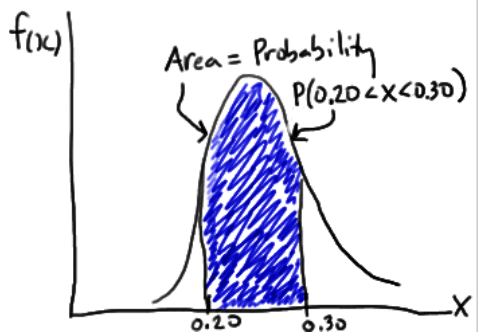
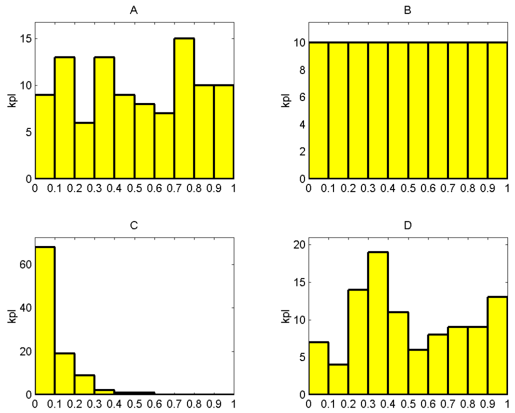
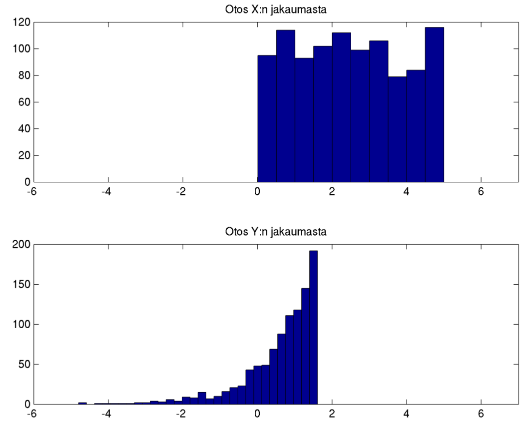
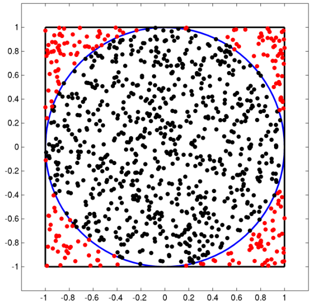
Luento 6 – sekalaisia aiheita



- Samoin jakautunut ei tarkoita yhtäsuurta
- Diskreetin summan jakauma: 2 näkökulmaa, analyysi ja empiria
- Jatkuvien jakaumien käsittelyn kertausta
- Empiirinen jakauma histogrammina: Mitä siitä näkee, millainen se voi olla?
- Miten muunnoksen jakaumaa käsitellään?
- Monte Carlo –integrointi







Samoin jakautunut ei tarkoita yhtäsuurta

- 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 ei tarkoita yhtäsuurta

- **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) = ?

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

  X, Y ~ Bin(10, 0.5) tai ”=” jonka päällä pieni *d* (lähinnä todennäköisyysteoriassa).

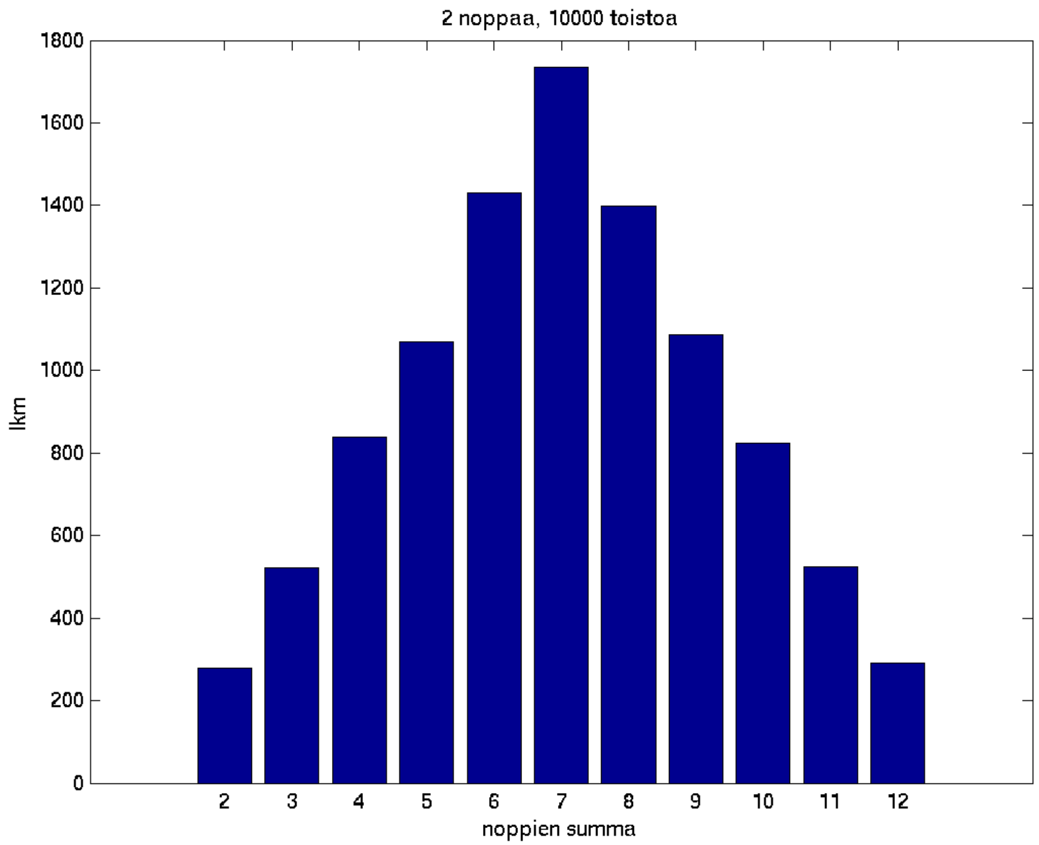
Esim 1: Kruunien määrät kahdessa eri toistokokeessa.
Esim 2: Kruunien ja klaavojen määrät yhdessä toistokokeessa.

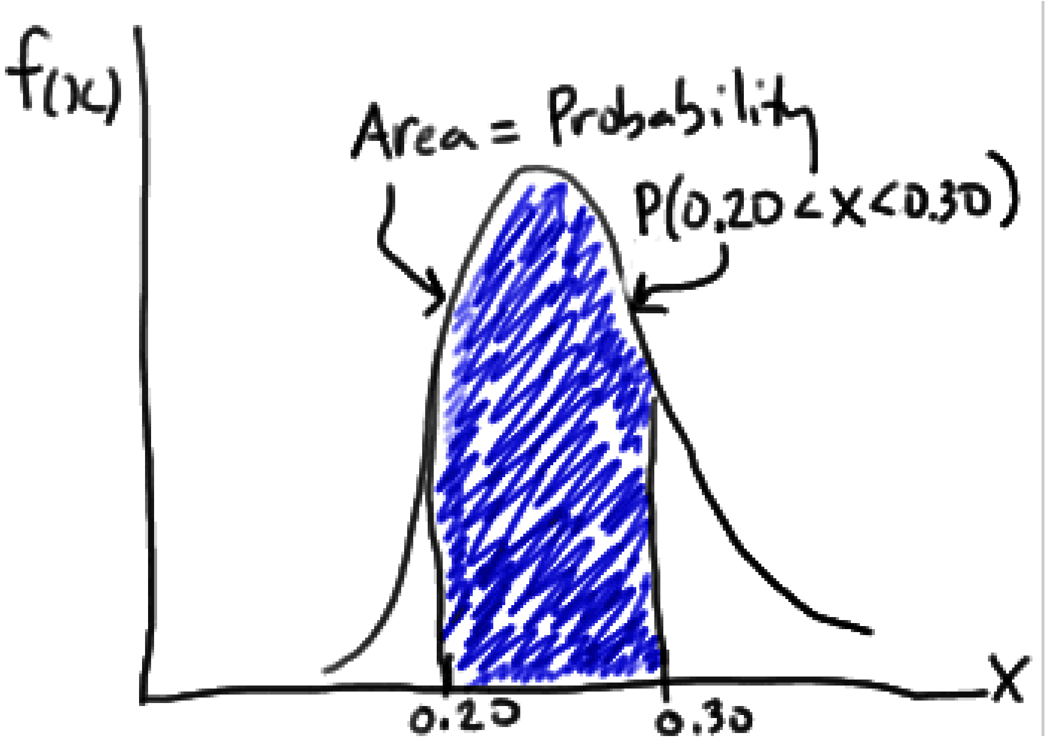
Kaksi noppaa, eri summien tn:t

Noppien summa	Kyseisen tyyppiset toteumat perusjoukossa	Toteumia 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)
```





JATKUVA SATUNNAISMUUTTUJA

Tiheysfunktio

- Jos X:n tiheysfunktio on f,
  niin f(x) **ilmaisee verrannollisuuskertoimen**:
  tn osua (lyhyelle) välille (x:n lähellä) on välin pituus kertaa f(x).

- Huom: f(x) **ei ole todennäköisyys**, että X=x
- Esim. ennuste lämpötilasta, joka tasajakautunut välillä (20, 30)
- Koska f(x) ei ole todennäköisyys, se voi olla suurempikin kuin 1. (Mitä tämä tarkoittaa?)

Kertymäfunktio

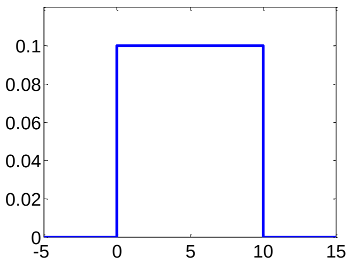
- Vaikka tiheysfunktio antaa **intuitiivisemman** käsityksen jakaumasta (mihin jakauma keskittyy), kertymäfunktio on keskeinen **laskennallinen** työkalu jakaumien käsittelyssä.
- F(x) vastaa kysymykseen
  ”mikä on tn, että $X \leq x$”.
- Esim. lämpötilalle F(25) = ½ tarkoittaa, että on ½ todennäköisyys, että lämpötila on enintään 25 astetta.

- Kertymäfunktion on pakko olla kasvava – miksi?
- Kertymäfunktio kasvaa arvojoukossa vasemmalta 0:sta oikealle 1:een – miksi?
- Kertymäfunktio on yleensä tf:n integraali, ja tf on kertymäfunktion derivaatta – miksi?

