

Todennäköisyyskäsitteen juurista ennen valistuksen aikaa

'Tiltu' –luento 17.9.2015

Elja Arjas

Luennolla käsiteltävät asiat

- Varhaishistoriaa
- Ensimmäiset idut
- Todennäköisyyslaskennan syntyvaiheet
- Käytännön sovelluksista
- Alkutahdeista eteenpäin: 'suurten lukujen laki'
- Kohti valistuksen aikaa: induktiopäätelyn ongelma

* * *

- Luento perustuu etupäässä Ian Hackingin kirjaan '*Emergence of Probability*', Cambridge University Press (1975). (Kirjasta tehdyt lainaukset ovat vapaita ja osin myös lyhennettyjä käännöksiä.)

Varhaishistoriaa (1)

- '*talus*', lautapelit, vedonlyönti, oraakkelit, mytologia, ...
- Varsinainen todennäköisyyden käsite näyttää puuttuvan ennen 1600-luvun puoltaväliä
- Latinan sanan '*probabilis*' merkitys: 'hyväksyttävä' / 'tunnustettujen auktoriteettien suosittelu' (mielipide). Esimerkkejä: '*Such a fact is probable but undoubtedly false*', '*probable doctor*'.
- Sanan käyttö ei edellyttänyt asiasta saatua puoltavaa näyttöä ('evidence'); riitti, että jokin huomattava tai luotettavana pidetty taho oli esittänyt mahdollisen perusteen sitä koskevalle mielipiteelle (mm. jesuiittojen edustama *probabilismi* suhteessa 1500-luvulla löydettyihin vanhoihin teksteihin).

Varhaishistoriaa (2)

- Vielä vuonna 1640 englantilainen J. Wilkins (1614-1672) pohti mahdollisuutta, että kuussa olisi elämää, kirjoittamalla näin:
"That 'tis probable there may be inhabitants in this other world but of what kind they are is uncertain."
- Perustelu oli suunnilleen tällainen: Joidenkin auktoriteettien mukaan elämä kuussa ei ole mahdotonta, ja jos se kerran on mahdollista, kaikkivoipa Jumala on varmasti istuttanut elämää minne vain voi.
- Sama Wilkins esitti kuitenkin uskomattoman - mutta sittemmin toteutuneen - ennustuksen kirjoituksessaan *'Mercury, or the Secret and Swift messenger, showing how a man with privacy and speed may communicate his thoughts to a friend at a distance'* (1641).

Varhaishistoriaa (3)

- Todennäköisyyskäsitteen kasvualustana olivat 'pehmeät tieteet' ('*low sciences*'), mm. alkemia, astrologia, geologia, lääketiede, joilta puuttui toimiva tieteellinen käsitteistö.
- Niiden puuttuessa vedottiin 'merkkeihin': "*When a man hath a great disease or feebleness and a cold sweat breaketh out about the nose, that is a very deadly sign.*"
- 'Doctrine of signatures', Paracelsus (1493-1541). Kilpailevat käsitykset tautien lääkitsemisestä: lääkitys 'samanlaisilla' vs. 'vastakkaisilla'.

Varhaishistoriaa (4)

- 'Kovilla tieteenaloilla' ('*high sciences*'), mm. tähtitiede, geometria, mekaniikka, optiikka, pyrittiin suoraan demonstraatioon, johon ei tarvittu todennäköisyyden käsitettä eikä siihen pohjautuvaa matemaattista teorianmuodostusta.
- Francis Bacon (1561-1626), Galileo Galilei (1564-1642), Descartes (1596-1650).

