

Bondor Cosmina

Variabile aleatoare, distributii de probabilitate

A ALWAYS

S SEEK

K KNOWLEDGE

Obiectivele cursului

- eșantionare
 - un eșantion reprezentativ al populației țintă într-un studiu
- distribuția de probabilitate a unei variabile aleatoare
- caracteristicile distribuției normale

Populația



- din punct de vedere
 - statistic
 - o colecție de elemente care au aceeași caracteristică
 - in domeniul sănătății
 - pacienți
 - unități spitalicești

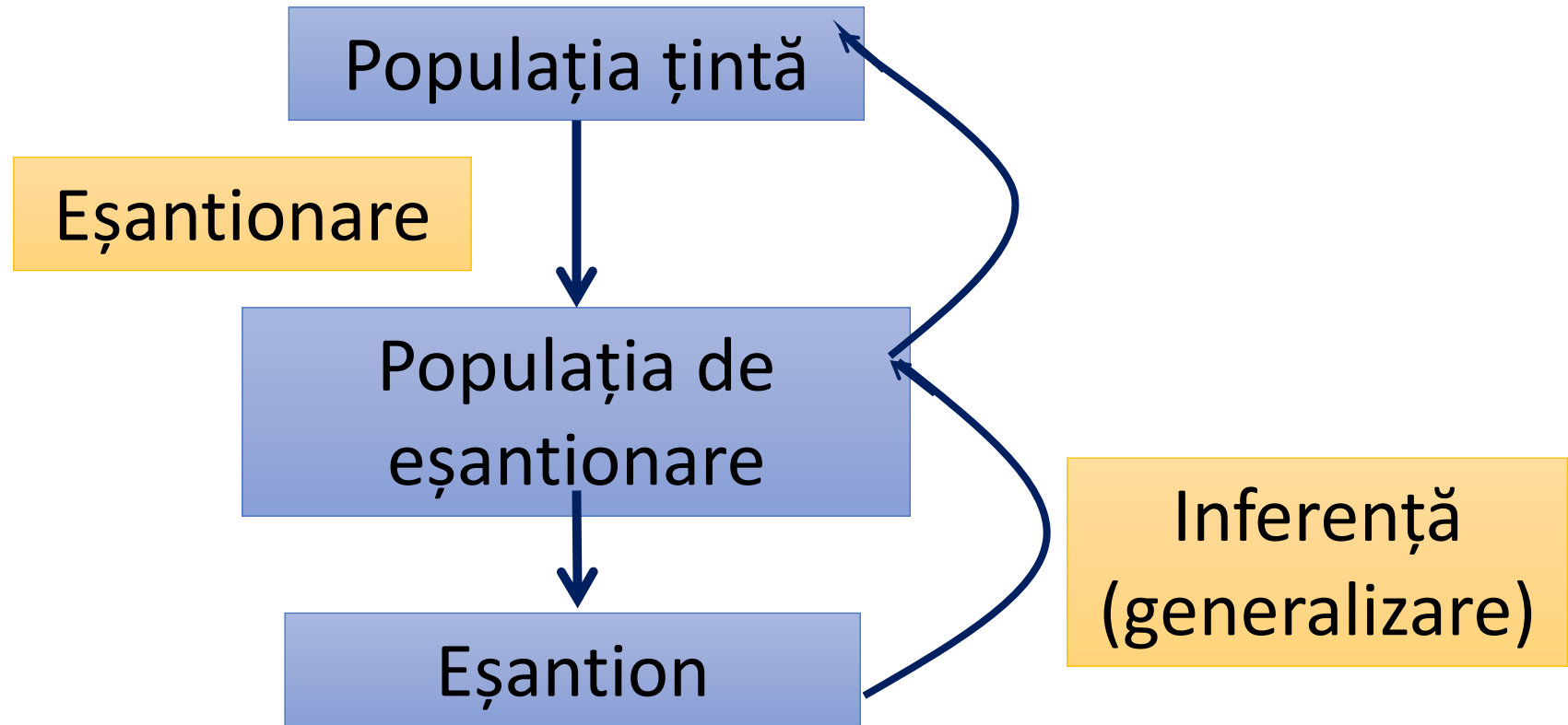
Eroare (bias) de selecție

- Ex.
- obiectivul - numărul de fracturi în populația generală într-un an
 - selecție de indivizi de la clubul de ski
- obiectivul – numărul de persoane cu infertilitate
 - selecție indivizi care vin la laborator să testeze infertilitatea
 - aceștia suspectează că sunt infertili