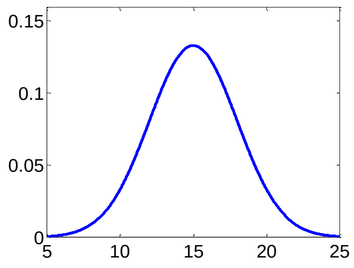
Jatkuvien jakaumien käsittelyä

Tiheys- ja kertymäfunktio; E(X)

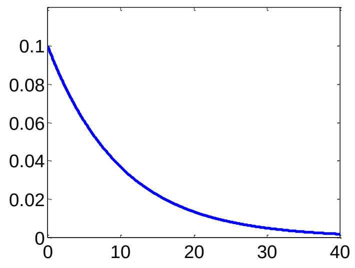
- Käydään läpi muutamia jatkuvia jakaumia



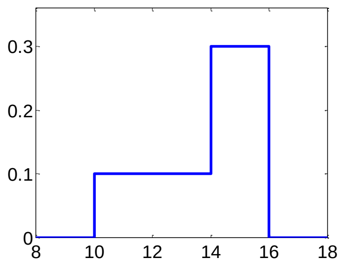
Tas(*a*, *b*)



N(μ, σ^2)



Exp(λ)

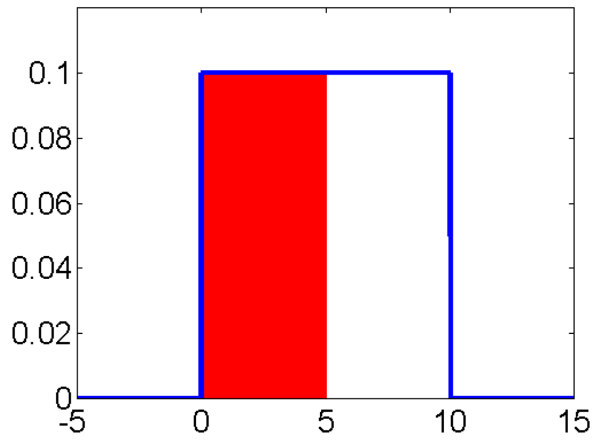


Mielivaltainen tiheysfunktio

- Miten lasketaan tietyn **välin todennäköisyys**
  - Tiheydestä: integroimalla
  - Kertymästä: vähennyslaskulla

- Miten lasketaan tai todetaan jakauman **odotusarvo**
  - Tiheydestä: integroimalla

Tasajakauma



- Bussin odotusaika **$X \sim$ Tas(0, 10)**

- Tiheys $f(x) = 1/10$ kun $0<x<10$
- Kertymä $F(x)= x/10$ kun $0<x<10$

Huom. $F'(x)=f(x)$

- Välien todennäköisyydet integroimalla tiheyttä (vakiota) ko. välin yli, **tai suoraan kertymäfunktiosta:**

  $P(2 < X < 5) = F(5) - F(2) = 0.5 - 0.2 = 0.3$

  $P(X > 7) = 1 - F(7) = 1.0 - 0.7 = 0.3$

  $P(X < 5) = F(5) = 0.5$

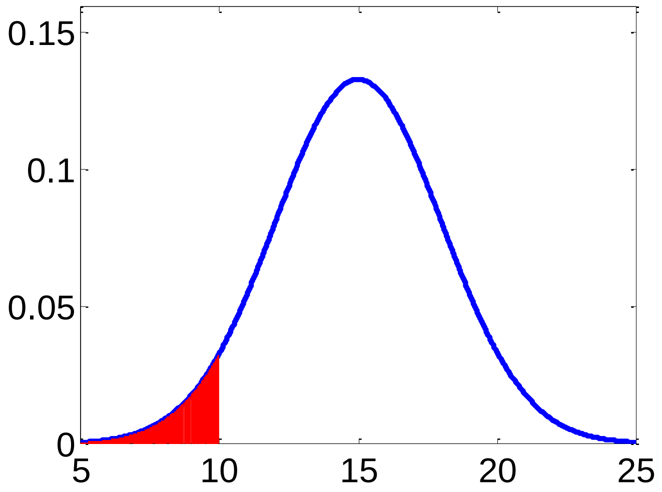
- Tasajakauman odotusarvolle on tunnettu lauseke (= päätepisteiden keskiarvo)

  $E(X) = (0+10) / 2 = 5.0$

  Mutta se voidaan myös laskea raa'asti integraalina:

  $$E(X) = \int_{-\infty}^{\infty} x f(x)dx = \int_{0}^{10} x \cdot 0.1 dx = 5.0$$

Normaalijakauma



- Bussin matka-aika **$X \sim N(15, 3^2)$**

  Parametrit μ=15, σ=3

- Tiheys $f(x) = c \cdot \exp(- \ldots)$
- Kertymä $F(x) = \Phi((x - 15) / 3)$

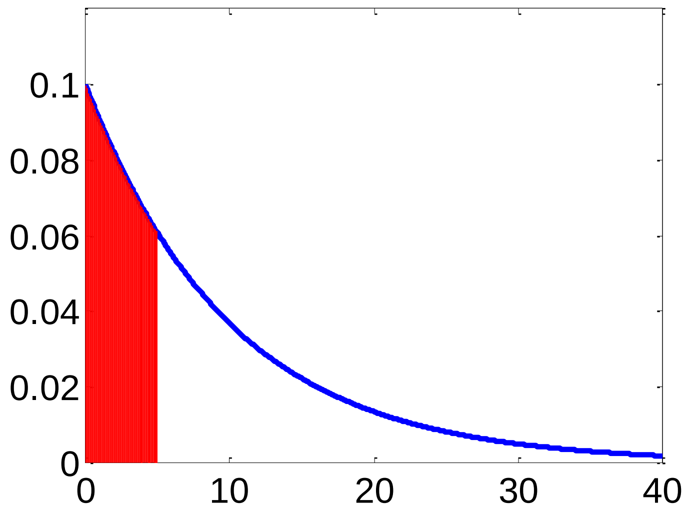
- Tiheysfunktiota hankala integroida joten käytetään kertymäfunktiota. Välien todennäköisyydet suoraan kertymäfunktiosta:

$P(10 < X < 15)$	$= F(15)$	$- F(10)$	$= 0.500 - 0.048$	$= 0.452$
$P(X > 20)$	$= 1$	$- F(20)$	$= 1.000 - 0.952$	$= 0.048$
$P(X < 10)$	$= F(10)$		$= 0.048$	

- Normaalijakauman odotusarvolle on tunnettu lauseke (= eka parametri μ)

  $E(X) = 15$

Exp-jakauma



- Lampun kestoaika **$X \sim$ Exp(0.1)**

  Parametri λ=0.1

- Tiheys $f(x)$ = 0.1 exp(–0.1x) kun x>0
- Kertymä $F(x)$= 1 – exp(–0.1x) kun x>0 Huom. F'(x)=f(x)

- Välien todennäköisyydet integroimalla tiheyttä ko. välin yli,
  tai suoraan kertymäfunktiosta:

  P(5 < X < 10) = F(10)− F(5) = 0.632 − 0.394 = 0.239

  P(X > 20) = 1 − F(20) = 1.000 − 0.865 = 0.135

  P(X < 5) = F(5) = 0.394

- Eksponenttijakauman odotusarvolle on tunnettu lauseke (= 1 / λ)

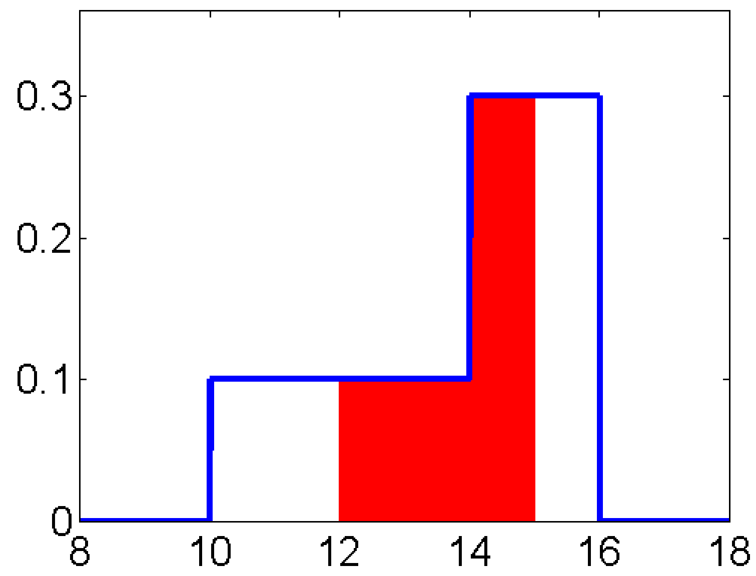
  E(X) = 1 / 0.1 = 10

  Sen voisi laskea myös integraalina (tarvitaan osittaisintegrointia)

Yleinen tiheysfunktio

- Meteorologi kertoo, että huomisaamun lämpötilalla *X* on tiheysfunktio

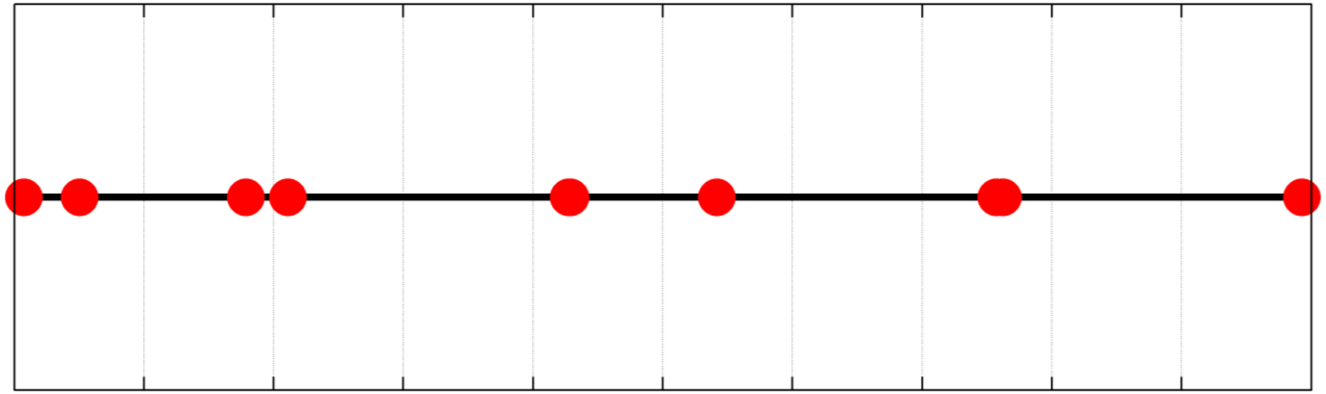
$$f(x) = \begin{cases} 0.1 & \text{kun } 10 \le x < 14 \\ 0.3 & \text{kun } 14 \le x < 16 \\ 0 & \text{muualla} \end{cases}$$



- Todennäköisyydet ja odotusarvo saadaan integroimalla

$$P(12 < X < 15) = \int_{12}^{15} f(x)dx = \int_{12}^{14} 0.1\,dx + \int_{14}^{15} 0.3\,dx = 0.2 + 0.3 = 0.5$$

$$E(X) = \int_{10}^{16} x \cdot f(x)dx = \int_{10}^{14} x\,0.1\,dx + \int_{14}^{16} x\,0.3\,dx = 4.8 + 9.0 = 13.8$$

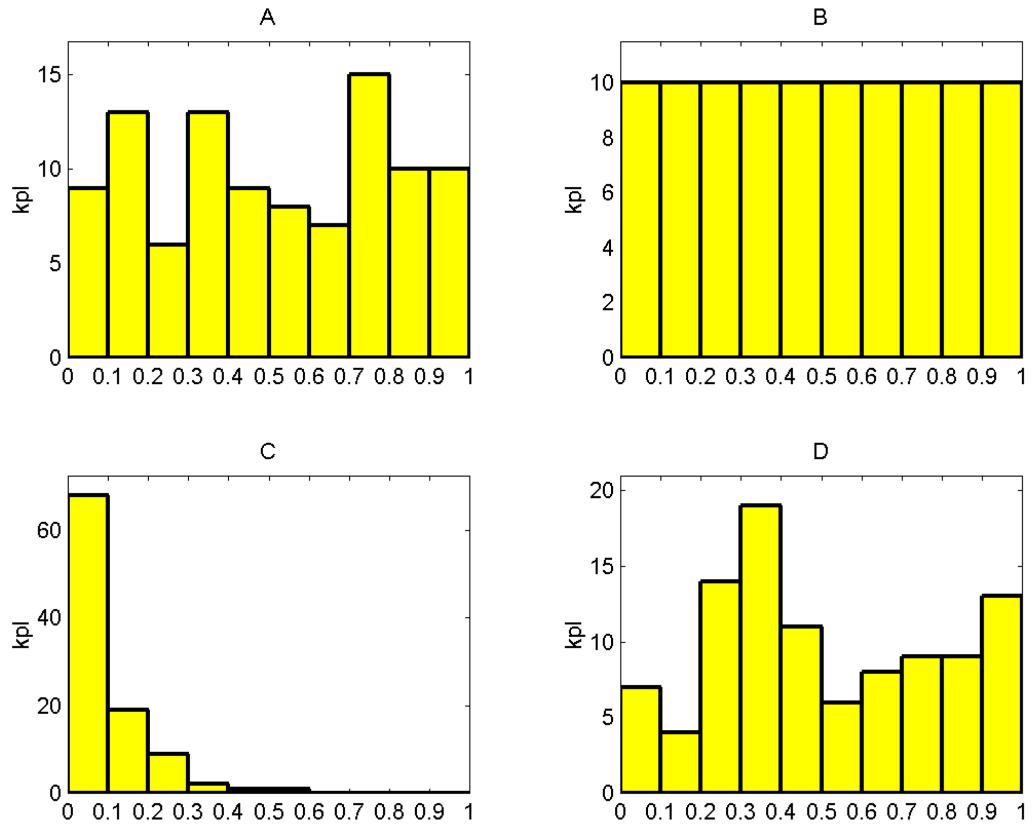


RIIPPUMATON OTOS TASAJAKAUMASTA

Tunnistustehtävä

Kukin näistä histogrammeista on piirretty sadasta luvusta.

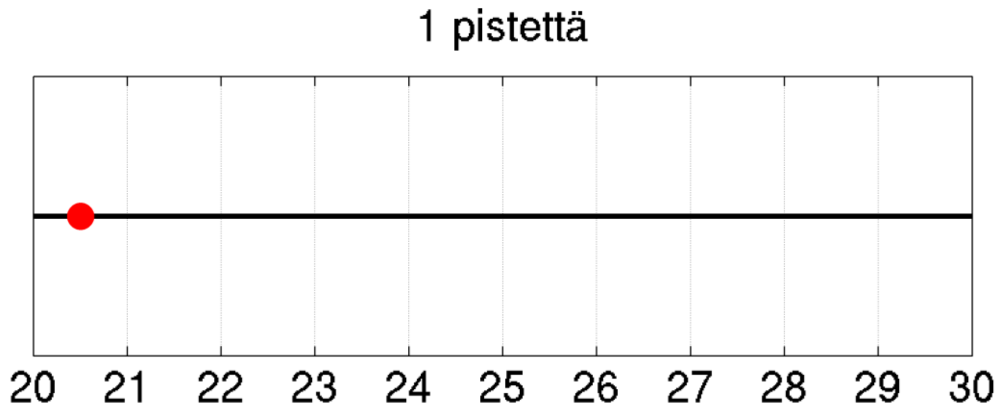
Mikä histogrammeista on syntynyt siten, että luvut on arvottu jakaumasta **Tas(0, 1)** ?