Varhaishistoriaa (5)

- Sama asenne säilyi fysiikassa pitkään: *“If your experiment needs statistics, you ought to have done a better experiment.”* (Ernest Rutherford, 1871-1937)
- Nyt tilanne on - ainakin osittain - toinen: Moderni fysiikka, erityisesti kvanttimekaniikka, edellyttää todennäköisyyskäsitteen käyttöä. Vastaavasti tilastollisten menetelmien merkitys korostuu laajojen fysikaalisten mittausaineistojen analyysissä esimerkiksi ilmastotutkimuksessa ja astrofysiikassa.
- Katso myös: 'Higgs boson' verkkosivulla <http://www.tonyohagan.co.uk/academic/>

Ensimmäiset idut (1)

- Herätteenä todennäköisyyttä koskeville tarkasteluille mm. vanha ja ratkaisemattomana pysynyt 'keskeytetyn pelin' ongelma:
"Kahden joukkueen pallopelissä voittoon tarvitaan 60 pistettä, ja jokaisesta maalista saa 10 pistettä. Voitosta saa palkintona 10 dukaattia. Peli keskeytyy tilanteessa, jossa toinen joukkue on saanut 50 pistettä ja toinen 30 (20?) pistettä. Kuinka paljon palkintopotista kuuluu silloin kummallekin joukkueelle." (Luca Pacioli 1494).
- Ratkaisuyrityksiä mm. Tartaglia (1556) ja Peverone (1558).

Ensimmäiset idut (2)

- Toinen ongelmaryhmä liittyi kombinatoriikkaan, esimerkkinä:
Pitkäaikainen kokemus osoittaa, että heitettäessä kolmea noppaa, kannattaa paremmin veikata silmälukujen summaksi lukuja 10 ja 11 kuin lukuja 9 ja 12.
- Galilei esitti ongelmalle kombinatoriikkaan perustuvan ratkaisun. Ratkaisun avain oli 'oikean' symmetriaperiaatteen valinta.
- Huom. Yhteys 1900-luvun hiukkasfysiikan *Maxwell-Boltzmann*, *Bose-Einstein* ja *Fermi-Dirac* -symmetriaperiaatteisiin.

Ensimmäiset idut (3)

- Näissä tarkasteluissa tulee esiin todennäköisyyskäsitteen dualismi: jako *episteemiseen* ja *aleatoriseen* tulkintaan.
- Galilein päättely edellä liittyi ilmeisellä tavalla eri tulosvaihtoehtojen suhteellisiin frekvensseihin.
- Toisaalta jako-ongelmassa pyrkimyksenä on arvioida - ja verrata toisiinsa - niitä eri mahdollisuuksia, joilla peli olisi voinut jatkua ja lopuksi johtaa peliin sijoitettujen panosten voittamiseen tai häviämiseen.

Ensimmäiset idut (4)

- Jo noin sata vuotta ennen Galileita Cardano (1501-1576) oli kirjoittanut:
”Pystyn yhtä hyvin heittämään nopalla tuloksen 1 tai 3 tai 5 kuin tuloksen 2 tai 4 tai 6. Jos noppa on kunnollinen, vedonlyönnissä tulee noudattaa näitä suhteita, ja jos se ei ole, niiden tulee vastata poikkeamaa yhtäsuuruudesta.”
- Vrt. Karl Popperin (1902-1994) esittämä **propensiteetin** käsite: (tässä nopan objektiivinen fysikaalinen) ’taipumus tuottaa pitkissä koesarjoissa stabiileja suhteellisia frekvenssejä.’
- Huom. Pierre-Simon Laplace (1749-1827), ja 1900-luvun kuluessa mm. J. M. Keynes, F. P. Ramsey, H. Jeffreys, B. de Finetti, J. L. Savage, E. T. Jaynes ja D. Lindley ovat puolustaneet todennäköisyyskäsitteen **episteemistä** tulkintaa. Heidän suhtautumisensa **loogisen/objektiivisen** todennäköisyyden mahdollisuuteen kuitenkin vaihteli.

Todennäköisyyslaskennan syntyvaiheet (1)

- Leibnizin myöhemmin (1676) antaman kuvauksen mukaan:

”Chevalier de Méré, joka oli sekä uhkapeluri että filosofi, antoi matemaatikoille ajankohtaista askarreltavaa esittämällä heille joitakin kysymyksiä. Niiden tarkoituksena oli saada selville, miten peliin sijoitetut panokset olisi jaettava pelaajien kesken tilanteessa, jossa peli olisi keskeytynyt sen jossakin vaiheessa. Hänen onnistui houkutella ystävänsä Pascal tutkimaan näitä asioita.

