Johdatus todennäköisyyslaskentaan
Kevät 2014

Luento 8 / 13

Jukka Kohonen

Matematiikan ja tilastotieteen laitos

Helsingin yliopisto

Monen tapahtuman riippumattomuus

(Tuominen s. 40)

- Monta tapahtumaa ovat riippumattomat, jos **mitkään** tapahtumat (edes **yhdessä**) eivät auta arvioimaan minkään muiden tapahtumien todennäköisyyksiä.

- Yleinen tulokaava, määritelmä 1.8.6
- Voisi esittää vastaavat ehdolliset todennäköisyydet, esim.

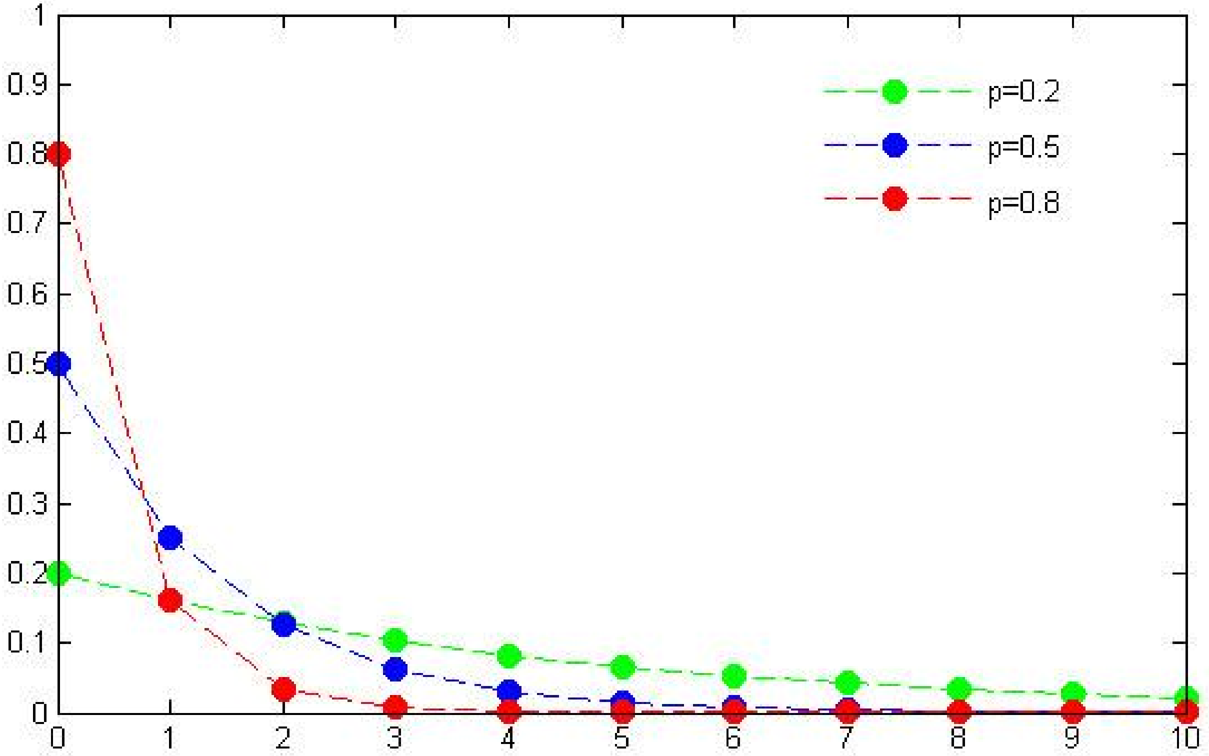
  $P(A \mid B) = P(A \mid C) = P(A \mid (B \cap C)) = P(A)$

  jne.

- Monen tapahtuman riippumattomuus on erittäin vahva oletus (paljon vahvempi kuin että tapahtumat ovat pareittain riippumattomat) – esimerkki: pariteetti

Monen tapahtuman riippumattomuus

- Monen tapahtuman riippumattomuus voi merkitä, että kaikkien tapahtumien **leikkausta** (ts. että kaikki tapahtumat tapahtuvat) **pidetään hyvin epätodennäköisenä**
- **Tämä ei välttämättä ole realistinen oletus!**
- Esim. lentokoneen ohjauksen 3 hydraulijärjestelmää
  - jokaisen vikaantumis-tn = 1/1000 (arvio tai tilasto)
  - vikaantumiset riippumattomia (oletus)
  - tn, että kaikki vikaantuvat kertolaskulla (laskutoimitus)
  - Eräällä lennolla kaikki vikaantuivat (havainto)



DISKREETTI SATUNNAISMUUTTUJA

Satunnaismuuttujan käsite

- Tähän asti on puhuttu **tapahtumista** ja niiden todennäköisyyksistä.
- Tapahtumissa on kyllä vilahdellut numeerisia suureita, esim. tapahtuma ”X=3”, missä X on toistokokeessa tulevien onnistumisten määrä.
- Tämmöinen X on **satunnaismuuttuja**.

