Introduction to Probability with MATLAB
Spring 2014

Lecture 11 / 12

Jukka Kohonen
Department of Mathematics and Statistics
University of Helsinki

LAW OF LARGE NUMBERS

Is sample mean ≈ expected value?

- The density (for continuous) or point probabilities (for discrete) can be approximated by an empirical histogram (drawn from a random sample).

- Can we also approximate other things about the distribution – for example its expected value?
- What about the sample mean, ie. arithmetic average of the sample values?
  MATLAB: **`sum(x)/n`** tai **`mean(x)`**

- The law of large numbers proves that for this is a good approximation (for large n)

Example: Bernoulli trials

Bernoulli trieals, ***n*** trials with success probability ***p***

Z_n = number of successes

f_n = relative frequency of successes = Z_n / n

We know **$Z_n \sim \mathrm{Bin}(n, p)$.** Thus $E(Z_n)=np$ and $E(f_n) = p$.

But we are **not likely** to hit the expected value **exactly.**

This probability **decreases** as *n* increases. E.g. if 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

Is it at least close?

Z_n = number of successes

f_n = relative frequency of successes = Z_n / n

ε = our requirement for accuracy (we can choose this freely!)

We shall say that f_n <u>is nearly right</u>, if $| f_n - p | < \varepsilon$

Prob. of being nearly right **increases** as *n* increases – at least seem so:
E.g. p=0.3 and **ε=0.01**, so we are requiring that $0.29 < f_n < 0.31$:

n	*f_n has to be here*	*Z_n has to be here*	**P(nearly right) ≈**
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)	0.9999999999946

P(nearly right) = sum of the binomial probabilities...

INEQUALITIES FOR LARGE DEVIATIONS

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.

Markov's inequality

- If we know certainly that **$X \geq 0$**, and we know **E(X),** we have an **upper bound** for the **tail probability** starting from any point ***a***

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

We don't need to know the exact distribution of X !!!

Proof sketch:

E(X) is a sum (or integral, for continuous X).
If the tail probability were very large, then those tail values would contribute much to this sum, so E(X) would be large.

Bounding a sum or integral from above/below

- Easy observations:
  - If for all terms $a_i \leq b_i, \quad \forall i,$
    then for the sums $\sum a_i \leq \sum b_i$
  - If at every point $f(x) \leq g(x), \quad \forall x,$
    then also $\int f(x)dx \leq \int g(x)dx$

Markov's inequality, discrete X

Assume for simplicity that X takes only integer values

- **Assume X ≥ 0**, and **E(X) known.**
- Denote
  point probabilities p_k $= P(X = k)$
  tail probabilities q_a $= P(X \geq a) = p_a + p_{a+1} + \ldots$

$$E(X) = \sum_{k=0}^{\infty} k\, p_k$$ Expected value is a sum

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

$$\geq \left(\sum_{k=0}^{\infty} 0\, p_k \right) + \left(\sum_{k=a}^{\infty} a\, p_k \right)$$ Every term **bounded from below**. First sum is zero. Second sum has *a* as common factor.

$$= 0 + a \cdot q_a$$

$$= a \cdot q_a,$$

Thus $q_a \leq E(X)/a$, which is Markov's inequality.

Chebysev's inequality

- If both $\mu = E(X)$, $\sigma = D(X)$ are known,
  tail probability has a tighter upper bound:

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

- That is: Probability that *X* differs from its mean "more than *k* standard deviations", is at most $1/k^2$

- *X* need not be nonnegative (as in Markov)
- Tighter bound since the denominator has *k* **squared**

- Proof sketch: Markov's inequality for $Y=(X-\mu)^2$

Example: Exp tail probabilities, different methods

- $X \sim \text{Exp}(0.1)$. Compute or estimate $P(X \geq a)$ for different values of a and with different methods.
- $E(X)=10$, $D(X)=10$, we know the cdf: $F(x)=1-\exp(-0.1x)$

	Markov	Chebysev	Exact cdf
$P(X \geq 10)$	< 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
$P(X \geq 200)$	$< 1/20$	$< 1/361 = 0.0028$	$2.1 \cdot 10^{-9}$

- Chebysev tighter bounds, especially for large deviations
- Exact cdf can be much tighter **if known**!
- But inequalities can be ”good enough” for some purposes.
- For example, from Chebysev we can prove the Law of large numbers.

Chebysev → Law or large numbers

- (X_1, X_2, ...) independent random variables with same mean μ and variance σ^2
- **Pick arbitrary accuracy requirement** $\varepsilon > 0$

Partial sum S_n $= X_1 + \ldots + X_n$

- has mean $= n\mu$ why?
- has variance $= n\sigma^2$ why?

