



Autor: Bondor Cosmina-loana

Probabilități



ALWAYS



SEEK



KNOWLEDGE

Objective

Semnificația termenului Probabilitate

Metode de eșantionare

Distribuții de probabilitate

Exemple

Probabilitate

- Definiții de bază
 - experiment
 - trial
 - rezultat, eveniment
- Reguli despre combinarea probabilităților a două evenimente:
AND, OR, NON
- Evenimente: mutual exclusive, imposibile, complementare, certe, independente-dependente

Probabilitatea



Un **experiment** repetat de mai multe ori,

Fiecare repetare se numește trial

Din fiecare trial pot rezulta unul sau mai multe rezultate (evenimente)

$$\text{Probabilitatea (eveniment)} = \frac{\text{numărul de apariții ale acelui eveniment}}{\text{numărul total de trialuri}}$$

Exemplu 1 – Ebola virus

- Ebola – infecție acută cu risc mare de mortalitate.
- 2014 Outbreack CDC. 2014 Ebola Outbreak in West Africa - Case Counts. Nov.2015. Available at:
<http://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/case-counts.html>

Țara	Cazuri	Decese	Supraviețuiri	Probabilitatea decesului	Probabilitatea supraviețuirii
Siera Leone	14.122	3.955			

Exemplu 1 – Ebola virus

Țara	Cazuri	Decese	Supraviețuiri	Probabilitatea decesului	Probabilitatea supraviețuirii
Siera Leone	14.122	3.955	=14.122-3.955		

Exemplu 1 – Ebola virus

Țara	Cazuri	Decese	Supraviețuiri	Probabilitatea decesului	Probabilitatea supraviețuirii
Siera Leone	14.122	3.955	10.167	=3.955/14.122	

Exemplu 1 – Ebola virus

Țara	Cazuri	Decese	Supraviețuiri	Probabilitate a decesului	Probabilitatea supraviețuirii
Siera Leone	14122	3955	10167	0.28	

Exemplu 1 – Ebola virus

Deces \longleftrightarrow Supraviețuire

Evenimente complementare

Nr. de decese + Nr. de supraviețuitori = Total

Decese (%) + Supraviețuire (%) = 100 %

Exemplu 1 – Ebola virus

Supraviețuire (%) = 100 % - decese (%)

Țara	Cazuri	Decese	Supraviețuiri	Probabilitatea decesului	Probabilitatea supraviețuirii
Siera Leone	14.122	3.955	10.167	0,28	=1 – 0,28 or =10.167/14.122

Exemplu 1 – Ebola virus - $P(A)$ and $P(\text{non}A)$

Decese (%) + Supraviețuire (%) = 100 %

$$P(A) + P(\text{non}A) = 1 \Rightarrow P(\text{non}A) = 1 - P(A)$$

Țara	Cazuri	Decese	Supraviețuiri	Probabilitate a decesului	Probabilitatea supraviețuirii
Siera Leone	14.122	3.955	10.167	0,28	0,72

Exemplu 1 – Ebola virus - P(A) and P(nonA) - Home Exercises

Decese (%) + Supraviețuire (%) = 100 %

$$P(A) + P(\text{non}A) = 1$$

Țara	Cazuri	Decese	Supraviețuiri	Probabilitate a decesului	Probabilitatea supraviețuirii
Guinea	3.805	2.536	?	?	?
Siera Leone	14.122	3.955	10.167	0,28	0,72
Liberia	10.672	4.808	?	?	?
Others	36	15	?	?	?
Total	28.635	11.314	?	?	?

Exemplu 2 – $P(A \text{ și } B)$ – Grupa de sânge pentru 1000 de persoane

Grupa sanguină	Frecvența	$P(\text{Grupa sanguină})$	Femei	$P(\text{Femei})$	$P(\text{Femei și grupa sanguină})$
O	400		200		
A	450		200		
B	142		96		
AB	8		4		
Total	1000		500		

Exemplu 2 – P(A și B) – Grupa de sânge pe genuri

Grupa sanguină	Frecvența	P(Grupa sanguină)	Femei	P (Femei)	P (Femei și grupa sanguină)
A	400 450	=450/1.000	200 200		
	142		96		
	8		4		
Total	1.000				

