Johdatus todennäköisyyslaskentaan
Kevät 2014

Luento 12 / 13

Jukka Kohonen
Matematiikan ja tilastotieteen laitos
Helsingin yliopisto

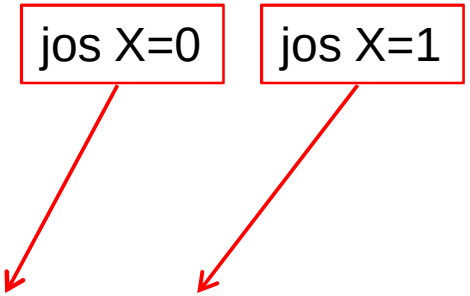
VARIANSSI (JATKUU)

Epäreilun kolikon varianssi

Diskreetti satunnaismuuttuja joukossa {0, 1},

$P(X=0) = q$

$P(X=1) = p$ $\quad p+q=1$



Odotusarvo: $\mu = E(X) = q \cdot 0 + p \cdot 1 = \mathbf{p}$

Poikkeama $(X-\mu)$ on joko $-p$ tai $(1-p)$

Neliöpoikkeama $Z = (X-\mu)^2$ on joko $(-p)^2$ tai $(1-p)^2$

$\mathrm{Var}(X) = E(Z) = q \cdot (-p)^2 + p \cdot (1-p)^2$

$= qp^2 + pq^2$

$= \mathbf{pq}$.

Summan varianssi, riippumattomat muuttujat

Olkoon S = X + Y.

Osataan tietysti laskea E(S) = E(X) + E(Y)

Lasketaan auki Var(S).

Kaava sievenee kummasti, jos pystytään käyttämään riippumattomien muuttujien tulokaavaa E(XY) = E(X) E(Y)

→Lause 3.2.5(iii) (sivu 83)

Sovellutus: Binomijakauman varianssi ***npq***

Helppoja hajonnan ominaisuuksia

Tuominen s. 83: Lause 3.2.5

- Eräissä jakauman muunnoksissa on helppo(?) päätellä, mitä hajonnalle tapahtuu

- Vakion **lisääminen**: Hajonta ei muutu
- Vakiolla **kertominen**: Hajonnalle sama kerroin (tai itseisarvo, jos vakio on negatiivinen)

- Kahden **riippumattoman** sm:n **summa**: **varianssit** voi laskea yhteen
  (hajontoja ei voi laskea yhteen – hajonta on varianssin neliöjuuri)

Eräiden jakaumien tunnuslukuja

Jakauma	Odotusarvo	Hajonta
Tas(0, 1)	1/2	$1 / \sqrt{12}$
Tas(a, b)	(a+b) / 2	($b-a$) / $\sqrt{12}$
N(0, 1)	0	1
N(μ, σ^2)	μ	σ
Exp(1)	1	1
Exp(λ)	$1 / \lambda$	$1 / \lambda$

Näissä kaikissa ”yleinen jakauma” saadaan ”perusjakaumasta” jollain venytyksellä (vakiokerroin) ja siirrolla (vakiolisäys), ja tämä näkyy tunnusluvuissakin edellä opitulla tavalla.

Hajonta on yleensä helpommin ymmärrettävä tunnusluku kuin varianssi (varianssia käytetään lähinnä eräiden laskukaavojen, kuten summakaavan takia)

Vakiokerroin ja vakiolisäys

- Puhelinpalvelun jonotusaika minuutteina $X \sim \text{Tas}(0, 10)$

  $E(X) = (0+10)/2 \quad = 5.00$ min

  $D(X) = 10 \cdot \sqrt{(1/12)} \quad = 2.89$ min

  D(X) voidaan esim. laskea integroimalla; tai muistaa kaava Tas(a,b):lle;
  tai päätellä Tas(0,1)-jakauman hajonnasta, joka on $1/\sqrt{12} \approx 0.289$

- **Vakiokerroin**: jonotus maksaa 0.20 €/min, hinta $Y = 0.2 \cdot X$

  $E(Y) = 0.2 \cdot E(X) \quad = 0.2 \cdot 5 \quad = 1.00$ € sama kerroin

  $D(Y) = 0.2 \cdot D(X) \quad = 0.2 \cdot 2.89 \quad = 0.58$ € sama kerroin

- **Vakiolisäys**: puheluaika 3 min, kokonaisaika $Z = X+3$

  $E(Z) = E(X) + 3 \quad = 5 + 3 \quad = 8.00$ min sama lisäys

  $D(Z) = D(X) \quad = 2.89$ min hajonta ei kasva!

Odotusarvon ja varianssin kaavoja

Muunnos	**Odotusarvo** (Lause 3.1.1)	**Varianssi** (Lause 3.2.5)
vakion lisäys	$E(X + b) = E(X) + b$	$Var(X + b) = Var(X)$
vakiokerroin	$E(aX) = a \cdot E(X)$	$Var(aX) = a^2 \cdot Var(X)$
sm:ien summa	$E(X+Y) = E(X) + E(Y)$	$Var(X+Y) = Var(X) + Var(Y)$ *jos* $X \perp\!\!\!\perp Y$
sm:ien tulo	$E(XY) = E(X) \cdot E(Y)$ *jos* $X \perp\!\!\!\perp Y$	
yleinen muunnos $g(X)$	$\mathrm{E}[g(X)] = \int g(x) f(x) dx$ (Lause 3.1.8)	

SUURTEN LUKUJEN LAKI

Mihin pyritään

- Olemme empiirisesti huomanneet, että otoksen histogrammi muistuttaa jakauman tiheysfunktiota.

- Voisiko otoksen avulla arvioida myös tunnuslukuja, esim. jakauman odotusarvoa?
- Ilmeinen ehdokas olisi otoksen aritmeettinen keskiarvo. MATLAB: `sum(x)/n` tai `mean(x)`

- Todistetaan lause, jonka mukaan otoskeskiarvo on ”pitkän päälle” todennäköisesti hyvä arvio odotusarvolle.
  - Yritetään tehdä tämä yleisesti (monille eri jakaumille)
  - Ajatellaan ensin erikoistapausta: Bernoullin lausetta

Kohti Bernoullin lausetta

Tuominen s. 93

Toistokoe, ***n*** toistoa tn:llä ***p***

Z_n = onnistumisten lukumäärä

f_n = onnistumisten osuus = Z_n / n

Tiedetään vanhastaan, että: $Z_n \sim \mathbf{Bin}(n, p)$

Siispä onnistumisten määrän odotusarvo on ***np*** ja osuuden odotusarvo ***p***.

Myös jakaumien moodit ovat suunnilleen tässä kohdassa (Tuominen s. 51)

Mutta **täsmälleen** odotusarvoon ei osuta helposti (harj. 6:8)

Osumisen tn jopa **pienenee**, kun *n* kasvaa. Esim. kun *p*=0.3:

n	np	$P(Z_n = np) \approx$
10	3	0.27
100	30	0.087
1 000	300	0.028
1 000 000	300 000	0.000 87

Osutaanko edes lähelle?

