Johdatus todennäköisyyslaskentaan
Kevät 2015

Luento 5 / 13

Jukka Kohonen

Matematiikan ja tilastotieteen laitos

Helsingin yliopisto

KOMBINATORIIKKA
SUURTEN LUKUMÄÄRIEN LASKEMISEN TAITO

Kombinatoriikkaa

- Symmetrisessä mallissa tapahtuman tn on helppo laskea, jos osataan laskea suotuisat alkeistapaukset.
- Joskus, varsinkin yhdistetyssä satunnaiskokeessa, alkeistapauksia on **hyvin monta** (tuhansia, miljardeja, enemmän).
- Kombinatoriikassa lasketaan näitä suuria lukumääriä.

Kombinatoriikkaa

Kymmenen henkilöä, numeroituna 1,…,10, asettuu jonoon umpimähkään. Millä tn he ovat aakkosjärjestyksessä?

Alkeistapaus on tietty järjestys, esim.

(2, 5, 1, 10, 3, 4, 7, 9, 5, 8).

Eri alkeistapauksia on hyvin monta. **Luetteleminen** ei kannata.

Osaamme kuitenkin **laskea**, että niitä on 10! = 3 628 800 kappaletta.

Isot symmetriset tn-avaruudet

Pääkysymykset ovat:

1. Mitkä ovat **alkeistapaukset**? Montako niitä on?
   - voiko ne esim. **luetella** kaikki, tai
   - voiko **kuvitella** niiden luettelemisen jonkinlaisen järjestyksen tai valintaprosessin kautta
2. Mitkä ovat tapahtuman **suotuisat alkeistapaukset**? Millaisia ne ovat? Montako niitä on?

Kun näihin on vastattu, niin tapahtuman *A* tn on

$$\frac{n(A)}{n(\Omega)} \quad \begin{array}{l} \leftarrow \textbf{suotuisien alkeistapausten määrä} \\ \leftarrow \textbf{kaikkien alkeistapausten määrä} \end{array}$$

Helpommin sanottu kuin tehty

Perusjoukko voi olla melko suuri tai **hyvin suuri**

- Kahden erilaisen kolikon heittotuloksia 4 kpl

  $\{(0,0), (0,1), (1,0), (1,1)\}$

- Kahden erilaisen nopan heittotuloksia 36 kpl

  $\{(1,1), (1,2), ..., (1,6), (2,1), (2,2), ..., (6,6)\}$

- Kolmen erilaisen nopan heittotuloksia 216 kpl

  ei jaksa luetella

- Korttipakan järjestyksiä $52! \approx 8 \cdot 10^{67}$

  enemmän kuin maapallossa on atomeita

Tuloperiaate

- Tarkasteltavat oliot (alkeistapaukset) voidaan muodostaa tekemällä ***k* peräkkäistä valintaa**
- *i*:nnessä vaiheessa valitaan n_i:stä vaihtoehdosta siten, että **lukumäärä ei riipu siitä aiemmin valittiin**

→ Erilaisia olioita on
$n_1 \cdot n_2 \cdot \ldots \cdot n_k$ kappaletta.

Kolmen nopan heitto

- Alkeistapauksina järjestetyt jonot (a,b,c), missä a, b ja c ovat kokonaislukuja joukosta {1,2,3,4,5,6}
  - 1. luku voidaan valita 6 tavalla $n_1 = \mathbf{6}$
  - 2. luku voidaan valita 6 tavalla $n_2 = \mathbf{6}$
  - 3. luku voidaan valita 6 tavalla $n_3 = \mathbf{6}$
- Alkeistapauksia on $6 \cdot 6 \cdot 6 = 216$ kpl

- Tässä oli $n_1 = n_2 = n_3$

  eli joka vaiheessa oli **sama määrä vaihtoehtoja.**

  Kertolaskun voisi laskea suoraan potenssina 6^3.

Kolmen ihmisen jono (permutaatio, s. 23)

- A, B, C asettuvat jonoon. Kukin ihminen on jonossa vain yhdessä kohtaa! Jono (A,A,A) eli AAA on mahdoton.
- Tuloperiaate, mutta **vaihtoehdot vähenevät** vaihe vaiheelta (joka kerta 1:llä)
- 1. paikassa voi olla A, B tai C $n_1 = \mathbf{3}$
- 2. paikassa jompikumpi jäljelle jääneistä $n_2 = \mathbf{2}$
- 3. paikassa on viimeiseksi jäänyt $n_3 = \mathbf{1}$
- Jonoja on 3·2·1 = 6 erilaista. Luetellaan ne:

{ABC, ACB,
BAC, BCA,
CAB, CBA}

Osajono (variaatio, Tuominen s. 26)

Kymmenestä ihmisestä ABCDEFGHIJ kolme tulee huoneeseen jonossa.

