Johdatus todennäköisyyslaskentaan
Kevät 2015

Luento 12 / 13

Jukka Kohonen
Matematiikan ja tilastotieteen laitos
Helsingin yliopisto

ODOTUSARVON OMINAISUUKSIA

Odotusarvo on lineaarinen

Tuominen s. 78

- Vakiolla kertominen $E(aX) = a\,E(X)$
- Sm:ien summa $E(X+Y) = E(X) + E(Y)$

Huom! Summakaava pätee aina, kaikille satunnaismuuttujille – riippuvillekin!

(Kunhan odotusarvot ovat olemassa ja äärelliset)

Miksi pätee? Todistus perustuu siihen, että

- odotusarvo on (painotettu) summa, ja
- yhteenlaskun vaihdannaisuuteen.

Muut laskutoimitukset

- **Bad news:**
  Juuri mitään muuta laskutoimitusta ei odotusarvolla yleensä voikaan tehdä ”noin vain” (siis vaihtamalla järjestystä)
- Esim.
  - $E(X^2) \neq E(X)^2$ (yleensä; noppaesimerkki)
  - $E(1/X) \neq 1 / E(X)$ (yleensä; noppaesimerkki)
  - $E(\sin X) \neq \sin(E(X))$ (yleensä)
  - $E(XY) \neq E(X)\, E(Y)$ (yleensä)
  - jne.

- Sellaisen odotusarvon voi kyllä laskea, mutta muilla keinoin, esim. suoraan odotusarvon määritelmästä (painotettu summa)

Eri jakaumilla voi olla sama odotusarvo

- Esim.
  - X ~ symmetrinen joukossa {1, 6}
  - Y ~ symmetrinen joukossa {1, 2, 3, 4, 5, 6}
  - Z ~ Geom(7/9)
  - U ~ Tas(3.3, 3.7) (jatkuva tasajakauma kapealla välillä!)
  - W ~ Exp(2/7)
- Jakaumat ovat hyvin erilaiset, mutta
  E(X) = E(Y) = E(Z) = E(U) = E(W) = 7/2 = 3.5

Toisaalta *samoin jakautuneilla* sm:illa on tietysti sama odotusarvo.
(koska odotusarvo lasketaan jakaumasta: mahdollisista arvoista ja pistetodennäköisyyksistä)

VARIANSSI

Jakauman leveys

- Usein halutaan tunnusluku, joka kertoo miten tiiviisti mahdolliset arvot (eli X:n jakauma) **keskittyvät** odotusarvon lähelle
  - esim. ”koko arvojoukon leveys” ei oikein kerro tätä asiaa (vrt. tasa- ja kolmiojakauma)
- Eräs ratkaisu: **varianssi** (Tuominen s. 82)

  $$\mathbf{Var(X) = E((X - \mu)^2)}$$

  ja varianssin neliöjuuri eli **hajonta**
  - varianssilla paljon käteviä **laskusääntöjä**
  - varianssi myös **tunnetaan** useille tutuille jakaumille
- Varianssille käytetään vaihtelevasti merkintöjä Var(X) ja $D^2(X)$, ne tarkoittavat samaa.

Varianssin idea

- Oletetaan tunnetuksi $E(X) = \mu$.
- Tiedetään, ettei X aina (eikä edes kovin usein) osu odotusarvoonsa. (Ehkä ei koskaan)
- Katsotaan, paljonko se meni huti (= **poikkeama**) ja korotetaan poikkeama toiseen (= **neliöpoikkeama**)
- Jokaiseen X:n mahdolliseen arvoon liittyy vastaava neliöpoikkeama $(X-\mu)^2$
- X:n mahdollisilla arvoilla on todennäköisyydet, joten voidaan laskea, miten neliöpoikkeama on jakautunut. Lasketaan sen odotusarvo ja nimitetään sitä varianssiksi.
- Suuri varianssi siis merkitsee, että X usein poikkeaa paljon odotusarvostaan
- Pienimillään varianssi voisi olla 0, jos X on aina $= \mu$

Varianssin laskentaa

- Lasketaan suoraan määritelmän perusteella
  - Reilun kolikon varianssi
  - Epäreilun kolikon varianssi
- Sitten huomataan kirjasta (s. 82) lause 3.2.3
  - Miksi se pitää paikkansa?
  - Miten sitä käytetään?
- Lauseen 3.2.3 avulla laskenta usein helpompaa, koska lausekkeet sievempiä

Reilun kolikon varianssi

Diskreetti satunnaismuuttuja joukossa {0, 1},

$P(X=0) = ½$

$P(X=1) = ½$

Odotusarvo: $\mu = E(X) = ½ \cdot 0 + ½ \cdot 1 =$ **½**

Poikkeama	$(X-\mu)$	on joko $+½$ tai $-½$
Neliöpoikkeama	$Z = (X-\mu)^2$	on varmasti ¼
Var(X) =	E(Z)	= ¼

Epäreilun kolikon varianssi

Diskreetti satunnaismuuttuja joukossa {0, 1},

$P(X=0) = q$

$P(X=1) = p \qquad p+q=1$

jos X=0

jos X=1

Odotusarvo: $\mu = E(X) = q \cdot 0 + p \cdot 1 =$ **p**

Poikkeama $(X-\mu)$ on joko $-p$ tai $(1-p)$

Neliöpoikkeama $Z = (X-\mu)^2$ on joko $(-p)^2$ tai $(1-p)^2$

$\mathrm{Var}(X) = E(Z)$

$$= q \cdot (-p)^2 + p \cdot (1-p)^2$$

$$= qp^2 + pq^2$$

$$= \mathbf{pq}.$$

Lause 3.2.3

- Merkitään $\mu = E(X)$
- Varianssin määritelmästä lähdetään laskemaan binomin neliötä auki...

$$
\begin{aligned}
\textbf{Var(X)} &= E(\,(X-\mu)^2\,) \\
&= E(X^2 - 2X\mu + \mu^2\,) \\
&= E(X^2) - E(2X\mu) + E(\mu^2) \\
&= E(X^2) - 2\mu\mu + \mu^2 \\
&= \mathbf{E(X^2) - \mu^2}
\end{aligned}
$$

Nopanheitto lauseella 3.2.3

Arvojoukko {1,2,3,4,5,6}, symmetrisesti

$$
\begin{aligned}
E(X) &= 3.5 \\
\mathrm{Var}(X) &= E(X^2) && - [E(X)]^2 \\
&= (1+4+9+16+25+36) / 6 && - 3.5^2 \\
&= (1+4+9+16+25+36) / 6 && - 3.5^2 \\
&= 91/6 && - 3.5^2 \\
&\approx 2.92
\end{aligned}
$$