Z_n = onnistumisten lukumäärä (binomijakautunut)

f_n = onnistumisten osuus = $Z_n \,/\, n$

ε = osuuden tarkkuusvaatimus (voimme valita tämän vapaasti)

Sanomme, että f_n osuu lähelle, kun $|\, f_n - p \,| < \varepsilon$

Lähelle osumisen tn **kasvaa**, kun *n* kasvaa – ainakin siltä näyttää:
Esim. kun p=0.3 ja **ε=0.01**, lähelle osutaan kun $0.29 < f_n < 0.31$:

n	f_n oltava välillä	Z_n oltava välillä	P(lähellä) ≈
100	(0.29, 0.31)	{30}	0.087
1 000	(0.29, 0.31)	(290, 310)	0.488
10 000	(0.29, 0.31)	(2 900, 3 100)	0.970
100 000	(0.29, 0.31)	(29 000, 31 000)	0.999999999946

P(lähellä) = summa binomijakauman pistetodennäköisyyksistä

Bernoullin lause (Tuominen 3.5.4, s. 93)

- Valitaan mikä tahansa tarkkuusvaatimus **ε** > 0
- Bernoullin lause kertoo **raja-arvotuloksen:**

$$P\left(\left|f_n - p\right| < \varepsilon\right) \to 1$$

kun $n \to \infty$.

- Tämän todistaminen edellyttää, että pystytään sanomaan jotain binomijakauman todennäköisyydestä **isolla välillä**. Suoraviivainen lasku, **ptnf yhteenlasku** ei ole helppoa, kun termien määräkin kasvaa *n*:n mukana
- Tunnemme binomijakauman **varianssin** *npq*, mutta onko siitäkään apua? Miten hajonta liittyy todennäköisyyksiin?
- Käytetään yleistä matemaattista menetelmää: todistetaan joku **epäyhtälö**, kun ei osata todistaa tarkkaa yhtälöä
- Todistus perustuu kahteen epäyhtälöön: Markovin ja Tsebysevin. Tutustumme niihin seuraavaksi.

EPÄYHTÄLÖITÄ POIKKEAMIEN ARVIOIMISEEN

Epäyhtälöitä

- Jos jakaumaa ei tunneta tarkasti, mutta tunnetaan E(X) ja ehkä Var(X),
  niin **suurten poikkeamien todennäköisyyksiä**
  (jakauman ”**häntiä**”) voidaan arvioida erilaisilla epäyhtälöillä.

- Arviot ovat kuitenkin aika karkeita.

Markovin epäyhtälö (Tuominen s. 81)

- Jos varmasti **X ≥ 0**, ja **E(X) tunnetaan**,
  voidaan arvioida häntätodennäköisyyttä mistä tahansa kohdasta ***a*** oikealle (äärettömiin asti)

$$P(X \geq a) \leq E(X) / a$$

Jakaumasta ei tarvitse tietää muuta kuin em. asiat.

Todistusidea:

Odotusarvo on summa tai integraali.
Rajoitetaan sen termejä tai integroitavaa **alhaalta**
→ saadaan odotusarvolle **alaraja**
→ saadaan todennäköisyydelle **yläraja**.

Summan tai integraalin rajoittaminen

- Yleisiä matemaattisia havaintoja:

  – Jos termeittäin pätee $a_i \leq b_i, \quad \forall i,$
  pätee myös summille $\sum a_i \leq \sum b_i$

  – Jos pätee $f(x) \leq g(x), \quad \forall x,$
  pätee myös $\int f(x)dx \leq \int g(x)dx$

Markovin epäyhtälö, diskreetti X

Oletetaan yksinkertaisuuden vuoksi, että X on kokonaislukuarvoinen

- **Tiedetään, että $X \geq 0$**, ja **$E(X)$ tunnetaan.**
- Merkitään
  pistetodennäköisyydet **p_k** $= P(X = k)$
  häntätodennäköisyys **q_a** $= P(X \geq a) = p_a + p_{a+1} + \ldots$

$$E(X) = \sum_{k=0}^{\infty} k\, p_k$$

Odotusarvo esitetty summana

$$= \left(\sum_{k=0}^{a-1} k\, p_k \right) + \left(\sum_{k=a}^{\infty} k\, p_k \right)$$

Summa hajotettu kahteen osaan

$$\geq \left(\sum_{k=0}^{\infty} 0\, p_k \right) + \left(\sum_{k=a}^{\infty} a\, p_k \right)$$

Kaikkia termejä **rajoitettu alhaalta**.
Ensimmäinen osasumma häviää.
Toisessa otetaan a ulos summasta

$$= 0 + a \cdot q_a$$

$$= a \cdot q_a,$$

joten puolestaan **$q_a \leq E(X)/a$**, eli Markovin epäyhtälö pätee.

Markovin epäyhtälö, jatkuva *X*

- Todistus samaan tapaan kuin diskreetillä *X*:llä. Odotusarvo on integraali, jota rajoitetaan alhaalta, nytkin saadaan

$$E(X) \geq pa,$$

josta seuraa

$$p \leq E(x)/a.$$

Tšebyševin epäyhtälö (Tuominen s. 85)

- Jos ***μ* = E(*X*) ja *σ* = D(*X*) molemmat tunnetaan**, voidaan häntä-tn arvioida tehokkaammin:

$$P\left(\left|X-\mu\right| \geq k\sigma\right) \leq \frac{1}{k^2}$$

- Siis tn, että *X* poikkeaa odotusarvostaan ”yli *k*:n hajonnan verran”, on enintään $1/k^2$

- *X* ei tarvitse olla epänegatiivinen (kuten Markovissa)
- Tiukempi raja, koska nimittäjässä on toinen potenssi

- Todistuksessa Markov-epäyhtälöä sovelletaan neliöpoikkeamiin $(X-\mu)^2$, niiden odotusarvohan on = Var(X)

Esim: Exp-jakauman häntä-tn eri tavoilla

- Koneen elinikä $X \sim \text{Exp}(0.1)$. Arvioidaan todennäköisyyttä $P(X \geq a)$ a:n eri arvoilla ja eri menetelmillä.
- $E(X)=10$, $D(X)=10$, kertymäfunktio tunnetaankin $F(x)=1-\exp(-0.1x)$

	Markov	Tsebysev	Tarkasti kf:stä
$P(X \geq 10)$	< 1		0.368
$P(X \geq 20)$	$< 1/2$	< 1	0.135
$P(X \geq 30)$	$< 1/3$	$< 1/4 = 0.25$	0.050
$P(X \geq 40)$	$< 1/4$	$< 1/9 = 0.11$	0.018
$P(X \geq 200)$	$< 1/20$	$< 1/361 = 0.0028$	$2.1 \cdot 10^{-9}$

- Suurille poikkeamille Tšebyšev on tarkempi kuin Markov.
- Vielä paljon tarkempi on tarkka kertymäfunktio F, jos se tunnetaan!
- Mutta epäyhtälöillä saatetaan saada johonkin tarkoitukseen ”riittävä” arvio.
- Tšebyševistä on apua esim. suurten lukujen lain todistamisessa.

Tšebyšev → Suurten lukujen laki

Tuominen s. 91

- (X_1, X_2, ...) jono riippumattomia sm:ia, joilla sama odotusarvo $\boldsymbol{\mu}$ ja varianssi $\boldsymbol{\sigma^2}$
- valitaan mv. tarkkuusvaatimus $\boldsymbol{\varepsilon} > 0$

Osasumma S_n $= X_1+...+X_n$

- Osasumman odotusarvo $= \boldsymbol{n\mu}$ miksi?
- Osasumman varianssi $= \boldsymbol{n\sigma^2}$ miksi?

Keskiarvo $= (X_1+...+X_n) / n$

- Keskiarvon odotusarvo $= \boldsymbol{\mu}$ miksi?
- Keskiarvon varianssi $= \boldsymbol{\sigma^2 / n}$ miksi?

