Johdatus todennäköisyyslaskentaan
Kevät 2015

Luento 9 / 13

Jukka Kohonen

Matematiikan ja tilastotieteen laitos

Helsingin yliopisto

HUOMAUTUKSIA SATUNNAISMUUTTUJISTA

⚠ Samoin jakautunut ≠ sama

- Heitetään mustaa ja valkoista noppaa.
  Tulokset *M* ja *V*, kummallakin symmetrinen jakauma joukossa {1,..,6}
  Jakaumat ovat samat – esim. P(M=3) = P(V=3) = 1/6
  Muuttujat eivät ole samat – ei ole välttämättä M=V
  (Itse asiassa todennäköisyys P(M=V) on vain 1/6)

- Heitetään yhtä noppaa kolme kertaa.
  Tulokset M_1, M_2, M_3
  Nämäkin ovat samoin jakautuneet, mutta
  eivät sama satunnaismuuttuja
  – niillähän voi olla eri arvot (ja usein onkin)

⚠ Samoin jakautunut ≠ sama

- **Yhtäsuuruusmerkillä** merkitään, että satunnaismuuttujat ovat **arvoltaan** samat

  $X = Y$
- Jos sanomme tosiasiana, että X=Y, väitämme siis, että tiedämme arvojen olevan varmasti samat (tämä ei välttämättä tarkoita, että tiedämme **mikä** se arvo on)
- Jos emme tiedä, onko X=Y, voimme ehkä esittää ko. väitettä koskevia todennäköisyyksiä: $P(X=Y)$ = jotain

- **Samoin jakautuneisuus** ilmaistaan yleensä sanallisesti, tai ilmoittamalla että kummallakin on tietty sama jakauma

  $X, Y \sim \mathrm{Bin}(10, 0.5)$ tai "=" jonka päällä pieni *d* (lähinnä todennäköisyysteoriassa).

Esim 1: Kruunien määrät kahdessa eri toistokokeessa.
Esim 2: Kruunien ja klaavojen määrät yhdessä toistokokeessa.

Riippumattomat sm:t (Tuominen s. 67-74)

- Tunnemme käsitteen ”riippumattomat **tapahtumat**”
- **Satunnaismuuttujat** X ja Y ovat riippumattomat, jos:

  Jokainen väite muotoa ”X kuuluu tiettyyn joukkoon A” on riippumaton jokaisesta väitteestä ”Y kuuluu tiettyyn joukkoon B”

Tällöin esim. $P(X=3 \mid Y=4) = P(X=3)$

ja yleisemmin $P(X=a \mid Y=b) = P(X=a)$ kaikilla a, b

ts. tieto Y:n arvosta ei muuta käsitystä X:n arvosta. Riippumattomien sm:ien ominaisuuksiin palataan myöhemmin.

Esimerkki: *Kahden nopanheiton tulokset X ja Y.*

- **Usean** satunnaismuuttujan riippumattomuus määritellään samoin kuin usean tapahtuman, ts. tieto useastakaan muuttujasta ei anna tietoa muista

Esimerkki: *Kymmenen nopanheiton tulokset* $X_1, X_2, ..., X_{10}$.

Riippumattomuus ei ole sama kuin samoin jakautuneisuus

X = nopanheiton tulos ja Y = kolikonheiton tulos.

Riippumattomia, mutta **eri jakauma**

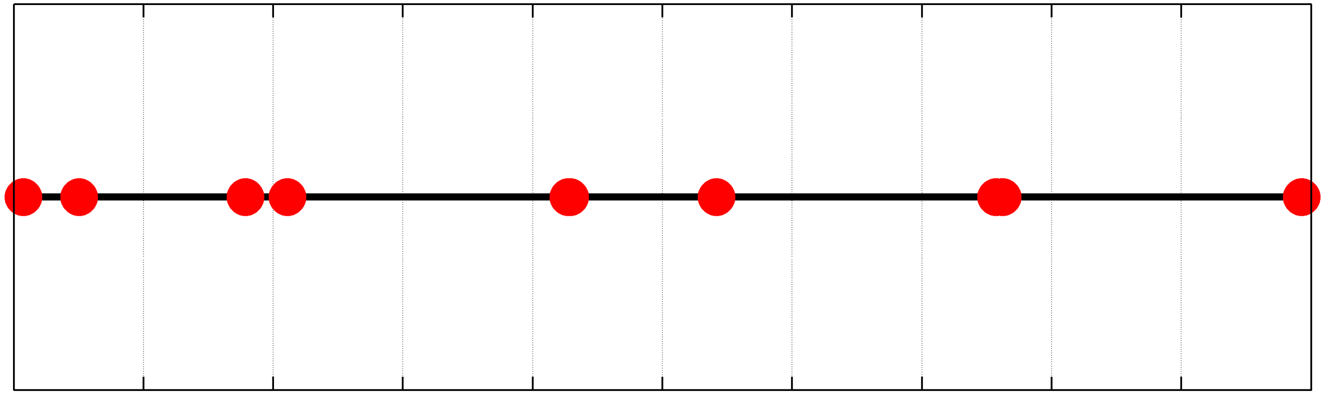
(Vrt. Elmerin tennispelit, M-peleissä eri voittotn kuin I-peleissä)