Selecție aleatoare

- Fiecare individ din populație are aceeași probabilitate de a fi selectat în eșantion
- Ex. Mergeți la primărie. Luați toate CNP-urile – extrageți aleator





Populație țintă – populația la care se dorește generalizarea rezultatelor studiului
Populația de eșantionare – populația din care a fost extras eșantionul

- Când folosim termenul „eșantion” în contextul cercetării medicale
 - vom presupune că eșantionul a fost selectat aleator într-un mod corect



Metode de eșantionare

Probabilistică: fiecare subiect din populație are o probabilitate cunoscută de a fi selectat

- Eșantionare simplu randomizată
 - Subiecților li se atribuie un număr
 - Se extrag numere **aleatorii** din listă
- Eșantionare sistematică
 - tot **al k-lea individ** se alege pentru a fi inclus în eșantion
- Eșantionare stratificată
 - Populația este împărțită în straturi după **însușiri care nu sunt echiprobabile, dar care pot influența obiectivul studiului**, se extrage aleator din fiecare strat
- Eșantionare de tip cluster
 - Cluster= **arie delimitată geografic**
 - Delimitarea clusterelor, selectarea aleatorie a clusterelor
 - Selectare aleatorie a subiecților din fiecare cluster selectat

Variabila aleatoare,
distribuția de probabilitate

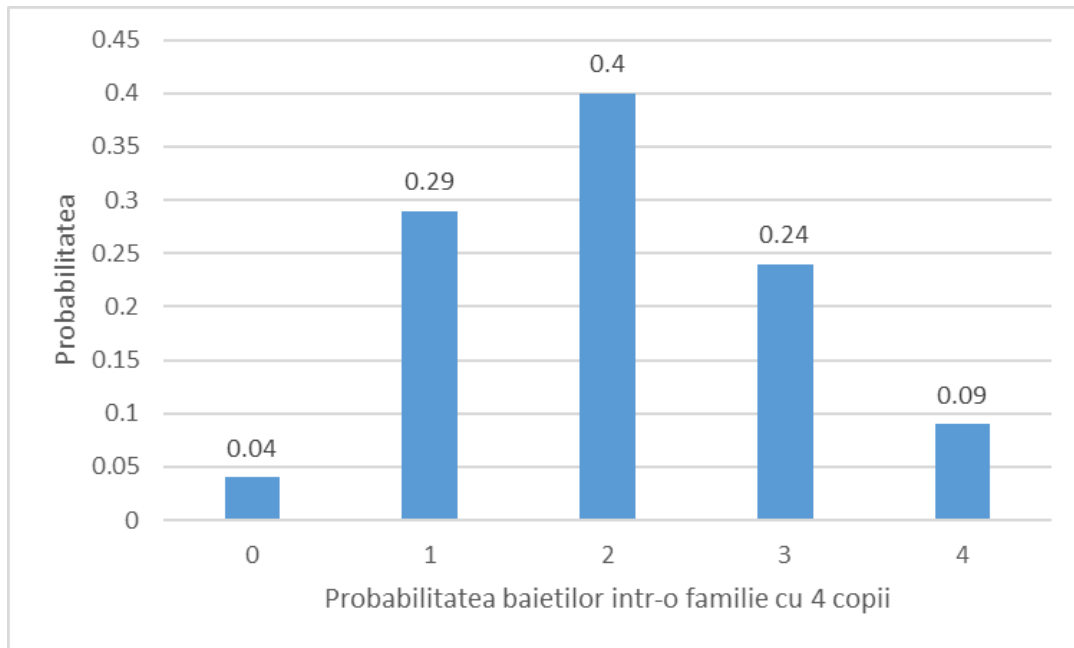
Distribuție de probabilitate

distribuție de frecvențe = distribuție de probabilitate

Cum calculăm distribuția de probabilitate?