Tšebyšev → Suurten lukujen laki

Tuominen s. 91

valitaan mv. tarkkuusvaatimus $\varepsilon > 0$

Otoskeskiarvo $= (X_1 + ... + X_n) / n$

- Otoskeskiarvon odotusarvo $= \mu$
- Otoskeskiarvon varianssi $= \sigma^2 / n$
- Otoskeskiarvon hajonta $= \sigma / \mathrm{sqrt}(n)$ miksi?

Otoskeskiarvon **hajonta pienenee** *n*:n kasvaessa.

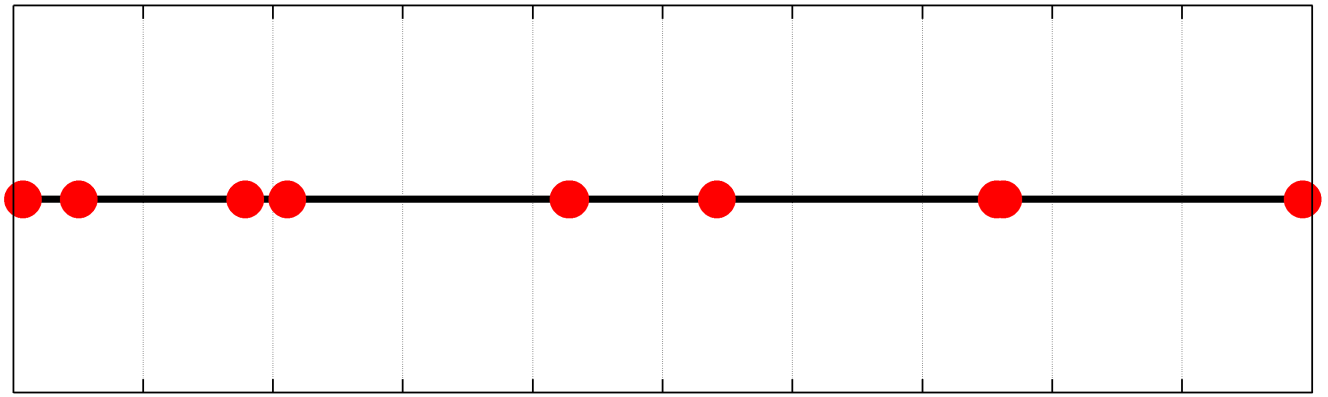
Tällöin kiinteä tarkkuusvaatimus **ε** merkitsee yhä suurempaa **kerrointa (*k*)** hajonnalle, jolloin Tsebysevin epäyhtälön perusteella **tn niin suureen poikkeamaan** odotusarvosta **pienenee**.

On siis yhä todennäköisempää (tn → 1), että otoskeskiarvo osuu alle **ε**:n päähän odotusarvostaan eli ***μ***:stä. = Suurten lukujen laki

SLL erikoistapaus: Bernoullin lause

- $(X_1, X_2, ...)$ jono riippumattomia **indikaattorimuuttujia** toistokokeelle
- indikaattorien osasumma = onnistumisten **lukumäärä**
- lukumäärä on tunnetusti binomijakautunut,
  $E(S_n) = np$
  $Var(S_n) = npq$

- Indikaattorien keskiarvo = onnistumisten **osuus**
- SLL näille indikaattoreille → tn, että onnistumisosuus on ”lähellä” odotusarvoaan, lähenee rajatta 1:tä
- Bernoullin lause kertoo, miksi **todennäköisyydellä** on jotain tekemistä toistokokeessa toteutuvan **osuuden** kanssa.

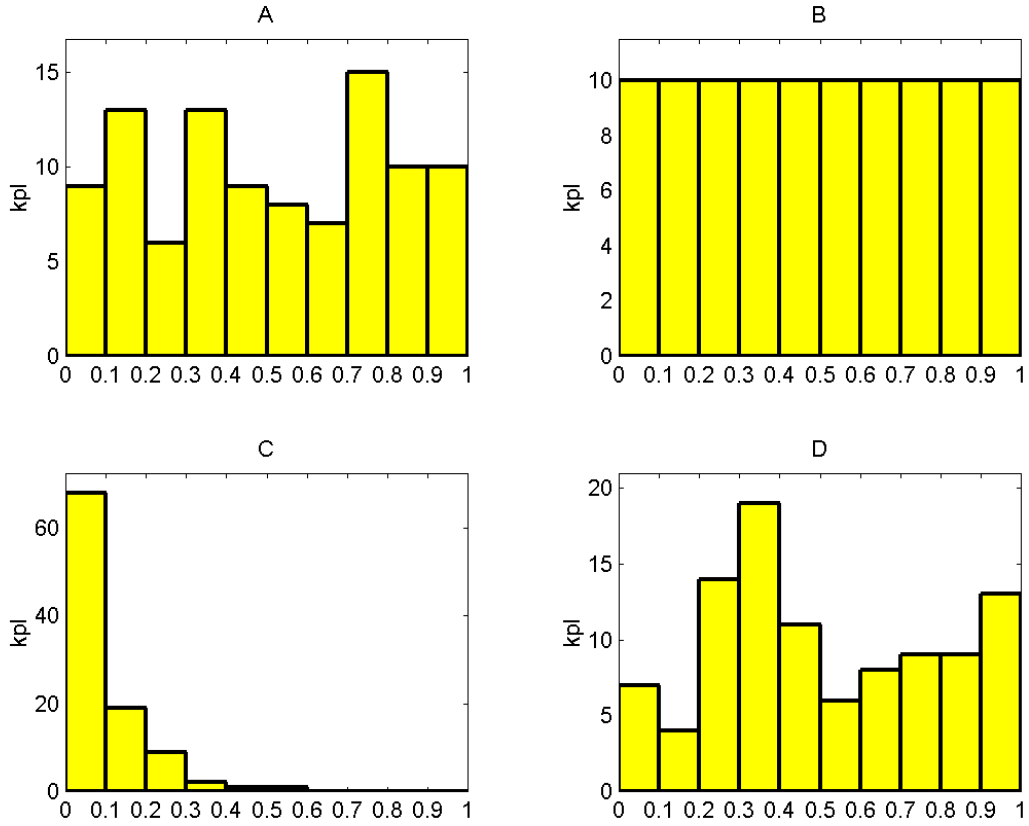


RIIPPUMATON OTOS TASAJAKAUMASTA

Tunnistustehtävä

Kukin näistä histogrammeista on piirretty sadasta luvusta.

Mikä histogrammeista on syntynyt siten, että luvut on arvottu jakaumasta **Tas(0, 1)** ?