Montako erilaista jonoa on?

- 1. ihminen on joku 10:stä
- 2. ihminen on joku 9:stä (ei sama kuin 1. ihminen)
- 3. ihminen on joku 8:sta (ei kumpikaan aiemmista kahdesta)

Laskeva tulo $10 \cdot 9 \cdot 8 = (10)_3 = 720$

= alkaa 10:stä, kolme tekijää, pienenee aina 1:llä

Huom 1. Kussakin vaiheessa valinta<u>vaihtoehdot</u> riippuvat siitä, mitä aiemmin valittiin, mutta vaihtoehtojen <u>määrä</u> ei riipu

Huom 2. Jos osajonoon tulee kaikki 10 ihmistä, saadaan permutaatio

Vertaa alkeistapauksia

	3 noppaa	**3 ihmisen jono (permutaatio)**	**3 ihmisen osajono 10:stä (variaatio)**
Vaihtoehtoja vaiheittain	6 · 6 · 6	3 · 2 · 1	10 · 9 · 8
Lyhyempi merkintä	6^3 **potenssi**	$(3)_3 = 3!$ **kertoma**	$(10)_3$ **laskeva tulo**
Yhteensä	216 kpl	6 kpl	720 kpl
Alkeistapausten luettelo = perusjoukko	{111, 112, 113, ... 121, 122, 123, ... 211, 212, ... 611, ..., 666}	{ABC, ACB, BAC, BCA, CAB, CBA}	{ABC, ABD, ABE, ..., ABJ, ACB, ACD, ..., ACJ, ... BAC, BAD, ... DAB, ..., EAB, ..., JIH}

Entäs ne suotuisat

- **Perusjoukon** koon laskeminen on usein helppoa.
  - Usein joku kombinatoriikan perusmenetelmä (permutaatiot tms.)

- **Suotuisien** alkeistapausten laskeminen voi olla ”hiukan” vaikeampaa.
  - Pitää olla tarkkana, että kaikki mahdollisuudet tulee laskettua (ja jokainen vain kerran).
  - Usein ei onnistu suoraan yhdellä lausekkeella (kertoma, binomikerroin tms) vaan joudutaan käyttämään tuloperiaatetta ja (joskus hyvinkin yksityiskohtaista) päättelyä ja vaihtoehtojen läpikäyntiä.

Kaksi noppaa, P(molemmat parillisia)?

- **Alkeistapaukset** = 2-jonot joukosta {1,...,6}, niitä on 36 kpl: {(1,1),(1,2),...,(6,5),(6,6)}
- **Suotuisat** alkeistapaukset: Eka noppa on joku joukosta {2,4,6}, samoin toinen, suotuisia ovat siis {(2,2),(2,4),(2,6),...,(6,6)}
  - **Tuloperiaate**, suotuisia on 3·3 = 9 kpl

- Todennäköisyys = **9 / 36** = 1/4

Kaksi noppaa, P(summa = 6)?

- Samat 36 alkeistapausta kuin edellä
- **Suotuisia** ovat {(1,5),(2,4),(3,3),(4,2),(5,1)} eli **5** kpl
- Todennäköisyys = 5/36

