Johdatus todennäköisyyslaskentaan
Kevät 2015

Luento 10 / 13

Jukka Kohonen

Matematiikan ja tilastotieteen laitos

Helsingin yliopisto

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.

Sovellusesimerkki

- Pitkulainen esine (pituus = L) nähdään kaukaa satunnaisessa kulmassa (θ).
  - Nähty projektio = $X = L \cdot \cos \theta$
  - Projektion jakauma: $X \sim ?$
  - Pitää siis osata ratkaista θ:n muunnoksen jakauma.

- **Käänteisongelma:** Sama pitkulainen esine (pituus <u>tuntematon</u>) nähdään monta kertaa satunnaisissa kulmissa.
  - Tunnetaan nähtyjen projektioiden jakauma
  - Mikä on pituuden jakauma?
  - Sovelletaan esim. Bayesin kaavaa.

- Epäsäännöllisen muotoinen asteroidi nähdään satunnaisissa kulmissa.
  - Tunnetaan nähtyjen (asteroidin heijastamien) kokonaiskirkkauksien jakauma ja/tai kehitys ajan kuluessa (kun asteroidi pyörii ja liikkuu)
  - Mikä on asteroidin muoto?

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, mutta esim. syksyn kurssilla *Todennäköisyyslaskenta*).

3. Osataan **simuloida** $X_1 ... X_n$ 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.

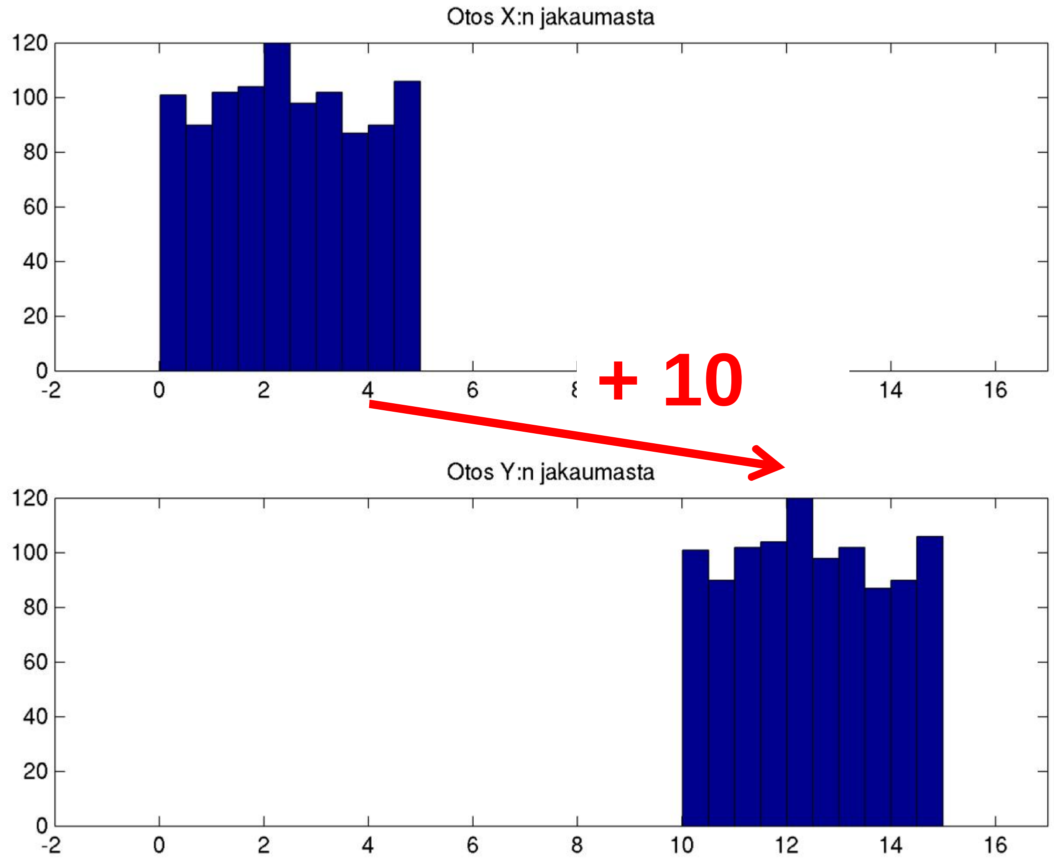
Muunnos: vakion lisääminen

X ~ Tas(0, 5)
Y = X **+ 10**

Y ~ ?