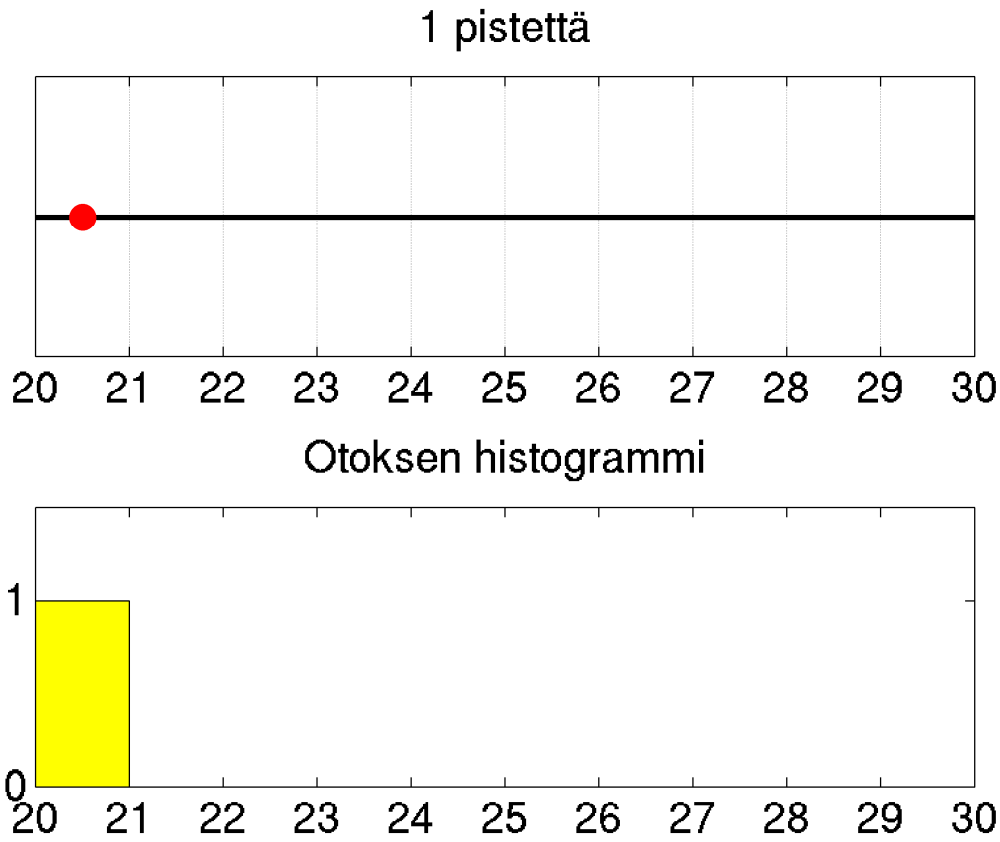
Otos tasajakaumasta

Arvotaan riippumattomia satunnaislukuja
$X_1, X_2, X_3 \ldots \sim$ Tas(20, 30).

Miten ne sijoittuvat?

Piirretään myös 10 pylvään histogrammi, ei **jakaumasta** (se on tasainen), vaan **otoksesta**, ts. mille väleille arvotut luvut osuivat.

(Histogrammi on yhteenveto siitä, missä pisteet *suunnilleen* ovat.)

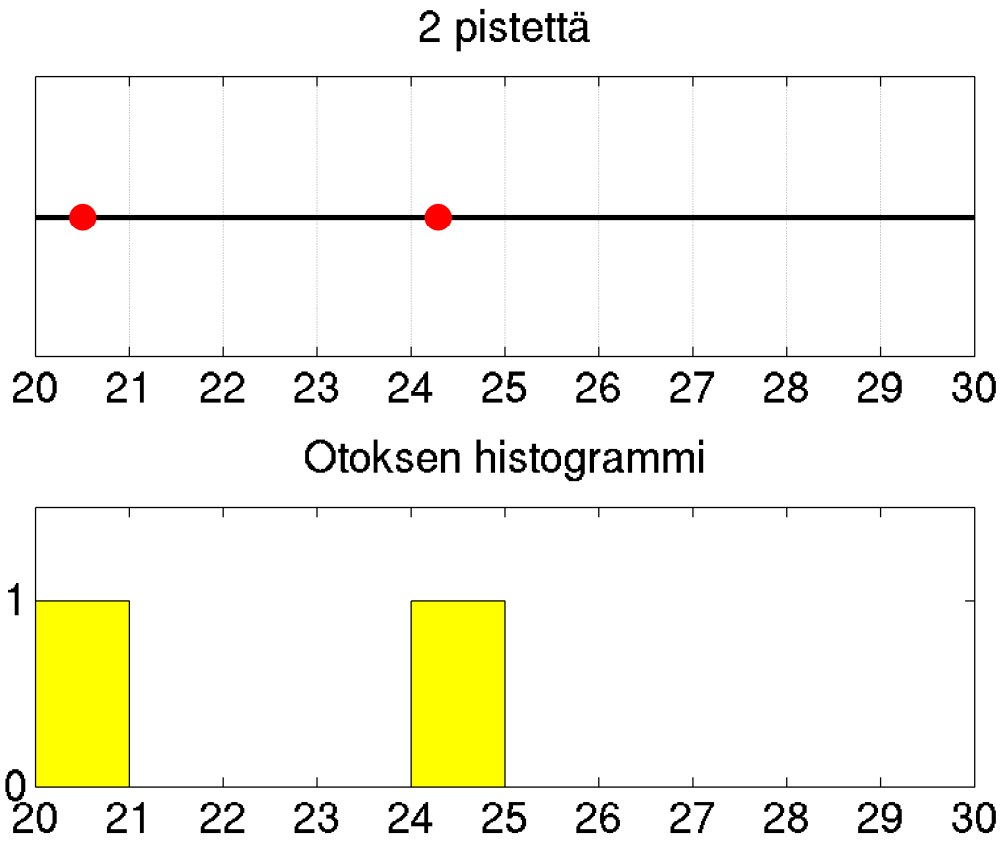


Otos tasajakaumasta

Arvotaan riippumattomia satunnaislukuja
$X_1, X_2, X_3 \ldots \sim$ Tas(20, 30).

Miten ne sijoittuvat?

Piirretään myös 10 pylvään histogrammi, ei **jakaumasta** (joka on tasajakauma), vaan **otoksesta**, ts. mille väleille arvotut luvut osuivat.