Sample mean $= (X_1 + \ldots + X_n) / n$

- has mean $= \mu$ why?
- has variance $= \sigma^2 / n$ why?

Chebysev→ LLN

Pick arbitrary accuracy requirement $\varepsilon > 0$

Sample mean $= (X_1+...+X_n) / n$

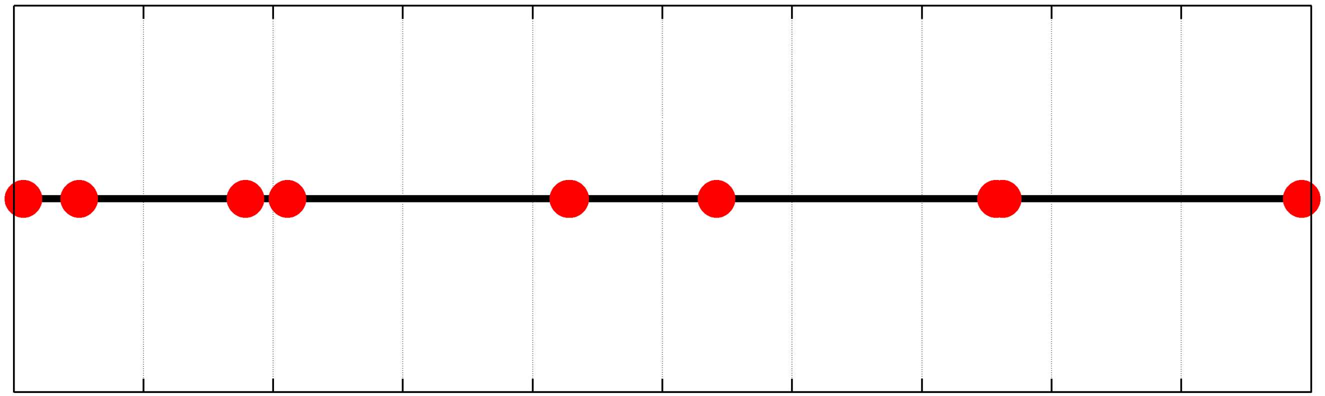
- Mean of the sample mean $= \mu$
- Std dev of the sample mean $= \sigma / \mathrm{sqrt}(n)$

Std deviation **decreases** as *n* increases.

This means that our requirement ε is an ever larger multiple ***k*** of the standard deviation. Applying Chebysev means that the tail probability for such a large multiple decreases.

As n increases, the probability that the sample mean is within ε of the true expected value μ increases (towards 1, as n goes to infinity).

This is the (weak) law of large numbers.

