

# 1 Multinomikoe

Kun toistokokeessa on joka kerralla 2 vaihtoehtoa (“onnistuminen” ja “epäonnistuminen”), koko koesarjan onnistumisten lukumäärä on binomijakautunut, ja sen pistetodennäköisyydet saadaan binomikertoimien avulla.

Kun vaihtoehtoja on enemmän kuin 2, saadaan samaan tapaan lukumäärille ns. multinomijakauma, jonka pistetodennäköisyydet saadaan multinomikertoimien avulla.

**Esimerkki.** Suuressa populaatiossa on kolmen puolueen A, B ja C kannattajia osuuksin  $p = 0.5$ ,  $q = 0.3$  ja  $r = 0.2$ . Poimitaan populaatiosta umpimähkään  $n = 10$  henkilön otos. Millä todennäköisyydellä saadaan otos, jossa puolueiden kannattajien lukumäärät ovat  $a$ ,  $b$  ja  $c$  (missä  $a + b + c = 10$ )?

**Ratkaisu.** Kun populaatio on suuri ja otos pieni, voimme ajatella otannan tapahtuvan “takaisinpanolla” eli otoksen kukin henkilö on, toisista riippumatta, puolueen A kannattaja todennäköisyydellä  $p$ , puolueen B tn:llä  $q$  ja puolueen C tn:llä  $r$ . Ajatellaan otosta 10 henkilön jonona, jossa tarkastellaan vain heidän puoluekantojaan. Jonosta voidaan sitten laskea lukumäärät, jotka ovat satunnaismuuttujia  $(N_A, N_B, N_C)$ .

Esimerkiksi jonon  $(A, A, A, A, A, A, A, A, A, A)$  (eli 10 A:ta) todennäköisyys on riippumattomuuden nojalla tulo

$$p \times p \times p \times p \times p \times p \times p \times p \times p \times p = p^{10}.$$

Toisaalta esim. jonon  $(A, A, A, B, B, B, C, C, C, C)$  (jossa on 3 A:ta, 3 B:tä ja 4 C:tä) tn on riippumattomuuden nojalla

$$p \times p \times p \times q \times q \times q \times r \times r \times r \times r = p^3 q^3 r^4.$$

Mutta tapahtuma  $(N_A = 3 \wedge N_B = 3 \wedge N_C = 4)$  (ts. että saadaan tällaiset lukumäärät) voi tapahtua monessa eri järjestyksessä; järjestysten määrä on multinomikerroin

$$\binom{10}{3, 3, 4} = \frac{10!}{3!3!4!}.$$

Tapahtuman yksittäiset jonot ovat keskenään yhtä todennäköiset (vrt. toistokoe), joten sen tn saadaan tulona

$$P(N_A = 3 \wedge N_B = 3 \wedge N_C = 4) = \binom{10}{3, 3, 4} p^3 q^3 r^4.$$

Vastaavalla päättelyllä todetaan yleisemmin

$$P(N_A = a \wedge N_B = b \wedge N_C = c) = \binom{n}{a, b, c} p^a q^b r^c. \quad (1)$$

(Samankaltaisella kaavalla voidaan laskea todennäköisyydet myös tilanteessa, jossa on enemmän kuin 3 vaihtoehtoa.)

Mahdollisia lukumääriä ei 10 henkilön otoksessa ja 3:lla puolueella ole kovin monta:  $a$  on jokin luvuista  $0, \dots, 10$ , ja kun  $a$  on valittu, niin  $b$  on jokin luvuista  $0, \dots, (10 - a)$ . Sitten  $c$ :ssä ei olekaan enää valinnanvaraa, vaan  $c = 10 - a - b$ . Seuraavassa taulukossa on kaikki 66 vaihtoehtoa lueteltu, siten että todennäköisyydet on järjestetty laskevaan järjestykseen.