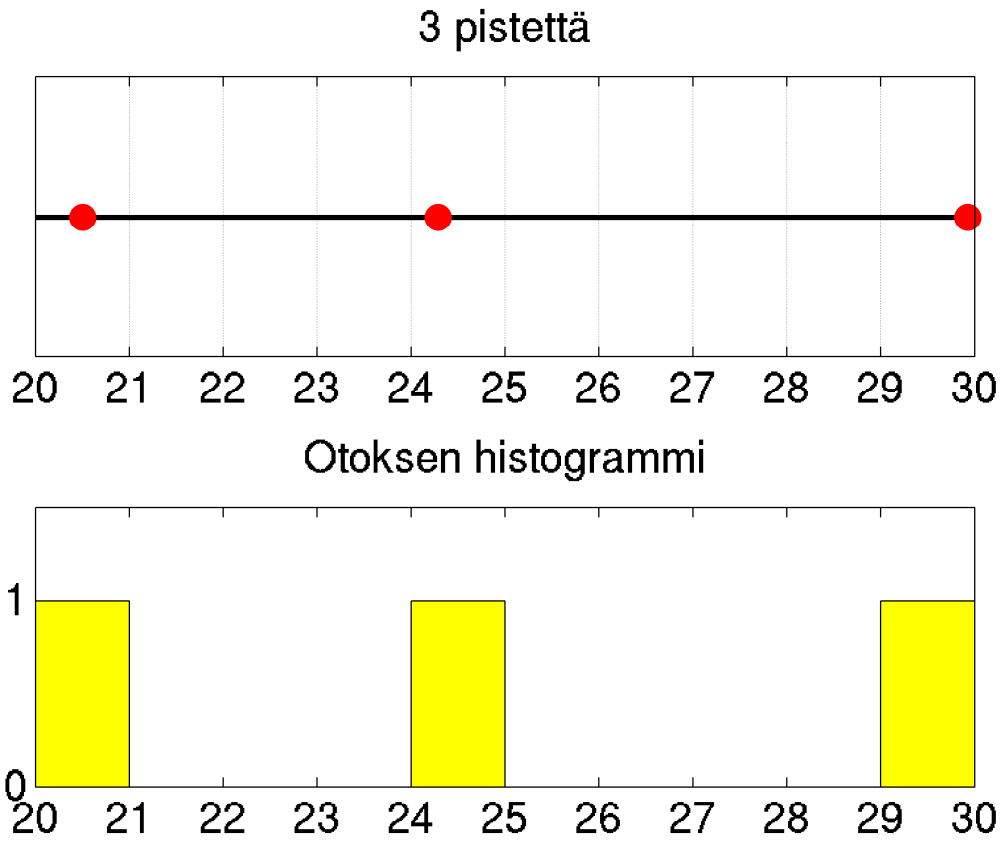


Otos tasajakaumasta

Arvotaan riippumattomia satunnaislukuja
$X_1, X_2, X_3 \ldots \sim$ Tas(20, 30).

Miten ne sijoittuvat?

Piirretään myös 10 pylvään histogrammi, ei **jakaumasta** (joka on tasajakauma), vaan **otoksesta**, ts. mille väleille arvotut luvut osuivat.



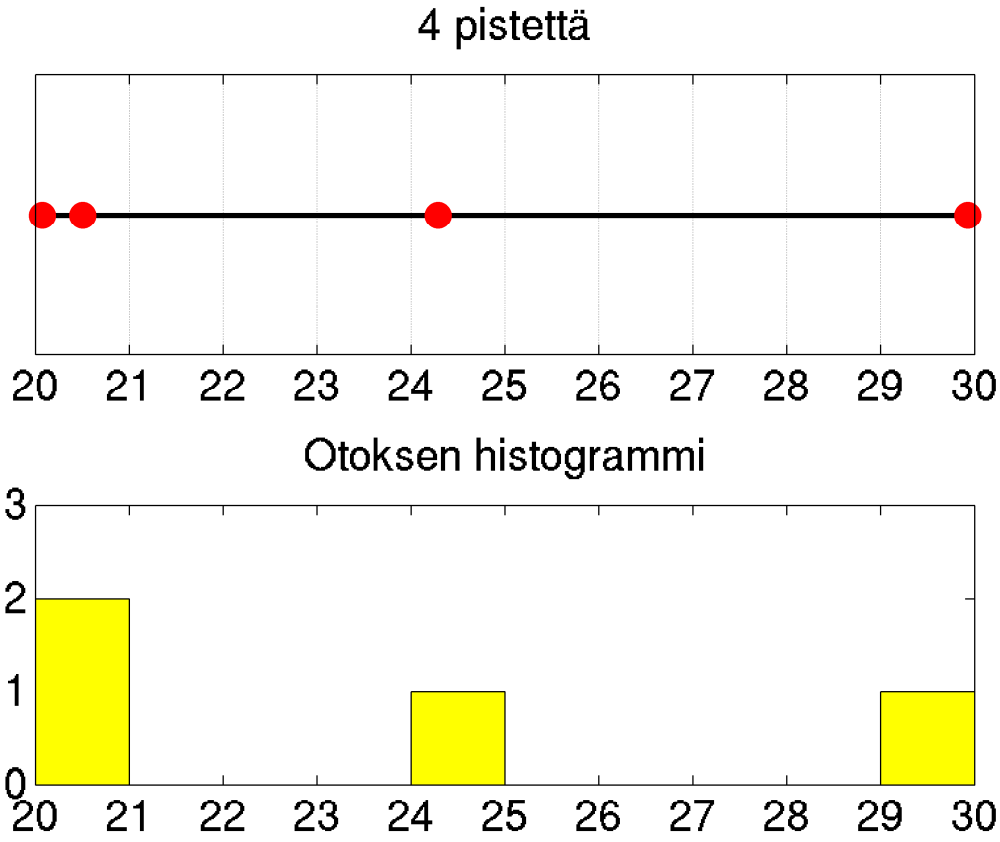
Otos tasajakaumasta

Arvotaan **riippumattomia** satunnaislukuja
$X_1, X_2, X_3 \ldots \sim$ Tas(20, 30).

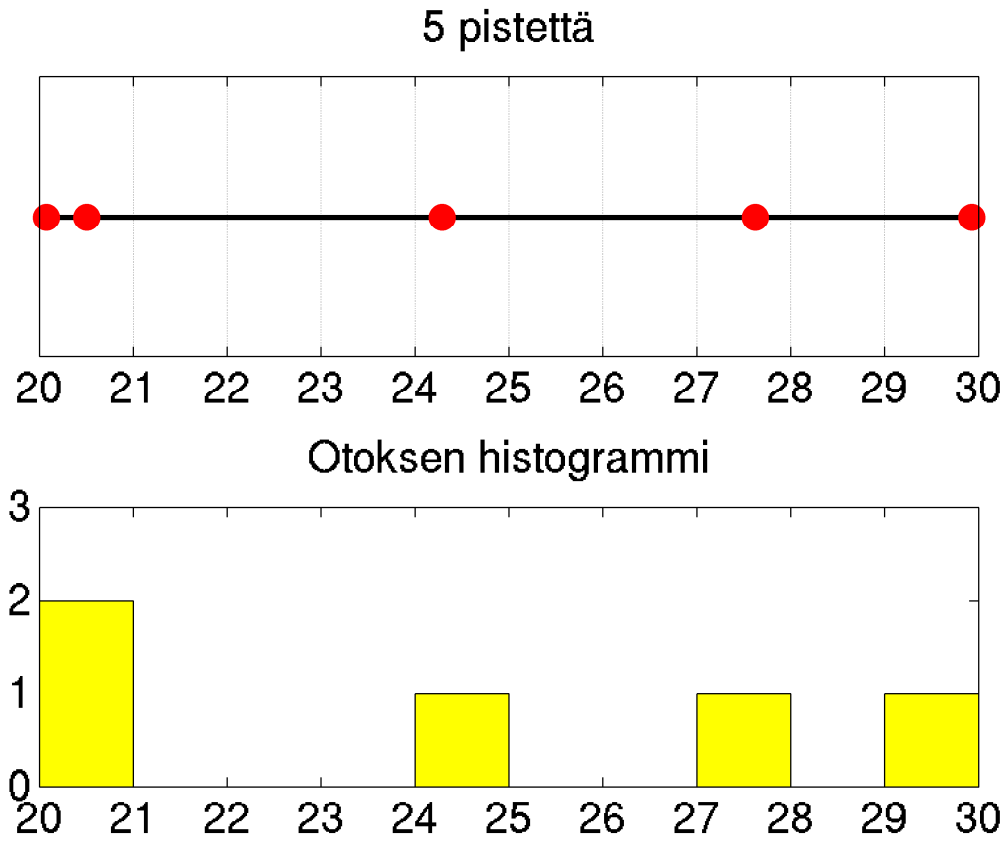
Riippumattomuus:
Aiemmat pisteet eivät vaikuta myöhempiin.