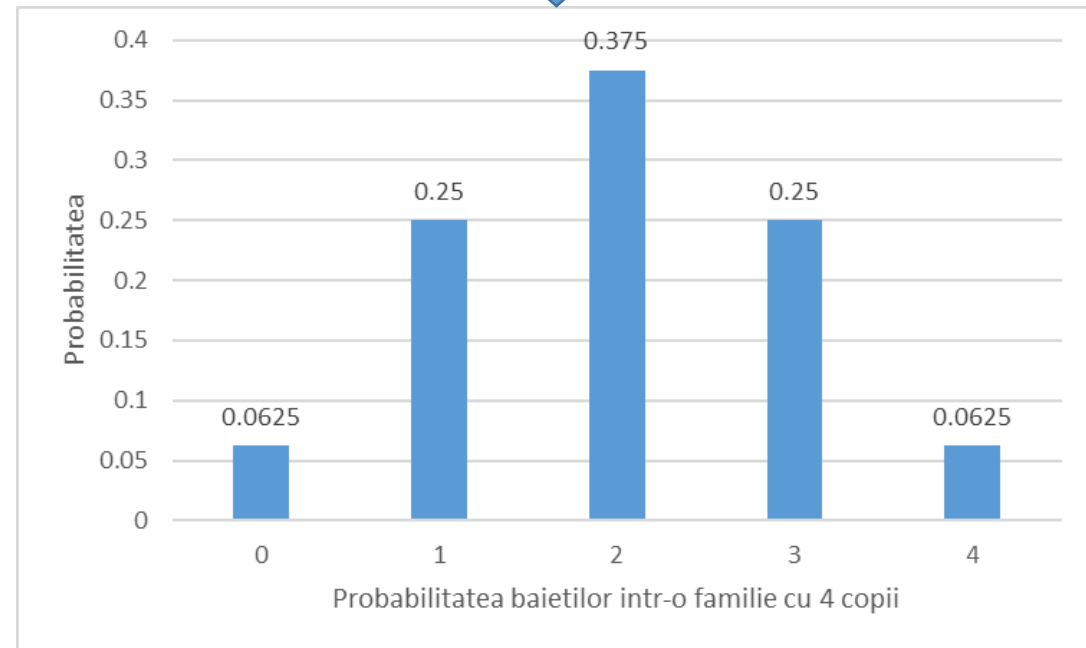
- Empiric

- eșantion, apoi inferență



- Teoretic

- Formulă
 - Aproximare cu o distribuție de probabilitate teoretică cunoscută



Cum aflăm distribuția de probabilitate?



Dacă variabila nu e continuă sau infinită

Empiric – experiment – distribuție de frecvențe

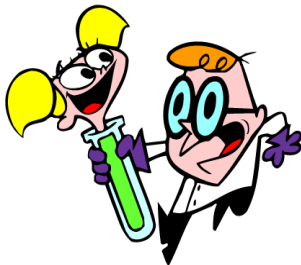


Dacă e ∞ sau continuă?

! suntem norocoși - găsim

Formulă

Regulă



Dacă nu suntem norocoși:

Modelăm (aproximăm) după o distribuție teoretică de probabilitate (cunoscută – una la care am fost norocoși)

