

4. Todennäköisyyslaskennan kertausta

Sisältö

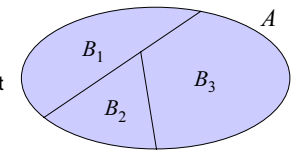
- Peruskäsitteet
- Diskreetit satunnaismuuttujat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttujat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttujat

Otosavaruus, alkeistapaus, tapahtuma

- **Otosavaruus** Ω (sample space) on kaikkien mahdollisten **alkeistapausten** ω (sample) muodostama joukko, $\omega \in \Omega$
 - **Esim. 0.** Rahanheitto: $\Omega = \{H, T\}$
 - **Esim. 1.** Nopanheitto: $\Omega = \{1, 2, 3, 4, 5, 6\}$
 - **Esim. 2.** Asiakkaiden lkm jonossa: $\Omega = \{0, 1, 2, \dots\}$
 - **Esim. 3.** Asiakkaan palveluaika (esim. minuutteina): $\Omega = \{x \in \mathbb{R} \mid x > 0\}$
- **Tapahtumat** A, B, C, \dots (events) ovat otosavaruuden Ω mitallisia osajoukkoja, $A, B, C, \dots \subset \Omega$
 - **Esim. 1.** "Nopanheitossa parillinen luku": $A = \{2, 4, 6\}$
 - **Esim. 2.** "Jono tyhjä": $A = \{0\}$
 - **Esim. 3.** "Asiakkaan palvelu kestää yli 3 minuuttia": $A = \{x \in \mathbb{R} \mid x > 3.0\}$
- Merkitään \mathcal{F} :llä kaikkien tapahtumien A joukkoa, $A \in \mathcal{F}$
 - **Varma tapahtuma:** otosavaruus $\Omega \in \mathcal{F}$ itse
 - **Mahdoton tapahtuma:** tyhjä joukko $\emptyset \in \mathcal{F}$

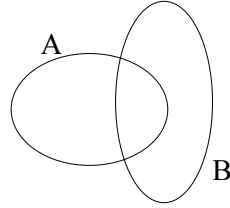
Tapahtumien yhdistely

- **Yhdiste** (union) "A tai B": $A \cup B = \{\omega \in \Omega \mid \omega \in A \text{ tai } \omega \in B\}$
- **Leikkaus** (intersection) "A ja B": $A \cap B = \{\omega \in \Omega \mid \omega \in A \text{ ja } \omega \in B\}$
- **Komplementti** (complement) "ei A": $A^c = \{\omega \in \Omega \mid \omega \notin A\}$
- Tapahtumat A ja B ovat **toistensa poissulkevia** (disjoint), jos
 - $A \cap B = \emptyset$
- Kokoelma tapahtumia $\{B_1, B_2, \dots\}$ muodostaa tapahtuman A **osituksen** (partition), jos
 - (i) $B_i \cap B_j = \emptyset$ kaikilla $i \neq j$
 - (ii) $\cup_i B_i = A$
 - **Esim. 1.** Nopanheitossa parittomat ja parilliset luvut osittavat koko otosavaruuden: $B_1 = \{1, 3, 5\}$ ja $B_2 = \{2, 4, 6\}$



Todennäköisyys

- Tapahtuman A **todennäköisyyttä** (tn, probability) merkitään $P(A)$:lla, $P(A) \in [0,1]$
 - Todennäköisyysmitta P on siis ns. joukkofunktio, $P: \mathcal{F} \rightarrow [0,1]$
- **Ominaisuuksia:**
 - (i) $0 \leq P(A) \leq 1$
 - (ii) $P(\emptyset) = 0$
 - (iii) $P(\Omega) = 1$
 - (iv) $P(A^c) = 1 - P(A)$
 - (v) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
 - (vi) $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$
 - (vii) kokoelma $\{B_i\}$ on tapahtuman A ositus $\Rightarrow P(A) = \sum_i P(B_i)$
 - (viii) $A \subset B \Rightarrow P(A) \leq P(B)$