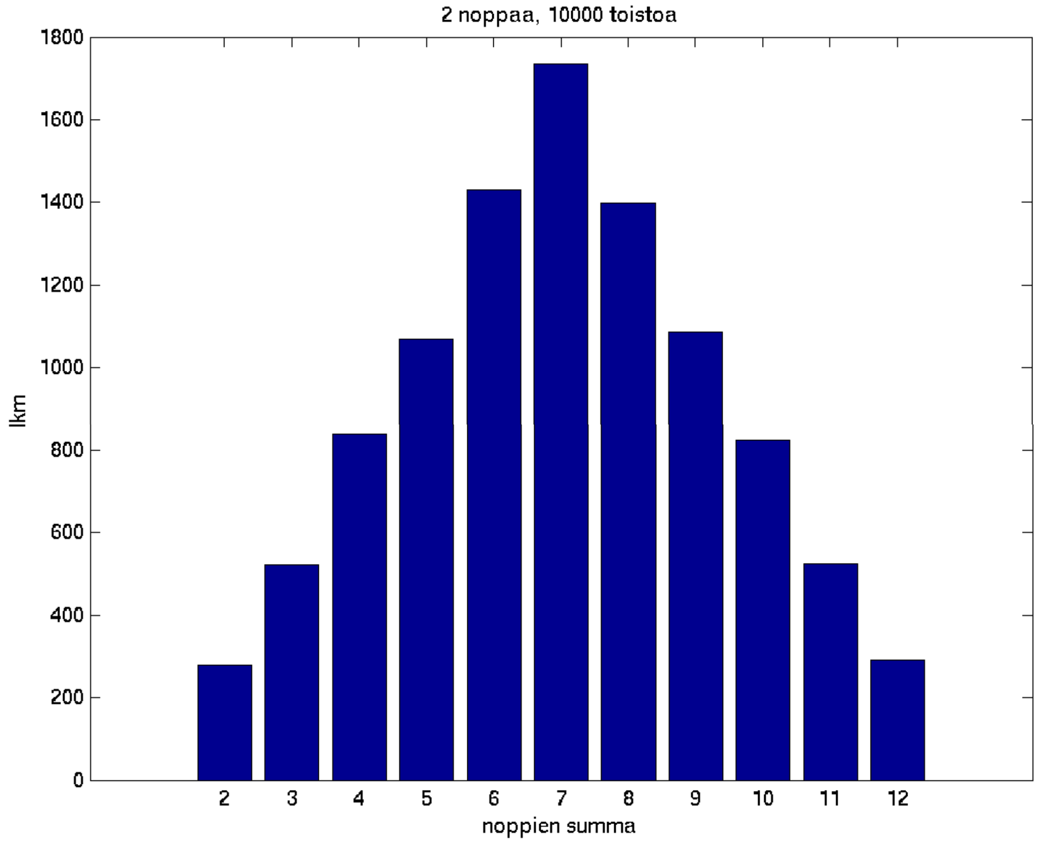
- Huom. Noppien summa on jokin 11 luvusta {2,3,4,...,12}, mutta nämä mahdollisuudet eivät ole keskenään yhtä todennäköiset!

Kaksi noppaa, eri summien tn:t

Noppien summa	Suotuisat alkeistapaukset	Suotuisia kpl	tn
2	”11”	1	1 / 36
3	”12”, ”21”	2	2 / 36
4	”13”, ”22”, ”31”	3	3 / 36
5	”14”, ”23”, ”32”, ”41”	4	4 / 36
6	”15”, ”24”, ”33”, ”42”, ”51”	5	5 / 36
7	”16”, ”25”, ”34”, ”43”, ”52”, ”61”	6	6 / 36
8	”26”, ”35”, ”44”, ”53”, ”62”	5	5 / 36
9	”36”, ”45”, ”54”, ”63”	4	4 / 36
10	”46”, ”55”, ”64”	3	3 / 36
11	”56”, ”65”	2	2 / 36
12	”66”	1	1 / 36
yhteensä		36	1

Kaksi noppaa, summan kokeellinen jakauma

```
n = 10000;
a = noppa(n);
b = noppa(n);
s = a+b;
pylvas(s)
```



10 ihmisen jono, onko A **ennen** B:tä?

- Jonoja on 10! = 3.6 miljoonaa kpl, ikävä läpikäydä yksitellen
- Millaisia ovat suotuisat jonot?
- Jonossa 10 paikkaa.
  - A on jossain 10 paikasta
  - B on jossain, jäljellä 9 paikkaa, **mutta niistä vain osa on suotuisia**
  - Muut 8 ovat jossain (jäljellä 8, 7, ..., 1 paikkaa)
- Jatkuu ...

Onko A ennen B:tä (jatkuu)

- **Jos** A on 1. paikalla, B saa olla missä vain (9 vaihtoehtoa)
- **Jos** A on 2. paikalla, B saa olla paikalla 3...10 (8 vaihtoehtoa)
- ...
- **Jos** A on 9. paikalla, B täytyy olla 10. paikalla (1 vaihtoehto)
- **Jos** A on 10. paikalla, B ei voi olla missään (0 vaihtoehtoa)

Summaperiaate: A:n ja B:n paikat voidaan valita 9+8+7+6+5+4+3+2+1+0 = 45 tavalla.