X = 1. jaetun kortin arvo ja Y = 2. jaetun kortin arvo.

Molemmilla sama jakauma, mutta **riippuvia**

Toistokokeessa tulokset $X_1,...,X_n$ ovat *riippumattomat* ***ja*** *samoin jakautuneet*

Merk. **iid** ”independent and identically distributed”

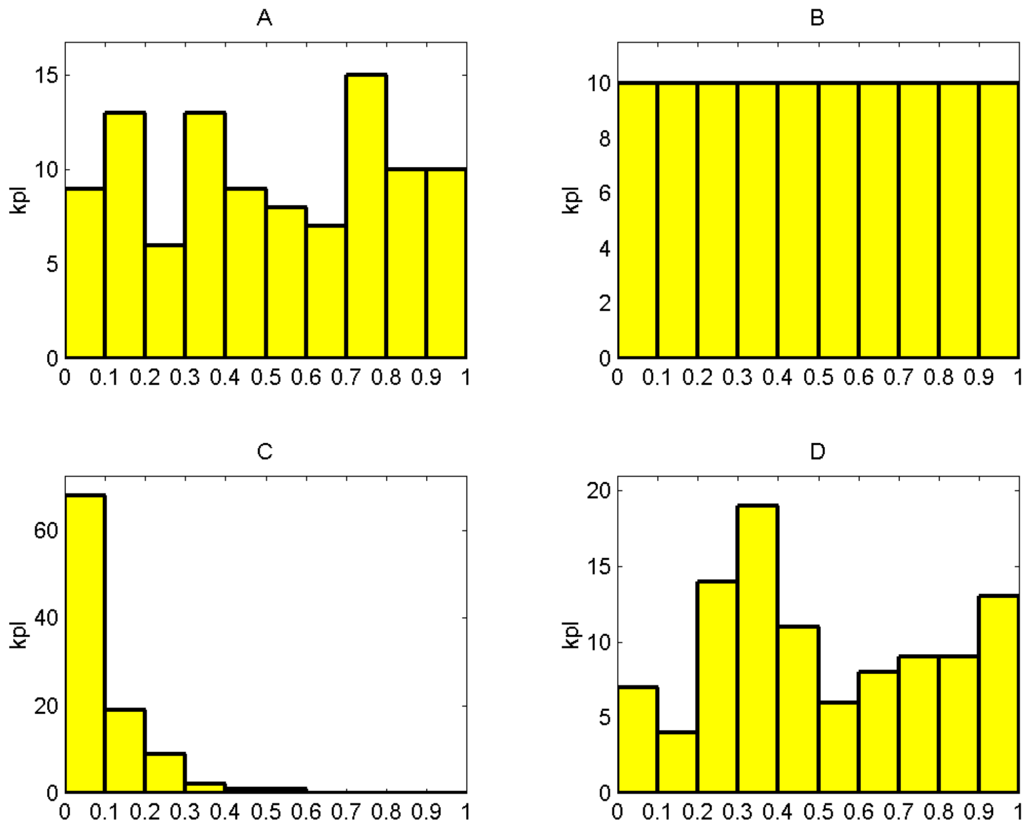


RIIPPUMATON OTOS TASAJAKAUMASTA

Tunnistustehtävä

Kukin näistä histogrammeista on piirretty sadasta luvusta.