5

Ehdollinen todennäköisyys

- Oletetaan, että tapahtumalle B : $P(B) > 0$
- **Määr.** Tapahtuman A **ehdollinen todennäköisyys** (conditional probability) ehdolla B on

$$P(A | B) = \frac{P(A \cap B)}{P(B)}$$

- Seuraus:

$$P(A \cap B) = P(B)P(A | B) = P(A)P(B | A)$$

6

Kokonaistodennäköisyyden kaava

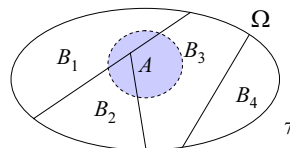
- Olkoon kokoelma $\{B_i\}$ otosavaruuden Ω ositus
- Tällöin kokoelma $\{A \cap B_i\}$ on tapahtuman A ositus, joten (kts. kalvo 5)

$$P(A) \stackrel{(vii)}{=} \sum_i P(A \cap B_i)$$

- Oletaan lisäksi, että $P(B_i) > 0$ kaikilla i . Tällöin (kts. kalvo 6)

$$P(A) = \sum_i P(B_i)P(A | B_i)$$

- Tätä kutsutaan **kokonaistodennäköisyyden kaavaksi**



7

Bayesin kaava

- Olkoon kokoelma $\{B_i\}$ otosavaruuden Ω ositus
- Oletetaan, että $P(A) > 0$ ja $P(B_i) > 0$ kaikilla i . Tällöin (kts. kalvo 6)

$$P(B_i | A) = \frac{P(A \cap B_i)}{P(A)} = \frac{P(B_i)P(A | B_i)}{P(A)}$$

- Näin ollen, kokonaistodennäköisyyden kaavan nojalla (kts. kalvo 7),

$$P(B_i | A) = \frac{P(B_i)P(A | B_i)}{\sum_j P(B_j)P(A | B_j)}$$

- Tätä kutsutaan **Bayesin kaavaksi**
 - tn:iä $P(B_i)$ kutsutaan tapahtumien B_i **a priori** todennäköisyyksiksi
 - tn:iä $P(B_i | A)$ taas sanotaan tapahtumien B_i **a posteriori** todennäköisyyksiksi (ehdolla, että tapahtuma A tapahtui)

8

Tilastollinen riippumattomuus

- **Määr.** Tapahtumat A ja B ovat **riippumattomia** (independent), jos

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

- Seuraus:

$$P(A | B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$$

- Vastaavasti:

$$P(B | A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$$

9

Satunnaismuuttujat

- **Määr.** Reaaliarvoinen **satunnaismuuttuja** X (sm, random variable) on mitallinen kuvaus otosavaruudesta Ω reaalilukujen joukkoon \mathfrak{R} , $X: \Omega \rightarrow \mathfrak{R}$
 - jokaiseen alkeistapaukseen $\omega \in \Omega$ liitetään reaaliluku $X(\omega)$
- **Mitallisuus** (measurability) tarkoittaa, että kaikki tyyppiä

$$\{X \leq x\} := \{\omega \in \Omega \mid X(\omega) \leq x\} \subset \Omega$$

olevat otosavaruuden joukot kuuluvat tapahtumien joukkoon \mathcal{F} , ts.

$$\{X \leq x\} \in \mathcal{F}$$

- Tapahtuman " $X \leq x$ " todennäköisyys on siten $P\{X \leq x\}$

10

Esimerkki

- Rahaa heitetään kolme kertaa peräkkäin
- Otosavaruus:

$$\Omega = \{(\omega_1, \omega_2, \omega_3) \mid \omega_i \in \{H, T\}, i=1,2,3\}$$

- Olkoon X satunnaismuuttuja, joka kertoo klaavojen (T = tails) lkm:n näissä kolmessa heitossa:

| ω | HHH | HHT | HTH | THH | HTT | THT | TTH | TTT |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| $X(\omega)$ | 0 | 1 | 1 | 1 | 2 | 2 | 2 | 3 |

11

Tapahtuman indikaattori

- Olkoon $A \in \mathcal{F}$ mielivaltainen tapahtuma
- **Määr.** Satunnaismuuttujaa 1_A , joka määritellään kaavalla

$$1_A(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \omega \notin A \end{cases}$$

sanotaan tapahtuman A **indikaattoriksi** (indicator)