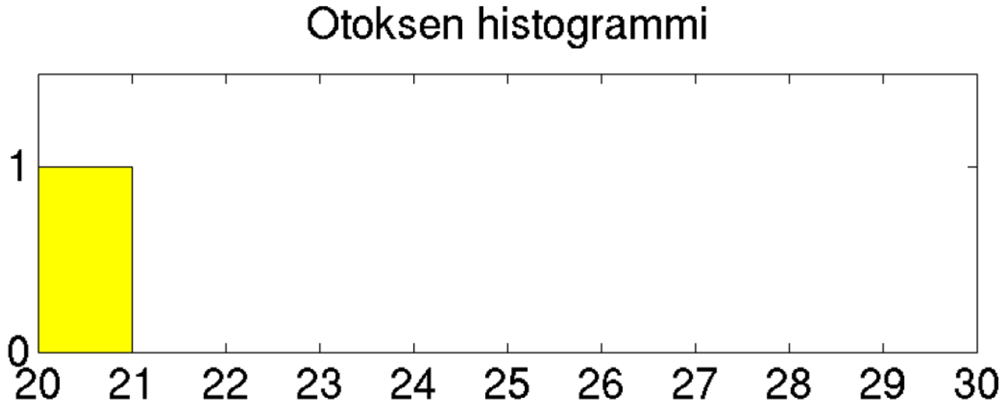
Otos tasajakaumasta

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

Miten ne sijoittuvat?



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



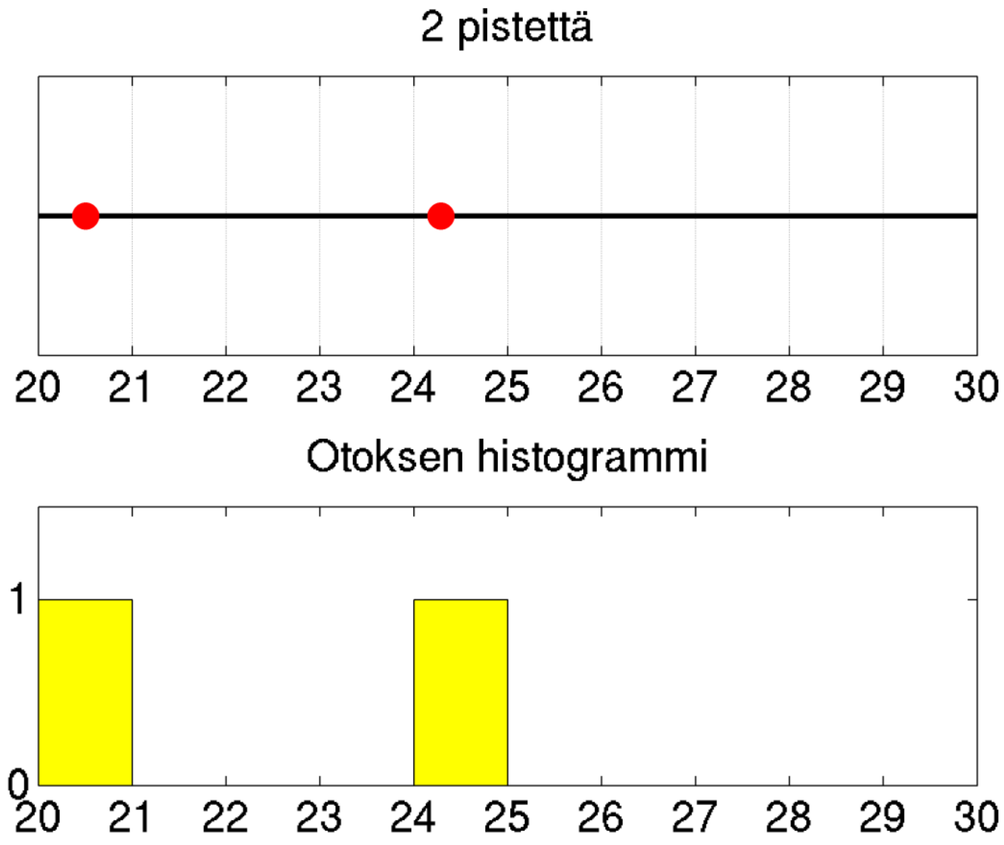
(Histogrammi on yhteenveto siitä, missä pisteet *suunnilleen* ovat.)

Otos tasajakaumasta

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

Miten ne sijoittuvat?

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

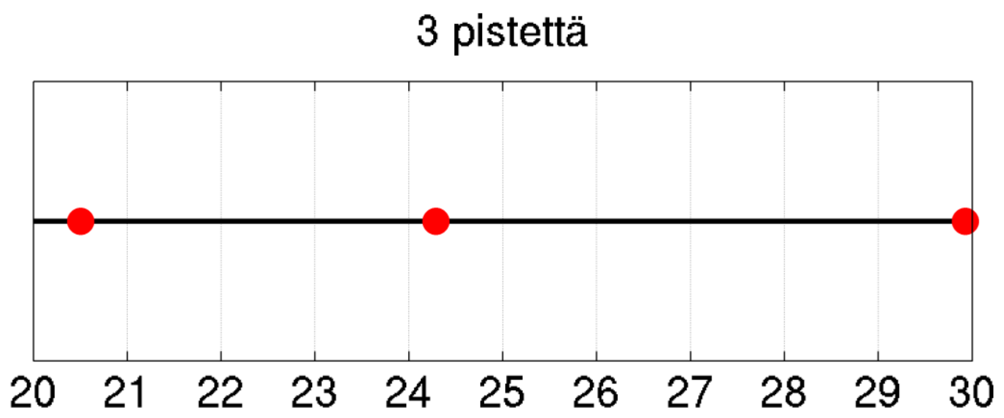


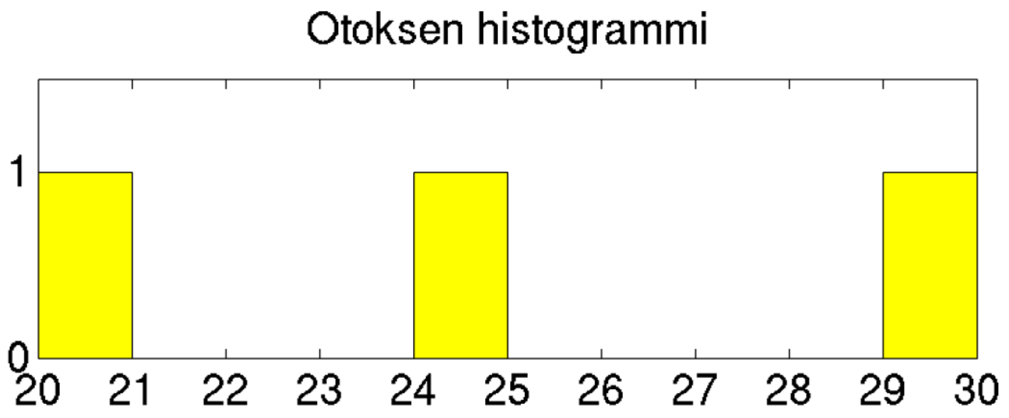
Otos tasajakaumasta

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

Miten ne sijoittuvat?

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





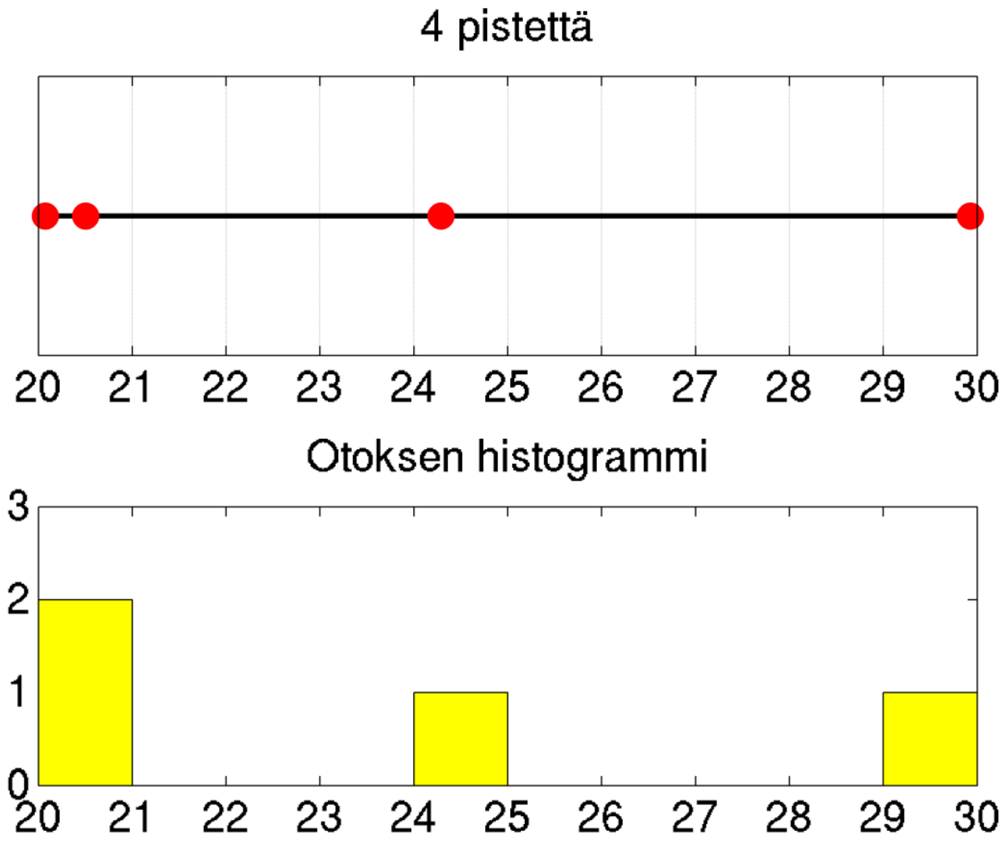
Otos tasajakaumasta

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

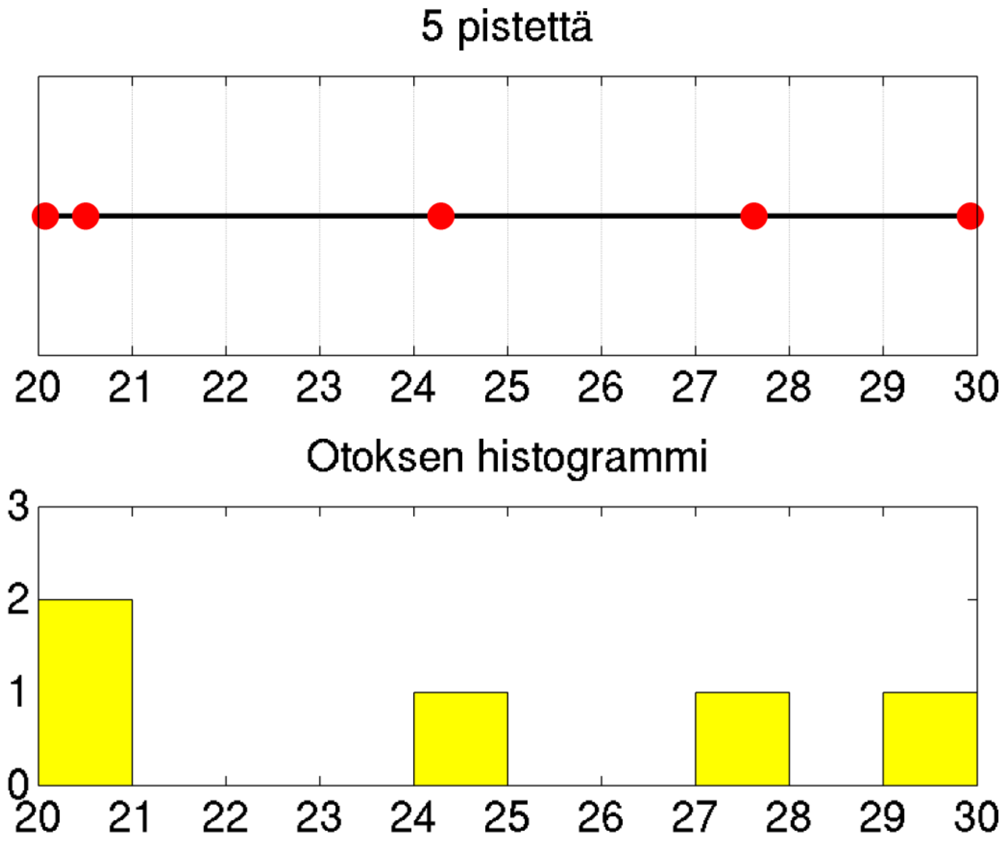
Riippumattomuus:
Aiemmat pisteet eivät vaikuta myöhempiin.