Ongelmasta tuli yleisesti tunnettu ja Huygens kirjoitti siitä kirjan nimeltä 'De Aleae'. Muutkin oppineet kiinnostuivat tästä aiheesta. Joitakin aksioomia asetettiin. Eläkeläinen de Witt käytti niitä pienessä hollanniksi painetussa kirjassaan, joka käsitteli elinkoron laskemista.”

Todennäköisyyslaskennan syntyvaiheet (2)

- Blaise Pascal (1623-1662) oli kertomuksen varsinainen sankari.
- Esimerkki de Méré'n Pascalille esittämistä ongelmista:
'Kuinka monta kertaa kahta noppaa pitää heittää, jotta 'kahden kuutosen' esiintyminen ainakin kerran olisi vähintään yhtä mahdollista kuin se, ettei niin tapahdu?'
- De Méré'n käsityksen mukaan vastaus olisi ollut 24, Pascalin antama vastaus oli 25. De Méré'n mukaan ongelman eri ratkaisut johtivat 'aritmeettiseen ristiriitaan'.

Todennäköisyyslaskennan syntyvaiheet

(3)

- Pascal on tunnettu mm. seuraavista todennäköisyyslaskentaan liittyvistä tarkasteluista:
 - 'Pascalin kolmio' (binomikertoimet),
 - 'Pascalin veto' (*Pensées*, ilm. 1670, suom. *Mietteitä*), rationaalinen perustelu uskoa (kristillisen uskonkäsityksen mukaisen) Jumalan olemassaoloon: *infini – rien*; ensimmäinen tunnettu esimerkki päätösteoreettisesta tarkastelusta.

Todennäköisyyslaskennan syntyvaiheet

(4)

- Pierre de Fermat'n (1601-1665) osuus todennäköisyyslaskennan kehityksessä lienee ollut Pascalia vähäisempi. Hyvin tunnettu on joka tapauksessa hänen kirjeenvaihtonsa Pascalin kanssa, jossa tämä oli pyytännyt 'tarkistamaan tehtyjen laskujen pätevyuden.'

(Huom. Englanninkielisessä Wikipediassa sanotaan, että ratkaisun em. 'kahden kuutosen ongelmaan' esitti nimenomaan Fermat. Hackingin mukaan ratkaisija oli Pascal.)

Käytännön sovelluksista (1)

- Mielenkiintoinen – ja myös aikanaan käytännön kannalta tärkeä – ongelma oli koettaa löytää järkevä peruste ns. elinkoron määrittämiselle. Tässä järjestelyssä, joka oli eräänlainen käänteinen vakuutus, henkilö maksoi tietyn summan rahaa esimerkiksi kaupungille, joka vastineeksi sitoutui maksamaan maksajan elatuksen tämän kuolemaan saakka (kts. esim. <http://fi.wikipedia.org/wiki/Elinkorko>).
- Ongelman lähinnä aleatoriseksi tulkittavaa komponenttia edusti tällöin henkilön jäljellä oleva elinaika, ts. elinkoron maksuvuosien lukumäärä, episteemistä taas kyseisenä ajankohtana ja lähitulevaisuudessa vallitsevan korkotason ja kuolleisuuden arviointi. Toisaalta molempia kysymyksiä voidaan hyvin tarkastella pelkästään episteemiseltä kannalta.

Käytännön sovelluksista (2)

- Elinkorkojen määrittäminen tapahtui pitkään puhtaasti 'mutu'-perusteilla. Järjestelmää käytettiin yhtenä varainkeruun muotona, mutta epärealistiset laskelmat maksuista ja velvoitteista johtivat joskus merkittäviin tappioihin.
- Systemaattisia tietoja kuolleisuudesta alettiin kerätä Lontoossa v. 1603 riehuneen ruttoepidemian aikana. Näitä tietoja ei kuitenkaan ryhdytty käyttämään ennen 1660-lukua, jolloin John Graunt (1620-1674) ja William Petty (1623-1687) ryhtyivät selvittämään asiaa. He eivät tällöin olleet tietoisia esimerkiksi Pascalin ja Huygensin esittämistä uusista matemaattisista menetelmistä.