- Selvästikin:

$$P\{1_A = 1\} = P(A)$$

$$P\{1_A = 0\} = P(A^c) = 1 - P(A)$$

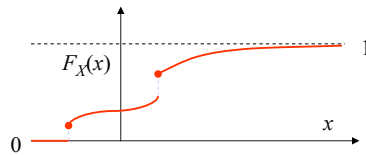
12

Kertymäfunktio

- **Määr.** Sm:n X **kertymäfunktio** (kf, cumulative distribution function) on kuvaus $F_X: \mathbb{R} \rightarrow [0,1]$, joka määritellään kaavalla

$$F_X(x) = P\{X \leq x\}$$

- Kf määrää täydellisesti ko. sm:n **jakauman** (distribution)
 - so. tn:t $P\{X \in B\}$, missä $B \subset \mathbb{R}$ ja $\{X \in B\} \in \mathcal{F}$
- **Ominaisuuksia:**
 - (i) F_X on kasvava
 - (ii) F_X on oikealta jatkuva
 - (iii) $F_X(-\infty) = 0$
 - (iv) $F_X(\infty) = 1$



13

Satunnaismuuttujien tilastollinen riippumattomuus

- **Määr.** Sm:t X ja Y ovat **riippumattomia**, jos kaikilla x ja y

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\}$$

- **Määr.** Sm:t X_1, \dots, X_n ovat **täydellisesti riippumattomia**, jos kaikilla i ja x_i

$$P\{X_1 \leq x_1, \dots, X_n \leq x_n\} = P\{X_1 \leq x_1\} \cdots P\{X_n \leq x_n\}$$

14

Riippumattomien satunnaismuuttujien maksimi ja minimi

- Olkoot sm:t X_1, \dots, X_n **täydellisesti riippumattomia**
- Merkitään $X^{\max} := \max\{X_1, \dots, X_n\}$. Tällöin

$$\begin{aligned} P\{X^{\max} \leq x\} &= P\{X_1 \leq x, \dots, X_n \leq x\} \\ &= P\{X_1 \leq x\} \cdots P\{X_n \leq x\} \end{aligned}$$

- Merkitään $X^{\min} := \min\{X_1, \dots, X_n\}$. Tällöin

$$\begin{aligned} P\{X^{\min} > x\} &= P\{X_1 > x, \dots, X_n > x\} \\ &= P\{X_1 > x\} \cdots P\{X_n > x\} \end{aligned}$$

15

Sisältö

- Peruskäsitteet
- Diskreetit satunnaismuuttujat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttujat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttujat

16

Diskreetit satunnaismuuttujat

- **Määr.** Joukkoa $A \subset \mathfrak{R}$ sanotaan **diskreetiksi** (discrete), jos se on
 - äärellinen, $A = \{x_1, \dots, x_n\}$, tai
 - numeroituvasti ääretön, $A = \{x_1, x_2, \dots\}$.
- **Määr.** Sm X on **diskreetti**, jos on olemassa sellainen diskreetti joukko $S_X \subset \mathfrak{R}$, että

$$P\{X \in S_X\} = 1$$

- Seuraus:
 - $P\{X=x\} \geq 0$ kaikilla $x \in S_X$
 - $P\{X=x\} = 0$ kaikilla $x \notin S_X$
- Joukkoa S_X sanotaan sm:n X **arvojoukoksi**

17

Pistetodennäköisyysfunktio

- Olkoon sm X diskreetti
- Sm:n X jakauman määräävät **pistetodennäköisyydet** p_i

$$p_i := P\{X = x_i\}, \quad x_i \in S_X$$

- **Määr.** Sm:n X **pistetodennäköisyysfunktio** (ptnf, probability mass function) $p_X: \mathfrak{R} \rightarrow [0,1]$ määritellään kaavalla

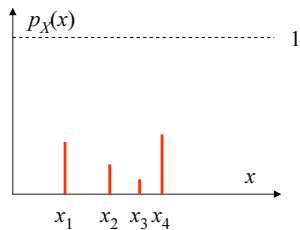
$$p_X(x) := P\{X = x\} = \begin{cases} p_i, & x = x_i \in S_X \\ 0, & x \notin S_X \end{cases}$$

- Kf on tässä tapauksessa seuraava porrassfunktio:

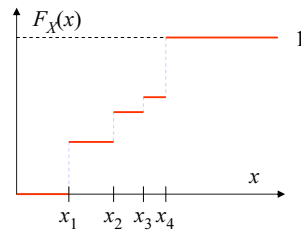
$$F_X(x) = P\{X \leq x\} = \sum_{i: x_i \leq x} p_i$$

18

Esimerkki



pistetodennäköisyysfunktio (ptnf)



kertymäfunktio (kf)

$$S_X = \{x_1, x_2, x_3, x_4\}$$

19

Diskreettien satunnaismuuttujien riippumattomuus

- Diskreetit sm:t X ja Y ovat riippumattomia, jos ja vain jos kaikilla $x_i \in S_X$ ja $y_j \in S_Y$