$P(A \text{ și } B) = P(A) * P(B)$ – pentru evenimente independente (care nu sunt cauză-efect)

Grupa sanguină	Frecvența	P(Grupa sanguină)	Femei	P (Femei)	P (Femei și grupa sanguină)
A	450	0.45	200	=200/450	
Total	1.000				

$P(A \text{ și } B) = P(A) * P(B)$ – pentru evenimente independente

Grupa sanguină	Frecvența	P(Grupa sanguină)	Femei	P (Femei)	P (Femei și grupa sanguină)
A	400 450 142 8	0,45	200 96 4	0,44	$= P(femei)$ $* P(Grupa A)$ $= \mathbf{0,45} * \mathbf{0,44}$
Total	1000				

Exemplu 2 – P(A și B) – Distribuția grupei sanguine – Exerciții temă

Grupa sanguină	Frecvența	P(Grupa sanguină)	Femei	P (Femei)	P (Femei și grupa sanguină)
O	400	?	200	?	?
A	450	0,45	200	0,44	0,20
B	142	?	96	?	?
AB	8	?	4	?	?
Total	1000	?	500	?	?

$$P(A \text{ sau } B) = P(A) + P(B) - P(A \text{ și } B)$$

Grupa sanguină	Frecvența	P(Grupa sanguină)
O	400	0,40
A	450	0,45
B	142	0.142
AB	8	0,008
Total	1.000	1,0

$P(\text{Grupa 0 sau Grupa A}) = ?$

$= P(\text{Grupa 0}) + P(\text{Grupa A}) - P(\text{Grupa 0 și Grupa A}) =$

$$= \frac{400}{1000} + \frac{450}{1000} - 0 = \frac{950}{1000} = 0.95$$

Grupa 0 și Grupa A sunt mutual exclusive:

$$P(\text{Grupa 0 and Grupa A}) = 0$$

Două evenimente sunt mutual exclusive dacă nu pot avea loc în același timp

Rezumat



Oricare două evenimente A, B:

$$P(A \text{ sau } B) = P(A) + P(B) - P(A \text{ și } B)$$

Dacă A, B două evenimente independente:

$$P(A \text{ și } B) = P(A) * P(B)$$

Dacă A, B două evenimente dependente:

$$P(A \text{ și } B) = P(A \text{ dependent } B) * P(B)$$

$$P(A \text{ și } B) = P(A|B) * P(B)$$

Rezumat



- Dacă A, B sunt evenimente **mutual exclusive**

$$P(A \cap B) = P(A \text{ și } B) = 0$$

- Dacă B este un eveniment **complementar** lui A

$$P(B) = P(\text{eveniment complementar lui A}) = P(\text{non A}) = 1 - P(A)$$

- Dacă A este un eveniment **cert** (se întâmplă la fiecare trial)

$$P(\text{eveniment cert}) = 1$$

- Dacă A este un eveniment **imposibil**

$$P(\text{eveniment imposibil}) = 0$$

Teorema lui Bayes



Evenimente dependente (cauză și efect)

Probabilitatea condiționată - probabilitatea unui rezultat care depinde de un rezultat anterior.

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}.$$

Implicată în luarea deciziilor medicale - teste diagnostice.

- A depend de B adică mai întâi B apoi A
- ! Asta nu înseamnă că Non B ~~cauzează~~ non A
- Non A cauzează non B

De ce să studiem eșantioanele în locul întregii populații?

Se studiază eșantioane în locul întregii populații

Mai repede

Mai puțin costisitor

Mai puțin periculos

Concluzii mai precise

Cercetarea clinică

Generalizarea rezultatelor:

de la **Grup de pacienți** → la **toți subiecții similari**

adică

de la **Eșantion** → **Populație**

Exemplu

1000 votanți
la alegerile
parlamentare

Cum ați
votat?

55 % PSD

Când putem generaliza?

Generalizare: Probabilitatea ca să învingă PSD în alegeri este de 0,55

Selecția eșantionului influențează rezultatul găsit!

- Dacă selectăm numai persoane din **Teleorman** atunci pe eșantionul de 1000 de persoane vom găsi **95%** frecvența votului cu PSD.
- Dacă selectăm numai **persoane tinere** atunci pe eșantionul de 1000 de persoane vom găsi **35%** frecvența votului cu PSD.

