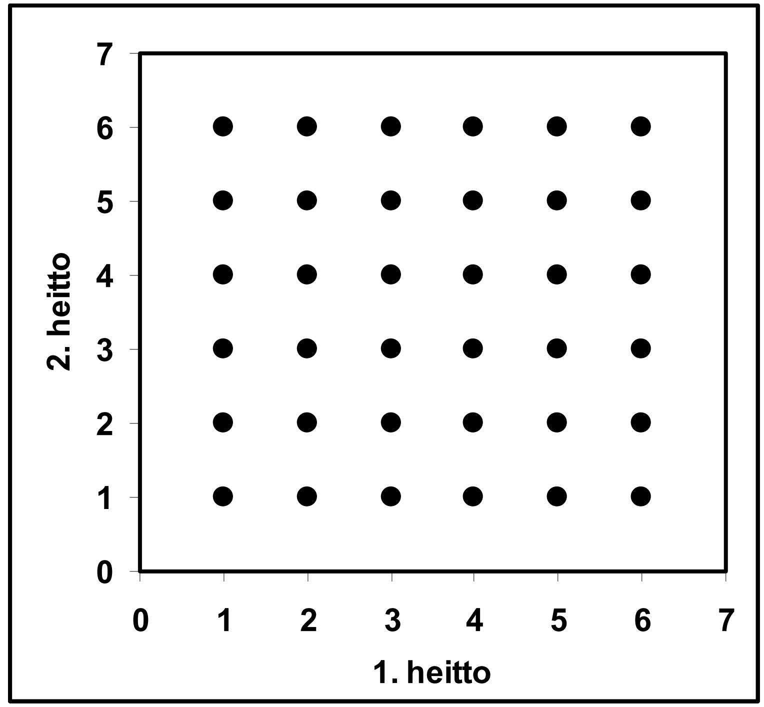
	1	**2**	**3**	**4**	**5**	**6**	**Yht**
12	0	0	0	0	0	1/6	**1/6**
11	0	0	0	0	1/6	1/6	**2/6**
10	0	0	0	1/6	1/6	1/6	**3/6**
9	0	0	1/6	1/6	1/6	1/6	**4/6**
8	0	1/6	1/6	1/6	1/6	1/6	**5/6**
7	1/6	1/6	1/6	1/6	1/6	1/6	**6/6**
6	1/6	1/6	1/6	1/6	1/6	0	**5/6**
5	1/6	1/6	1/6	1/6	0	0	**4/6**
4	1/6	1/6	1/6	0	0	0	**3/6**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

Diskreetin 2-ulotteisen jakauman ehdolliset odotusarvot: 1. esimerkki nopanheitosta 1/2

- Heitetään *virheetöntä* noppaa kaksi kertaa:

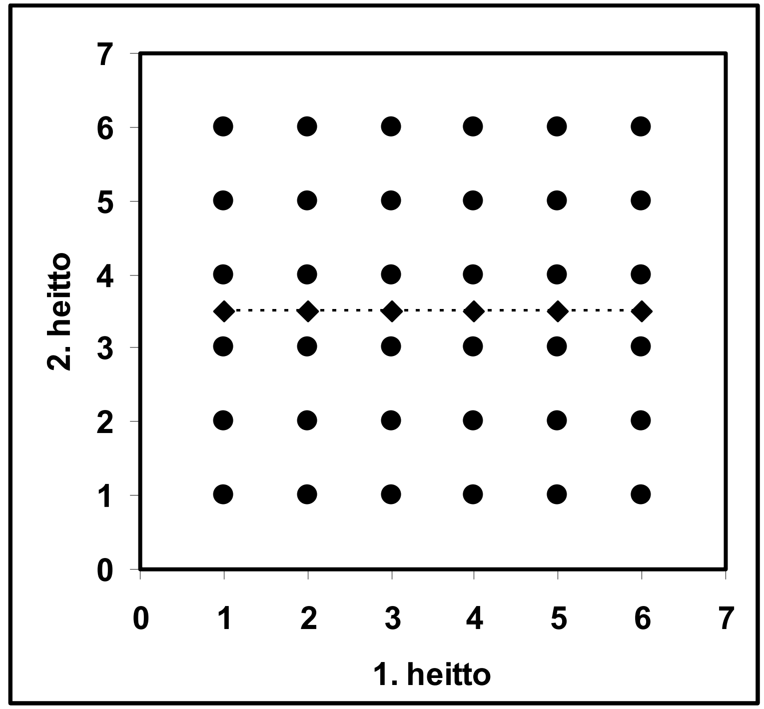
  X = tulos 1. heitosta
  Y = tulos 2. heitosta