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}$$

20

Odotusarvo

- **Määr. Sm:n X odotusarvo** (mean, expectation) määritellään kaavalla

$$\mu_X := E[X] := \sum_{x \in S_X} P\{X = x\} \cdot x = \sum_{x \in S_X} p_X(x)x = \sum_i p_i x_i$$

- Huom. 1. Odotusarvo on (hyvin) määritelty vain, jos $\sum_i p_i |x_i| < \infty$
- Huom. 2. Jos $x_i \geq 0$ ja $\sum_i p_i x_i = \infty$, niin voidaan merkitä $E[X] = \infty$

- **Ominaisuuksia:**

- (i) $c \in \mathfrak{R} \Rightarrow E[cX] = cE[X]$
- (ii) $E[X + Y] = E[X] + E[Y]$
- (iii) X ja Y riippumattomia $\Rightarrow E[XY] = E[X]E[Y]$

21

Varianssi

- **Määr. Sm:n X varianssi** (variance) määritellään kaavalla

$$\sigma_X^2 := D^2[X] := \text{Var}[X] := E[(X - E[X])^2]$$

- Kätevä kaava (todista!):

$$D^2[X] = E[X^2] - E[X]^2$$

- **Ominaisuuksia:**

- (i) $c \in \mathfrak{R} \Rightarrow D^2[cX] = c^2 D^2[X]$
- (ii) X ja Y riippumattomia $\Rightarrow D^2[X + Y] = D^2[X] + D^2[Y]$

22

Kovarianssi

- **Määr. Sm:ien X ja Y välinen kovarianssi** (covariance) määr. kaavalla

$$\sigma_{XY}^2 := \text{Cov}[X, Y] := E[(X - E[X])(Y - E[Y])]$$

- Kätevä kaava (todista!):

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$$

- **Ominaisuuksia:**

- (i) $\text{Cov}[X, X] = \text{Var}[X]$
- (ii) $\text{Cov}[X, Y] = \text{Cov}[Y, X]$
- (iii) $\text{Cov}[X + Y, Z] = \text{Cov}[X, Z] + \text{Cov}[Y, Z]$
- (iv) X ja Y riippumattomia $\Rightarrow \text{Cov}[X, Y] = 0$

23

Muita jakaumaan liittyviä tunnuslukuja

- **Määr. Sm:n X hajonta** (standard deviation):

$$\sigma_X := D[X] := \sqrt{D^2[X]}$$

- **Määr. Sm:n X variaatiokerroin** (coefficient of variation):

$$c_X := C[X] := \frac{D[X]}{E[X]}$$

- **Määr. Sm:n X k :s momentti** (moment), $k = 1, 2, \dots$:

$$\mu_X^{(k)} := E[X^k]$$

24

Riippumattomien satunnaismuuttujien keskiarvo

- Olkoot sm:t X_1, \dots, X_n riippumattomia ja samoin jakautuneita (IID) odotusarvonaan μ ja varianssinaan σ^2
- Merkitään näiden sm:ien keskiarvoa (sample mean) seuraavasti:

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

- Tällöin (todista!)

$$E[\bar{X}_n] = \mu$$

$$D^2[\bar{X}_n] = \frac{\sigma^2}{n}$$

$$D[\bar{X}_n] = \frac{\sigma}{\sqrt{n}}$$

25

Suurten lukujen laki (SLL)

- Olkoot sm:t X_1, \dots, X_n riippumattomia ja samoin jakautuneita (IID) odotusarvonaan μ ja varianssinaan σ^2
- **Heikko suurten lukujen laki:** kaikilla $\varepsilon > 0$

$$P\{|\bar{X}_n - \mu| > \varepsilon\} \rightarrow 0$$

- **Vahva suurten lukujen laki:** todennäköisyydellä 1

$$\bar{X}_n \rightarrow \mu$$

- **Seuraus:** Suurilla n :n arvoilla

$$\bar{X}_n \approx \mu$$

26

Sisältö

- Peruskäsitteet
- Diskreetit satunnaismuuttujat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttujat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttujat