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

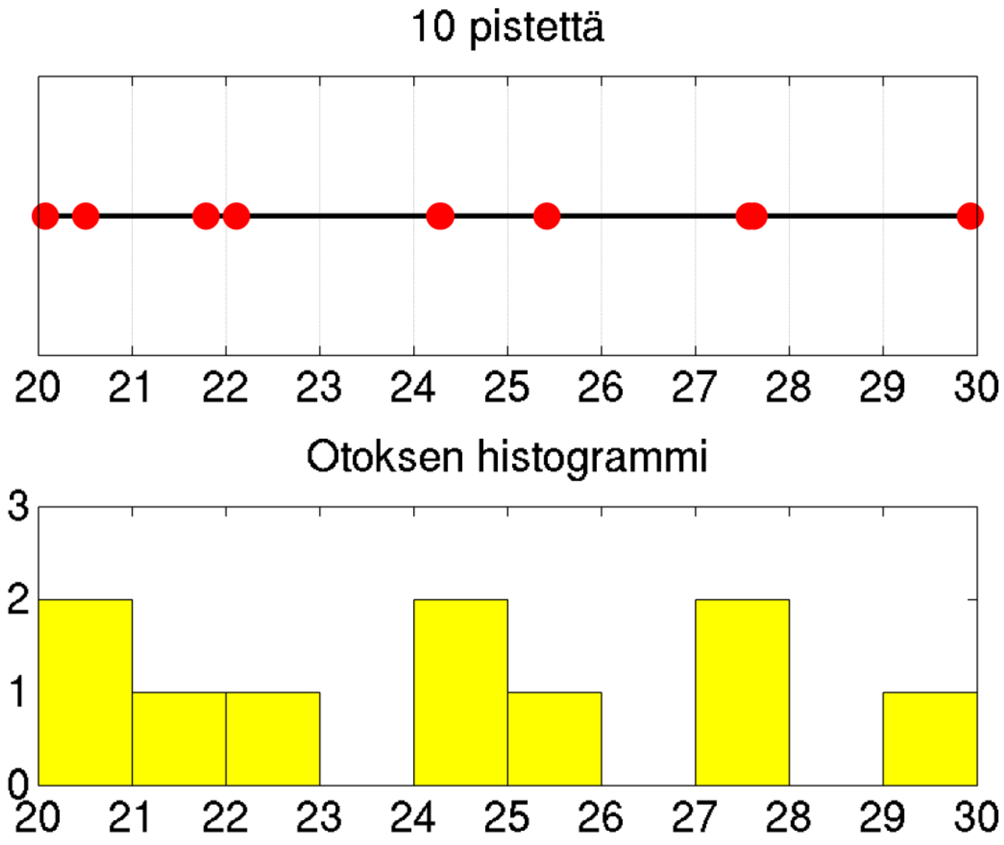
(ei edes lisää eikä vähennä ko. tapahtuman tn:ää, joka on jatkuvasti 1/10)



Otos tasajakaumasta



Otos tasajakaumasta

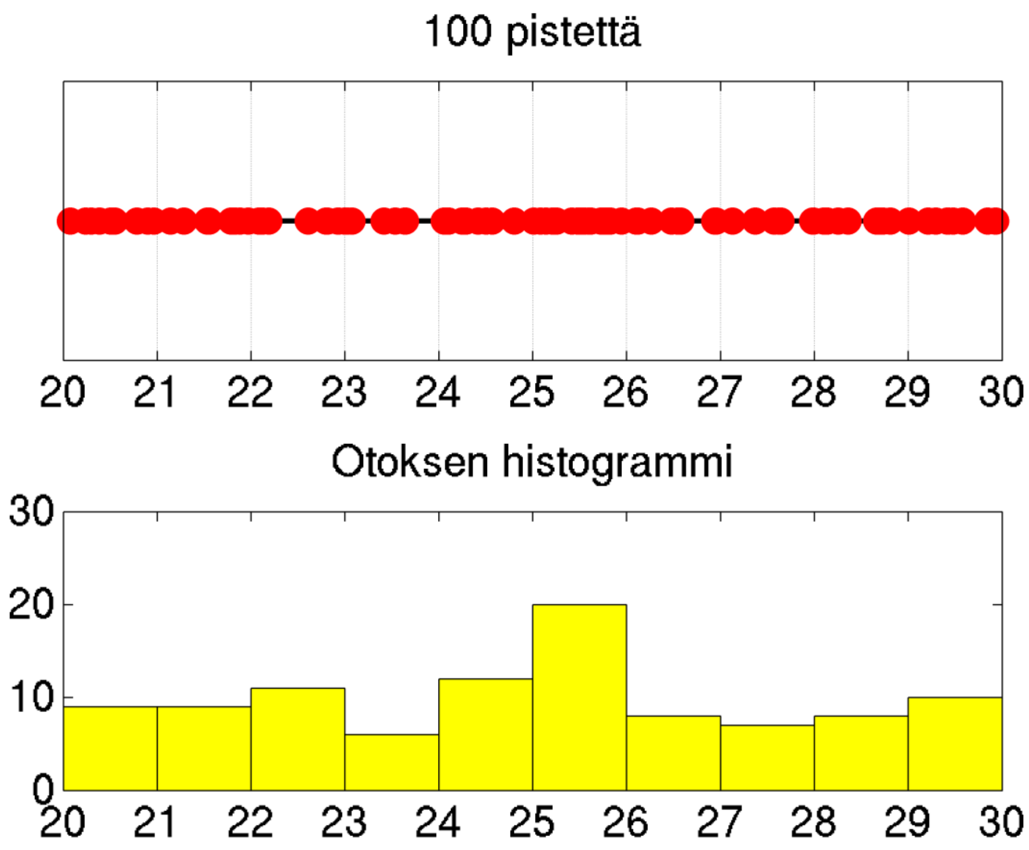


Otos tasajakaumasta

100 pistettä:

Kunkin pylvään
korkeuden odotusarvo
on 10,   (miksi?)

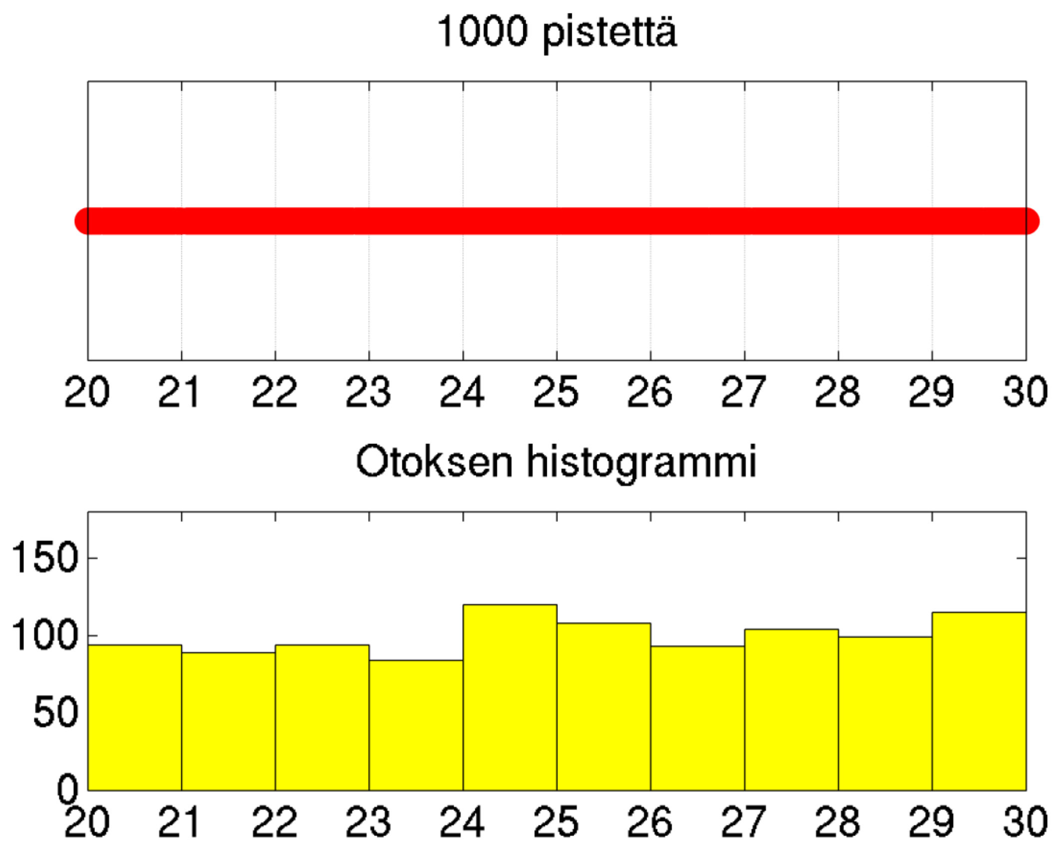
mutta toteutuneet korkeudet
vaihtelevat melkoisesti.



Otos tasajakaumasta

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

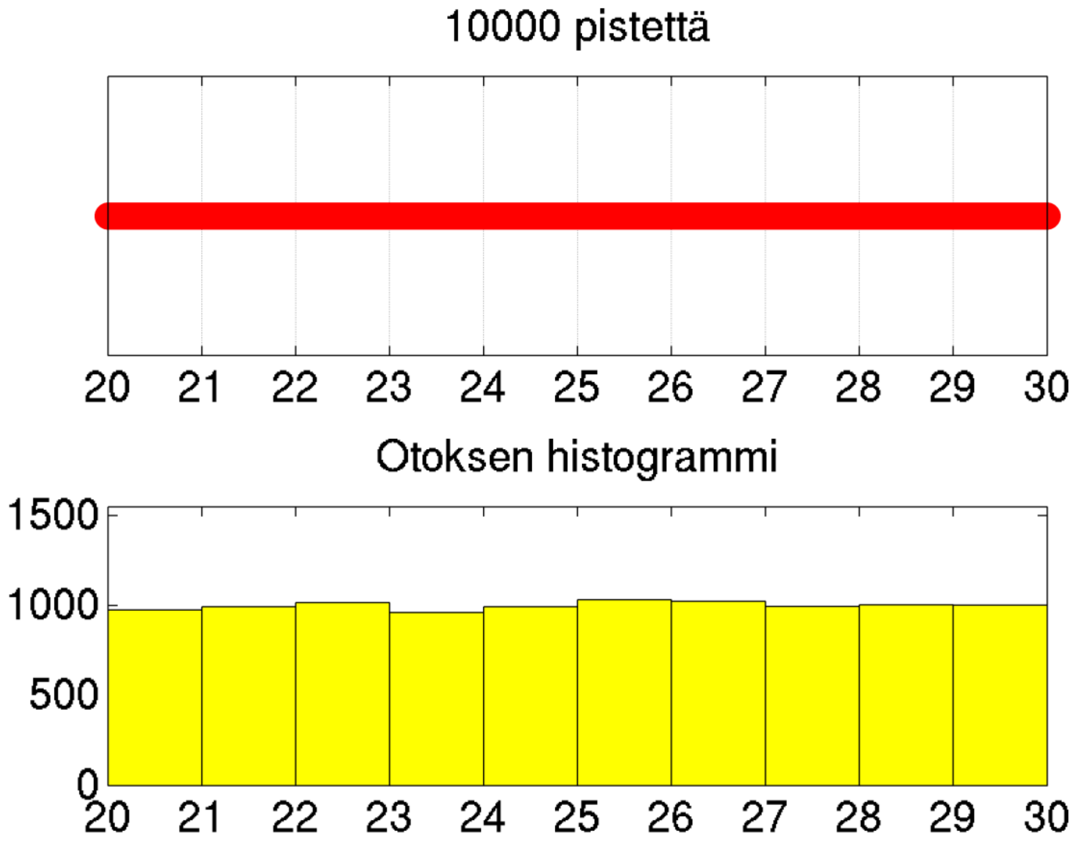
mutta histogrammista näemme
osapuilleen
pisteiden jakauman.



Otos tasajakaumasta

10 000 pistettä:

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



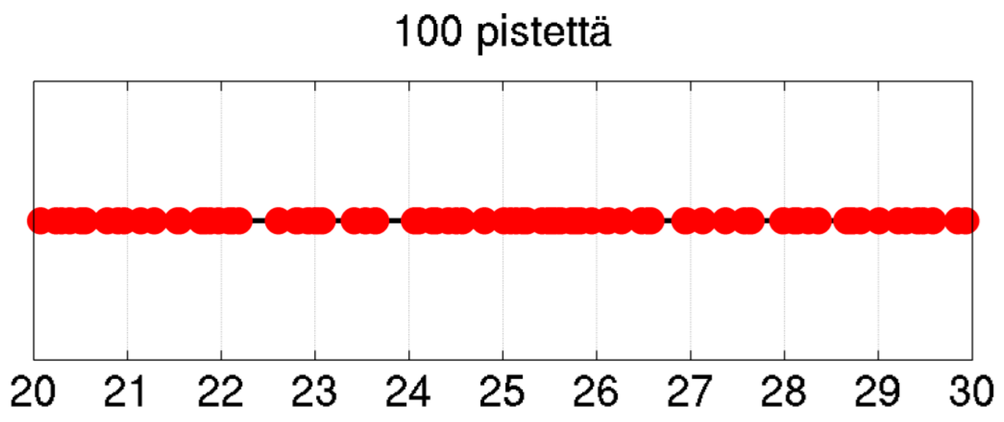
Pylväiden korkeuksien jakauma

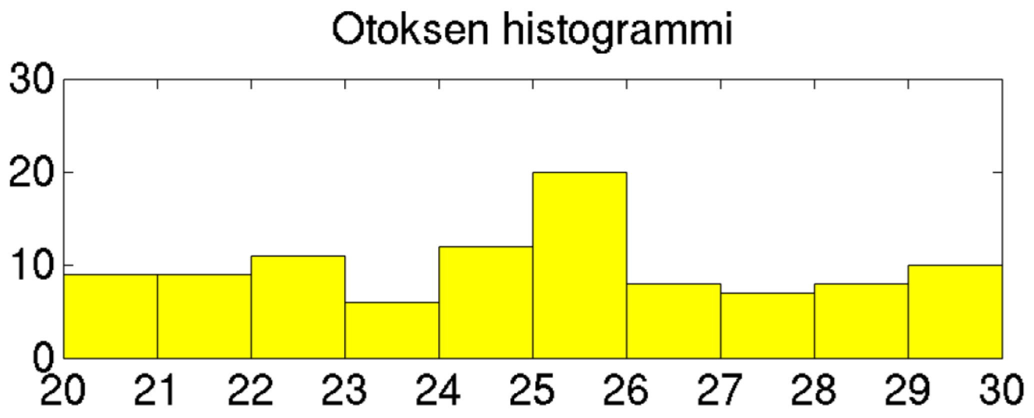
Mennään takaisin 100 pisteeseen.