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



Diskreetin 2-ulotteisen jakauman ehdolliset odotusarvot: 1. esimerkki nopanheitosta 2/2

- Satunnaismuuttujan
  X (= tulos 1. heitosta)
  ehdollisia odotusarvoja satunnaismuuttujan
  Y (= tulos 2. heitosta)
  arvojen suhteen on merkitty katkoviivan yhdistämillä *vinoneliöillä*.
- 1. nopanheiton tuloksen tuntemisesesta *ei ole hyötyä* 2. heiton tuloksen *ennustamisessa*, koska 2. heiton tulos *ei riipu* 1. heiton tuloksesta.

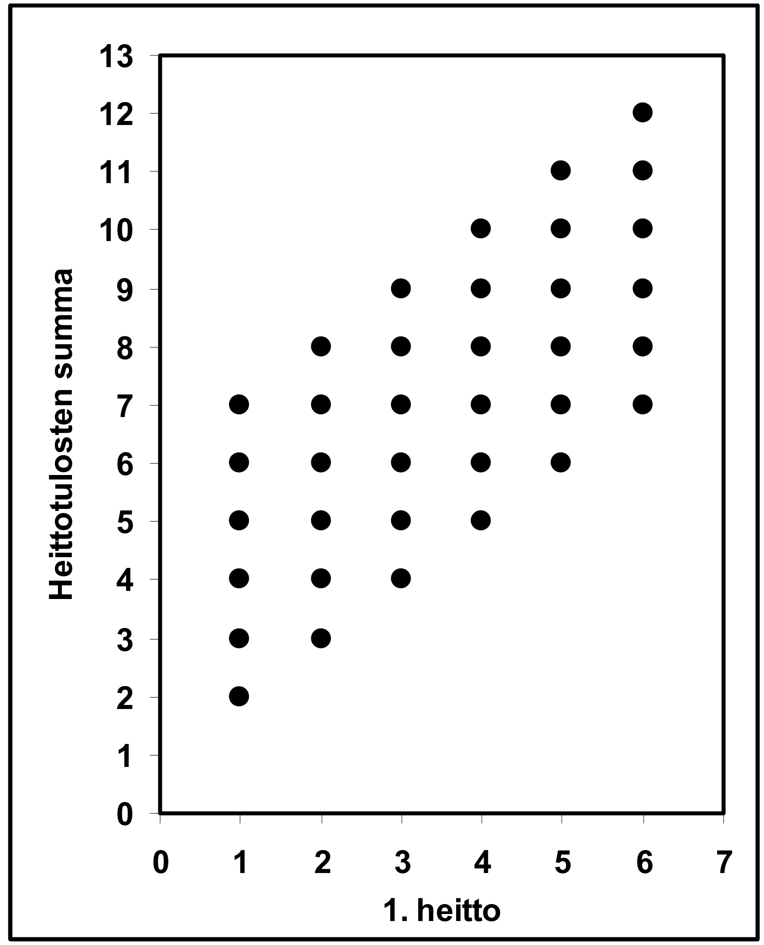


Diskreetin 2-ulotteisen jakauman ehdolliset odotusarvot: 2. esimerkki nopanheitosta 1/2

- 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.