27

Bernoulli-jakauma

$$X \sim \text{Bernoulli}(p), \quad p \in (0,1)$$

- kuvaa yksittäistä satunnaiskoetta, jonka tuloksena joko **onnistuminen** (1) tai **epäonnistuminen** (0); vrt. rahanheitto
- onnistuminen tn:llä p (ja epäonnistuminen tn:llä $1 - p$)
- Arvojoukko: $S_X = \{0, 1\}$
- Pistetodennäisyydet:

$$P\{X = 0\} = 1 - p, \quad P\{X = 1\} = p$$

- Odotusarvo: $E[X] = (1 - p) \cdot 0 + p \cdot 1 = p$
- Toinen momentti: $E[X^2] = (1 - p) \cdot 0^2 + p \cdot 1^2 = p$
- Varianssi: $D^2[X] = E[X^2] - E[X]^2 = p - p^2 = p(1 - p)$

28

Binomijakauma

$$X \sim \text{Bin}(n, p), \quad n \in \{1, 2, \dots\}, p \in (0, 1)$$

- onnistumisten lkm n :ssä perättäisessä ja toisistaan riippumattomassa satunnaiskokeessa; $X = X_1 + \dots + X_n$ (missä $X_i \sim \text{Bernoulli}(p)$)
- n = satunnaiskokeiden lkm
- p = onnistumisen tn yksittäisessä satunnaiskokeessa
- Arvojoukko: $S_X = \{0, 1, \dots, n\}$
- Pistetodennäköisyydet:

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}$$

$$n! = n \cdot (n-1) \cdot \dots \cdot 2 \cdot 1$$

$$P\{X = i\} = \binom{n}{i} p^i (1-p)^{n-i}$$

- Odotusarvo: $E[X] = E[X_1] + \dots + E[X_n] = np$
- Varianssi: $D^2[X] = D^2[X_1] + \dots + D^2[X_n] = np(1-p)$ riippumattomuus!

Geometrinen jakauma

$$X \sim \text{Geom}(p), \quad p \in (0, 1)$$

- peräkkäisten onnistumisten lkm ennen ensimmäistä epäonnistumista (sarjassa peräkkäisiä ja toisistaan riippumattomia satunnaiskokeita)
- p = onnistumisen tn yksittäisessä satunnaiskokeessa
- Arvojoukko: $S_X = \{0, 1, \dots\}$
- Pistetodennäköisyydet:

$$P\{X = i\} = p^i (1-p)$$

- Odotusarvo: $E[X] = \sum_i i p^i (1-p) = p/(1-p)$
- Toinen momentti: $E[X^2] = \sum_i i^2 p^i (1-p) = 2(p/(1-p))^2 + p/(1-p)$
- Varianssi: $D^2[X] = E[X^2] - E[X]^2 = p/(1-p)^2$

Geometrisen jakauman unohtavuusominaisuus

- Geometrisella jakaumalla on ns. **unohtavuusominaisuus** (memoryless property): kaikilla $i, j \in \{0, 1, \dots\}$

$$P\{X \geq i + j \mid X \geq i\} = P\{X \geq j\}$$

- Todista!
 - Ohje: Todista ensin, että $P\{X \geq i\} = p^i$

Geometrisesti jakautuneiden satunnaismuuttujien minimi

- Olkoot sm:t $X_1 \sim \text{Geom}(p_1)$ ja $X_2 \sim \text{Geom}(p_2)$ **riippumattomia**
- Tällöin

$$X^{\min} := \min\{X_1, X_2\} \sim \text{Geom}(p_1 p_2)$$

ja

$$P\{X^{\min} = X_i\} = \frac{1-p_i}{1-p_1 p_2}, \quad i \in \{1, 2\}$$

- Todista!
 - Ohje: Kts. kalvo 15

Poisson-jakauma

$$X \sim \text{Poisson}(a), \quad a > 0$$

– binomijakauman rajatapaus, kun $n \rightarrow \infty$ ja $p \rightarrow 0$ siten, että $np \rightarrow a$

- Arvojoukko: $S_X = \{0, 1, \dots\}$
- Pistetodennäköisyydet:

$$P\{X = i\} = \frac{a^i}{i!} e^{-a}$$

- Odotusarvo: $E[X] = a$
- Toinen momentti: $E[X(X-1)] = a^2 \Rightarrow E[X^2] = a^2 + a$
- Varianssi: $D^2[X] = E[X^2] - E[X]^2 = a$

