12.1. Tilastollinen todennäköisyys

Luvut 6 – 11 ovat kuuluneet klassisen todennäköisyyden kenttään. Klassisen todennäköisyyden teorian soveltaminen edellytti, että ilmiöllä oli joukko tulosmahdollisuuksia eli alkeistapauksia, joista jokainen oli yhtä mahdollinen.

Klassinen todennäköisyys ei kata koko todennäköisyyslaskennan aluetta, sillä toistaiseksi ei ole esitetty sellaisia kysymyksiä, kuten esimerkiksi

- millä todennäköisyydellä MB-merkkinen henkilöauto joutuu korvausvastuulliseen liikenneonnettomuuteen kalenterivuoden aikana
- millä todennäköisyydellä 2 v täyttänyt Saimaan norppa täyttää 5 v
- millä todennäköisyydellä rahtialus haaksirikkoutuu alkavalla maailmanympärysmatkallaan?

Esitetyn tyyppiset kysymykset kuuluvat tilastollisen (empiirisen) todennäköisyyden kenttään.

****

MÄÄRITELMÄ 8

Tapahtuman A tilastollisella todennäköisyydellä tarkoitetaan n:n toistokokeen sarjassa sitä raja-arvoa, jota tapahtuman A suhteellinen frekvenssi lähenee, kun n rajattomasti kasvaa eli

$$P(A) = \lim_{n\to\infty} \frac{f_A}{n}$$

****

Matemaattisesti täsmällisen lähtökohdan todennäköisyydelle muotoili ensimmäisenä, jos historiatietoon voi luottaa, venäläinen Andrei Kolmogorov vuonna 1933. Kuten tasogeometriassa lähdetään liikkeelle tietyistä aksiomeista, niin menetellään myös todennäköisyyslaskennan aksiomaattisessa mallissa. Eihän esimerkiksi geometriassa määritellä käsitteitä piste, suora ja taso kovinkaan tarkasti, mutta luultavasti hyvin monet joitakin vuosia kouluja käyneet mieltävät, mistä on kyse.

Todennäköisyyslaskennan aksiomaattisessa mallissa sanotaan alkeistapausten $a_1, a_2, a_3, \dots, a_n$ muodostamaa joukkoa otosavaruudeksi E, jonka jokainen osajoukko A on tapahtuma. Kuhunkin tapahtumaan A liitetään määrätty reaaliluku funktion P(A) välityksellä asettaen määritelmä, jossa olevat joukko-opilliset symbolit tarkoittavat seuraavaa:

$A \subseteq B$, joukko A on joukon B osajoukko (mahdollisesti A = B).
$A \cap B = \varnothing$, A:n ja B:n leikkaus on tyhjä, eli joukot ovat erilliset
$P(A \cup B) = P(A) + P(B)$, A:n ja B:n unionin, yhdisteen todennäköisyys on A:n ja B:n todennäköisyyksien summa, jos joukot A ja B ovat erilliset.

****

MÄÄRITELMÄ 9

Olkoon E otosavaruus ja P joukkofunktio, joka liittää jokaiseen E:n osajoukkoon A reaaliluvun P(A). Tämä luku, siis P(A), on tapahtuman A todennäköisyys, jos

- $0 \leq P(A) \leq 1$
- Jos $A \subseteq E$ ja $B \subseteq E$ ja $A \cap B = \varnothing$, niin $P(A \cup B) = P(A) + P(B)$
- P(E) = 1

****

Määritelmässä ei puhuta alkeistapausten todennäköisyydestä mitään. Määritelmä sallii vaikka sen, että 6 − kanttisen nopan heitossa silmälukujen todennäköisyyksiksi voidaan valita mitkä tahansa ei-negatiiviset reaaliluvut p_1, p_2, p_3, p_4, p_5 ja p_6 joiden summa on 1. Mitkä arvot kulloinkin valitaan, riippuu nopasta ja yleensä tilanteesta; lukujen tulee vastata todellisuutta.

Esim. 1 Markolla oli norsunluinen noppa, johon oli asetettu metallinen paino. Heitettäessä tällä nopalla 3000 heittoa, saatiin tulokset

x	1	2	3	4	5	6
f	149	330	341	333	341	1506

a) Heitetään tätä noppaa kerran. Kiinnitä todennäköisyydet P(1), P(3) ja P(6)
b) Millä todennäköisyydellä yksittäisen heiton tulos on parillinen ja millä pariton?
c) Millä todennäköisyydellä neljän heiton sarjassa saadaan ainakin kolme kuutosta?

a) On luonnollista määritellä ao. nopan yhden heiton todennäköisyydet tulosten suhteellisina frekvensseinä, koska heittojen lukumäärä n on suurehko:

$$P(1) = \frac{149}{3000} = 0.04966..... \approx 0.05$$

$$P(3) = \frac{341}{3000} = 0.113666.... \approx 0.114$$

$$P(6) = \frac{1506}{3000} = 0.502$$

b) P(parillinen) = P(2) + P(4) + P(6) = $\frac{330}{3000} + \frac{333}{3000} + 0.502 \approx 0.723$

P(pariton) = 1– P(parillinen) ≈ 1 – 0.723 = 0.277.

c) Olkoon kuutosten lukumäärä k. Binomitodennäköisyyden avulla

$$P(k = 3 \text{ tai } k = 4) = P(k = 3) + P(k = 4) = \binom{4}{3} \cdot 0.502^3 \cdot 0.498 + 0.502^4 =$$

$$= 0.25199... + 0.0635... \approx 0.316.$$

Esim. 2 Suomen 1950-luvun kuolleisuus- ja elinikätilastojen perusteella on laadittu oheinen taulukko, josta ilmenee, kuinka moni tuhannesta elä-vänä syntyneestä poikalapsesta tulee elämään yli 20- , 40- , 60- tai 80-vuotiaaksi.

ikä	0	20	40	60	80
elossa	1000	940	890	690	170
f/n	1	0.94	0.89	0.69	0.17

Vastaa tilaston perusteella, millä todennäköisyydellä

a) vastasyntynyt poikalapsi elää yli 40-vuotiaaksi
b) 20-vuotta täyttänyt mies elää yli 80-vuotiaaksi
c) 40-vuotias mies kuolee ennen kuin täyttää 60 vuotta.

a) P(vastasynt. täyttää 40 v) = $\dfrac{\text{elossa olevat 40} - \text{vuotiaat}}{\text{vastasyntyneet}} = \dfrac{890}{1000} = 0.89$.

b) Perusjoukkona nyt 20 v. täyttäneiden lukumäärä 940 ja kun 80 ikäisiä on 170, niin P(20 v elää 80 vuotiaaksi) = $\dfrac{170}{940} = 0.18085.... \approx 0.18$

c) Perusjoukkona nyt 40 täyttäneet (890), joista 690 täyttää 60 v, joten P(40-vuotias kuolee ennen 60 v ikää) = 1 − P(täyttää 60) = $1 - \dfrac{690}{890} = 0.2247.... \approx 0.22$.