Sen jälkeen loppujen 8 ihmisen paikat mielivaltaisesti 8! = 40320 tavalla.

Suotuisia jonoja (tuloperiaate) on 45 · 40320 = 1814400kpl

Todennäköisyys = 1814400 / 10! = **0.5**

Olisiko saman voinut päätellä lyhyemmin jonkin symmetrian nojalla?

Jos korttipakka sekoitetaan, millä tn pataässä on ennen ruutuviitosta?

Osajoukot (kombinaatiot s. 26)

4 ihmistä ABCD kättelevät toisiaan. Montako kättelyä tapahtuu? (2 ihmisen osajoukot)

Jos luetellaan 2-osajonot, $4 \cdot 3 = 12$ kpl

AB, AC, AD,

BC, BA, BD,

CA, CB, CD,

DA, DB, DC

Osajoukot (kombinaatiot s. 26)

4 ihmistä ABCD kättelevät toisiaan. Montako kättelyä tapahtuu? (2 ihmisen osajoukot)

Jos luetellaan 2-osajonot, $4 \cdot 3 = 12$ kpl

AB, AC, AD,

BC, BA, BD,

CA, CB, CD,

DA, DB, DC

Punaisella merkityt ovat sama osajoukko
$\{A,B\} = \{B,A\}$

Osajoukot (kombinaatiot s. 26)

4 ihmistä ABCD kättelevät toisiaan. Montako kättelyä tapahtuu? (2 ihmisen osajoukot)

Jos luetellaan 2-osajonot, 4·3=12 kpl

AB, AC, AD,

BC, BA, BD,

CA, CB, CD,

DA, DB, DC

Punaisella merkityt ovat sama osajoukko
{A,B} = {B,A}

Sinisellä merkityt ovat sama osajoukko
{A,C} = {C,A}

Jne: Jokaista 2-osajoukkoa vastaa kaksi eri 2-osajonoa. Osajoukkojen määrä on vain puolet eli 6 kpl

Osajoukot (kombinaatiot s. 26)

4 ihmistä ABCD kättelevät toisiaan. Montako kättelyä tapahtuu? (2 ihmisen osajoukot)

AB, AC, AD,
BC, BD,
CD

Karsittu luettelosta tuplat pois,
nyt kukin osajoukko on kerran

Osajoukkojen määrä on 6 kpl

Osajoukot yleisesti

n:stä alkiosta valitaan *k*-osajoukkoja (k-kombinaatioita):

valinta voidaan tehdä näin:

- Muodostetaan ensin kaikki *k*-osajonot: $(n)_k$ kpl
- Kutakin osajoukkoa vastaa *k*! eri osajonoa, ja valitaan niistä vain yksi

Osajoukkoja on siis näin monta (miksi?)

$$\frac{(n)_k}{k!} = \frac{n(n-1)...(n-k+1)}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

Vertailutaulukko (täydennetty)

	3 noppaa	**3 ihmisen jono (permutaatio)**	**3 ihmisen osajono 10:stä (variaatio)**	**3 ihmisen osajoukko 10:stä (kombinaatio)**
Vaihtoehtoja vaiheittain	$6 \cdot 6 \cdot 6$	$3 \cdot 2 \cdot 1$	$10 \cdot 9 \cdot 8$	$\frac{10 \cdot 9 \cdot 8}{3 \cdot 2 \cdot 1}$
Lyhyempi merkintä	6^3 **potenssi**	$(3)_3 = 3!$ **kertoma**	$(10)_3$ **laskeva tulo**	$\binom{10}{3}$ **binomikerroin**
Yhteensä	216 kpl	6 kpl	720 kpl	120 kpl
Alkeis-tapausten luettelo = perusjoukko	{111, 112, 113, ... 121, 122, 123, ... 211, 212, ... 611, ..., 666}	{ABC, ACB, BAC, BCA, CAB, CBA}	{ABC, ABD, ABE, ..., ABJ, ACB, ACD, ..., ACJ, ... BAC, BAD, ... DAB, ..., EAB, …, HIJ, HJI, …, JIH}	{{A,B,C}, {A,B,D}, ..., {H,I,J}}

Montako lottoriviä on olemassa?

- 39:stä numerosta valitaan 7-osajoukko

$$\binom{39}{7} = \frac{(39)_7}{7!} = \frac{39 \cdot 38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 15\,380\,937$$

Montako lottoriviä, joissa on kolmonen?