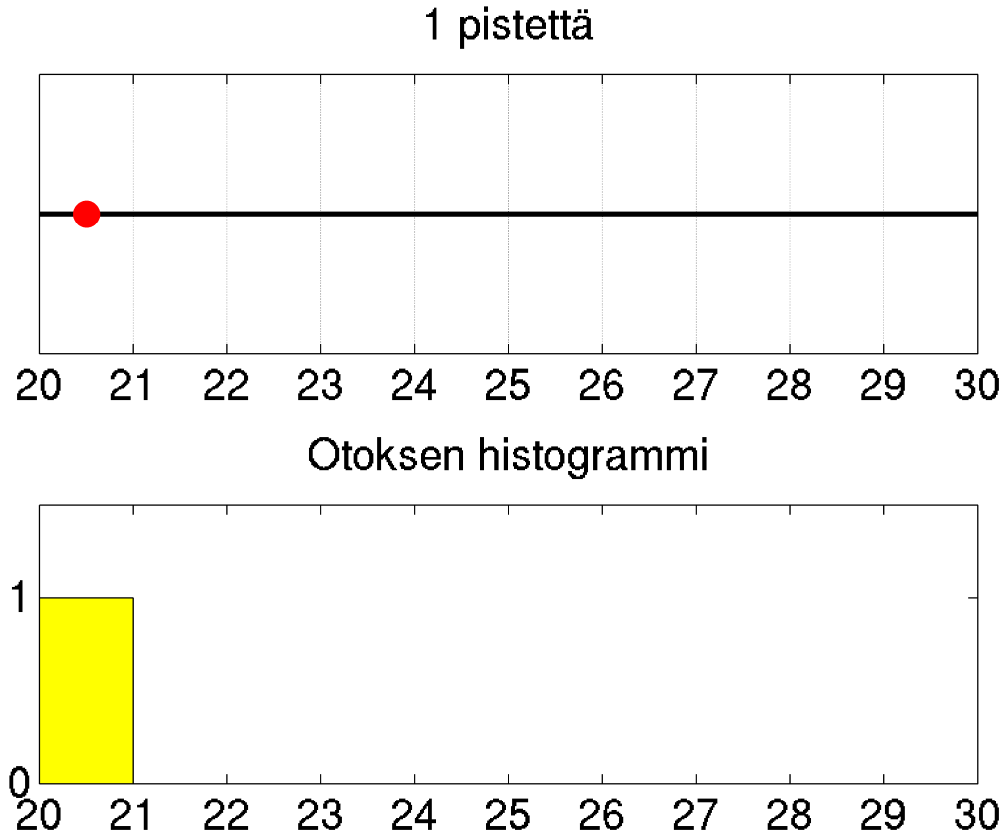
Mikä histogrammeista on syntynyt siten, että luvut on arvottu jakaumasta **Tas(0, 1)** ?



Otos tasajakaumasta

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

Miten ne sijoittuvat?



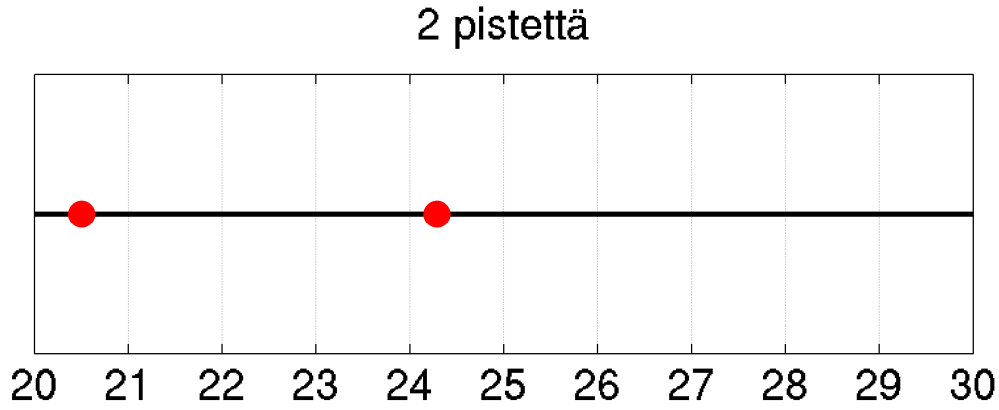
Piirretään myös 10 pylvään histogrammi, ei **jakaumasta** (se on tasainen), vaan **otoksesta**, ts. mille väleille arvotut luvut osuivat.