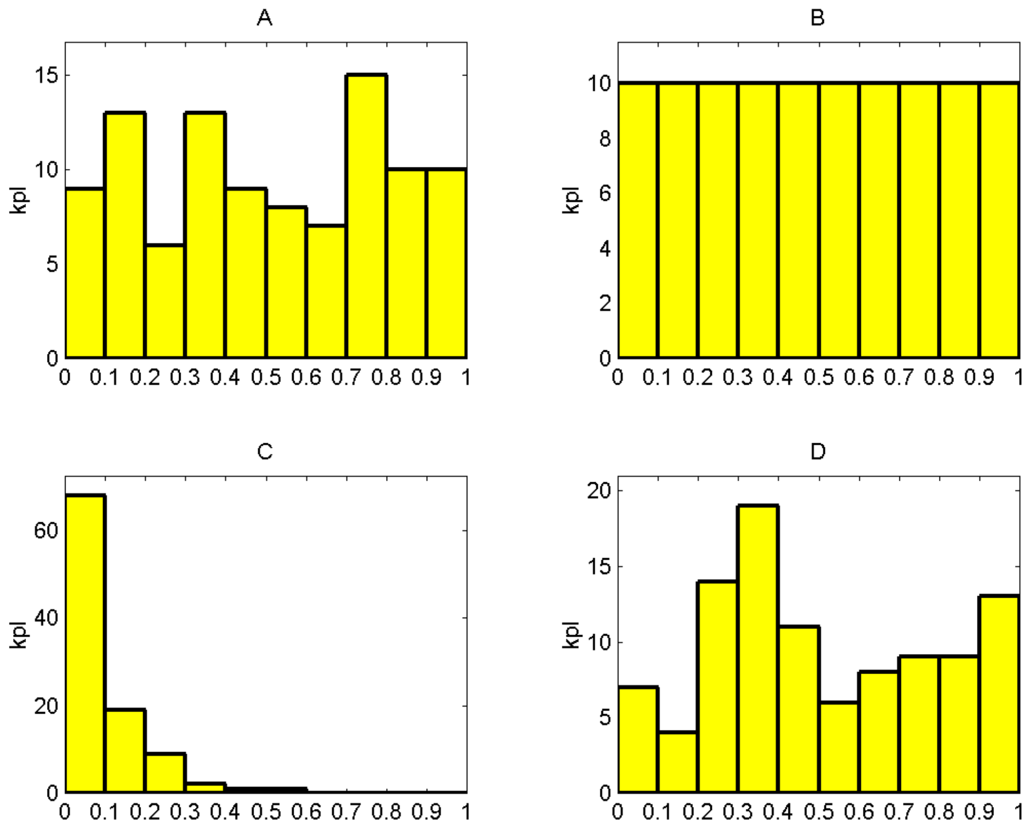


HISTOGRAMS OF RANDOM SAMPLES

Puzzle

We have four histograms of ten bins, from 100 numbers each.

Which histogram was the result of generating a random sample from the **U(0, 1)** uniform distribution?



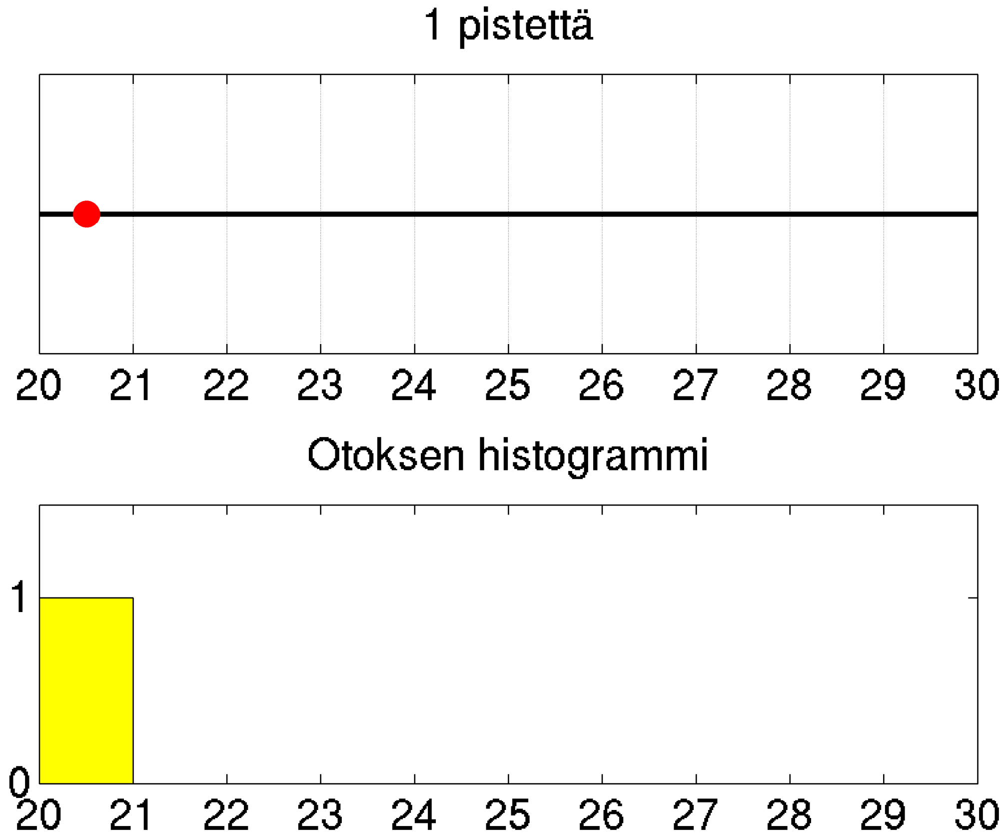
Sampling from U(0,1)

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

Miten ne sijoittuvat?