Legi de distribuție

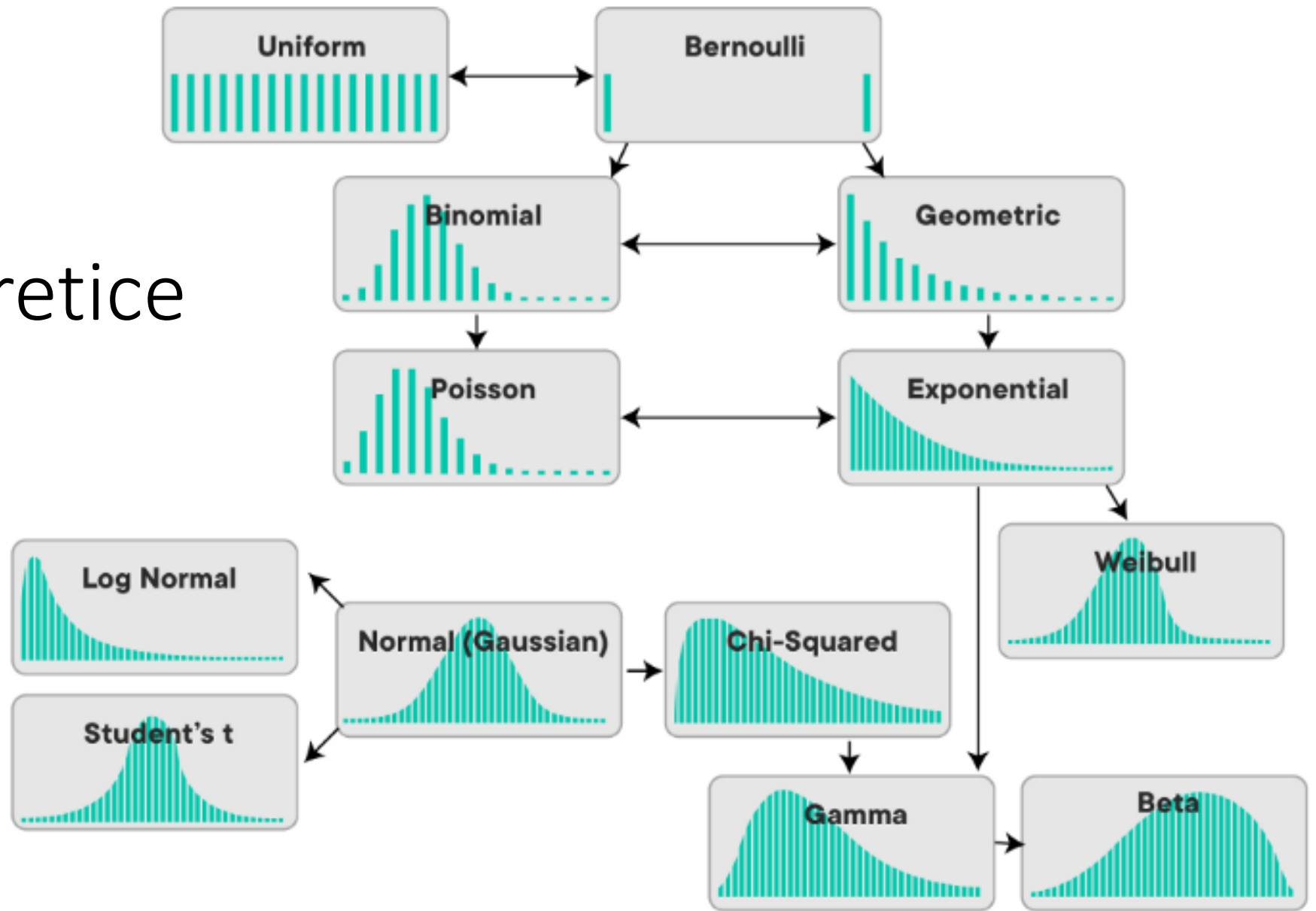
(distribuții de probabilitate)



Cele mai
cunoscute



Distribuții teoretice



- Au o funcție cunoscută, medie și deviație standard deductibilă



LEGEA NORMALĂ

- variabilă aleatoare continuă
- funcție de probabilitate - alură de clopot
 - curba normală
 - curba lui Gauss
- Această distribuție depinde de doi parametri:
 - media aritmetică μ
 - abaterea standard (varianța) σ
- densitate de probabilitate:

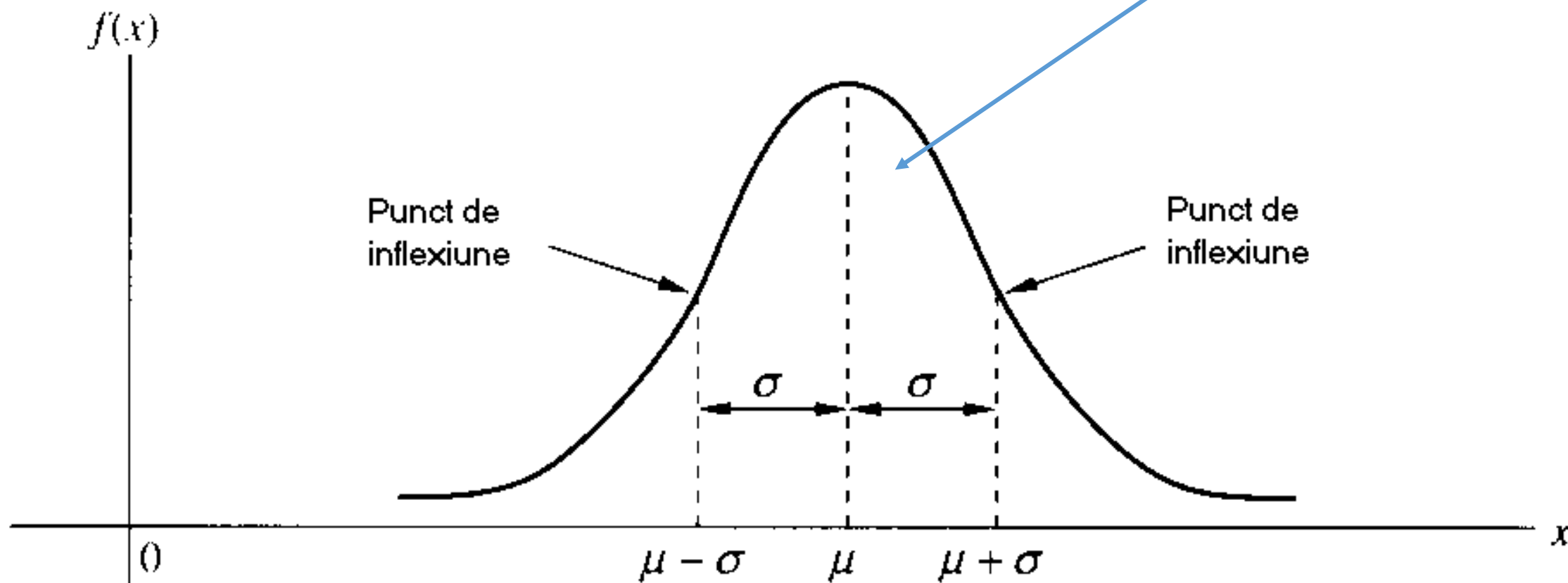
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