- Kolmonen on pakko ottaa mukaan osajoukkoon, siinä ei ole valinnanvaraa
- Lopuista 38:sta numerosta valitaan 6-osajoukko:

$$\binom{38}{6} = \frac{(38)_6}{6!} = \frac{38 \cdot 37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 2\,760\,681$$

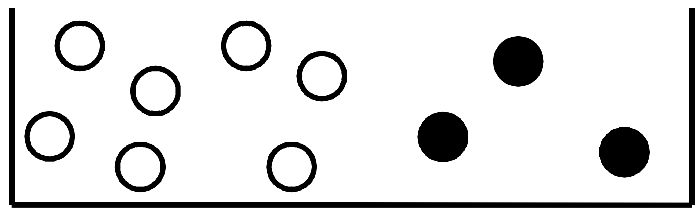
Montako lottoriviä, joissa on 3 ja 8?

- 3 ja 8 on pakko ottaa mukaan osajoukkoon, siinä ei ole valinnanvaraa
- Lopuista 37:sta numerosta valitaan 5-osajoukko:

$$\binom{37}{5} = \frac{(37)_5}{5!} = \frac{37 \cdot 36 \cdot 35 \cdot 34 \cdot 33}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 435\,897$$

Tn saada lotossa 4 oikein?

- Asiakas on valinnut jotkut 7 numeroa, esim. {2, 3, 10, 11, 26, 28, 33}
- Kone arpoo jotkut 7 numeroa
- Tn, että riveissä on tasan 4 samaa?
- Perusjoukossa on (39 yli 7) riviä, kone arpoo **yhden rivin**
- Suotuisia: Koneen pitää arpoa
  - Jotkut 4 numeroa asiakkaan 7:stä, (7 yli 4)
  - Jotkut 3 numeroa niistä 39-7 = 32:sta, joita asiakas ei valinnut (muuten tulee enemmän kuin 4 oikein)
  - Tuloperiaate: (7 yli 4) · (32 yli 3) suotuisaa riviä
- Tn = (7 yli 4) · (32 yli 3) / (39 yli 7)



OTANTA TAKAISINPANOLLA JA ILMAN

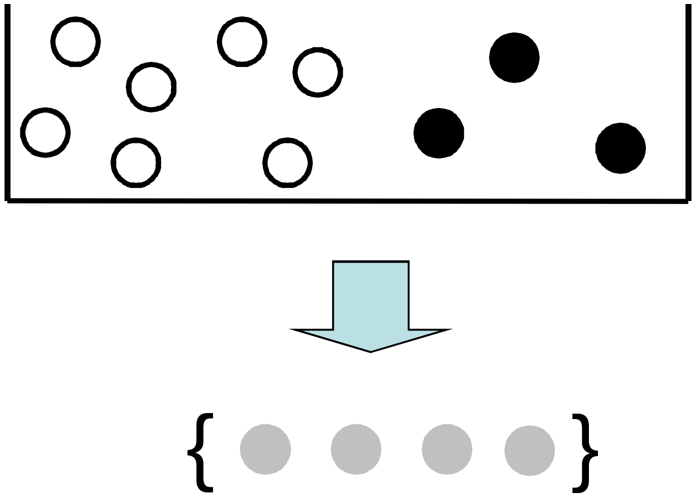
Otanta ”ilman takaisinpanoa”

- Populaatio, jossa N alkiota (palloa, ihmistä tms.), kahdenlaisia (”valkoinen”, ”musta”)
- Poimitaan umpimähkään (= symmetrisesti) n-osajoukko eli kombinaatio eli **otos**

- Merkitään tapahtuma $\boldsymbol{A_k}$ **= ”otoksessa on *k* valkoista palloa”**
- Tapahtumaan A_k sisältyy monta alkeistapausta (mitkä valkoiset, mitkä mustat pallot otoksessa)

N=10 palloa yhteensä

K=7 valkoista N-K=3 mustaa



otos n=4 palloa

Montako valkoista?