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

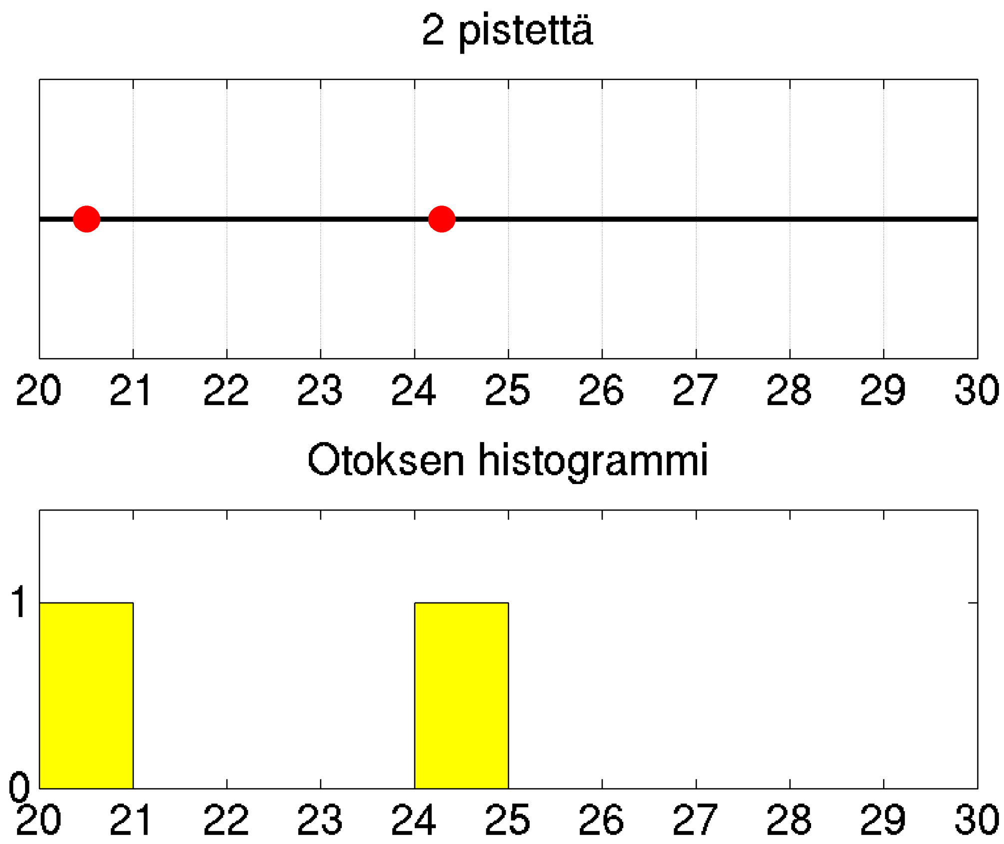
(Histogrammi on yhteenveto siitä, missä pisteet *suunnilleen* ovat.)



Sampling from U(0,1)

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

Miten ne sijoittuvat?



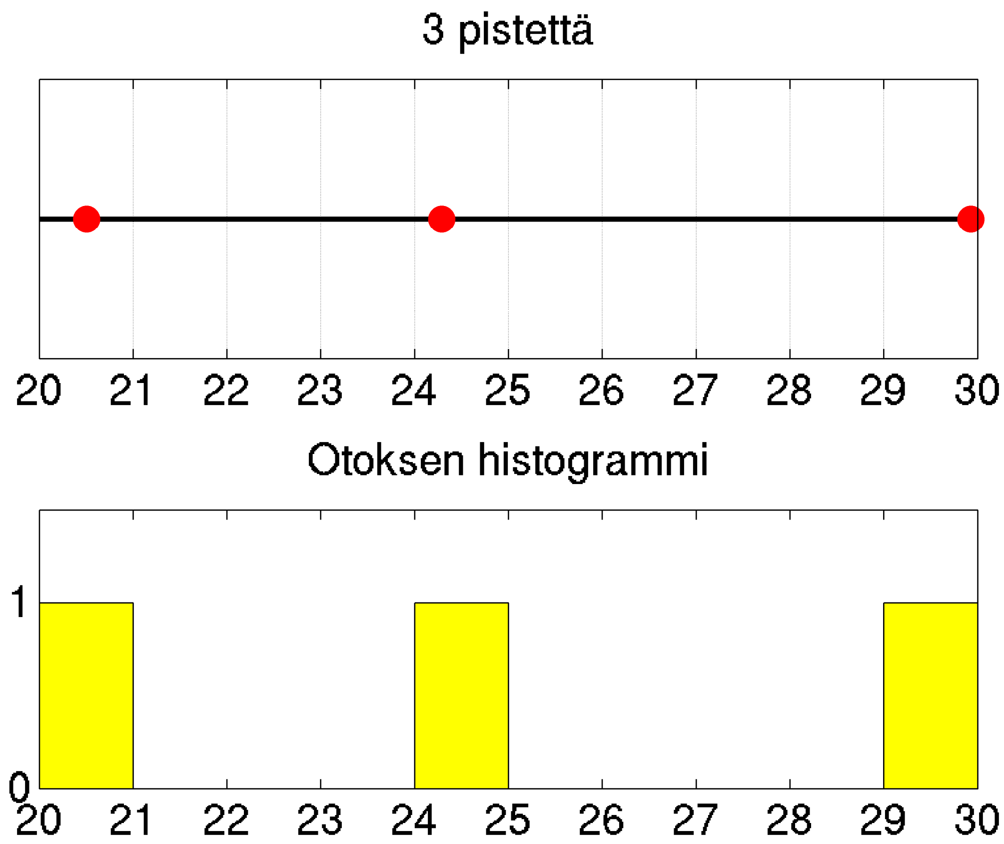
Piirretään myös 10 pylvään histogrammi, ei **jakaumasta** (joka on tasajakauma), vaan **otoksesta**, ts. mille väleille arvotut luvut osuivat.