Diskreetin 2-ulotteisen jakauman ehdolliset odotusarvot: 2. esimerkki nopanheitosta 2/2

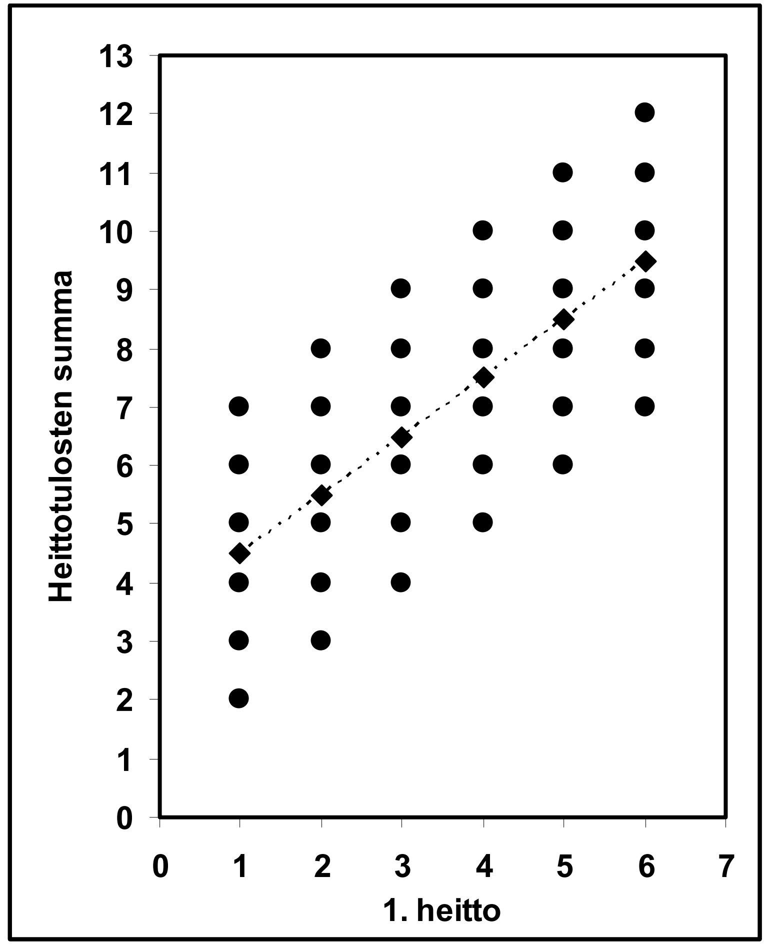
- Satunnaismuuttujan

  Z (= heittotulosten summa)

  ehdollisia odotusarvoja satunnaismuuttujan

  X (= tulos 1. heitosta)

  arvojen suhteen on merkitty katkoviivan yhdistämillä *vinoneliöillä*.
- 1. nopanheiton tuloksen tuntemisesta *on hyötyä* heitto-tulosten summan *ennustamisessa*, koska tulosten summa *riippuu* 1. heiton tuloksesta.



Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkki 1/2

- Valitaan luvut X ja Y *satunnaisesti* väliltä [0, 1] kahdessa vaiheessa:

  (1) Valitaan väliltä (0, 1) *satunnaisesti* luku X:

  $$X \sim \text{Uniform}(0, 1)$$

  Olkoon valittu luku x.

  (2) Valitaan väliltä $(0, x)$ *satunnaisesti* luku Y:

  $$Y \sim \text{Uniform}(0, x)$$

  Olkoon valittu luku y.

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

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkki 2/2

- Oletuksien mukaan satunnaismuuttujan X (*reuna*-) *jakauma* tunnetaan:

$$f_X(x) = 1 \qquad 0 < x < 1$$
$$f_X(x) = 0 \qquad \text{muulloin}$$

- Oletuksien mukaan satunnaismuuttujan Y *ehdollinen jakauma* ehdolla $X = x$ tunnetaan:

$$f_{Y|X}(y\,|\,x) = 1/x \qquad 0 < y < x$$
$$f_{Y|X}(y\,|\,x) = 0 \qquad \text{muulloin}$$

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin yhteisjakauma ja *Y*:n reunajakauma

- Satunnaismuuttujien X ja Y *yhteisjakauma* on

$$\begin{aligned} f_{XY}(x, y) &= f_{Y|X}(y \mid x) f_X(x) \\ &= 1/x \qquad\qquad 0 < y < x < 1 \end{aligned}$$