Cum facem selecția ca să găsim un rezultat corect?

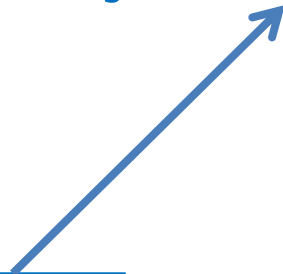
Condiția pentru a putea generaliza de pe eșantion
la populație



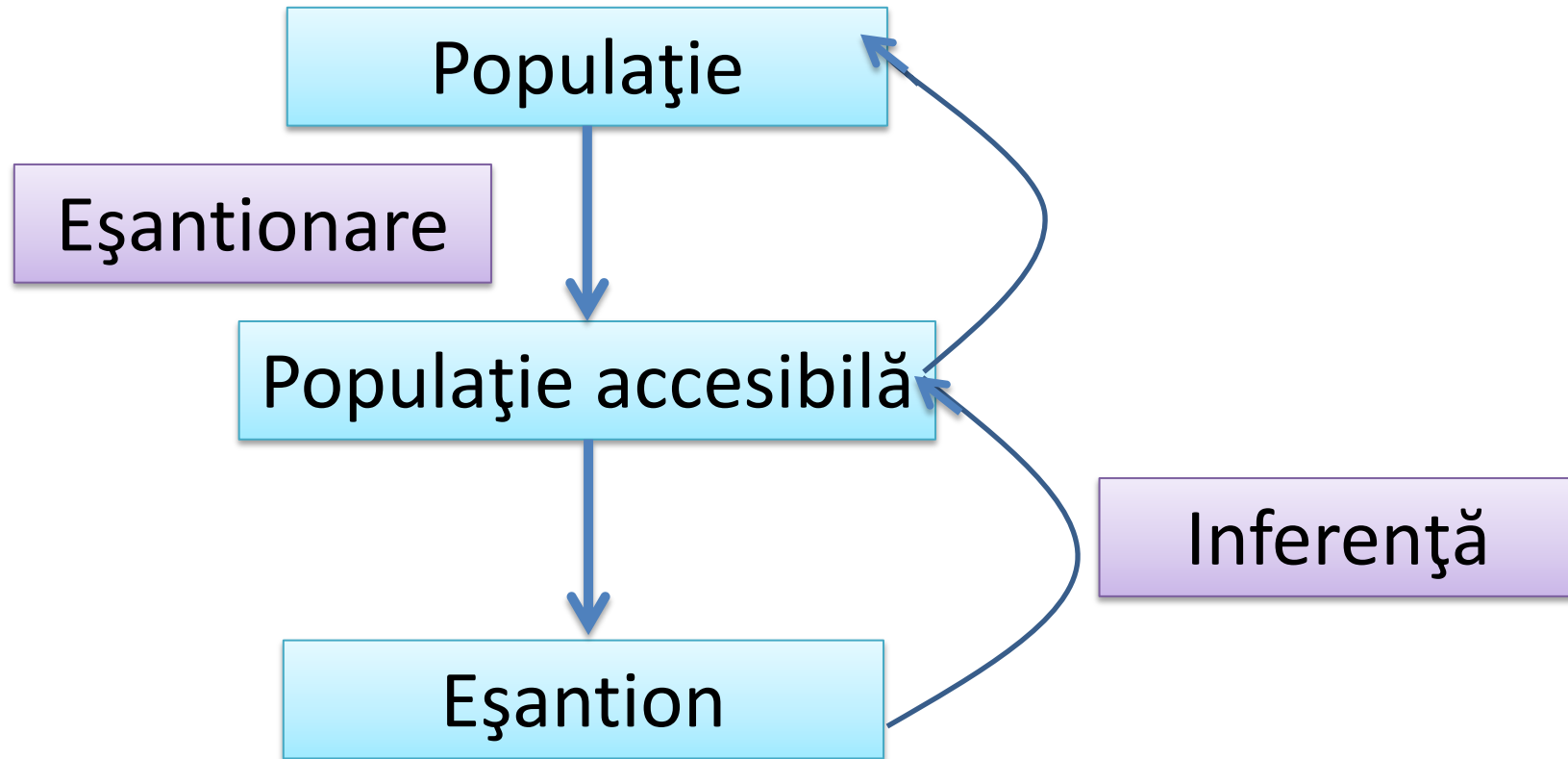
Selecția aleatoare

Care este populația: toți locuitorii țării

Populația



Eșantionare



De ce selecție aleatoare?