Ilmeisesti Y:in on tasajakautunut.

Millä parametreilla?



Muunnos: vakiolla kertominen

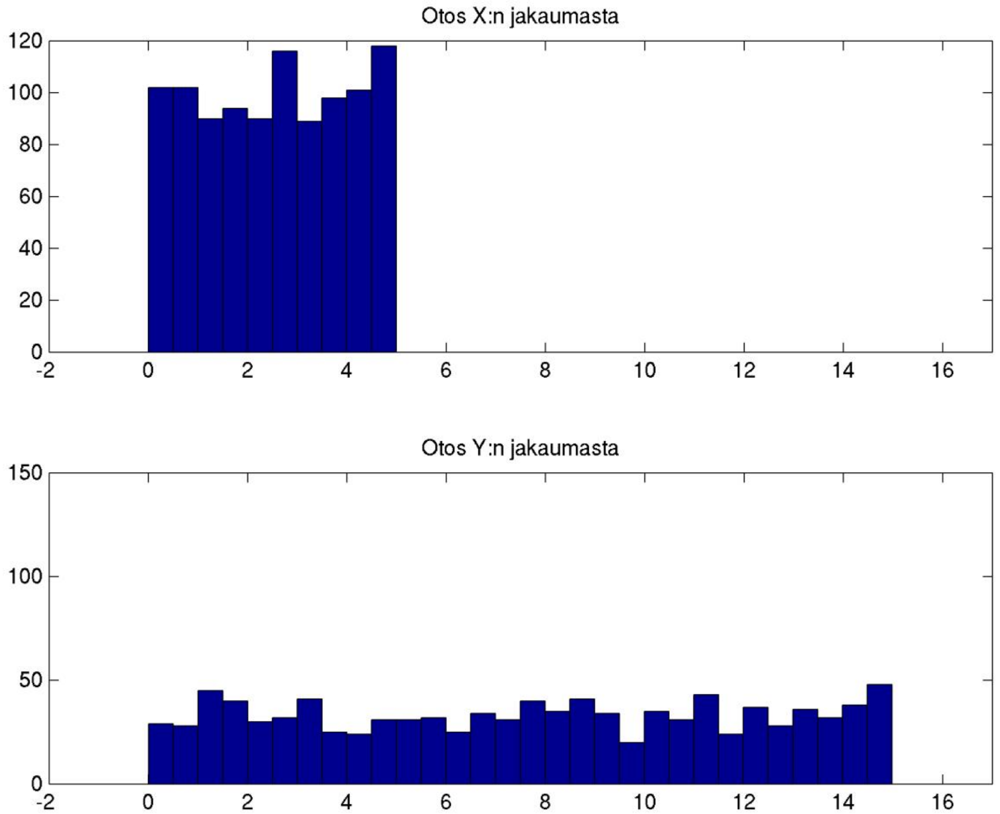
$X \sim \text{Tas}(0, 5)$
$Y = \mathbf{3} \cdot X$

$Y \sim \ ?$

Ilmeisesti Y:in on tasajakautunut.

Millä parametreilla?