Sampling from U(0,1)

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.



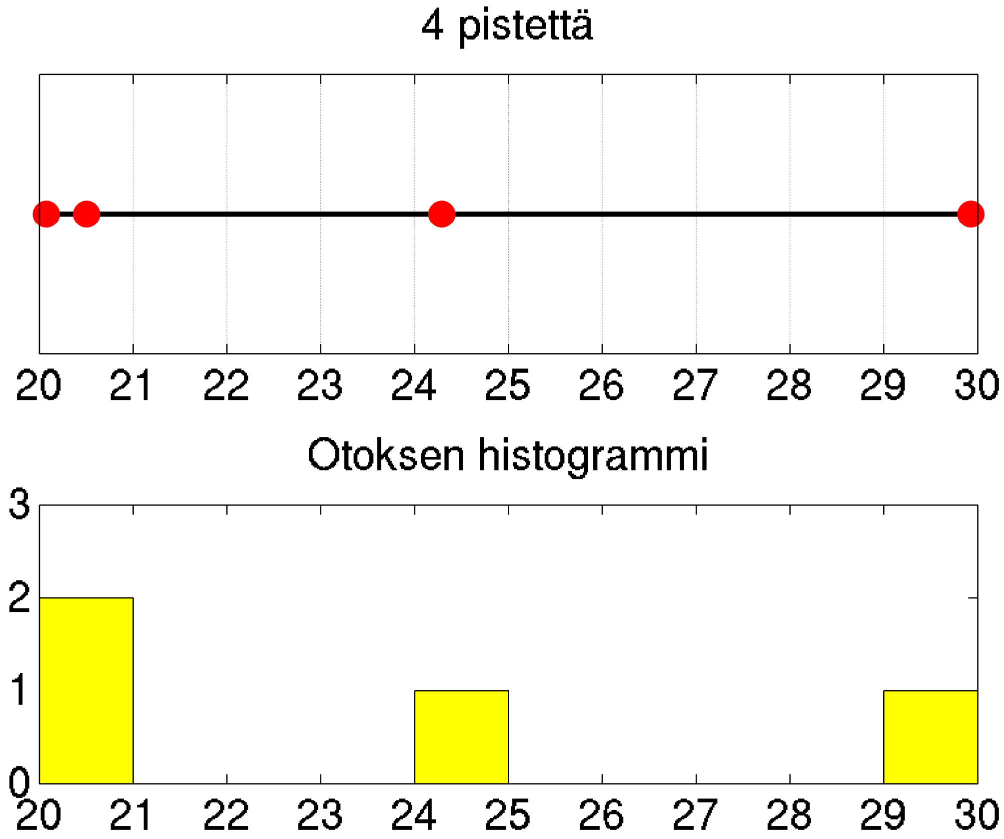
Sampling from U(0,1)