33

Esimerkki

- Oletetaan, että
 - paikalliskeskukseen on kytketty 200 tilaajaa
 - yksittäisen tilaajan ominaisliikenne on 0.01 erlangia
 - tilaajat toimivat toisistaan riippumattomasti
- Tällöin käynnissäolevien puhelujen lkm $X \sim \text{Bin}(200, 0.01)$
- Vastaava Poisson-approksimaatio: $X \approx \text{Poisson}(2.0)$
- Pistetodennäköisyyksien vertailua:

| | 0 | 1 | 2 | 3 | 4 | 5 |
|---------------|-------|-------|-------|-------|-------|-------|
| Bin(200,0.01) | .1326 | .2679 | .2693 | .1795 | .0893 | .0354 |
| Poisson(2.0) | .1353 | .2701 | .2701 | .1804 | .0902 | .0361 |

34

Poisson-jakauman ominaisuuksia

- (i) **Summa:** Olkoot sm:t $X_1 \sim \text{Poisson}(a_1)$ ja $X_2 \sim \text{Poisson}(a_2)$ riippumattomia. Tällöin

$$X_1 + X_2 \sim \text{Poisson}(a_1 + a_2)$$

- (ii) **Satunnaisotanta:** Olkoon $X \sim \text{Poisson}(a)$ alkioiden lkm (jossakin satunnaisen kokoisessa joukossa). Valitaan näistä alkiosta satunnainen osajoukko (jokainen yksittäinen alkio otetaan mukaan tn:llä p), jonka kokoa merkitään Y :llä. Tällöin

$$Y \sim \text{Poisson}(pa)$$

- (iii) **Satunnaislajittelu:** Olkoot sm:t X ja Y kuten yllä (ii). Merk. $Z = X - Y$. Tällöin Y ja Z ovat **riippumattomia** (ehdolla, että X :ä ei tunneta),

$$Z \sim \text{Poisson}((1-p)a)$$

35

Sisältö

- Peruskäsitteet
- Diskreetit satunnaismuuttujat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttujat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttujat

36

Jatkuvat satunnaismuuttajat

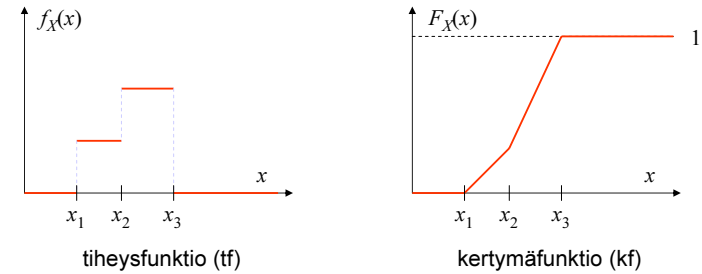
- **Määr.** Sm X on **jatkuva** (continuous), jos on olemassa sellainen integroitava funktio $f_X: \mathfrak{R} \rightarrow \mathfrak{R}_+$, että kaikilla $x \in \mathfrak{R}$ pätee

$$F_X(x) := P\{X \leq x\} = \int_{-\infty}^x f_X(y) dy$$

- Funktiota f_X sanotaan sm:n X **tiheysfunktioksi** (tf, probability density function)
 - Joukkoa S_X , missä $f_X > 0$, sanotaan sm:n X **arvojoukoksi**
- **Ominaisuuksia:**
 - (i) $P\{X=x\} = 0$ kaikilla $x \in \mathfrak{R}$
 - (ii) $P\{a < X < b\} = P\{a \leq X \leq b\} = \int_a^b f_X(x) dx$
 - (iii) $P\{X \in A\} = \int_A f_X(x) dx$
 - (iv) $P\{X \in \mathfrak{R}\} = \int_{-\infty}^{\infty} f_X(x) dx = \int_{S_X} f_X(x) dx = 1$

37

Esimerkki



$$S_X = (x_1, x_3)$$

38

Odotusarvo ja muita jakaumaan liittyviä tunnuslukuja

- **Määr.** Sm:n X **odotusarvo** (mean) määritellään kaavalla

$$\mu_X := E[X] := \int_{-\infty}^{\infty} f_X(x)x dx$$

- Huom. 1. Odotusarvo on (hyvin) määritelty vain, jos $\int_{-\infty}^{\infty} f_X(x)|x| dx < \infty$
- Huom. 2. Jos $S_X = \mathfrak{R}_+$ ja $\int_0^{\infty} f_X(x)x = \infty$, niin voidaan merkitä $E[X] = \infty$
- Jatkuvan sm:n odotusarvolla on samat ominaisuudet kuin diskreetin sm:n odotusarvolla (kts. kalvo 21)
- Muut jakaumaan liittyvät tunnusluvut (varianssi, kovarianssi,...) määritellään odotusarvon avulla täsmälleen samoin kuin diskreetin sm:n tapauksessa
 - Näin ollen myös näiden tunnuslukujen ominaisuudet säilyvät (kts. kalvot 22-24)