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

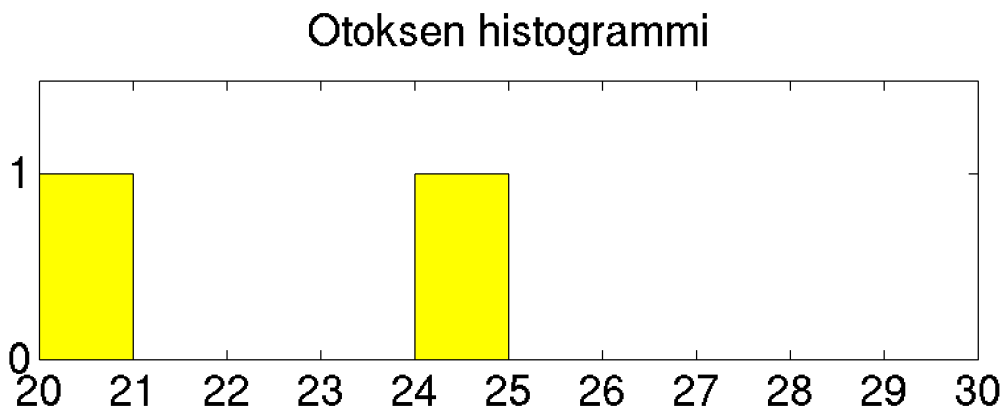
Otos tasajakaumasta

Arvotaan riippumattomia satunnaislukuja X_1, X_2, X_3 ... ~ Tas(20, 30).

Miten ne sijoittuvat?



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

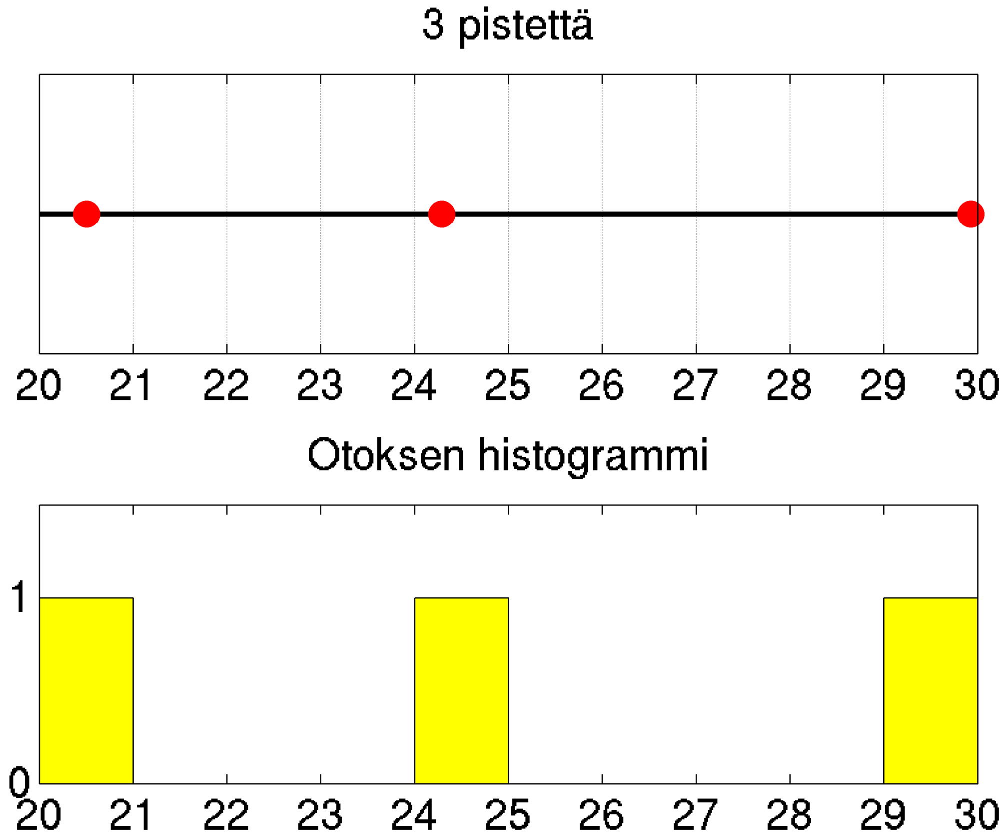


Otos tasajakaumasta

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

Miten ne sijoittuvat?