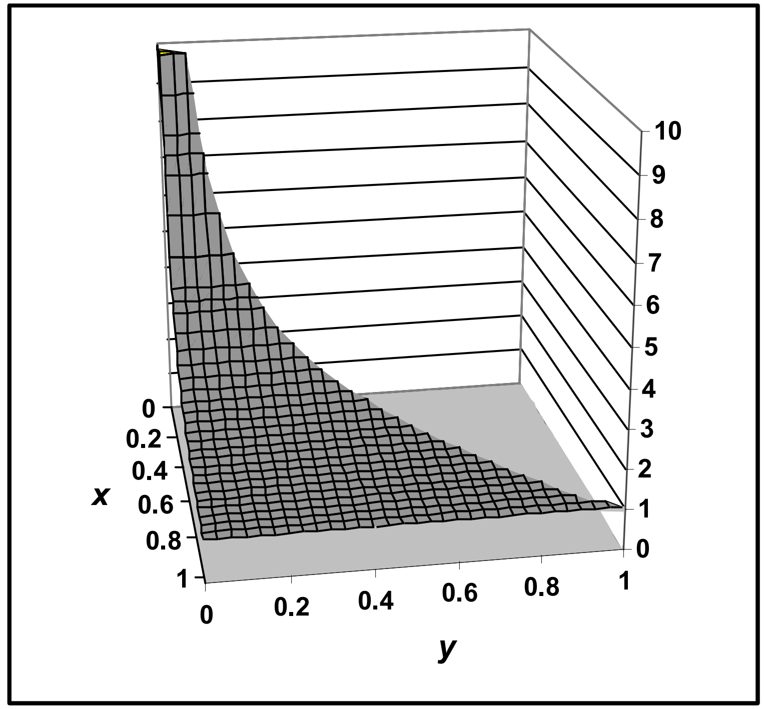
- Satunnaismuuttujan Y *reunajakauma* on

$$f_Y(y) = \int_y^1 f_{XY}(x, y)dx = \int_y^1 \frac{1}{x}dx = [\log x]_y^1 = -\log y$$

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin yhteisjakauman tiheysfunktio 1/2

- $X \sim \text{Uniform}(0, 1)$
- $Y \big| X = x \sim \text{Uniform}(0, x)$
- Kuva oikealla esittää satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktiota*

  $f_{XY}(x, y) = 1/x \text{ , } 0 < y < x < 1$



Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin yhteisjakauman tiheysfunktio 2/2

- $X \sim \text{Uniform}(0, 1)$
- $Y \,\big|\, X = x \sim \text{Uniform}(0, x)$
- Kuva oikealla esittää satunnaismuuttujien X ja Y *yhteisjakauman tiheysfunktion*

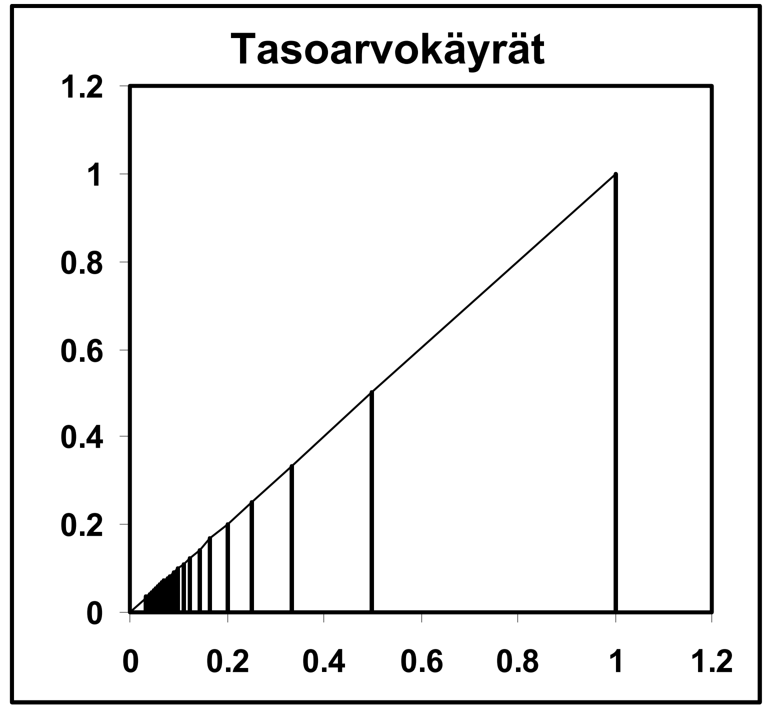
  $f_{XY}(x, y) = 1/x \;,\; 0 < y < x < 1$