39

Sisältö

- Peruskäsitteet
- Diskreetit satunnaismuuttajat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttajat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttajat

40

Tasajakauma

$$X \sim U(a, b), \quad a < b$$

- jatkuva vastine nopanheitolle (kaikki arvot "yhtä todennäköisiä")
- Arvojoukko: $S_X = (a, b)$
- Tiheysfunktio (tf):

$$f_X(x) = \frac{1}{b-a}, \quad x \in (a, b)$$

- Kertymäfunktio (kf):

$$F_X(x) := P\{X \leq x\} = \frac{x-a}{b-a}, \quad x \in (a, b)$$

- Odotusarvo: $E[X] = \int_a^b x/(b-a) dx = (a+b)/2$
- Toinen momentti: $E[X^2] = \int_a^b x^2/(b-a) dx = (a^2 + ab + b^2)/3$
- Varianssi: $D^2[X] = E[X^2] - E[X]^2 = (b-a)^2/12$

41

Eksponenttijakauma

$$X \sim \text{Exp}(\lambda), \quad \lambda > 0$$

- geometrisen jakauman jatkuva vastine ("epäonnistuminen" tn:llä $\approx \lambda dt$)
- $P\{X \in (t, t+h] | X > t\} = \lambda h + o(h)$, missä $o(h)/h \rightarrow 0$, kun $h \rightarrow 0$
- Arvojoukko: $S_X = (0, \infty)$
- Tiheysfunktio (tf):

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0$$

- Kertymäfunktio (kf):

$$F_X(x) := P\{X \leq x\} = 1 - e^{-\lambda x}, \quad x > 0$$

- Odotusarvo: $E[X] = \int_0^\infty \lambda x \exp(-\lambda x) dx = 1/\lambda$
- Toinen momentti: $E[X^2] = \int_0^\infty \lambda x^2 \exp(-\lambda x) dx = 2/\lambda^2$
- Varianssi: $D^2[X] = E[X^2] - E[X]^2 = 1/\lambda^2$

42

Eksponenttijakauman unohtavuusominaisuus

- Eksponenttijakaumalla on ns. **unohtavuusominaisuus** (memoryless property): kaikilla $x, y \in (0, \infty)$

$$P\{X > x + y | X > x\} = P\{X > y\}$$

- Todista!
 - Ohje: Todista ensin, että $P\{X > x\} = e^{-\lambda x}$
- Sovellus:
 - Oletetaan, että puhelujen pitoajat ovat eksponentiaalisesti jakautuneita odotusarvonaan h minuuttia.
 - Tarkastellaan puhelua, joka on jo kestänyt ajan x minuuttia. Unohtavuusominaisuuden nojalla tällä ei ole mitään merkitystä puhelun jäljellä olevan keston kannalta: keskimäärin tällainen puhelu kestää vielä h minuuttia (siis $x + h$ minuuttia kaikenkaikkiaan)!

43

Eksponentiaalisesti jakautuneiden satunnaismuuttujien minimi

- Olkoot sm:t $X_1 \sim \text{Exp}(\lambda_1)$ ja $X_2 \sim \text{Exp}(\lambda_2)$ **riippumattomia**.
- Tällöin

$$X^{\min} := \min\{X_1, X_2\} \sim \text{Exp}(\lambda_1 + \lambda_2)$$

ja

$$P\{X^{\min} = X_i\} = \frac{\lambda_i}{\lambda_1 + \lambda_2}, \quad i \in \{1, 2\}$$

- Todista!
 - Ohje: Kts. kalvo 15

44

Normeerattu normaalijakauma

$$X \sim N(0,1)$$

- riippumattomien ja samoin jakautuneiden (odotusarvona 0 ja varianssina 1) sm:ien "normeeratun" summan rajatapaus (kts. kalvo 48)