- Satunnaismuuttuja on jokin suure, jonka arvo ei ehkä ole vielä tunnettu, mutta siitä on jotain käsitystä (todennäköisyydet).
- Esitetään siis todennäköisyyksiä tapahtumille, jotka ovat muotoa ”X on jotain”.

Muodollinen määritelmä (2.1.2 s. 48)

- Tn-teoriassa ajatellaan muodollisesti, että **satunnaismuuttuja** (lyhenne: **sm**) on *funktio* perusjoukosta R:ään.

- Tämä tarkoittaa käytännössä sitä, että **perusjoukko (”mitä voi tapahtua”) on jaettu osiin, joissa X:llä on eri arvot.**
  (Kuva)
  Nämä osat ovat perusjoukon osajoukkoja, siis tapahtumia, ja niillä on todennäköisyydet.

- On siis esim. olemassa todennäköisyydet $P(X<5)$, $P(X=5)$, $P(2 < X < 3)$ jne.
- P:n sulkujen sisällä on, kuten ennenkin, jokin tapahtuma (väite), esimerkiksi yhtälö, epäyhtälö jne.

Diskreetti sm

- Sm on **diskreetti**, kun sillä on äärellinen tai numeroituva arvojoukko. (Esim. kokonaislukuja)
- Kuhunkin mahdolliseen arvoon liittyy **pistetodennäköisyys**, ts. todennäköisyys, **että sm saa juuri sen arvon**

- Tn, että *X* kuuluu johonkin **joukkoon**, saadaan todennäköisyyden additiivisuuden nojalla (ts. **laskemalla yhteen** pistetodennäköisyydet)

Matemaattista päättelyä

- Entä jos pistetodennäköisyyksiä ei tunneta?
- Harjoitus 4:9—11
  - X = suurin noppatuloksista kolmessa heitossa.
  - Varmasti X on joukossa $\{1,2,3,4,5,6\}$
  - Osataan laskea $P(X \leq 3)$ = eräs luku
    - Toisaalta tämä **on** $P(X=1) + P(X=2) + P(X=3)$
  - Osataan myös laskea $P(X \leq 4)$ = eräs luku
    - Tämä taas **on** $P(X=1) + P(X=2) + P(X=3) + P(X=4)$
  - **Nyt varmasti osataan laskea $P(X=4)$**

Eräitä diskreettejä jakaumia

- Diskreetti tasajakauma: X:llä on äärellinen arvojoukko, n kpl alkioita, joilla kullakin sama ptnf = $1/n$.

- Binomijakauma: onnistumisten määrä kolikonheitossa, tuttu ptnf binomikertoimen avulla

- Geometrinen jakauma
  - Tulkinta: Montako kertaa (toistokokeessa) joudutaan yrittämään, ennen kuin tulee yksi onnistuminen
  - Varoitus: Eri lähteissä eri käytäntöjä: Voi tarkoittaa
    (a) epäonnistuneiden yritysten määrää tai
    (b) yritysten määrää (ml. se onnistuminen), ts. 1 enemmän kuin (a)
  - Tällä kurssilla, Tuomisen kirjassa ja Matlabissa = (a)

Satunnaismuuttujan odotusarvo

Tuominen s. 54-55, 64-65, 78-

- Mahdollisten arvojen *painotettu* keskiarvo
  - painoina ko. arvojen todennäköisyydet!
- Odotusarvo ei ole ”todennäköisin arvo”
- Ei välttämättä edes mahdollinen arvo
  - Nopan odotusarvo = 3.5
- Odotusarvo kuvaa
  - Jakauman painopistettä: Mihin kohtaan reaaliakselia mahdolliset arvot **painottuvat**
  - Todennäköistä **keskiarvoa**, mikäli samoin jakautuneita satunnaismuuttujia on useita (vrt. vakuutusyhtiö)
  - Parasta mahdollista veikkausta X:n arvosta, jos väärän veikkauksen aiheuttama **haitta** (”kustannus”, ”tappio”) on verrannollinen **virheen suuruuden neliöön** (käsitellään ehkä myöhemmin)

HUOMAUTUKSIA SATUNNAISMUUTTUJISTA

Samoin jakautunut ≠ sama

- Heitetään mustaa ja valkoista noppaa.
  Tulokset *M* ja *V*, kummallakin symmetrinen jakauma joukossa {1,..,6}
  Jakaumat ovat samat – esim. P(M=3) = P(V=3) = 1/6
  Muuttujat eivät ole samat – ei ole välttämättä M=V
  (Itse asiassa todennäköisyys P(M=V) on vain 1/6)

- Heitetään yhtä noppaa kolme kertaa.
  Tulokset M_1, M_2, M_3
  Nämäkin ovat samoin jakautuneet, mutta
  eivät sama satunnaismuuttuja
  – niillähän voi olla eri arvot (ja usein onkin)

⚠ Samoin jakautunut ≠ sama

- **Yhtäsuuruusmerkillä** merkitään, että satunnaismuuttujat ovat **arvoltaan** samat