Merk.
Y_i = i:nnen pylvään korkeus
= i:nnelle jakovälille osuvien pisteiden lukumäärä

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

Tn, että histogrammi on täysin tasainen?
Huomaa, että pylväiden korkeudet <u>eivät ole riippumattomia</u>.

Tarvitaan ns. multinomijakauma.





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.

Pylväiden korkeuksien jakauma

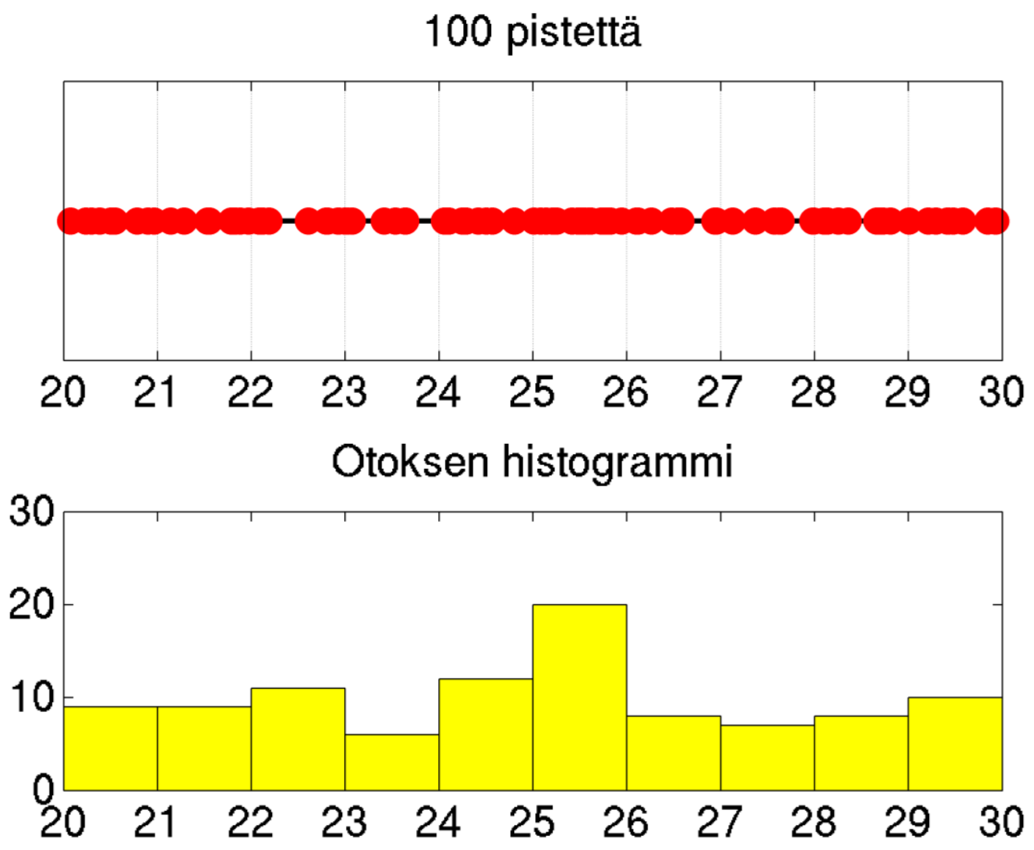
Merk.
Y_i = i:nnen pylvään korkeus
= i:nnelle jakovälille osuvien pisteiden lukumäärä

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

Tn, että histogrammi on täysin tasainen? Huomaa, että pylväiden korkeudet <u>eivät ole riippumattomia</u>.

Jos esim. ensimmäiseen väliin on osunut peräti 80 pistettä, niin ei muille väleille enää riitä kuin 20 pistettä: vahva riippuvuus.

Kyseessä on **multinomijakauma**.



Esim. 100 pistettä, 10 jakoväliä

- Jotta histogrammi olisi täysin tasainen, eri jakoväleille osuneiden pisteiden määrien (vrt. puolueiden kannattajat) täytyy olla täsmälleen (10,10,10,…,10).
- Tämän tn = $\binom{100}{10,10,\ldots,10} \cdot 0.1^{10} \cdot 0.1^{10} \cdot \ldots \cdot 0.1^{10} \approx 2.4 \cdot 10^{-8}$
- eli aika vähän.
- Toisaalta voitaisiin laskea tn saada melko tasainen histogrammi, esim. tn että jokaisen pylvään korkeus on välillä [7, 13]. Miten?
- Vaikka täysin tasainen histogrammi toteutuu harvoin (pienellä tn), ”melko” tasainen toteutuu suurella tn.

”Melko” tasaisen histogrammin toteutuminen

- Voitaisiin laskea tn saada melko tasainen histogrammi, esim. tn että jokaisen pylvään korkeus on välillä [7, 13]. Miten?
- Esim. käymällä läpi kaikki sellaiset lukujonot (pylväiden korkeuksia) jotka ovat näissä rajoissa, kuten (9,8,11,7,12,10,10,12,8,13) ja laskemalla joka kerta todennäköisyys multinomimallilla. Lasketaan yhteen.

- Ei onnistu helposti käsin, mutta tietokoneella kyllä (koodi luentojen liitteenä asiasta kiinnostuneille; ei kuulu kurssin vaatimuksiin).
- Kaikki 10 pylvästä ovat…
  - … tasakorkeat (10) tn:llä 2.3571e-08
  - … välillä [9,11] tn:llä 1.5528e-04
  - … välillä [8,12] tn:llä 0.0083
  - … välillä [7,13] tn:llä 0.0747

- TAI: Arvotaan otoksia ja tehdään niistä histogrammit. Katsotaan miten usein tuli melkein tasaisia. (Empiirinen lähestymistapa!)

Pohdittavaksi

- Otos n=100 tasajakaumasta välillä [20,30]
- Lasketaan lukumäärät

- Tn saada pylväät (10,10,10,10,10,10,10,10,10,10) on 2.3571e-08
- Tn saada pylväät (9,8,11,7,12,10,10,12,8,13) on 4.1796e-09

- Jälkimmäinen on **pienempi**. Miksi se kuitenkin herättää **vähemmän** epäilyksiä, ts. pystyt **helpommin** uskomaan, että se on tullut edellä kuvatusta satunnaisprosessista?

- Mahdollinen lähestymistapa ongelmaan: Mistä muusta prosessista lukujono on voinut tulla? Voiko asetelmaa kuvata bayes-menetelmillä?

MUUNNOKSEN Y=g(X) JAKAUMA

Satunnaismuuttujan muunnoksia

- Satunnaismuuttujille voi tehdä monenlaisia laskutoimituksia.
- Jos X = huominen sademäärä (millimetreinä), meitä ehkä kiinnostaa 1000 neliön tontille kertyvä litramäärä **$Y = 1000 \cdot X$**,
  joka on uusi satunnaismuuttuja.
  - (mutta ei ollenkaan riippumaton X:stä, vaan **täysin** riippuva: tieto X:n arvosta määrää täsmälleen Y:n arvon)
- Toisaalta vesivoimainsinööriä kiinnostaa useiden eri paikkojen sademäärien **summa** = paljonko vettä tulee kertymään voimalaan yhteensä – taas uusi sm

- Muunnoksen jakauma määräytyy tavallisten todennäköisyyslaskennan sääntöjen mukaan:
  P(Y on jotain) = P(X on sellainen, että Y on jotain)
- Usein kätevä työkalu tähän on kertymäfunktio.

Kertymäfunktion ideasta

- Vaikka tiheysfunktio antaa intuitiivisemman käsityksen jakaumasta (mihin jakauma keskittyy), kertymäfunktio on keskeinen **laskennallinen** työkalu jakaumien käsittelyssä.
- F(x) vastaa kysymykseen ”mikä on tn, että $X \leq x$”.
- Esim. lämpötilalle $F(25) = ½$ tarkoittaa, että on ½ todennäköisyys sille, että lämpötila on enintään 25 astetta.
- Jos tämän tapahtuman (”lämpötila on enintään 25 astetta”) voi esittää joidenkin muiden tapahtumien avulla (esim. luettelemalla vaihtoehtoja, joilla se voi toteutua), niin se voidaan laskea sitä kautta.
- Olennaista: kertymäfunktion arvo ilmaisee **erään tapahtuman todennäköisyyttä**, ja kaikki tutut todennäköisyyksien kaavat pätevät edelleen – esimerkiksi additiivisuus ja komplementtisääntö.

Erilaisia muunnostehtäviä

1. Tunnetaan **kertymäfunktio** F_X ja muunnosfunktio $Y = g(X)$.
   Mikä on Y:n kertymäfunktio?
   → Periaatteessa **helppo** tehtävä: Ratkaistaan $P(Y \leq y)$

2. Tunnetaan **tiheysfunktio** f_X ja muunosfunktio $Y = g(X)$.
   Mikä on Y:n tiheysfunktio?
   → **Ei aivan triviaalia**.
   On olemassa tiheysfunktion muuntokaava (ei käsitellä tällä kurssilla).

3. Osataan **simuloida** $X_1 \ldots X_n$ iid samasta jakaumasta kuin X,
   ja tunnetaan muunnosfunktio $Y = g(X)$.
   Miten simuloidaan Y:n jakaumasta?

   → **Todella helppoa**: lasketaan $Y_i = g(X_i)$

Kokeilemme muutamaa yksinkertaista muunnosta menetelmällä 3.

Muunnos: vakion lisääminen

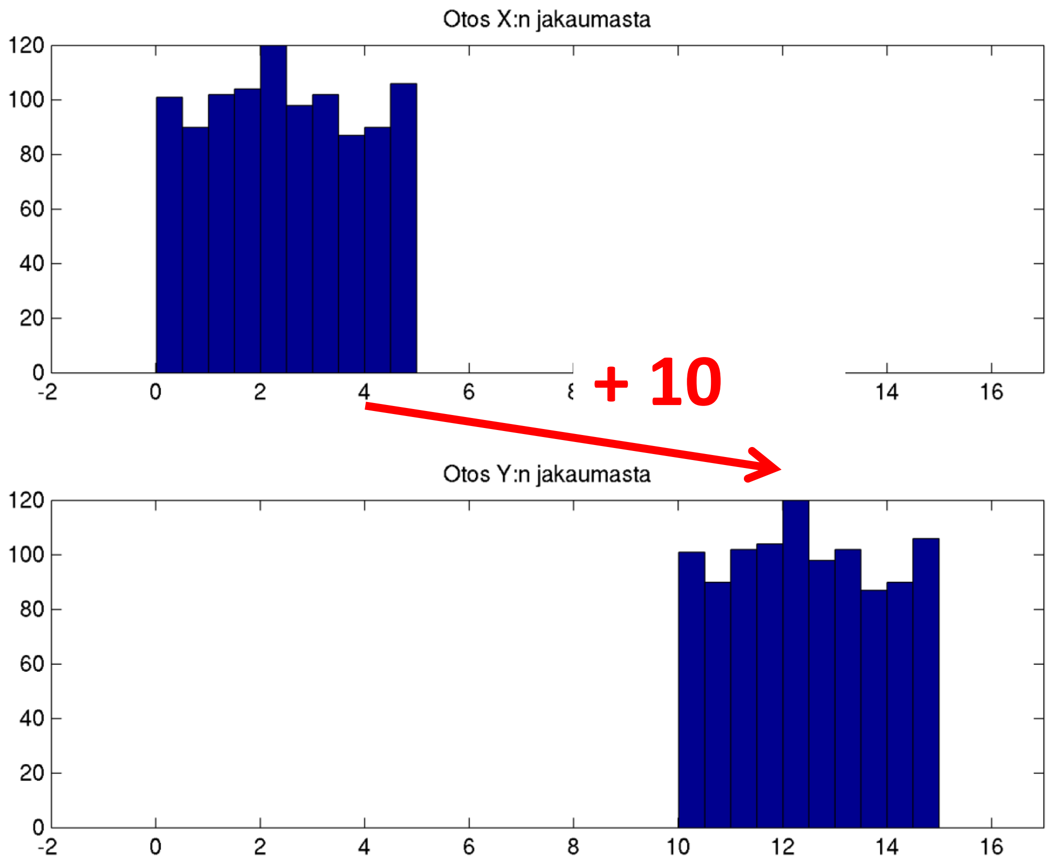
$X \sim Tas(0, 5)$

$Y = X$ **+ 10**

$Y \sim$?