1777–1855

LEGEA NORMALĂ

Aria de sub curbă este 1, ca la orice distribuție de probabilitate



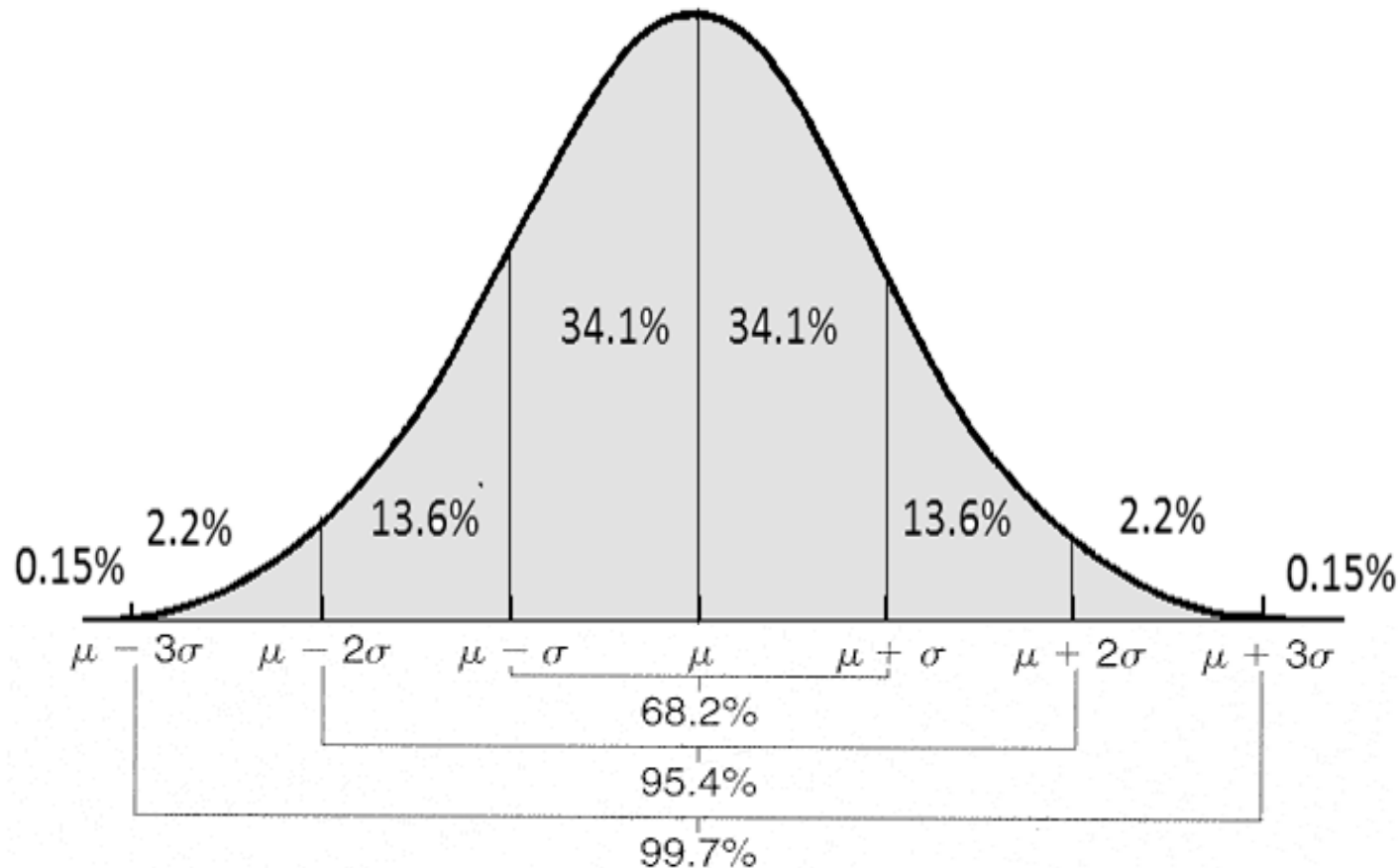
σ deviația standard (varianța) este distanța dintre medie și punctul de inflexiune (acolo unde curba se schimbă din concavă în convexă)

Proprietăți:

În intervalul medie \pm abatere standard - minim 68,2% din observații;

În intervalul medie ± 2 * abatere standard - minim 95,4% din observații;

În intervalul medie ± 3 * abatere standard - minim 99,7% din observații.





Aplicații: Cum este distribuția datelor?

Dacă aceste condiții sunt îndeplinite

- media \approx mediana \approx modulul
- simetria ≈ 0
- boltirea ≈ 0
- cvartilele 1 și 3 simetrice față de media aritmetică
- În intervalul $\text{medie} \pm \text{abatere standard}$ \ni minim 68,2% din observații;
- În intervalul $\text{medie} \pm 2 * \text{abatere standard}$ \ni minim 95,4% din observații;
- În intervalul $\text{medie} \pm 3 * \text{abatere standard}$ \ni minim 99,7% din observații,
- atunci distribuția datelor obținute empiric se apropie de distribuția normală

Distribuția paranormală

- Mulțumesc!!!