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



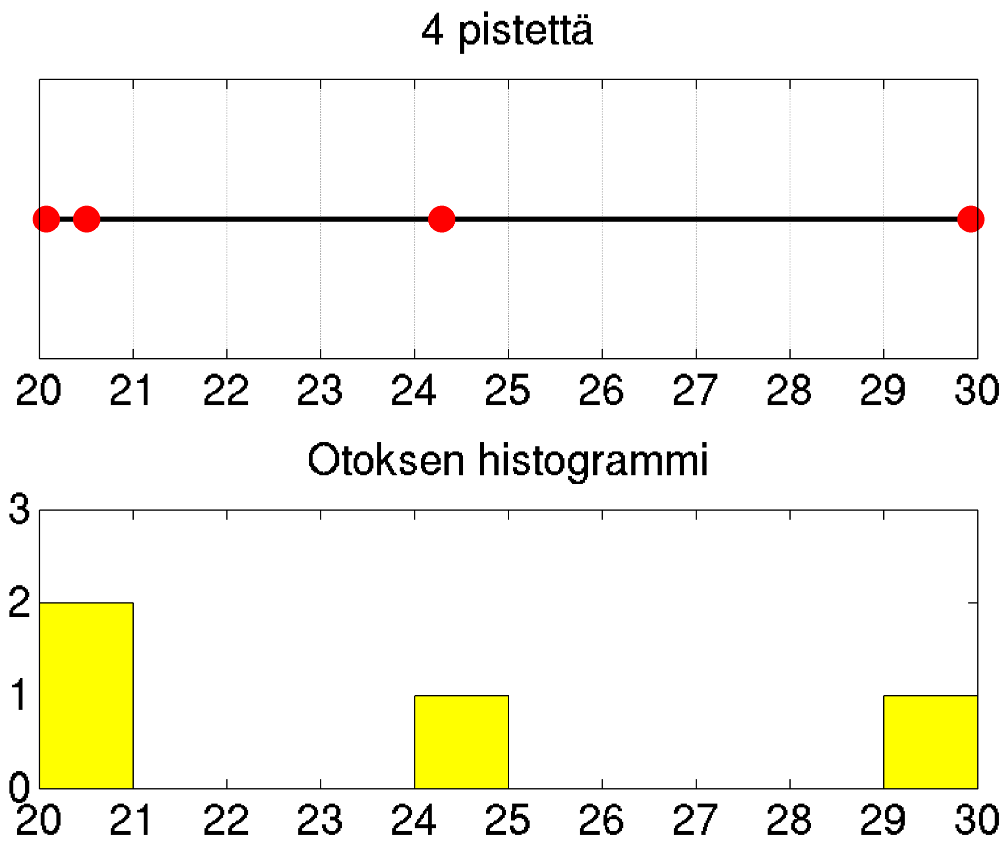
Otos tasajakaumasta

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

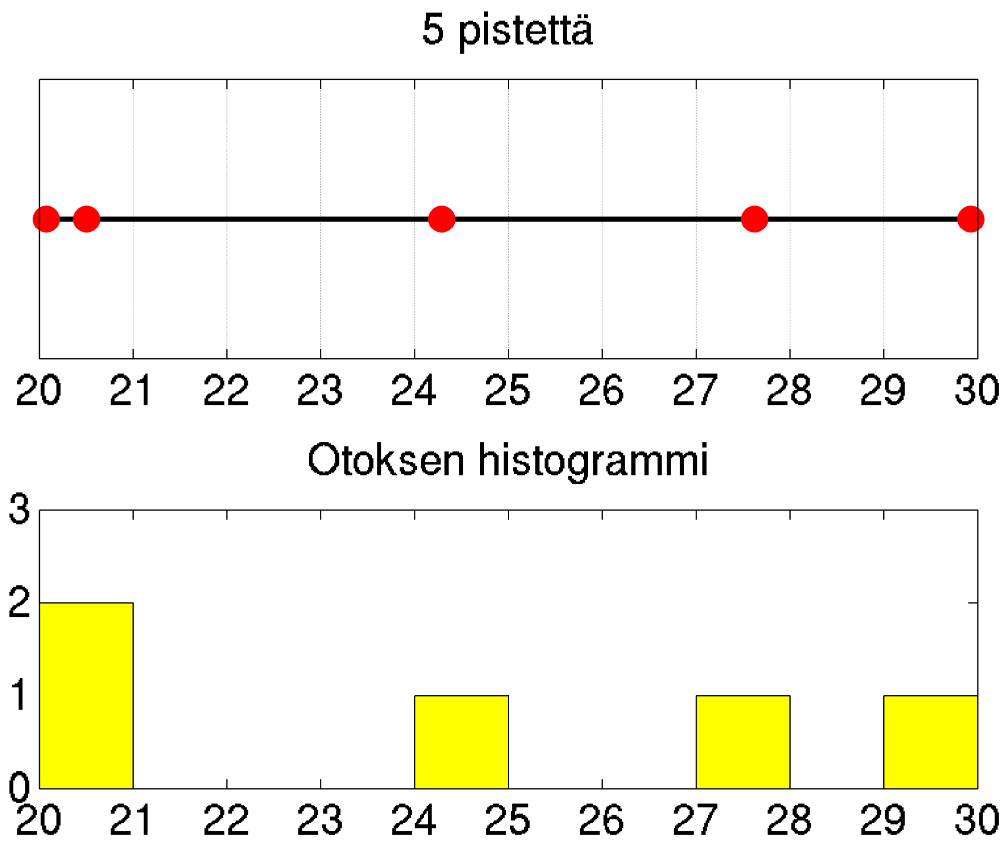
Riippumattomuus:
Aiemmat pisteet eivät vaikuta myöhempiin.