Ilmeisesti Y:in on tasajakautunut.

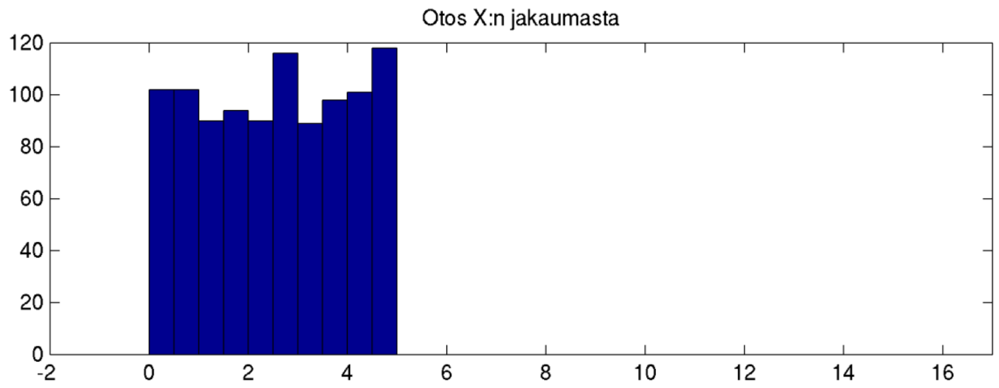
Millä parametreilla?

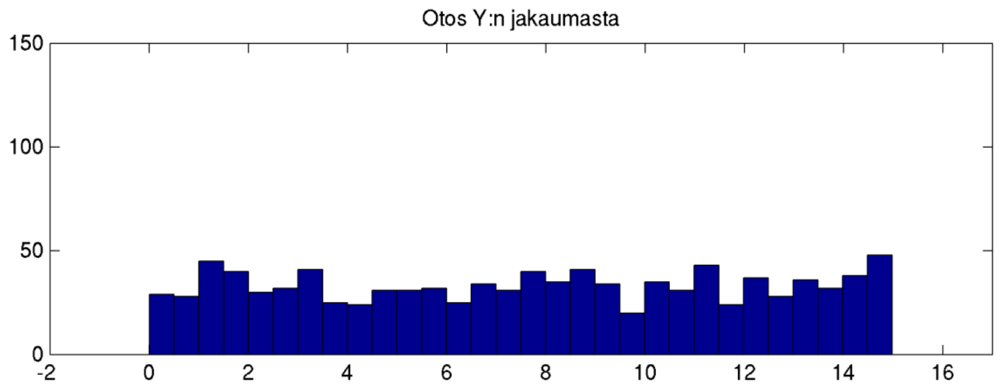


Muunnos: vakiolla kertominen

$X \sim \mathrm{Tas}(0, 5)$
$Y = \color{red}{\mathbf{3}} \cdot X$

$Y \sim$?





Ilmeisesti Y:in on tasajakautunut.

Millä parametreilla?

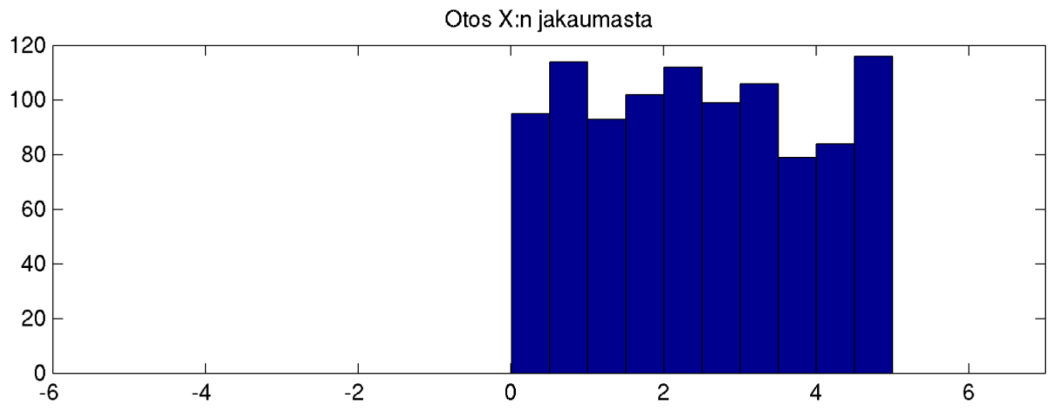
Huom. litistyminen, joka näkyy myös Y:n tiheysfunktiossa.

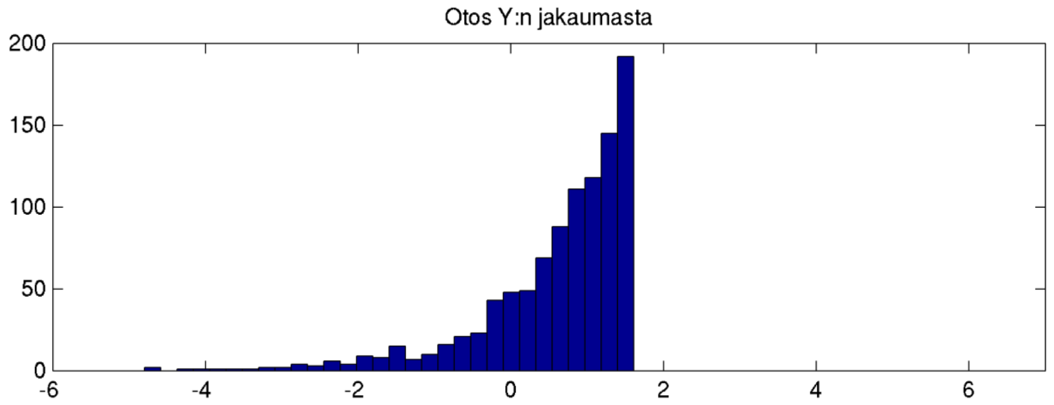
Muunnos: logaritmi

X ~ Tas(0, 5)
Y = **ln** X

Y ~ ?

Y:llä on ilmeisesti joku erikoisempi jakauma.





Kertymäfunktion muuntaminen

$X \sim \text{Tas}(0, 5)$
$Y = \ln X$
$Y \sim$?

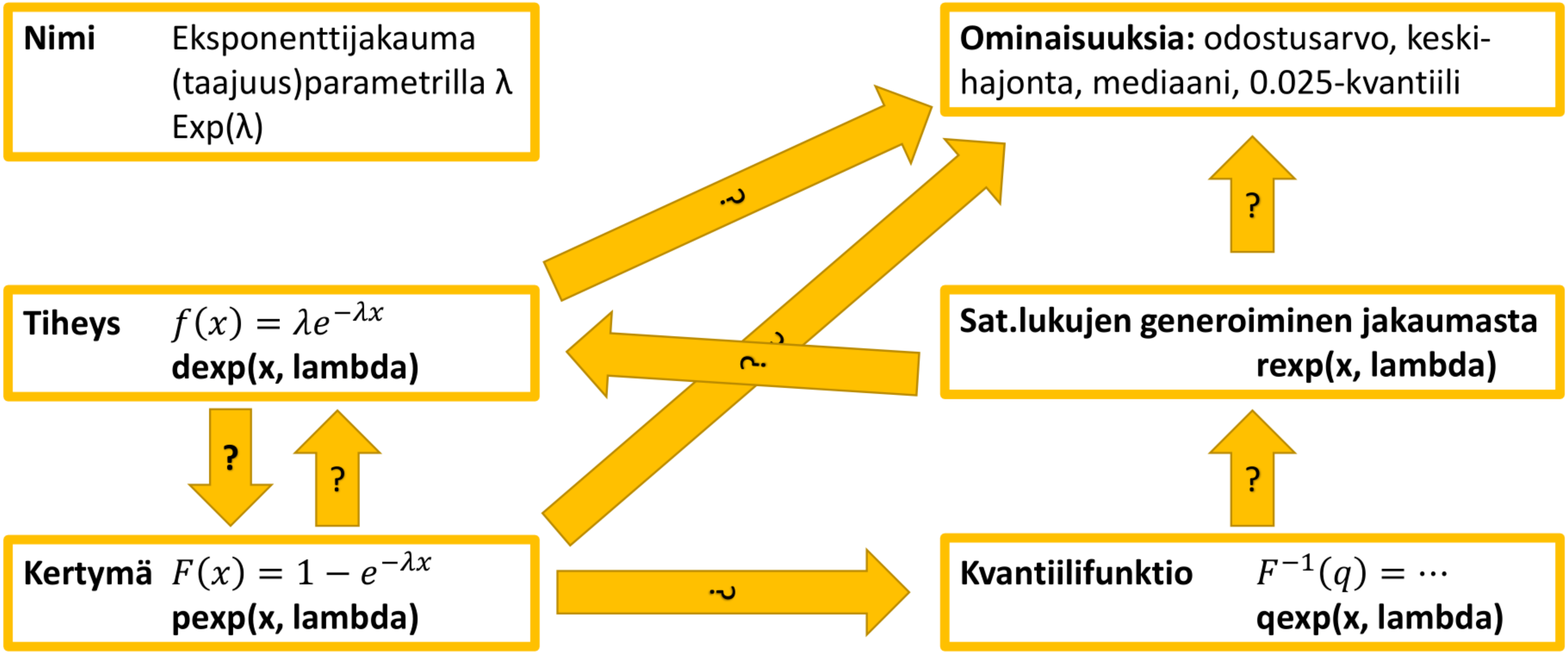
Ratkaistaan Y:n kertymäfunktio.

$F_Y(a)$	$= P(Y$	$\leq a)$	Kertymäfunktion määritelmä.
	$= P(\ln X$	$\leq a)$	Koska $Y = \ln X$.
	$= P(X$	$\leq \exp(a))$	Eksponenttifunktio aidosti kasvava.
	$= F_X(\exp(a))$		Kertymäfunktion määritelmä
	$= \exp(a) / 5$		Tasajakauman kertymäfunktio.

Derivoimalla saadaan Y:n tiheysfunktio.
$f_Y(a) \quad = (1/5)\exp(a)$

Huom. että Y:n jakauma sijoittuu välille $(-\infty, \ln 5)$, sillä sinne väli $(0, 5)$ kuvautuu logaritmimuunnoksessa!
Tulos tuntuu vastaavan otoshistogrammin muotoa.

Tn-jakauman käsitekartta – esim. eksponenttijak.



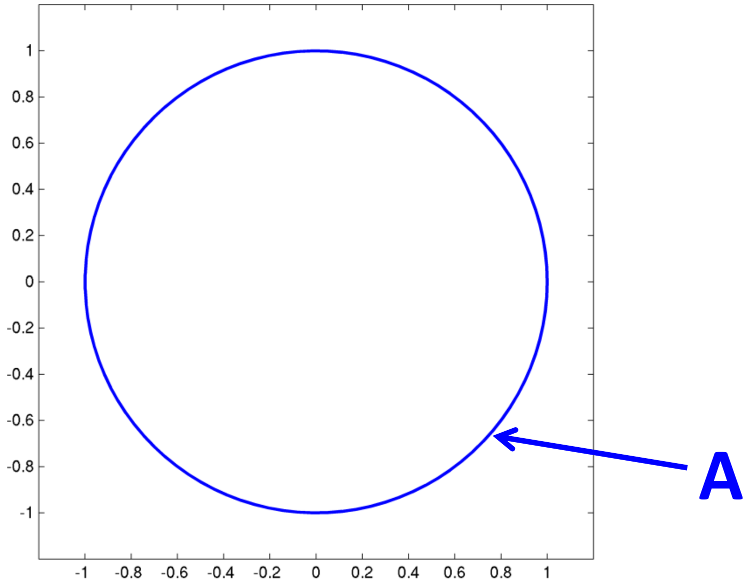
Harjoitus. Tutki, millaisin menetelmin päästään laatikosta toiseen. Tutki luentomateriaalia ja yritä sijoittaa oppimasi menetelmät kuvaan. Tutki harjoitustehtäviä: mitä menetelmiä käytettiin?