1. piste oli välillä (20,21),
se mitenkään estä 4. pistettä osumasta samalle välille

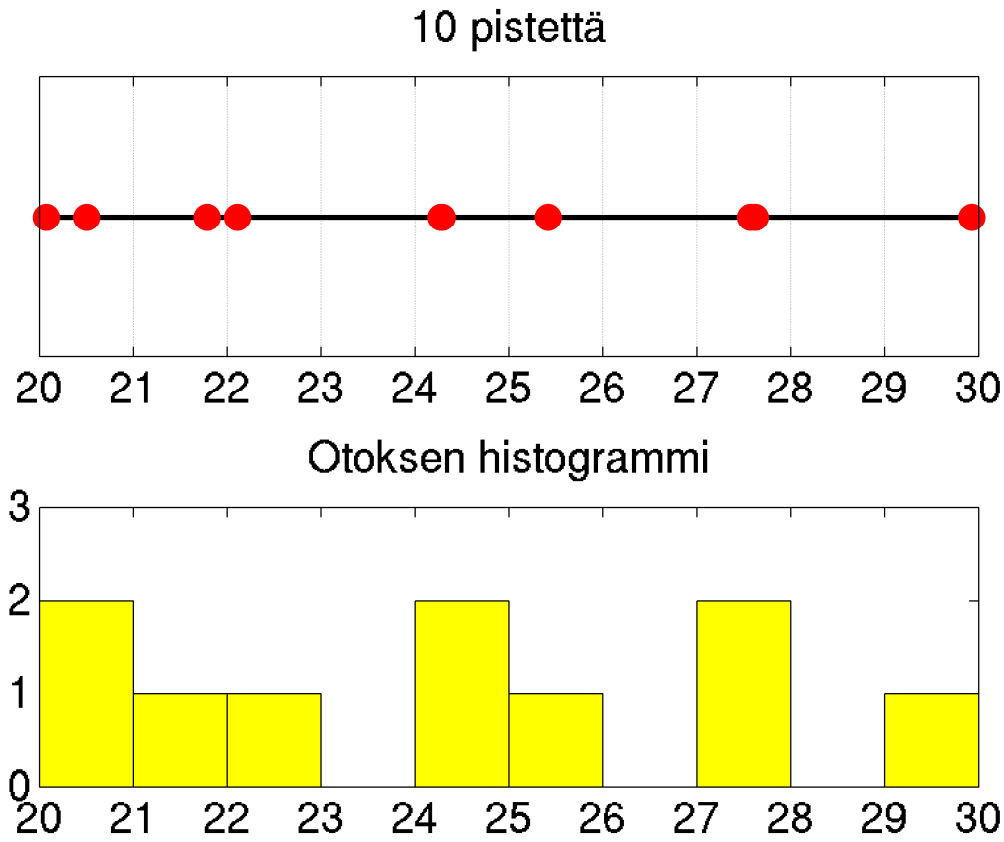
(ei edes lisää eikä vähennä ko. tapahtuman tn:ää, joka on jatkuvasti 1/10)



Otos tasajakaumasta



Otos tasajakaumasta

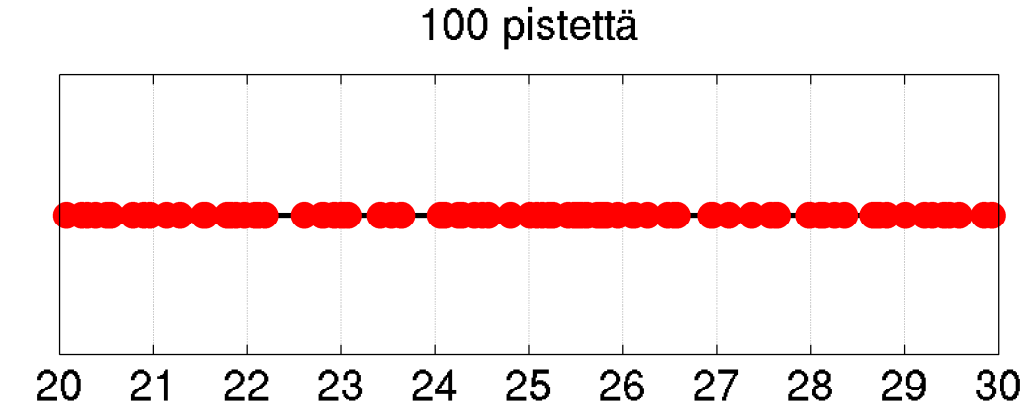


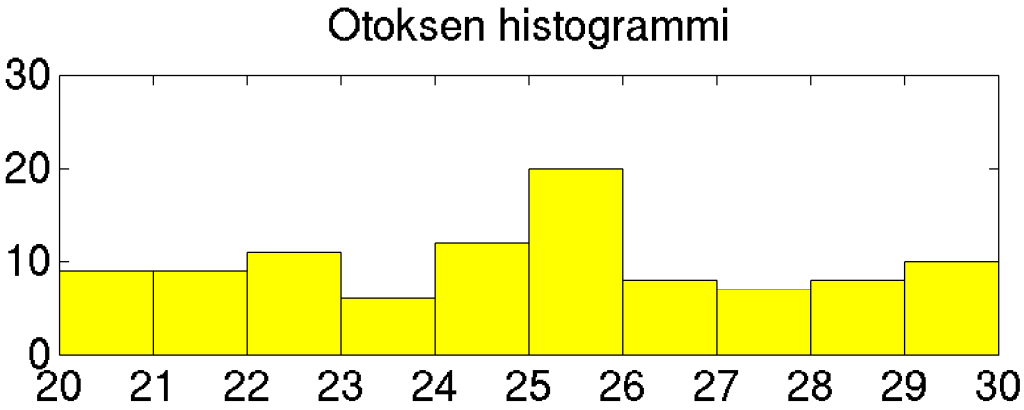
Otos tasajakaumasta

100 pistettä:

Kunkin pylvään
korkeuden odotusarvo
on 10, (miksi?)

mutta toteutuneet korkeudet
vaihtelevat melkoisesti.