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

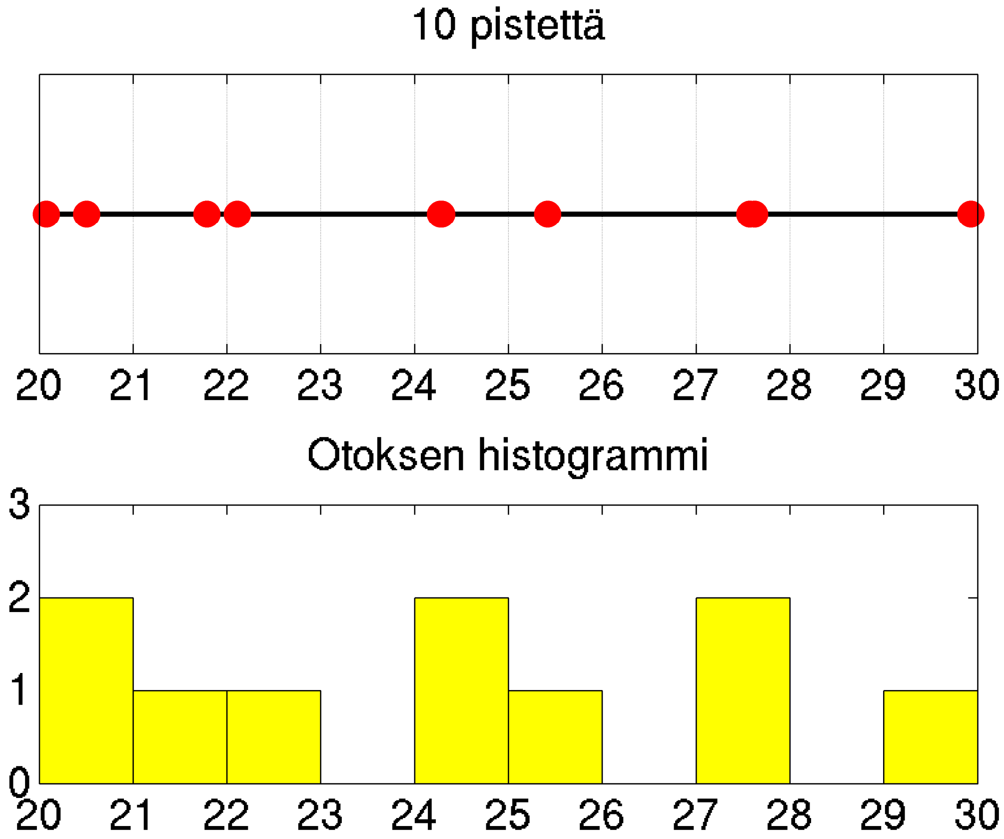
(ei edes lisää eikä vähennä ko. tapahtuman tn:ää, joka on jatkuvasti 1/10)