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



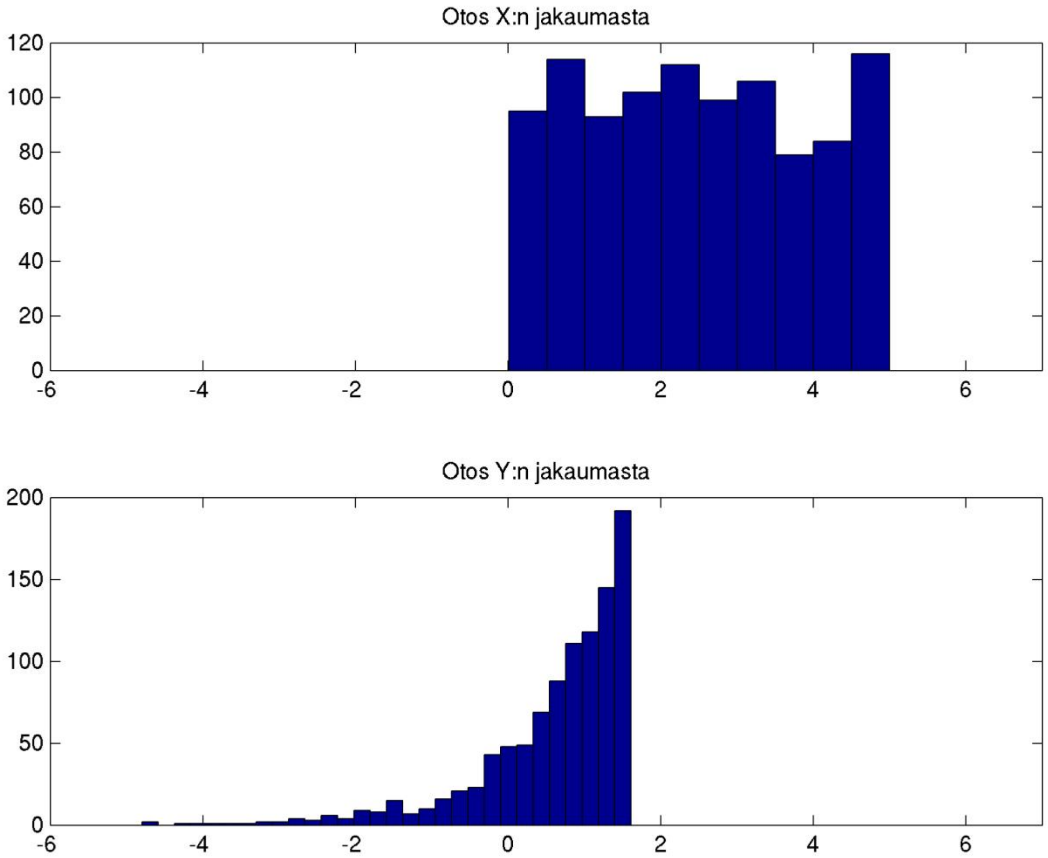
Muunnos: logaritmi

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

Y ~ ?

Y:llä on ilmeisesti joku erikoisempi jakauma.

Osataan kyllä laskea (Tuominen s. 74)



Kertymäfunktion muuntaminen

Tuominen s. 74

X ~ Tas(0, 5)
$Y = \ln X$
Y ~ ?

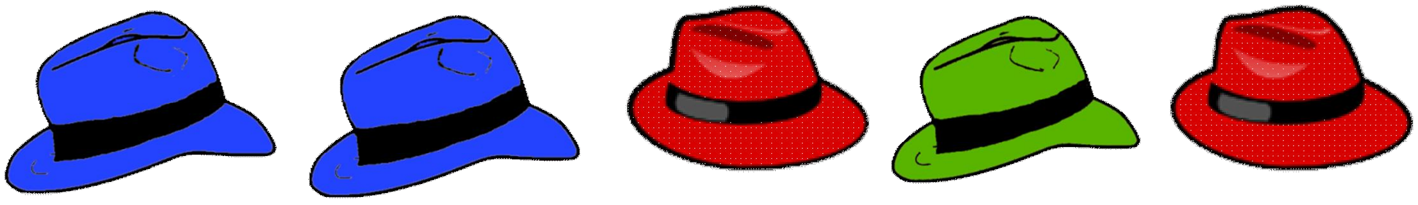
Ratkaistaan Y:n kertymäfunktio.

$F_Y(a)$	$= P(Y \leq a)$	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.