Suurten lukujen lain sovellutus:

Monte Carlo -integrointi

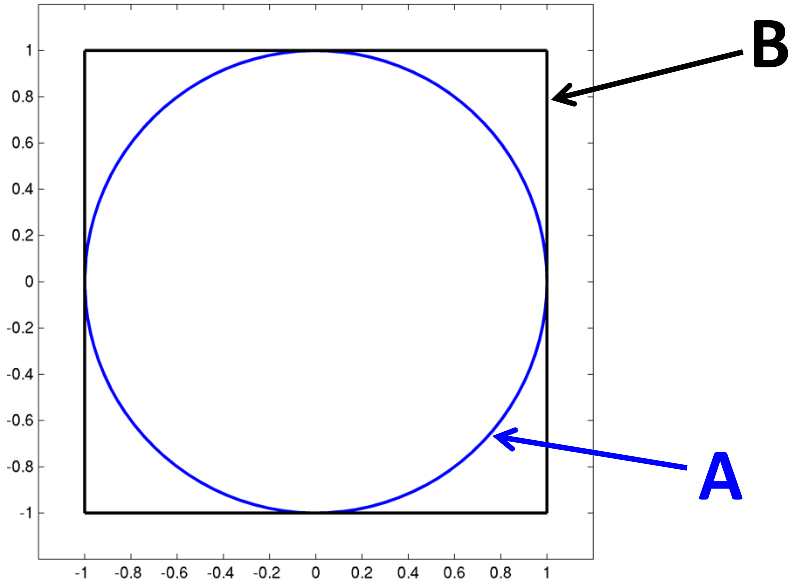
Tehtävässä ei todennäköisyyttä

- Mikä on mutkikkaan tasokuvion **A** pinta-ala?
  Osaamme vain testata, onko jokin piste sisällä vai ulkona:
  onko **sqrt(x^2 + y^2) < 1**



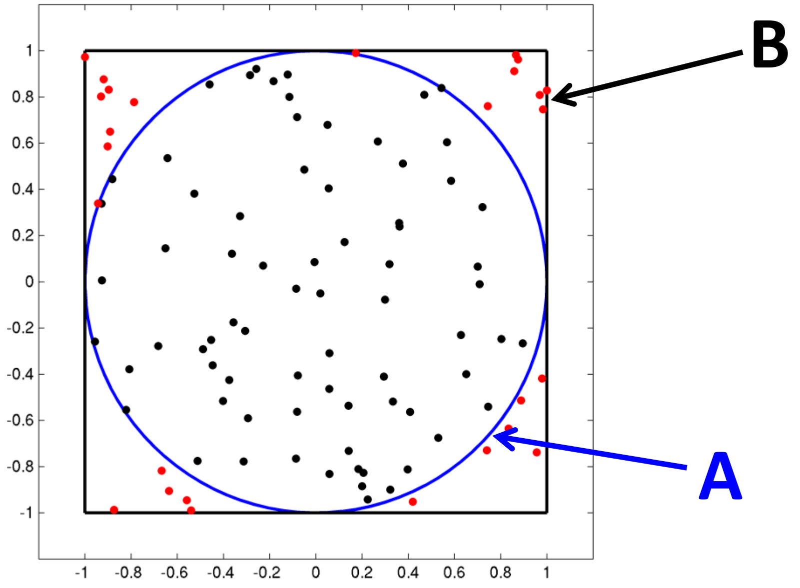
Muutetaan tehtävää

- Esimerkki: Mikä on mutkikkaan tasokuvion **A** pinta-ala?
  Osaamme vain testata, onko jokin piste sisällä vai ulkona.
- Ratkaisu: Piirrämme kuvion ympärille isomman (**B**),
  - jonka pinta-alan (= 4) tunnemme



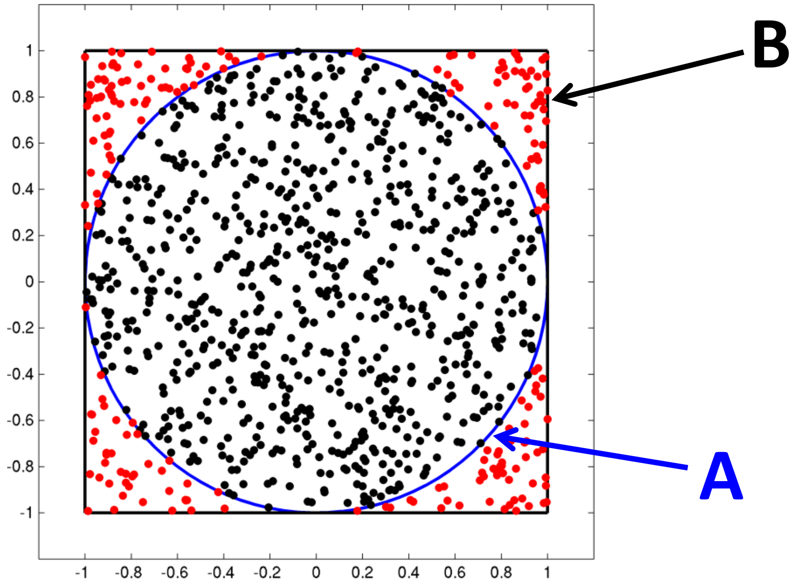
Muutetaan tehtävää

- Esimerkki: Mikä on mutkikkaan tasokuvion **A** pinta-ala?
  Osaamme vain testata, onko jokin piste sisällä vai ulkona.
- Ratkaisu: Piirrämme kuvion ympärille isomman (**B**),
  - jonka pinta-alan (= 4) tunnemme ja
  - josta osaamme arpoa (simuloida) pisteitä



Monte Carlo -integrointi

- Piste osuu kuvioon tn:llä $p = m(A) / m(B)$, m = pinta-ala
- n-kertainen toistokoe
- Suuren lukujen laki: osuus $f_n \approx p$
- Arvioimme, että $\mathbf{m(A)} = p\, m(B) \approx \mathbf{f_n\, m(B)}$



Monte Carlo -integrointi

n	**pisteitä *B*:ssä**	***m*(*B*) ≈**
100	80	3.200000
1 000	783	3.132000
10 000	7 849	3.139600
100 000	78 544	3.141760
1 000 000	785 132	3.140528

Samaa menetelmää voi periaatteessa soveltaa mielivaltaisessa *n*-ulotteisessa avaruudessa, esim. mikä on *n*-ulotteisen pallon tilavuus?

Keskeisen raja-arvolauseen perusteella voidaan arvioida integraalin likiarvon tarkkuutta (tunnetaan odotusarvo ja varianssi, approksimoidaan normaalijakautuneeksi). Yleisesti ottaen 100-kertaisella pistemäärällä saadaan yksi desimaali lisää tarkkuutta.

Varalla:

MONTY HALLIN ONGELMA: PALKINTOJA OVIEN TAKANA

Monty Hall,
amerikkalainen
tv-showisäntä

LET'S MAKE A DEAL

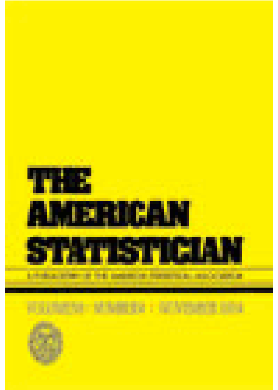
Let's Make a Dealissa osallistujille tarjottiin monenlaisia valintatilanteita:

Vaihdatko tunnetun, pienen palkinnon tuntemattomaan, **<u>mahdollisesti</u>** suureen palkintoon (auto)?

Tuntematon palkinto voikin osoittautua epätoivotuksi
(esim. vuohi tai laama, kuten kuvassa).

Kuvat: en.wikipedia.org, commons.wikimedia.org

”Monty Hallin ongelma”



- Tv-shown innoittama todennäköisyyspähkinä v. 1975 *The American Statistician* -lehdessä

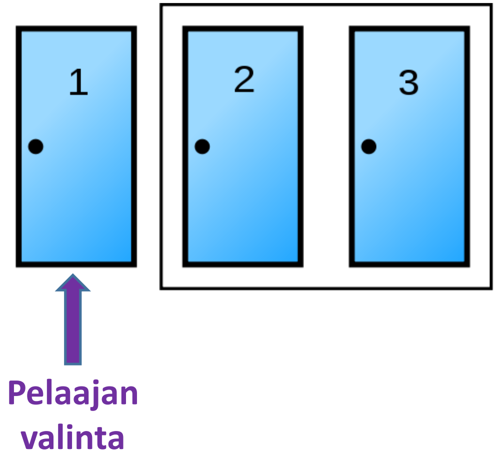
- Sittemmin tuli kuuluisaksi, kun vuonna 1990 viikkolehti *Paradessa* esiteltiin kyseinen ongelma ja sen oikea vastaus…

  … jonka jälkeen toimitukseen tuli **tuhansia** lukijakirjeitä, joissa oltiin eri mieltä vastauksesta!

Pelin säännöt

1. Aluksi on kolme suljettua ovea. Yhden takana on auto, kahdessa vuohi. Juontaja tietää missä auto on.

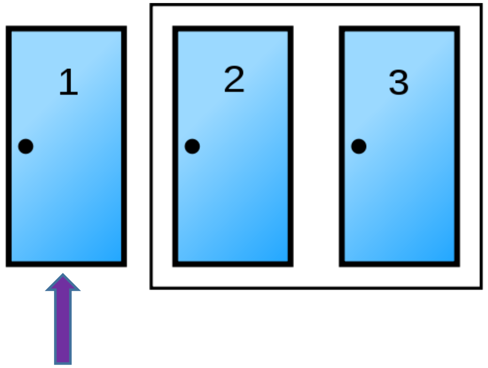
2. Pelaaja saa **valita** yhden oven (kuvassa ovi 1), <u>mutta ovea ei avata vielä.</u>



Pelin säännöt

1. Aluksi on kolme suljettua ovea. Yhden takana on auto, kahdessa vuohi. Juontaja tietää missä auto on.

2. Pelaaja saa **valita** yhden oven (kuvassa ovi 1), <u>mutta ovea ei avata vielä.</u>



3. Juontaja valitsee kahdesta <u>muusta</u> ovesta sellaisen, jossa on vuohi, ja **avaa** sen (kuvassa ovi 3).

4. Nyt jäljellä on kaksi suljettua ovea: pelaajan alun perin valitsema (1), ja toinen (2). Hän saa **vaihtaa** ovea, jos haluaa.

Millä tn:llä auto on oven 1 takana?
Entä oven 2?

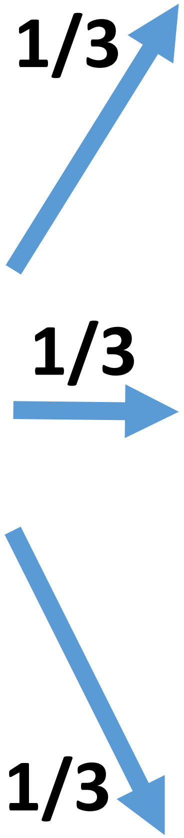
Analyysi

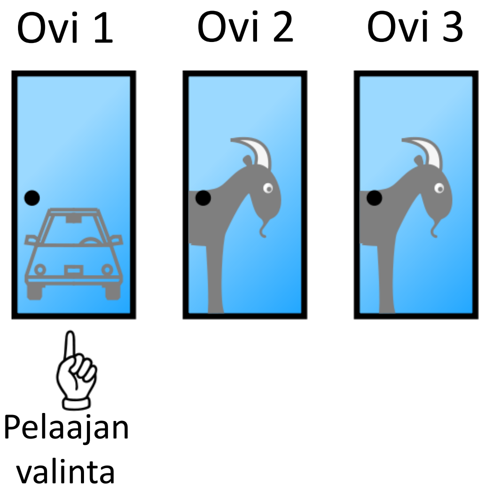
- Oletetaan, että pelaaja on valinnut oven 1.
  (Muut kaksi tapausta lasketaan aivan samaan tapaan.)

- Auto voi yhtä hyvin olla oven 1, 2 tai 3 takana, koska pelaajalla ei ole siitä mitään tietoa.

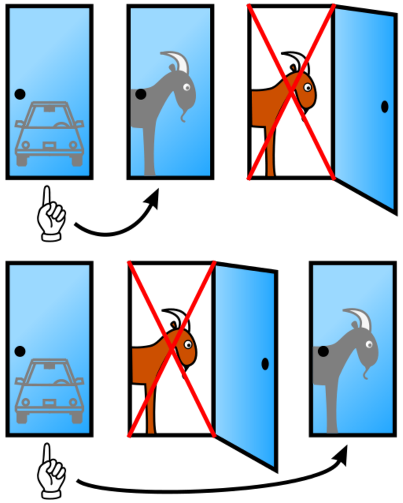
  Ts. kunkin mahdollisuuden todennäköisyys on 1/3.

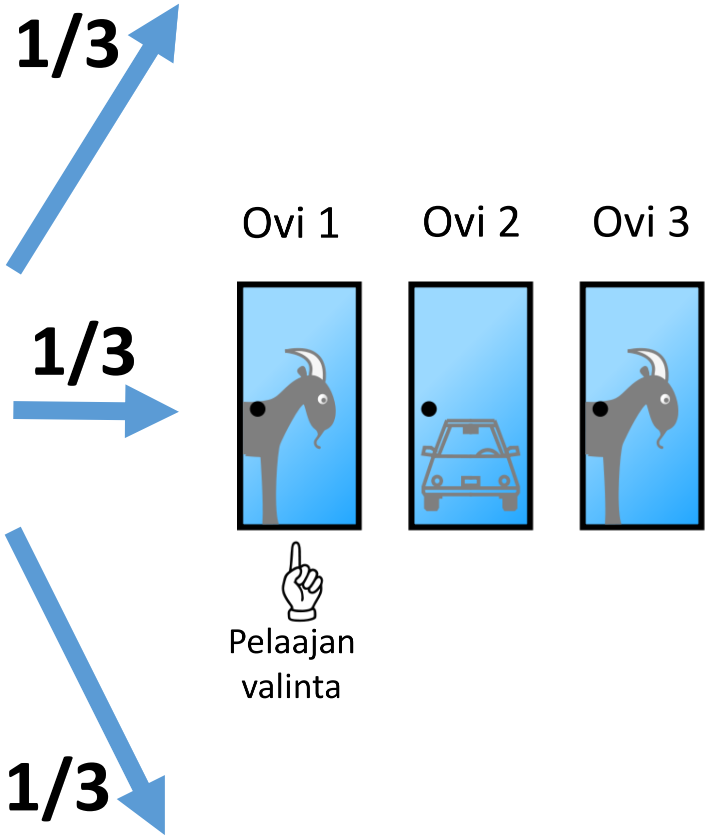
- Käsitellään nämä kolme mahdollista tapausta yksitellen.