Otos tasajakaumasta



Otos tasajakaumasta

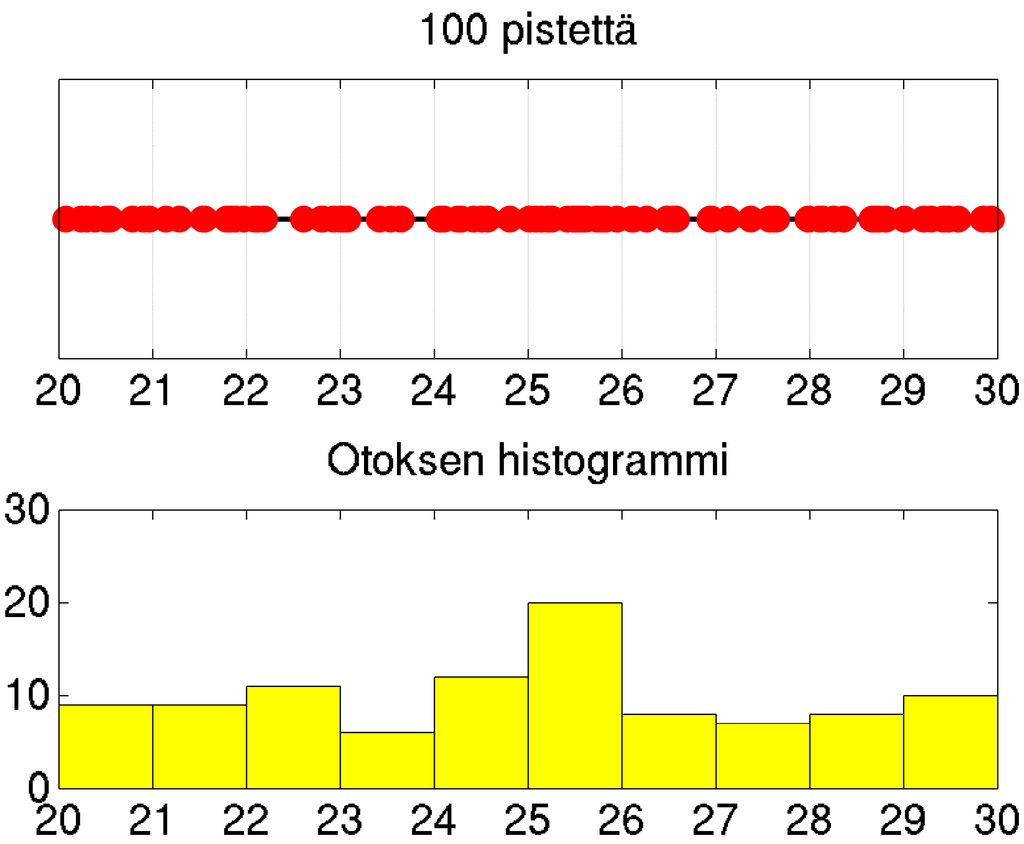


Otos tasajakaumasta

100 pistettä:

Kunkin pylvään **korkeuden odotusarvo** on 10,   (miksi?)

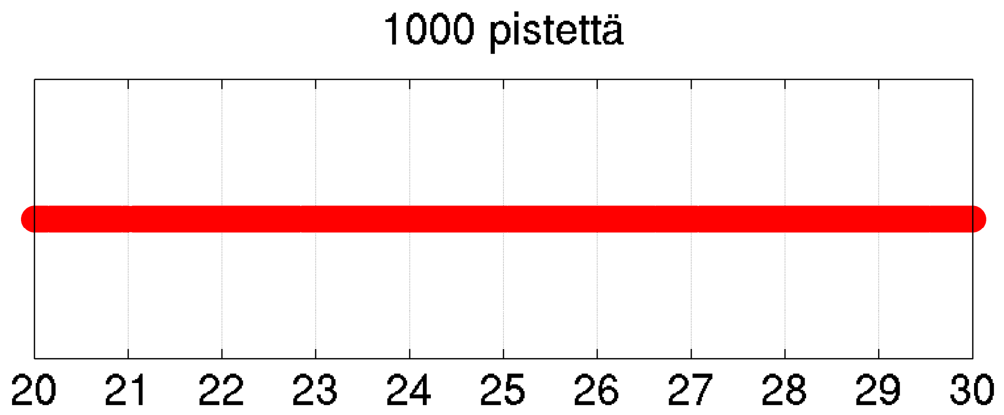
mutta toteutuneet korkeudet vaihtelevat melkoisesti.