- Reducerea / eliminarea erorilor experimentale = reducerea / eliminarea erorii (bias) de selecție
- Eșantionul ar trebui să fie reprezentativ pentru populație = să aibă aceeași distribuție a caracteristicilor importante ca și populația
- Unde: caracteristici importante = în legătură cu caracteristica studiată

Metode de eșantionare



- Eșantionare **aleatoare** - fiecare subiect are aceeași probabilitate de a fi selectat
- Eșantionarea **sistematică** - fiecare al k-lea subiect este selectat
- Eșantionarea **stratificată** - populația este împărțită în subgrupe și se selectează aleator din fiecare subgrupă (ex. după categorii de vârstă)

Metode de eșantionare



- Eșantionarea **cluster** - populația este împărțită în grupuri și se selectează un eșantion aleatoriu din fiecare grup (cluster = zone geografice)
- Eșantion **nonprobabilist** - probabilitatea ca un subiect să fie selectat este necunoscută

Metode de eșantionare

- Atunci când cercetătorii selectează eșantioane aleatorii și apoi fac măsurători, rezultatul este o **variabilă aleatoare**.

Distribuția de probabilitate

- Valorile unei variabile aleatoare pot fi rezumate într-o distribuție a frecvenței numită **distribuție de probabilitate**

Ex. Distribuția de probabilitate pentru numărul de băieți în familiile cu 4 copii

Număr de băieți	0	1	2	3	4	Total
Probabilitatea	0,0625	0,25	0,375	0,25	0,0625	1,00

- Cum calculăm distribuția de probabilitate?

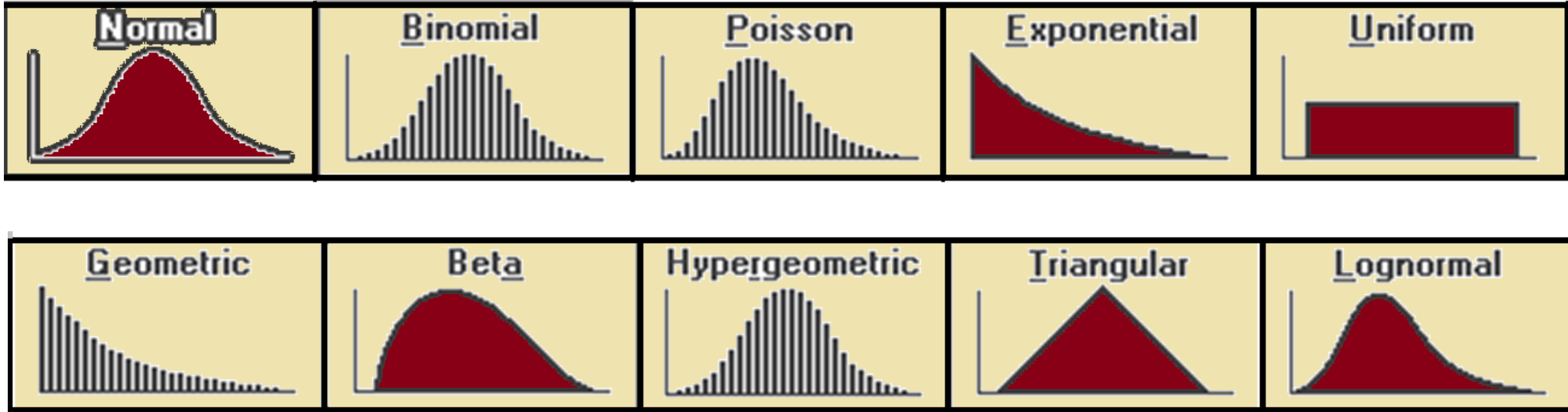
Cum calculăm distribuția de probabilitate?