  $X = Y$
- Jos sanomme tosiasiana, että X=Y, väitämme siis, että tiedämme arvojen olevan varmasti samat (tämä ei välttämättä tarkoita, että tiedämme **mikä** se arvo on)
- Jos emme tiedä, onko X=Y, voimme ehkä esittää ko. väitettä koskevia todennäköisyyksiä: $P(X=Y)$ = jotain

- **Samoin jakautuneisuus** ilmaistaan yleensä sanallisesti, tai ilmoittamalla että kummallakin on tietty sama jakauma

  $X, Y \sim \mathrm{Bin}(10, 0.5)$ tai ”=” jonka päällä pieni *d* (lähinnä todennäköisyysteoriassa).

Esim 1: Kruunien määrät kahdessa eri toistokokeessa.
Esim 2: Kruunien ja klaavojen määrät yhdessä toistokokeessa.

Riippumattomat sm:t (Tuominen s. 67-74)

- Tunnemme käsitteen ”riippumattomat **tapahtumat**”
- **Satunnaismuuttujat** X ja Y ovat riippumattomat, jos:

  Jokainen väite muotoa ”X kuuluu tiettyyn joukkoon A” on riippumaton jokaisesta väitteestä ”Y kuuluu tiettyyn joukkoon B”

Tällöin esim. $P(X=3 \mid Y=4) = P(X=3)$

ja yleisemmin $P(X=a \mid Y=b) = P(X=a)$ kaikilla a, b

ts. tieto Y:n arvosta ei muuta käsitystä X:n arvosta. Riippumattomien sm:ien ominaisuuksiin palataan myöhemmin.

Esimerkki: *Kahden nopanheiton tulokset X ja Y.*

- **Usean** satunnaismuuttujan riippumattomuus määritellään samoin kuin usean tapahtuman, ts. tieto useastakaan muuttujasta ei anna tietoa muista

Esimerkki: *Kymmenen nopanheiton tulokset* $X_1, X_2, \ldots, X_{10}$.

Riippumattomuus ei ole sama kuin samoin jakautuneisuus

X = nopanheiton tulos ja Y = kolikonheiton tulos.

Riippumattomia, mutta **eri jakauma**

(Vrt. Elmerin tennispelit, M-peleissä eri voittotn kuin I-peleissä)

X = 1. jaetun kortin arvo ja Y = 2. jaetun kortin arvo.

Molemmilla sama jakauma, mutta **riippuvia**

Toistokokeessa tulokset $X_1,\ldots,X_n$ ovat *riippumattomat* ***ja*** *samoin jakautuneet*

Merk. **iid** ”independent and identically distributed”

SATUNNAISMUUTTUJIEN MUUNNOKSIA

Satunnaismuuttujan muunnoksia

- Satunnaismuuttujille voi tehdä monenlaisia laskutoimituksia.
- Jos X kuvaa huomista sademäärää (millimetreinä), meitä ehkä kiinnostaa 1000 neliön tontille kertyvä litramäärä = $Y = 1000 \cdot X$, joka on uusi satunnaismuuttuja.
- (mutta ei ollenkaan riippumaton X:stä, vaan **täysin** riippuva: tieto X:n arvosta määrää täsmälleen Y:n arvon)
- Toisaalta vesivoimainsinööriä kiinnostaa useiden eri paikkojen sademäärien summa = paljonko vettä tulee kertymään voimalaan yhteensä – taas uusi sm
- Muunnoksen jakauma määräytyy tavallisten todennäköisyyslaskennan sääntöjen mukaan:
  P(Y on jotain) = P(X on sellainen, että Y on jotain)

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(2/7)
- Jakaumat ovat hyvin erilaiset, mutta
  E(X) = E(Y) = E(Z) = 7/2 = 3.5

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

Matlab ja jakaumat

- Matlabin Statistics toolboxissa on valtava valikoima sm-funktioita. (ks. **help stats**)
- Funktion nimet yleisesti
  - **xxxpdf** pistetodennäköisyys (diskreeteillä sm) tai tiheysfunktio (jatkuvilla sm)
  - **xxxcdf** kertymäfunktio
  - **xxxrnd** simuloidaan lukuja ko. jakaumasta
  - **xxxstat** laskee odotusarvon (ja varianssin)
- missä **xxx** on jakauman tunnus, esim.
  - **bino** binomijakauma
  - **geo** geometrinen jakauma
  - **unif** tasajakauma (jatkuva)
  - **exp** eksponenttijakauma
  - **norm** normaalijakauma

- Lasketaan binomijakauman pistetodennäköisyys (jolla on tuttu kaava)

  binopdf(50, 100, 0.5) → 0.0796
- Simuloidaan (arvotaan) binomijakautuneita lukuja

  binornd(100, 0.5, 1, 5) → [57 48 44 51 57]
- Näillä leikkiminen myös auttaa hahmottamaan eron "satunnaismuuttujan" ja sen pistetodennäköisyyden välillä.