$a$	$b$	$c$	$P$	$a$	$b$	$c$	$P$
5	3	2	0.0850500	2	7	1	0.0039366
6	2	2	0.0708750	9	0	1	0.0039063
6	3	1	0.0708750	3	7	0	0.0032805
4	4	2	0.0637875	5	0	5	0.0025200
5	4	1	0.0637875	1	6	3	0.0024494
5	2	3	0.0567000	1	5	4	0.0024494
4	3	3	0.0567000	3	1	6	0.0020160
7	2	1	0.0506250	2	2	6	0.0018144
4	5	1	0.0382725	1	4	5	0.0016330
3	4	3	0.0340200	1	7	2	0.0015746
7	1	2	0.0337500	10	0	0	0.0009766
6	1	3	0.0315000	4	0	6	0.0008400
3	5	2	0.0306180	2	8	0	0.0007381
4	2	4	0.0283500	1	3	6	0.0007258
6	4	0	0.0265781	1	8	1	0.0005905
7	3	0	0.0253125	2	1	7	0.0003456
3	3	4	0.0226800	0	6	4	0.0002449
8	1	1	0.0210938	0	7	3	0.0002100
5	5	0	0.0191362	1	2	7	0.0002074
5	1	4	0.0189000	0	5	5	0.0001960
8	2	0	0.0158203	3	0	7	0.0001920
3	6	1	0.0153090	0	8	2	0.0001181
2	5	3	0.0122472	0	4	6	0.0001089
2	4	4	0.0102060	1	9	0	0.0000984
4	6	0	0.0095681	0	3	7	0.0000415
2	6	2	0.0091854	0	9	1	0.0000394
3	2	5	0.0090720	1	1	8	0.0000346
4	1	5	0.0075600	2	0	8	0.0000288
7	0	3	0.0075000	0	2	8	0.0000104
8	0	2	0.0070313	0	10	0	0.0000059
9	1	0	0.0058594	1	0	9	0.0000026
2	3	5	0.0054432	0	1	9	0.0000015
6	0	4	0.0052500	0	0	10	0.0000001

Suurimman todennäköisyyden saivat (ei kovin yllättäen) lukumäärät (5,3,2), jolloin puolueiden osuudet otoksessa ovat samat kuin koko populaatiossa. Kuitenkin myös tästä hiukan poikkeavat lukumäärät ovat lähes yhtä todennäköisiä.

Jos meitä kiinnostaa tapahtuman ( $N_A = 5$ ) todennäköisyys, voisimme summata taulukosta kaikki ne rivit, joissa  $a = 5$ . Toisaalta sama voidaan laskea yksinkertaisemmin binomikoikeena (kahden vaihtoehdon toistokokeena): ajatellaan puoluetta A “onnistumisena” ja muita “epäonnistumisena”. Onnistumistodennäköisyys on  $p = 0.5$  ja siten todennäköisyys saada 5 onnistumista on

$$\binom{10}{5} p^5 (1-p)^5 \approx 0.2461.$$

Seuraavalla sivulla on esimerkin vuoksi MATLAB-ohjelma, jolla taulukon voi laskea.

```

% MATLAB-ohjelma: Esimerkki otosuosuuksista multinomikokeessa

function puolueotokset

n = 10;
p = 0.5;
q = 0.3;
r = 0.2;

taulukko = [];

% Käydään läpi (a,b,c)-parit joiden summa=10
for a=0:10
    for b=0:(n-a)
        c = n-a-b;

        % Lasketaan multinomitodennäköisyys
        P = multinomi(n,a,b,c) * p^a * q^b * r^c;

        % Lisätään rivi taulukkoon
        taulukko = [taulukko; P a b c];
    end
end

% Järjestetään laskevaan tn-järjestykseen
taulukko = flipud(sortrows(taulukko));

% Tulostetaan
for i=1:size(taulukko,1)
    fprintf('%.7f   %2d %2d %2d\n',...
        taulukko(i,1),taulukko(i,2),taulukko(i,3),taulukko(i,4));
end
summa = sum(taulukko(:,1));
fprintf('-----\n');
fprintf('%.7f\n',summa);

% Apufunktio: Multinomikerroin
function M=multinomi(n,a,b,c)
M = factorial(n) / (factorial(a)*factorial(b)*factorial(c));

```