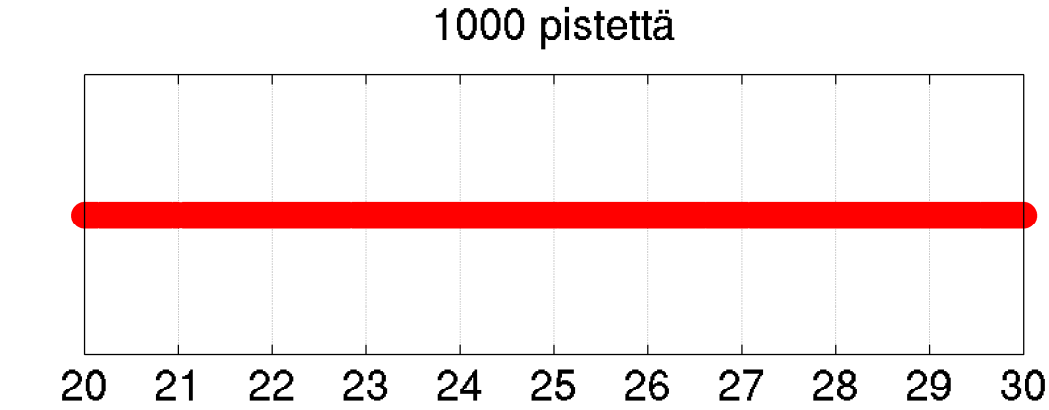


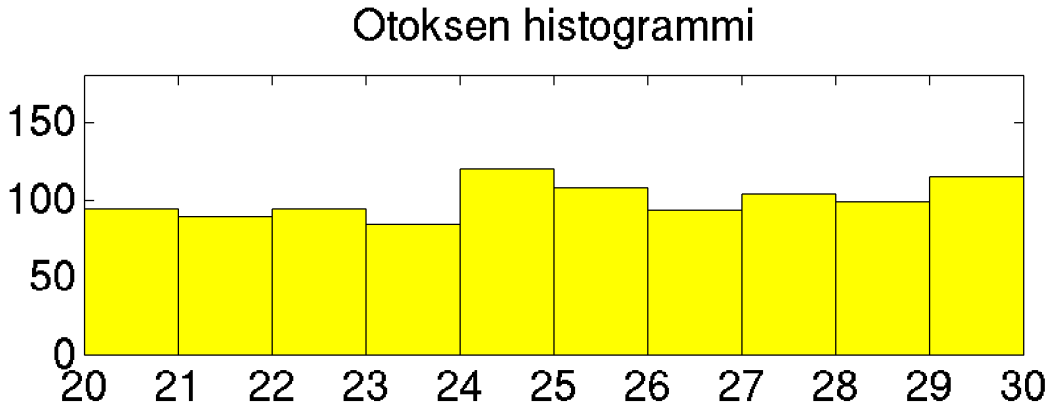


Otos tasajakaumasta

Yksittäisiä pisteitä on jo mahdoton erottaa (jakauma voisi olla joku muu ja näyttäisi samalta)

mutta histogrammista näemme osapuilleen pisteiden jakauman.

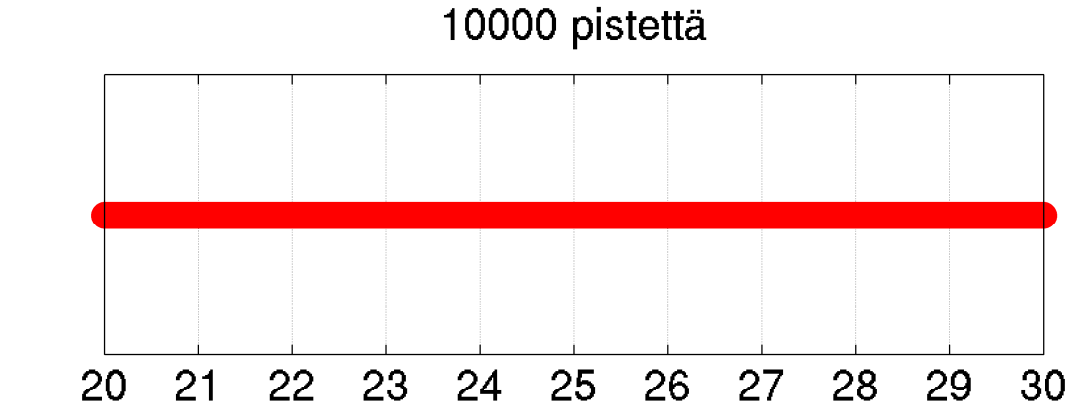


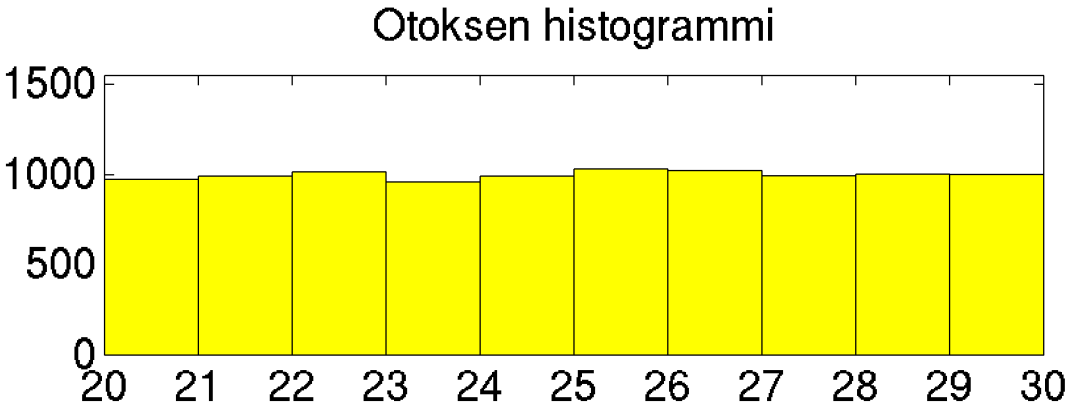


Otos tasajakaumasta

10 000 pistettä:

Melko tasaista.
Otos antaa jo hyvän käsityksen **jakauman** muodosta karkealla tasolla.