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

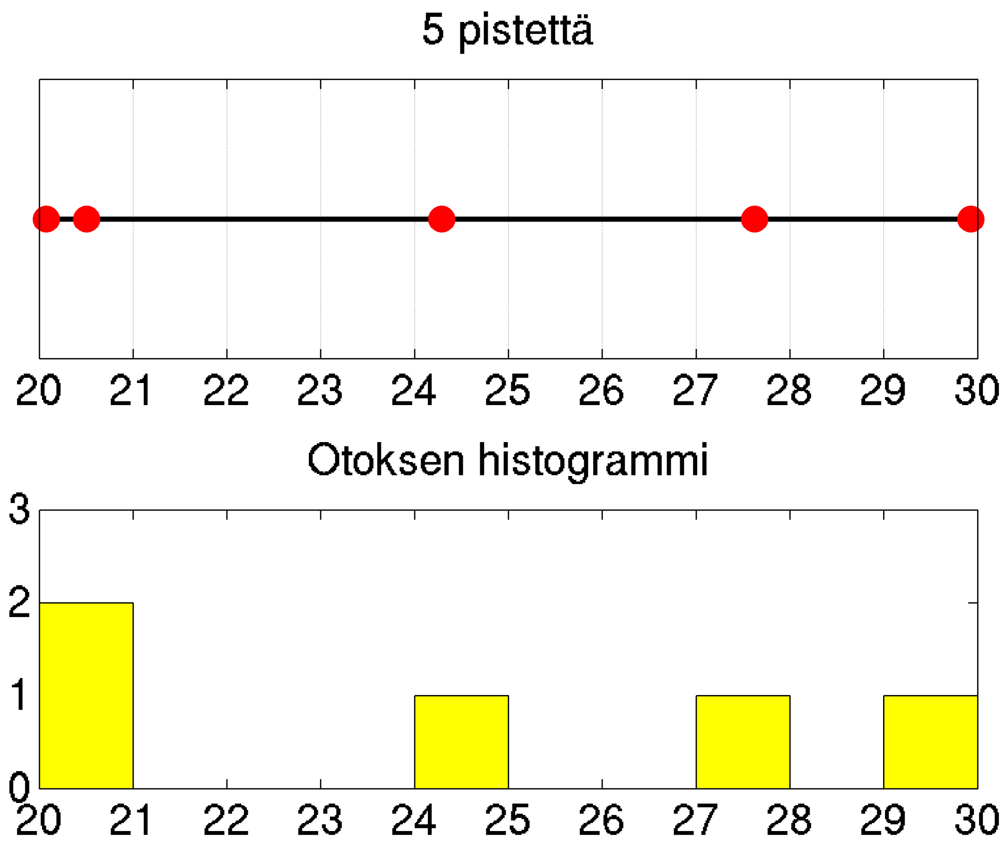
Riippumattomuus:
Aiemmat pisteet eivät vaikuta myöhempiin.

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

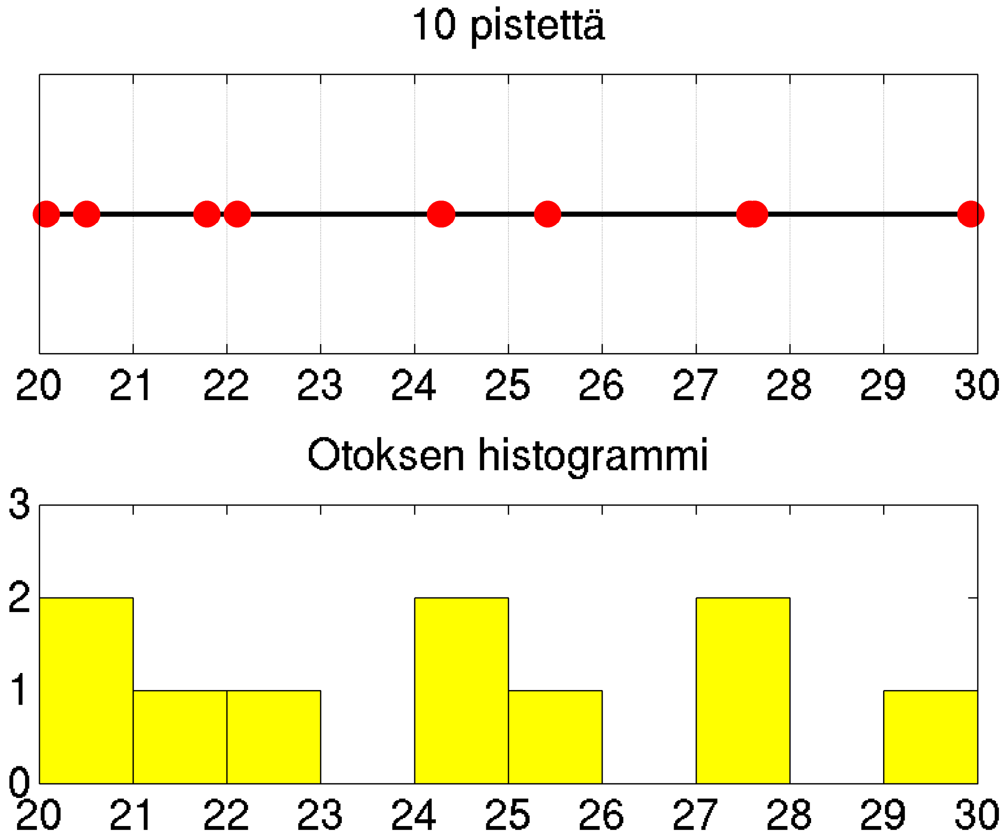
(ei edes lisää eikä vähennä ko. tapahtuman tn:ää, joka on jatkuvasti 1/10)



Sampling from U(0,1)