Suotuisat alkeistapaukset

k valkoista valitaan K:sta — n-k mustaa valitaan N-K:sta

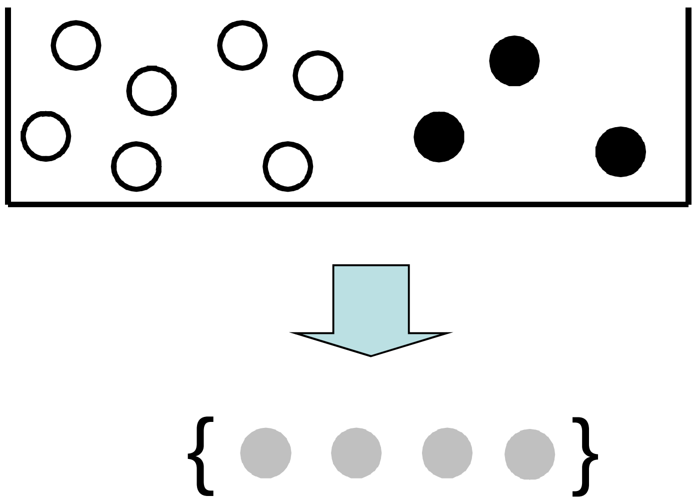
$$P(A_k) = \frac{\binom{K}{k} \times \binom{N-K}{n-k}}{\binom{N}{n}}$$

Kaikki alkeistapaukset = erilaiset n-otokset

Otanta ”ilman takaisinpanoa”

N=10 palloa yhteensä

K=7 valkoista *N*-*K*=3 mustaa



otos *n*=4 palloa

$$P(A_2) = \frac{\binom{7}{2} \times \binom{3}{2}}{\binom{10}{4}} = \frac{21 \times 3}{210} = 0{,}3$$

Samalla kaavalla

P(”0 valkoista”) = 0 (miksi?)

P(”1 valkoinen”) ≈ 0,033

P(”2 valkoista”) = 0,300

P(”3 valkoista”) = 0,500

P(”4 valkoista”) ≈ 0,167

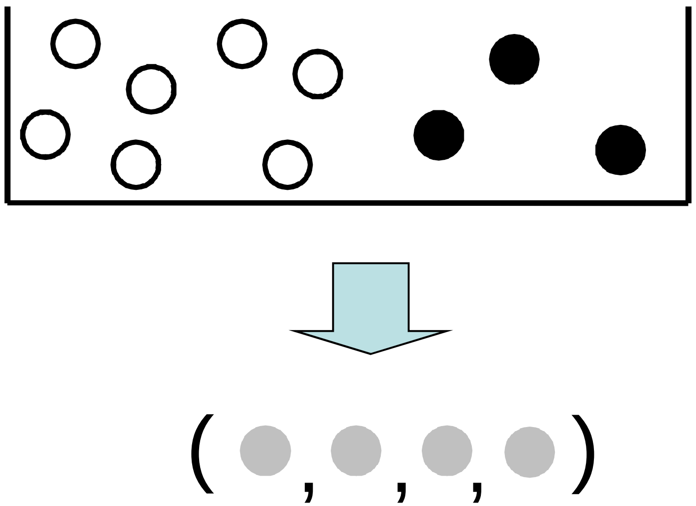
Otanta ”takaisinpanolla”

- Populaatio kuten edellä. Poimitaan umpimähkään yksi pallo, **pannaan takaisin**, otetaan umpimähkään toinen pallo jne.
- Ei välttämättä konkreettista ”takaisinpanoa”, olennaista on että **poimitaan joka kerta uudestaan samasta populaatiosta** (esim. puhelinluettelosta ihmisiä)

- Sama pallo voi tulla otokseen monta kertaa! Osajoukko ei mielekäs malli otokselle.
- Ajatellaan otos järjestetyksi n-jonoksi <u>jossa sama alkio saa esiintyä monta kertaa</u>.
- Montako valkoista palloa otoksessa?