  tasa-arvokäyriä

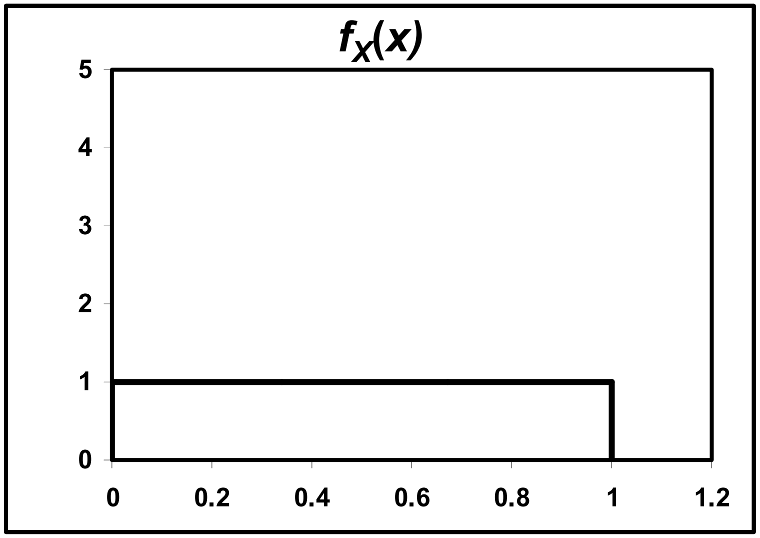
  $x = 1/k \;,\; 0 < y < x$

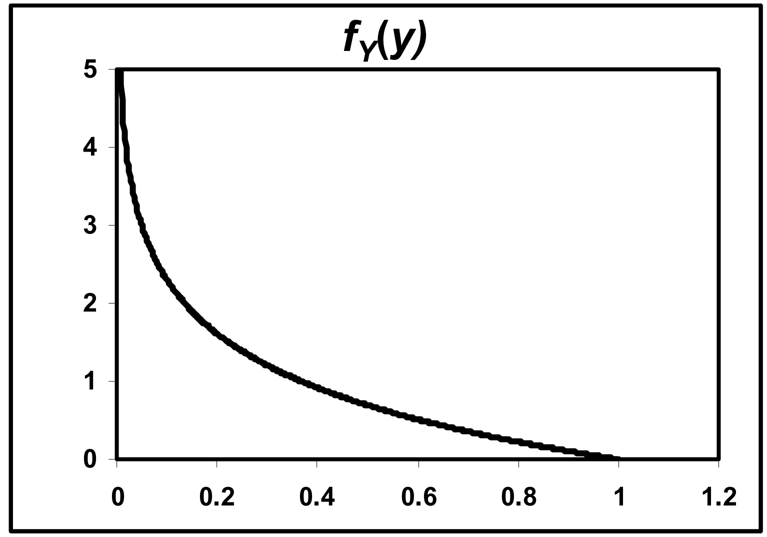
  kun $k = 1, 2, \ldots , 30.$
- Kuvaan on merkitty myös suora

  $y = x \;,\; 0 < x < 1$



Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin reunajakaumien tiheysfunktiot





- Kuvat yllä esittävät satunnaismuuttujien X ja Y *reunajakaumien tiheysfunktioita*:

$$f_X(x) = 1\ ,\ 0 < x < 1$$

$$f_Y(y) = -\log(y)\ ,\ 0 < y < x < 1$$

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin *Y*:n ehdollinen odotusarvo 1/2