Sampling from U(0,1)

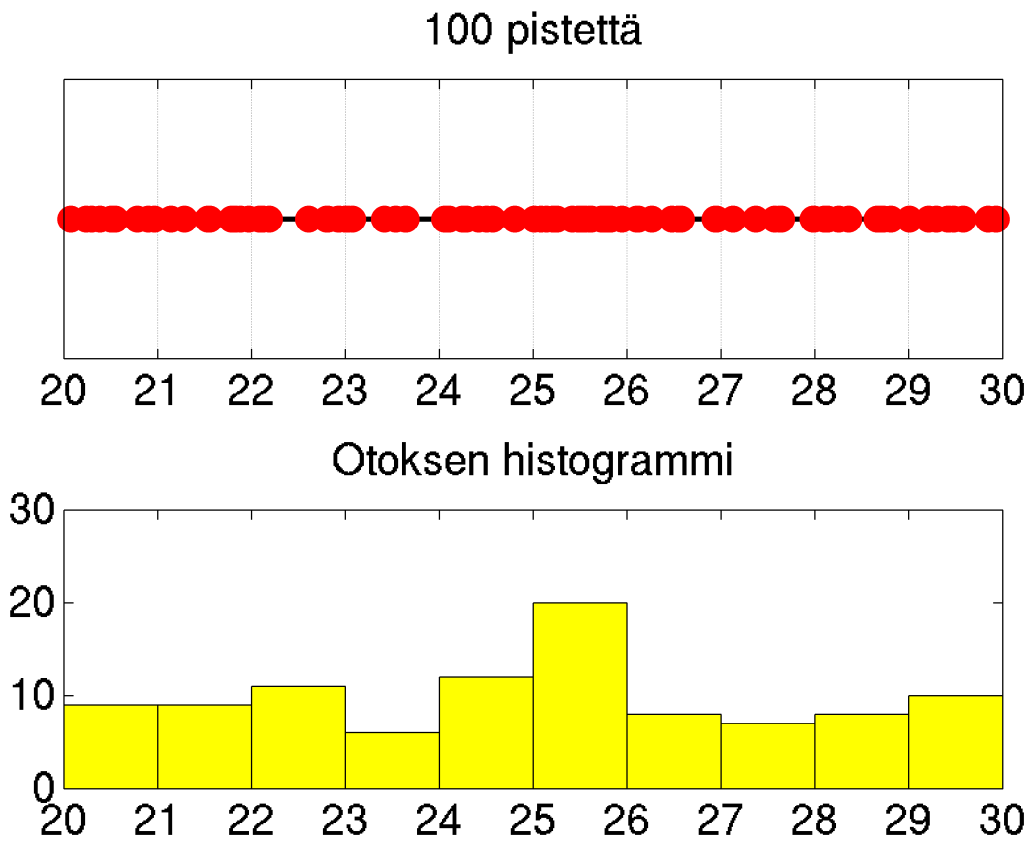


Sampling from U(0,1)

100 pistettä:

Kunkin pylvään
korkeuden odotusarvo
on 10,   (miksi?)

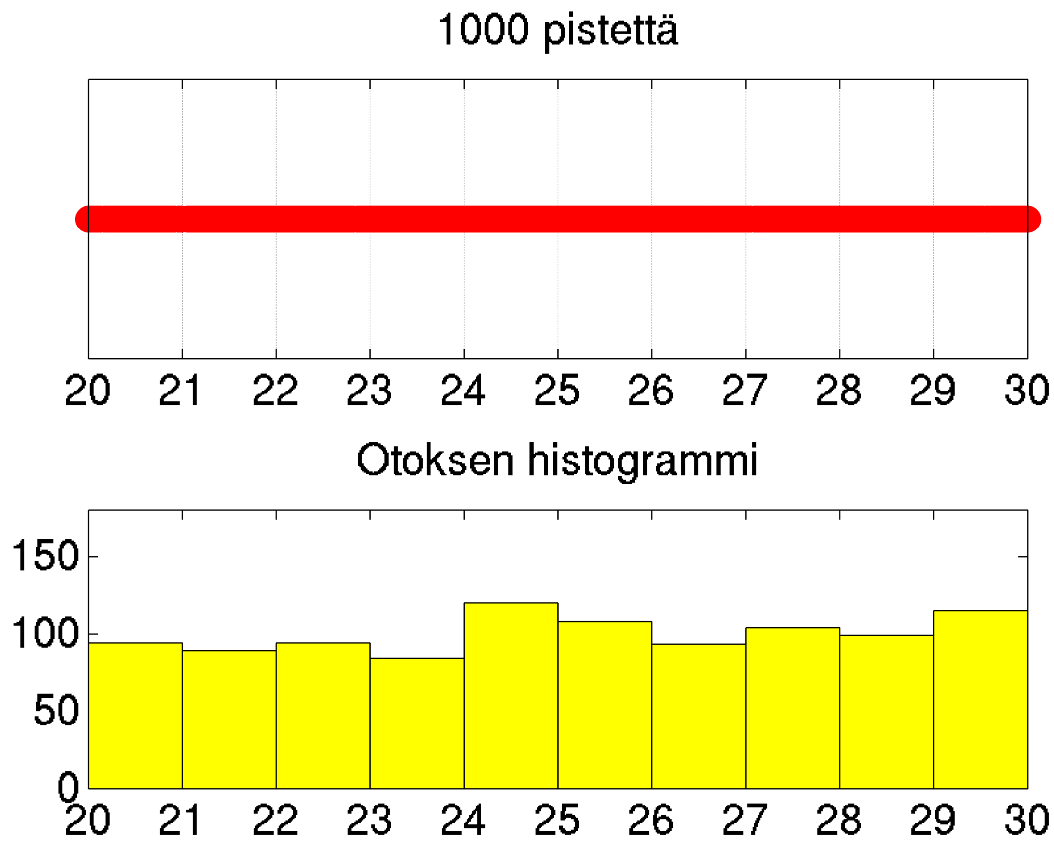
mutta toteutuneet korkeudet
vaihtelevat melkoisesti.