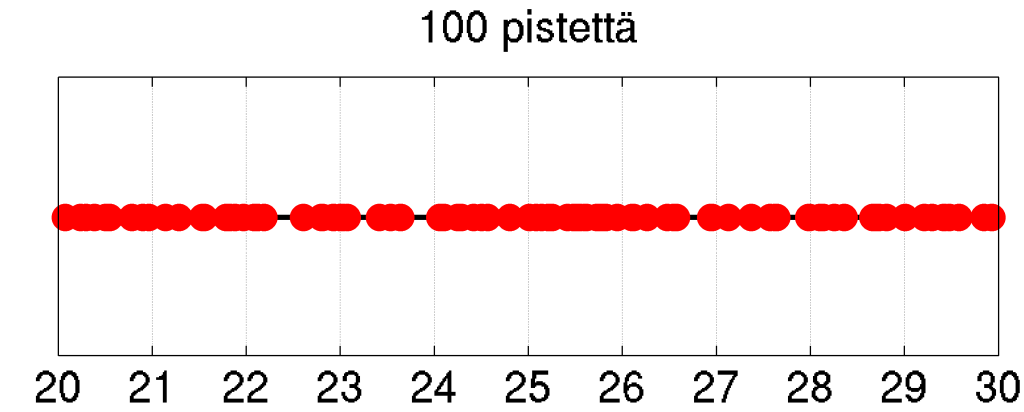


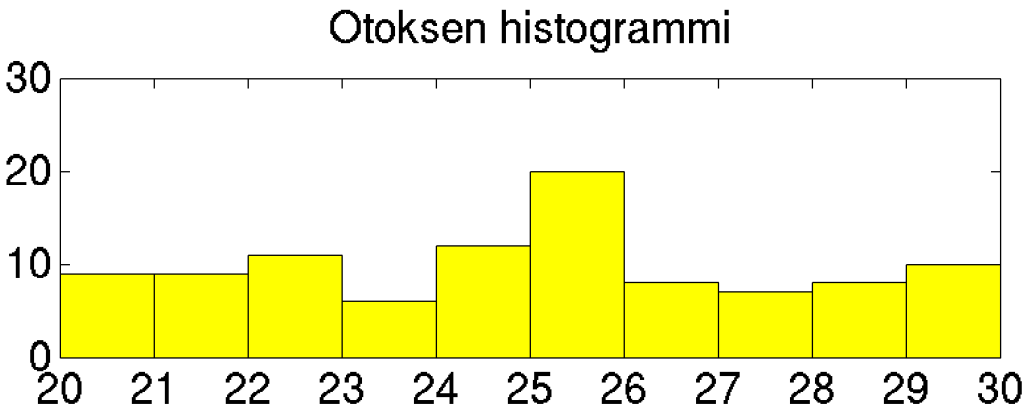
Pylväiden korkeuksien jakauma

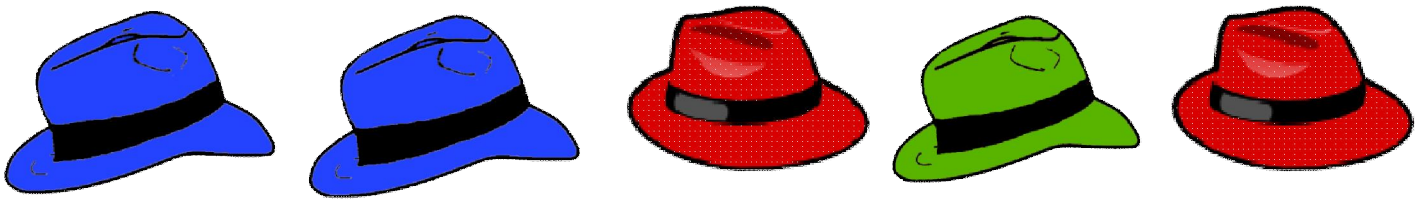
Mennään takaisin 100 pisteeseen.
Merk.
Y_i = *i*:nnen pylvään korkeus
= *i*:nnelle jakovälille osuvien pisteiden lukumäärä

Yksittäisen pylvään korkeus on **binomijakautunut. (miksi?)**

Tn, että histogrammi on täysin tasainen? Huomaa, että pylväiden korkeudet eivät ole riippumattomia. Tarvitaan ns. multinomijakauma.







MULTINOMIKERROIN JA MULTINOMIKOE

Multinomikerroin

Tuominen 34-36

10 ihmisellä on eriväriset hatut: 3 punaista, 2 sinistä ja 5 vihreää.
Ihmiset asettuvat jonoon.
Monessako eri järjestyksessä värit voivat olla? Emme välitä ihmisistä.

G B E H A J D F I C

Ihmisten jonoja on 10! erilaista. Mutta monessa eri jonossa on sama värijärjestys!

- Jos punaisten (B, E, C) järjestys vaihdetaan keskenään, saadaan sama värijono
- Jos sinisten (J, I) järjestys vaihdetaan keskenään, saadaan sama värijono
- Jos vihreiden (G, H, A, D, F) järjestys vaihdetaan keskenään, saadaan sama värijono

Tuloperiaate: sama värijono saadaan (3!) · (2!) · (5!) eri tavalla

Eri värijonoja on siis

$$\frac{10!}{(3!)\cdot(2!)\cdot(5!)} = \frac{3628800}{6\cdot 2\cdot 120} = 2520 = \binom{10}{3,\,2,\,5}$$

Uusi merkintä: **multinomikerroin**

”Monellako tavalla 10:stä alkiosta voi valita 3 kpl ensimmäiseen osajoukkoon, 2 kpl toiseen ja 5 kpl kolmanteen?” Vrt. binomikerroin.

Menikö oikein?

```
>> P = perms('PPPSSVVVVV');
>> size(P)
ans =
     3628800          10
```

Värit eri ihmisjonoissa
(monessa jonossa sama värijärjestys)

VVVVVSSPPP
VVVVVSSPPP
VVVVVSSPPP
VVVVVSSPPP
VVVVVSSPPP
...
PVVVVVSPSP
PVVVVVSSPP
PVVVVVSSPP

3 628 800 kpl

```
>> U = unique(P, 'rows');
>> size(U)
ans =
        2520          10
```

Erilaiset värijonot

VVVVVPPPSS
VVVVVPPSPS
VVVVVPPSSP
VVVVVPSPPS
VVVVVPSPSP
...
SSPPVVPVVV
SSPPVPVVVV
SSPPPVVVVV

2 520 kpl OK

Multinomikoe

Toistokokeen yleistys: Joka kerralla on **monta** poissulkevaa vaihtoehtoa, joiden tn:t tunnetaan. Kysytään montako kertaa toteutuu mikäkin vaihtoehto, kun kokeita on *n* kpl.

> 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 $a + b + c = 10$)?

- Suuri populaatio: approksimoimme otoksen ”takaisinpanolla”, kukin otoksen henkilö on toisista **riippumatta** A:n kannattaja tn:llä p jne.
- Alkeistapauksina mahdolliset 10-jonot puoluekantoja otoksessa – ei symmetriset! Esim. eräiden jonojen todennäköisyyksiä:

P(AAAAAAAAAA)	$= p^{10}$	$\approx 0.000\,977$	
P(AAAABBBBCC)	$= p^4 \cdot q^4 \cdot r^2$	$\approx 0.000\,020$	Miksi pienempi?!

Multinomikoe

Meitä ei kiinnosta, missä järjestyksessä puoluekantoja ilmaantuu otokseen, vain lukumäärät. **Lasketaan yhteen** alkeistapaukset, joissa lukumäärät ovat samat. Tapausten määrä = multinomikerroin!

P(AAAAAAAAAA)	$= p^{10}$	$\approx 0.000\ 977$	Näitä (10×A) on vain yksi
P(AAABABBBCC)	$= p^4 \cdot q^4 \cdot r^2$	$\approx 0.000\ 020$	Näitä alkeistapauksia (4×A, 4×B, 2×C) on $\binom{10}{4,4,2} = 3150$ kpl,
P(BBCAABBAAC)	$= p^4 \cdot q^4 \cdot r^2$	$\approx 0.000\ 020$	
P(AABCCAABBB)	$= p^4 \cdot q^4 \cdot r^2$	$\approx 0.000\ 020$	
...			
P(CCBBBBAAAA)	$= p^4 \cdot q^4 \cdot r^2$	$\approx 0.000\ 020$	tn yht. ≈ **0.064**

...

Multinomikoe

Tässä eri lukumäärien todennäköisyyksiä suuruusjärjestyksessä.

(a,b,c)	**tn**
(5,3,2)	0.085
(6,2,2)	0.071
(6,3,1)	0.071
(4,4,2)	0.064
(5,4,1)	0.064
…	…
(10,0,0)	0.000 977
…	…
(0,0,10)	0.000 000 102
yhteensä	**1**

Tn, että otososuudet ovat **täsmälleen** samat (50%, 30%, 20%) kuin populaatiossa, on vain 0.085

Mutta muutkaan todennäköiset osuudet eivät **paljon** poikkea

Tn saada ”pahasti pielessä” oleva otos on hyvin pieni

Multinomikoe ja multinomijakauma

- *n* riippumatonta koetta, jokaisessa 3 poissulkevaa vaihtoehtoa.
- Joka kerta vaihtoehtojen todennäköisyydet *p*, *q*, *r*.
- Tn, että toteutuneet lukumäärät ovat (*a*, *b*, *c*) on

$$\binom{n}{a,b,c} \cdot p^a \cdot q^b \cdot r^c$$

- Voimme sanoa, että lukumäärät (*a*, *b*, *c*) ovat yhdessä **multinomijakautuneet** parametrein *n* ja (*p*, *q*, *r*).
- Lukumäärät ovat satunnaismuuttujia, ja keskenään **riippuvia**!
  Jos esim. sattuu *a*=*n*, niin on pakko olla *b*=*c*=0. (miksi?)
- Jos vaihtoehtoja on > 3, kaava yleistyy ilmeisellä tavalla.
- Jos vaihtoehtoja on vain 2, kaava palautuu tuttuun binomijakaumaan.