- Oletetaan, että

$$X \sim \text{Uniform}(0,\,1)$$
$$Y \,\big|\, X = x \sim \text{Uniform}(0,\,x)$$

- Koska X ja $Y \,\big|\, X = x$ noudattavat tasaista jakaumaa, niin

$$\mathrm{E}(X) = \tfrac{1}{2}$$
$$\mathrm{E}(Y \,|\, X = x) = \tfrac{1}{2}x$$

- Satunnaismuuttujan Y *ehdollisen odotusarvon* $\mathrm{E}(Y \,|\, X)$ *odotusarvo* satunnaismuuttujan X suhteen on iteroidun odotusarvon lain mukaan

$$\mathrm{E}(Y) = \mathrm{E}_X\left[\mathrm{E}(Y \,|\, X)\right] = \tfrac{1}{2}\mathrm{E}(X) = \tfrac{1}{2}\cdot\tfrac{1}{2} = \tfrac{1}{4}$$

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin *Y*:n ehdollinen odotusarvo 2/2

- Koska toisaalta

$$\mathrm{E}(Y) = \int_0^1 y f_Y(y)\,dy$$

ja

$$f_Y(y) = -\log(y)$$

olemme sivutuloksena saaneet seuraavan integraalikaavan:

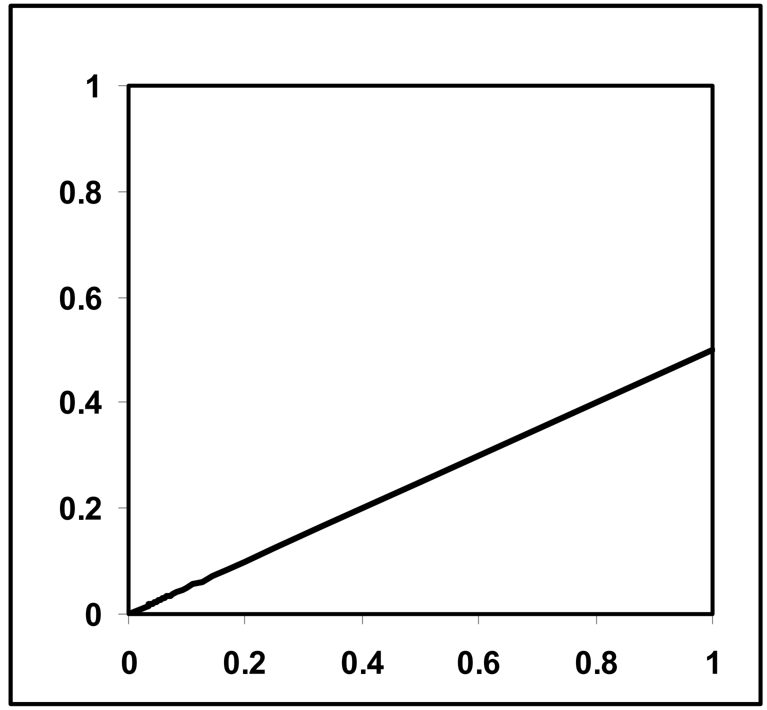
$$-\int_0^1 y\log(y)\,dy = \tfrac{1}{4}$$

Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin *Y*:n regressiokäyrä *X*:n suhteen 1/2

- $X \sim \text{Uniform}(0, 1)$
- $Y \big| X = x \sim \text{Uniform}(0, x)$
- Satunnaismuuttujan Y *ehdollinen odotusarvo* satunnaismuuttujan X suhteen on

  $\mathrm{E}(Y|X = x) = x/2$
- Siten muuttujan y *regressiokäyrä* muuttujan x suhteen on muotoa

  $y = x/2$



Jatkuvan 2-ulotteisen jakauman ehdolliset jakaumat: Esimerkin *Y*:n regressiokäyrä *X*:n suhteen 2/2

- Huomaa, että satunnaismuuttujan X saaman arvon tuntemisesta *on* tässä tapauksessa *hyötyä* satunnaismuuttujan Y saaman arvon *ennustamisessa*.