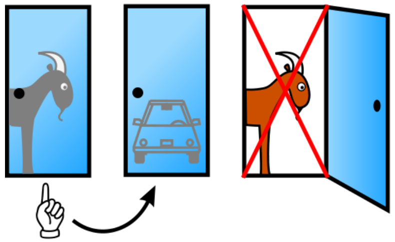


Auto on valitulla ovella 1.
Juontaja avaa joko oven 2 tai 3
ja paljastaa sieltä vuohen.
Kummassakin tapauksessa:
Vaihtamalla häviää.





Auto ovella 2
(ja pelaaja valitsi 1).
Juontajan pakko avata ovi 3.
Vaihtamalla voittaa auton.

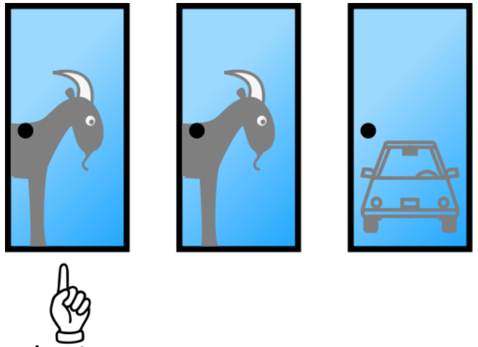


1/3

1/3

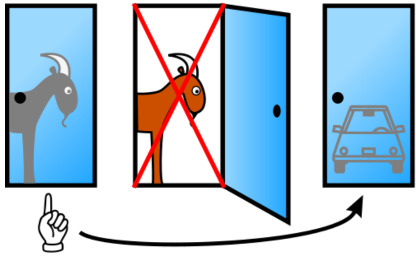
1/3

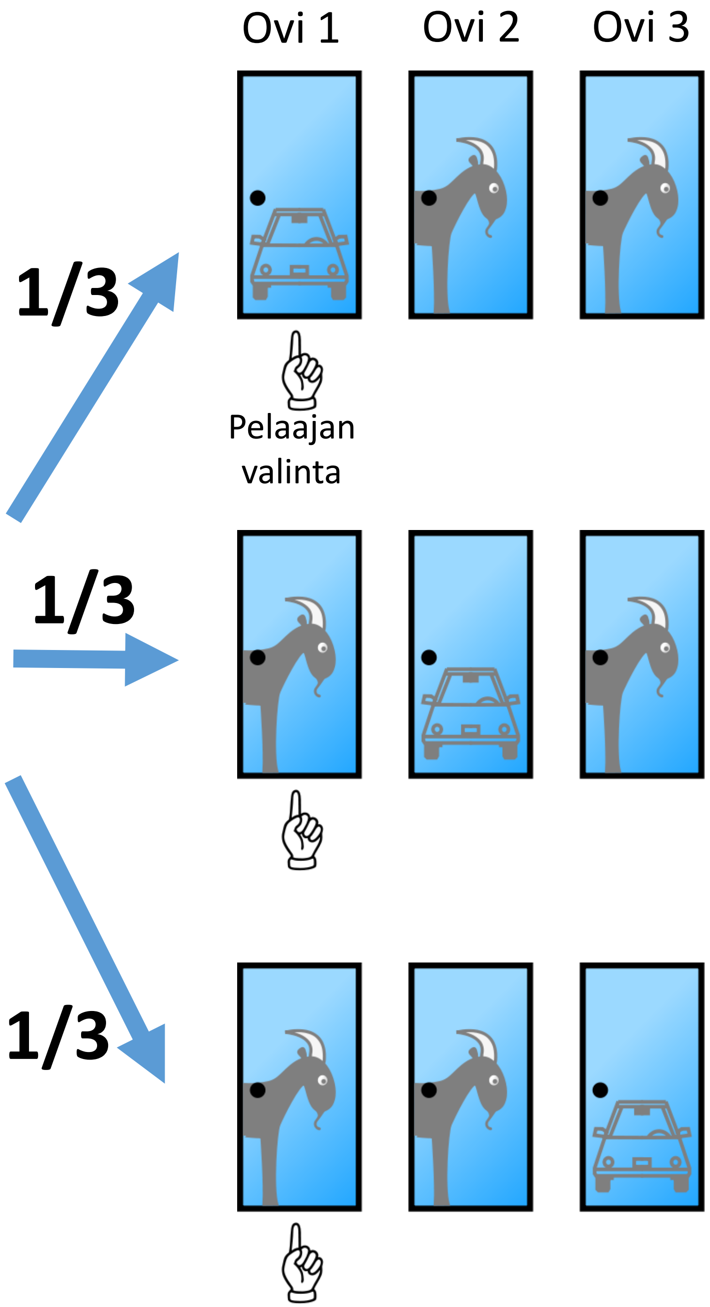
Ovi 1 Ovi 2 Ovi 3



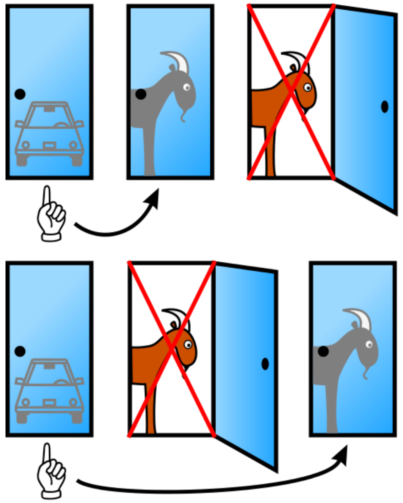
Pelaajan
valinta

Auto ovella 3
(ja pelaaja valitsi 1).
Juontajan pakko avata ovi 2.
Vaihtamalla voittaa auton.

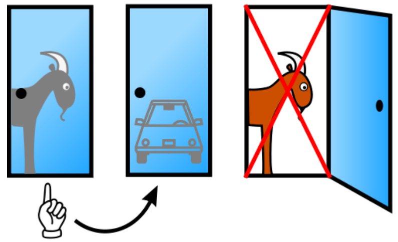




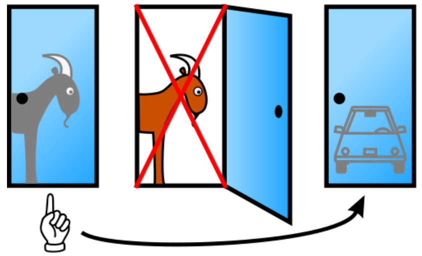
Auto ovella 1
(jonka pelaaja valitsi).
Juontaja avaa joko oven 2 tai 3
ja paljastaa vuohen.
Kummassakin tapauksessa:
Vaihtamalla saa vuohen.



Auto ovella 2
(ja pelaaja valitsi 1).
Juontajan pakko avata ovi 3.
Vaihtamalla voittaa auton.



Auto ovella 3
(ja pelaaja valitsi 1).
Juontajan pakko avata ovi 2.
Vaihtamalla voittaa auton.



Yhteenveto

Kun pelaaja valitsi alun perin oven 1:

Toden-näköisyys	Auto ovella	Juontaja avaa oven	Tulos, jos pysyy ovessa 1	Tulos, jos vaihtaa ovea
1/3	1	2 tai 3	**AUTO**	vuohi
1/3	2	3	vuohi	**AUTO**
1/3	3	1	vuohi	**AUTO**

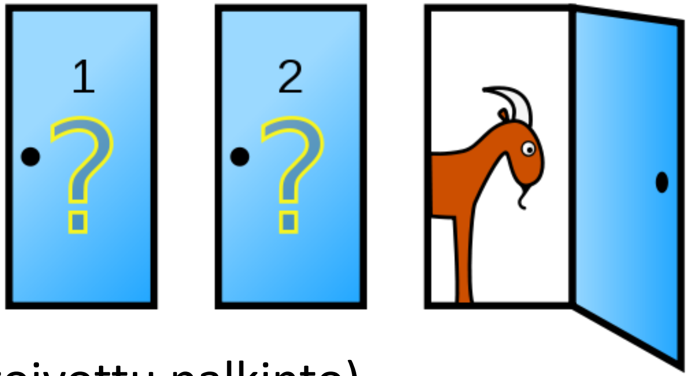
- Todennäköisyydellä 1/3 pelaaja valitsi oikean oven → saa auton **jos ei vaihda**.
  Todennäköisyydellä 2/3 pelaaja valitsi väärän oven → saa auton **jos vaihtaa**.

  Kannattaa siis vaihtaa ovea, sillä silloin palkintoauton saamisen todennäköisyys kasvaa 1/3:stä 2/3:een!

- Tulos saattaa vaikuttaa yllättävältä (mistä todisteena tuhannet lukijakirjeet), mutta se on laskennallisesti perusteltu. Se myös **vastaa empiirisiä havaintoja**, kun peliä pelataan toistuvasti.

Mysteeri?

Kun juontaja on avannut oven,
senhetkinen tilanne näyttää tältä.



Kaksi suljettua ovea ja yksi avonainen.

- Yhden suljetun oven takana on auto (= toivottu palkinto)
- Yhden suljetun oven takana on vuohi (= ei toivottu).
- Avonaisen oven takana on vuohi.

Saat valita kumman tahansa suljetun oven.

Eivätkö ovet 1 ja 2 ole aivan samassa asemassa (suljettuja)
ja auto siten **yhtä todennäköisesti** kumman tahansa takana?

Kyllä, jos muuta tietoa ei ole.
Mutta pelissä **on** muuta tietoa: Tieto siitä, miten tässä näkyvään valintatilanteeseen päädyttiin. Pelaajan ja juontajan toiminta (prosessi) johtaa epätasaisiin todennäköisyyksiin 1/3 ja 2/3.

Katso myös

http://fi.wikipedia.org/wiki/Monty_Hallin_ongelma

http://en.wikipedia.org/wiki/Monty_Hall_problem

(ja sieltä kirjallisuusviitteitä)

Kokeile peliä esim. osoitteessa
http://www.lintukoto.net/testit/ovi/
Pelin jälkeen näet tilaston yli 500 000 pelikerrasta:

- Pelaajista, jotka **vaihtoivat** ovea 66,7 % sai auton.
- Pelaajista, jotka **eivät vaihtaneet** 33,4 % sai auton.

Vaihtamisella oli merkitystä aivan kuten äsken laskettiin!

Muunnelmia

- Edellä lasketut todennäköisyydet perustuvat olennaisesti siihen, miten pelin oletettiin kulkevan, ts. miten valintatilanteeseen päädyttiin ja miksi.

- Jos pelin säännöt (prosessi) ovat erilaiset, voidaan taas saada näennäisesti samassa tilanteessa **eri todennäköisyydet** kuin edellä.

- Esim. juontaja avaakin **umpimähkään** jommankumman ovista 2 ja 3, ja avatun oven takana **sattuu** olemaan vuohi. Nytkin pelaajalle jää kaksi suljettua ovea, mutta nyt vaihtamisesta ei ole hyötyä!

- Tai juontaja tarjoaa vaihtomahdollisuuden **vain**, jos pelaaja on alun perin osunut autoon. Tällöin vaihtamisesta on haittaa!

- Nämä ja muita muunnelmia: ks. esim. engl. Wikipedian artikkelia *Monty Hall problem*.

52 kortin versio

- Ovien lukumäärän lisääminen voi helpottaa hahmottamista. Korvataan ovet pelikorteilla. Auton saa pataässällä.

- Sekoitetaan korttipakka ja pannaan kortit tarjolle.

- Pelaaja valitsee yhden korteista (mutta ei katso sitä).
  Todennäköisesti hänellä on väärä kortti! On vain 1/52 todennäköisyys, että hän osui umpimähkään pataässään.

- Jakaja katsoo kaikki muut 51 korttia ja poistaa (paljastaa) niistä 50 korttia, kuitenkin niin, että hän ei poista pataässää (jos sellainen tulee vastaan).

- Jäljelle jää kaksi korttia: Pelaajan alun perin valitsema, ja yksi muu kortti. Kannattaako pelaajan vaihtaa?

Kiitos!

- Kysymyksiä?

- (Jos aikaa on, pelataan Lintukodon Monty Hallia muutama kerta eri strategioilla ja kerätään tilastoa auton saamisesta.
  Strategiat: ”vaihda”, ”älä vaihda”, ”unohda tapahtumahistoria”)