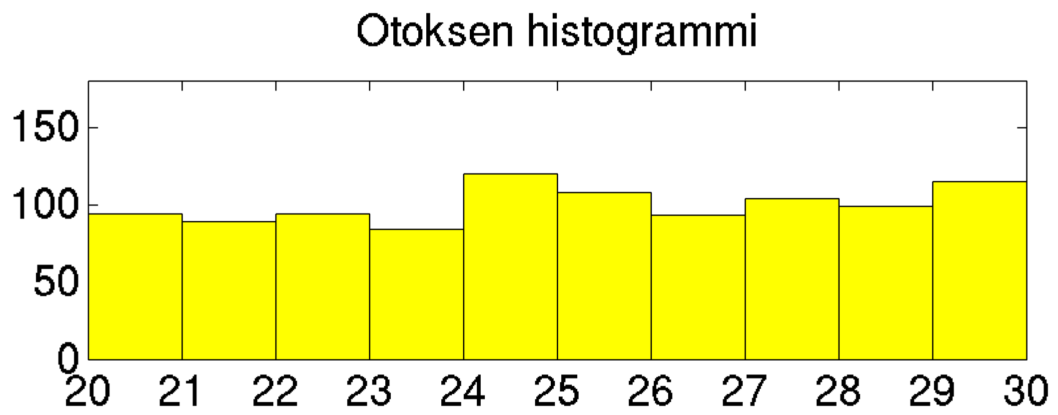


Otos tasajakaumasta

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

mutta histogrammista näemme osapuilleen pisteiden jakauman.

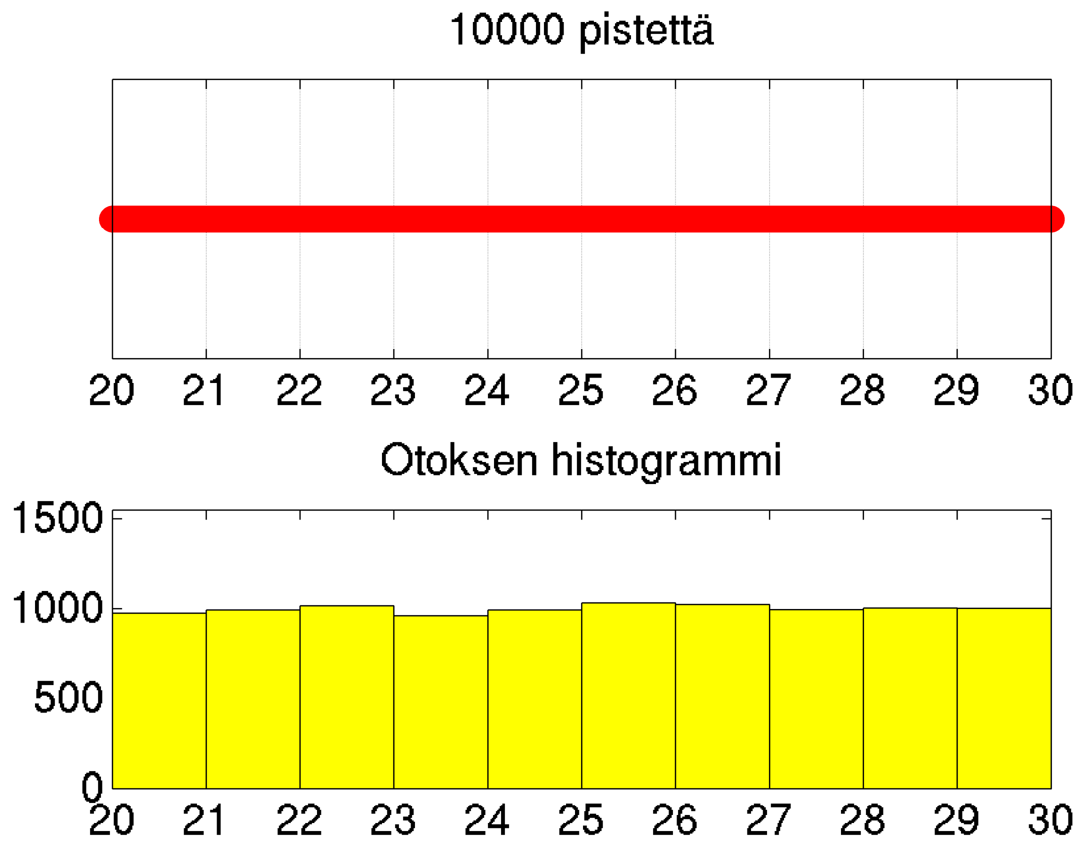




Otos tasajakaumasta

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 <u>eivät ole riippumattomia</u>.

Tarvitaan ns. multinomijakauma (josta lisää seuraavassa luennossa).