$f_Y(a) = (1/5) \exp(a)$

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



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**
"<u>Monellako tavalla</u> 10:stä alkiosta <u>voi valita</u> 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) $= p^{10}$ ≈ 0.000 977

P(AAAABBBBCC) $= p^4 \cdot q^4 \cdot r^2$ ≈ 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) $= p^{10}$ ≈ 0.000 977

Näitä (10×A) on vain yksi

P(AAABABBBCC) $= p^4 \cdot q^4 \cdot r^2$ ≈ 0.000 020

P(BBCAABBAAC) $= p^4 \cdot q^4 \cdot r^2$ ≈ 0.000 020

P(AABCCAABBB) $= p^4 \cdot q^4 \cdot r^2$ ≈ 0.000 020

. . .

P(CCBBBBAAAA) $= p^4 \cdot q^4 \cdot r^2$ ≈ 0.000 020

Näitä alkeistapauksia (4×A, 4×B, 2×C) on

$$\binom{10}{4,4,2} = 3150 \text{ kpl},$$

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.

Pylväiden korkeuksien jakauma

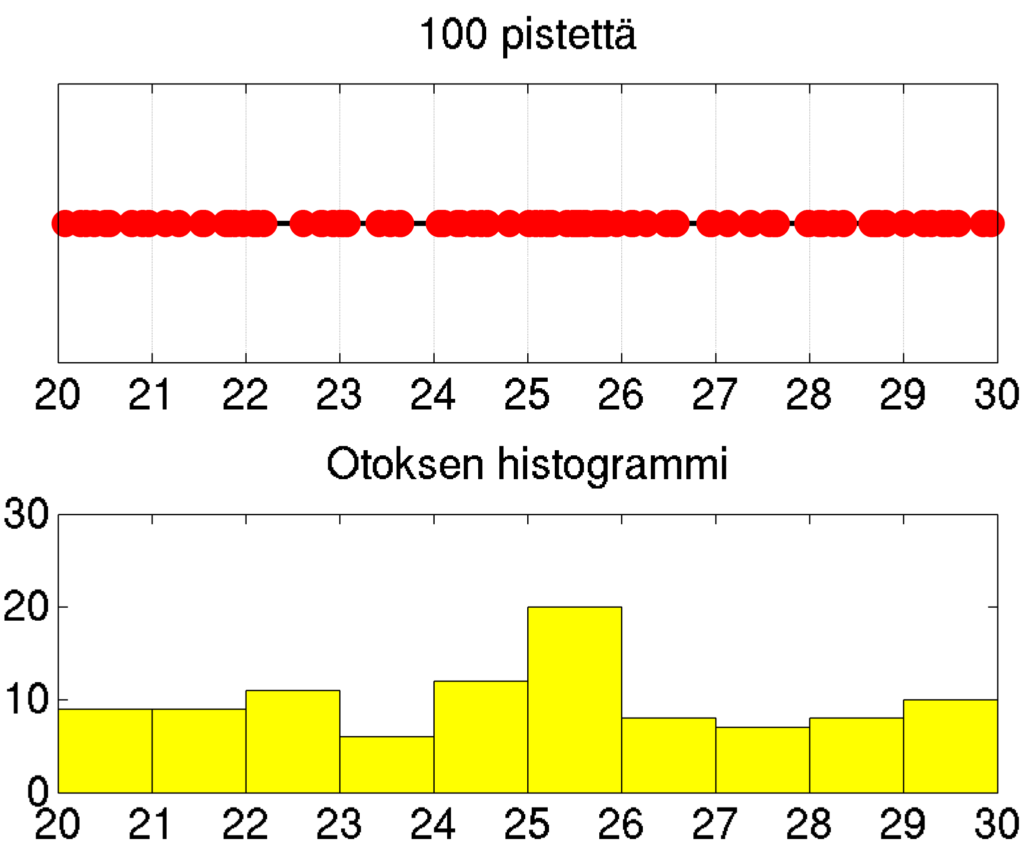
Merk.
Y_i = i:nnen pylvään korkeus
= i:nnelle jakovälille osuvien pisteiden lukumäärä

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

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

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.