Natural and Political
OBSERVATIONS
Mentioned in a following INDEX,
and made upon the
Bills of Mortality.

By *JOHN GRAUNT*,
Citizen of
LONDON.

With reference to the Government, Religion, Trade,
Growth, Age, Diseases, and the several Changes of the
said City.

— Non, nisi per Turis, labes.
Contracta pariter Libentia —

LONDON,
Printed by Tho: Keycroft, for John Atterlie, James Allcock,
and Tho: Dineen, at the Sign of the Bell in St. Paul's
Church-yard, MDCLXII.

Alkutahdeista eteenpäin: 'suurten lukujen laki' (1)

- Jacques Bernoullin (1654-1705) teoksessaan *Ars conjectandi* (ilm. 1713) todistama tulos, joka nykyään tunnetaan nimellä 'heikko suurten lukujen laki', on ensimmäinen varsinainen todennäköisyys-laskennan piiriin kuuluva teoreema.
- Tuloksen todistus edellyttää oletuksen, jonka mukaan 'toistettaessa olennaisesti samaa koetta useaan kertaan ja toisistaan (fysikaalisesti) riippumatta' mahdollisuus saada toinen kahdesta vaihtoehtoisesta koetuloksesta ('onnistuminen', merk. 'S') pysyy vakiona.

Kohti valistuksen aikakautta: induktiopäättelyn mahdollisuudesta (1)

- **Induktiivinen päättely** eli **induktio** lähtee liikkeelle yksittäisestä havaintojoukosta ja muodostaa niistä yleistyksen tai teorian.
- Arjen keskellä teemme induktiivisia yleistyksiä jatkuvasti. Oletus, jonka mukaan aurinko nousee huomenna on induktiivinen, usein tiedostamaton yleistys aikaisemmasta kokemuksestamme.

Kohti valistuksen aikakautta: induktiopäättelyn mahdollisuudesta (2)

- Skotlantilainen filosofi David Hume (1711-1776) esitti teoksessaan '*A Treatise of Human Nature*' (1739) induktiopäättelyn olevan mahdotonta. Hänen mukaansa siitä, mitä on ollut aikaisemmin, ei voi mitenkään päätellä, mitä tulee olemaan.
- Thomas Bayesin (1702?-1761) kirjoitus '*An essay towards solving a problem in the doctrine of chance*' (1763, Richard Pricen editoimana) voidaan nähdä vastauksena Hume'n kritiikkiin.
- Vähän myöhemmin Laplace käytti induktiopäätelyä menestyksellisesti mm. soveltaessaan Newtonin mekaniikan lakeja planeettojen liikeratoja koskeviin ongelmiin.

Lopuksi

- **Tilastotiede** koostuu systemaattisista menetelmistä, joiden avulla pyritään tekemään havaintoihin perustuvaa induktiivista päättelyä.
- **Todennäköisyyslaskenta** muodostaa tilastotieteen matemaattisen ytimen. Tämä korostuu erityisesti ns. Bayes-päättelyssä, jossa kaikki matemaattiset laskutoimitukset perustuvat todennäköisyyslaskentaan.
- Todennäköisyyslaskennan käyttö tilastotieteessä ei kuitenkaan ole luonteeltaan puhtaasti deduktiivista päättelyä, vaan siihen sisältyy merkittävä tulkinnallinen osuus. Erityisen tärkeää tässä suhteessa on asiayhteyteen soveltuvan **tilastollisen mallin** perusteltu valinta. Eri lähtökohdista voidaan hyvin päätyä erilaisiin lopputuloksiin. Tästä syystä (edes Bayes-paradigman mukainen) tilastotiede ei ole 'vain todennäköisyyslaskentaa'.