Sampling from U(0,1)

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

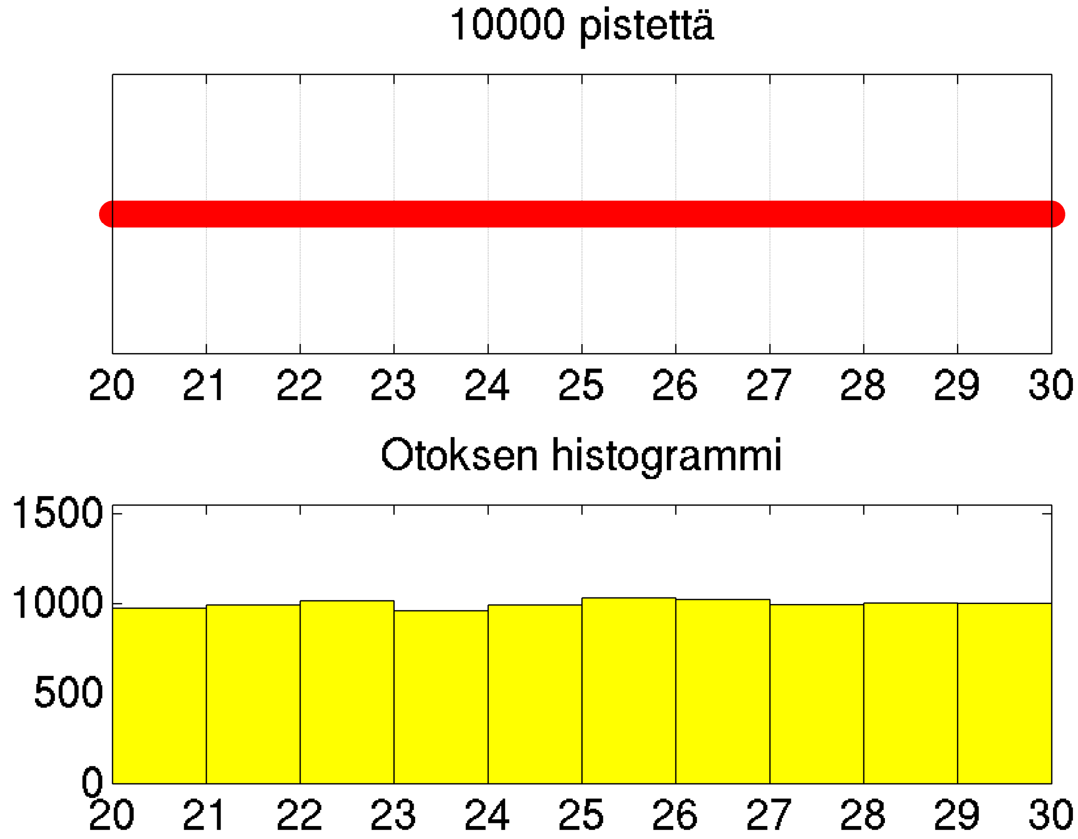
mutta histogrammista näemme osapuilleen pisteiden jakauman.



Sampling from U(0,1)

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.