N=10 palloa yhteensä

K=7 valkoista *N*-*K*=3 mustaa



Suotuisat alkeistapaukset

k:n valkoisen sijainti n-jonossa

k valkoista valitaan *K*:sta

n-*k* mustaa valitaan *N*-*K*:sta

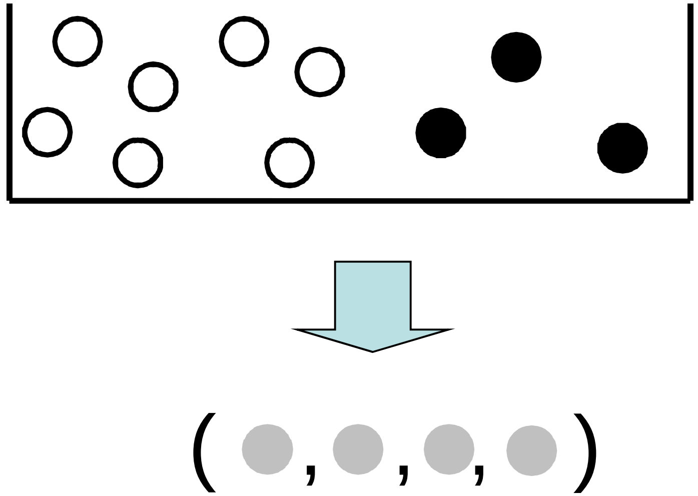
$$P(A_k) = \binom{n}{k} \frac{K^k (N-K)^{n-k}}{N^n}$$

Kaikki alkeistapaukset = erilaiset *n*-jonot. Joka kerta samat *N* vaihtoehtoa!

Otanta ”takaisinpanolla”

N=10 palloa yhteensä

K=7 valkoista *N*-*K*=3 mustaa



Suotuisat alkeistapaukset

2 valkoisen sijainti 4-jonossa

2 valkoista valitaan 7:stä

2 mustaa valitaan 3:sta

$$P(A_2) = \binom{4}{2} \frac{7^2 \times 3^2}{10^4} \approx 0{,}265$$

Kaikki alkeistapaukset
= erilaiset 4-jonot.
Joka kerta
samat 10 vaihtoehtoa!

Käytännössä ”otanta takaisinpanolla” on helpointa ajatella **toistokokeena**: joka nostolla valkoisen pallon tn on **sama,** nimittäin *K* / *N*.

Yleensä eri todennäköisyys

7 valkoista ja 3 mustaa, otos n=4

k	P(”k valkoista”) ilman takaisinpanoa	P(”k valkoista”) takaisinpanolla
0	0,000	0,008 (miksi > 0 ?)
1	0,033	0,076
2	0,300	0,265
3	0,500	0,412
4	0,167	0,240

Esimerkki: Hyvin suuri populaatio

- N = 5 000 000 (suomalaiset)
- K = 500 000 (helsinkiläiset)
- n = 3 (otoskoko) $n << K$ ja $n << N\text{-}K$

- Millä tn:llä otokseen osuu tasan yksi helsinkiläinen?
- Ilman takaisinpanoa

$$P(A_1) = \frac{\binom{500\,000}{1} \times \binom{4\,500\,000}{2}}{\binom{5\,000\,000}{3}} \approx 0{,}243\,000\,094$$

- Takaisinpanolla

$$P(A_1) = \binom{3}{1} \frac{500\,000^1 \times 4\,500\,000^2}{5\,000\,000^3} = 0{,}243\,000\,000$$

Lotossa P(tasan *k* oikein)

- Asiakas valitsee 7 numeroa. Kone arpoo 7 numeroa.
- Montako samaa tulee millä todennäköisyydellä?

= Otanta ilman takaisinpanoa. 39 palloa,

  - 7 palloa on ”mustia” (joiden numero on asiakkaan rivissä)
  - 32 palloa on ”valkoisia” (joiden numero ei ole asiakkaan rivissä)
  - Kone nostaa 7 palloa, montako niistä ”mustia”?

- Esim. P(asiakkaalla on viisi oikein) =

$$\frac{\binom{7}{5}\binom{32}{2}}{\binom{39}{7}}$$

Otoksessa oltava 5 mustaa palloa (7:stä) → $\binom{7}{5}$

Otoksessa oltava 2 valkoista palloa (32:sta) → $\binom{32}{2}$

Perusjoukko: 7 pallon otos 39:stä → $\binom{39}{7}$

P(tasan *k* oikein)

k	Suotuisia rivejä (kpl)		Tn
0	3 365 856		0,219
1	6 343 344		0,412
2	4 228 896		0,275
3	1258 600		0,081 8
4	173 600		0,011 3
5	10 416		0,000 677
6	224		0,000 014 6
7	1		0,000 000 065
yhteensä	15 380 937		1,000

Huom. Tuloksen ”7 oikein” pieni tn ei johdu siitä, että kyseinen rivi olisi muita epätodennäköisempi – jokaisen yksittäisen rivin tn on sama 1/ 15380937. Olennaista on eri tapahtumille suotuisien rivien **lukumäärä**. (Kuvittele kaikkien rivien luettelo ja jaa rivit **ryhmiin** ”*k* oikein”.)