- Arvojoukko: $S_X = (-\infty, \infty)$
- Tiheysfunktio (tf):

$$f_X(x) = \varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

- Kertymäfunktio (kf):

$$F_X(x) := P\{X \leq x\} = \Phi(x) := \int_{-\infty}^x \varphi(y) dy$$

- Odotusarvo: $E[X] = 0$ (tf symmetrinen!)
- Varianssi: $D^2[X] = 1$

45

Normaalijakauma

$$X \sim N(\mu, \sigma^2), \quad \mu \in \mathfrak{R}, \quad \sigma > 0$$

- jos $(X - \mu)/\sigma \sim N(0,1)$
- Arvojoukko: $S_X = (-\infty, \infty)$
- Tiheysfunktio (tf):

$$f_X(x) := F_X'(x) = \frac{1}{\sigma} \varphi\left(\frac{x-\mu}{\sigma}\right)$$

- Kertymäfunktio (kf):

$$F_X(x) := P\{X \leq x\} = P\left\{\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}\right\} = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

- Odotusarvo: $E[X] = \mu + \sigma E[(X - \mu)/\sigma] = \mu$ (tf symmetr. μ :n suhteen)
- Varianssi: $D^2[X] = \sigma^2 D^2[(X - \mu)/\sigma] = \sigma^2$

46

Normaalijakauman ominaisuuksia

- (i) **Lineaarimuunos:** Olk. $X \sim N(\mu, \sigma^2)$ ja $\alpha, \beta \in \mathfrak{R}$. Tällöin

$$Y := \alpha X + \beta \sim N(\alpha\mu + \beta, \alpha^2\sigma^2)$$

- (ii) **Summa:** Olkoot sm:t $X_1 \sim N(\mu_1, \sigma_1^2)$ ja $X_2 \sim N(\mu_2, \sigma_2^2)$ **riippumattomia**. Tällöin

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

- (iii) **Otoskeskiarvo:** Olkoot sm:t $X_i \sim N(\mu, \sigma^2)$, $i = 1, \dots, n$, **riippumattomia ja samoin jakautuneita** (IID) noudattaen normaalijakaumaa. Tällöin niiden keskiarvolle (vrt. kalvo 25) pätee

$$\bar{X}_n := \frac{1}{n} \sum_{i=1}^n X_i \sim N\left(\mu, \frac{1}{n}\sigma^2\right)$$

47

Keskeinen raja-arvolause (KRL)

- Olkoot sm:t X_1, \dots, X_n **riippumattomia ja samoin jakautuneita** (IID) odotusarvoaan μ ja varianssinaan σ^2 (ja lisäksi kolmas momentti olemassa)
- **Keskeinen raja-arvolause:**

$$\frac{1}{\sigma/\sqrt{n}} (\bar{X}_n - \mu) \xrightarrow{\text{i.d.}} N(0,1)$$

- **Seuraus:** Suurilla n :n arvoilla

$$\bar{X}_n \approx N\left(\mu, \frac{1}{n}\sigma^2\right)$$

48

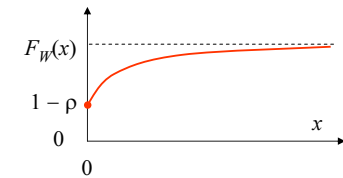
Sisältö

- Peruskäsitteet
- Diskreetit satunnaismuuttujat
- Diskreetit jakaumat (lkm-jakaumat)
- Jatkuvat satunnaismuuttujat
- Jatkuvat jakaumat (aikajakaumat)
- Muut satunnaismuuttujat

49

Muita satunnaismuuttujia

- Puhtaasti diskreettien ja jatkuvien sm:ien lisäksi on olemassa näiden sekamuotoja
- Esimerkki:
 - Merk. W :llä asiakkaan odotusaikaa $M/M/1$ jonossa. S_m :n W jakaumalla on ns. atomi nollassa (ts. $P\{W=0\} = 1 - \rho > 0$), mutta muuten jakauma on jatkuva



50

Sanastoa

- | | |
|--|--|
| • otosavaruus = sample space | • odotusarvo = mean (value) = expectation |
| • tapahtuma = event | • varianssi = variance |
| • todennäköisyys = probability | • kovarianssi = covariance |
| • ehdollinen tn = conditional probability | • hajonta = standard deviation |
| • riippumattomuus = independence | • variaatiokerroin = coefficient of variation |
| • satunnaismuuttuja = random variable | • suurten lukujen laki = law of large numbers |
| • indikaattori = indicator | • jatkuva = continuous |
| • jakauma = distribution | • tiheysfunktio = probability density function |
| • kertymäfunktio = cumulative distribution function | • unohtavaisuusominaisuus = memoryless property |
| • diskreetti = discrete | • keskeinen raja-arvolause = central limit theorem |
| • pistetodennäköisyysfunktio = probability mass function | |

51