Formulă

Regulă

Se modelează după o distribuție teoretică de probabilitate
(cunoscută)

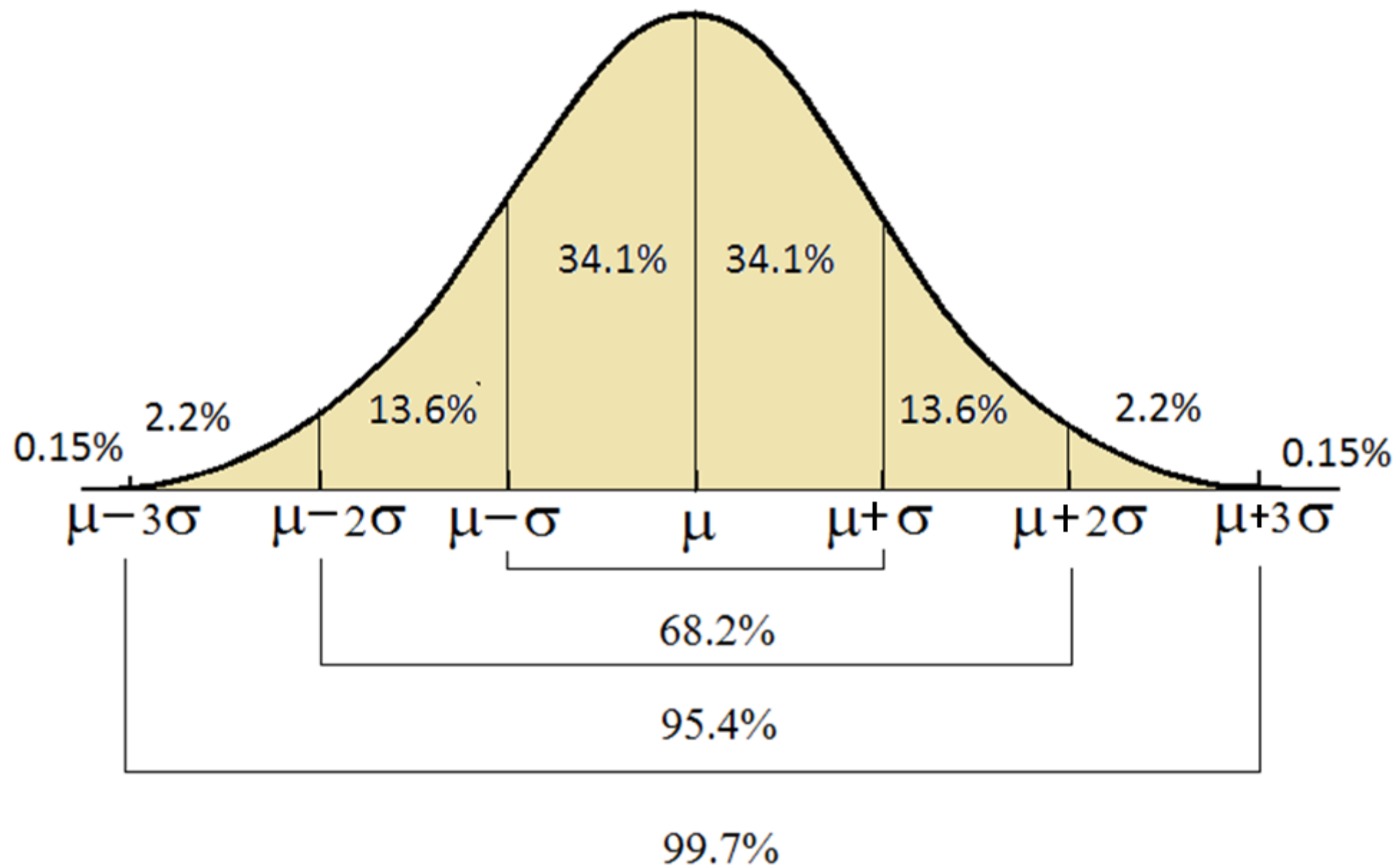
Distribuții teoretice



Caracteristici: au o ecuație cunoscută, medie deductibilă, deviație standard deductibilă și proprietăți cunoscute

Simboluri utilizate frecvent în inferența statistică

Caracteristica	Simbolul pentru parametru calculat pe populație	Simbolul pentru parametru calculat pe eșantion
Media aritmetică	μ	\bar{X}
Deviația standard	σ	s
Proporția	π	p



Distribuția normală

- Variabila aleatoare X se numește distribuită normal $N(\mu, \sigma)$ dacă distribuția depinde de doi parametri: media μ și deviația standard σ și are formula:

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

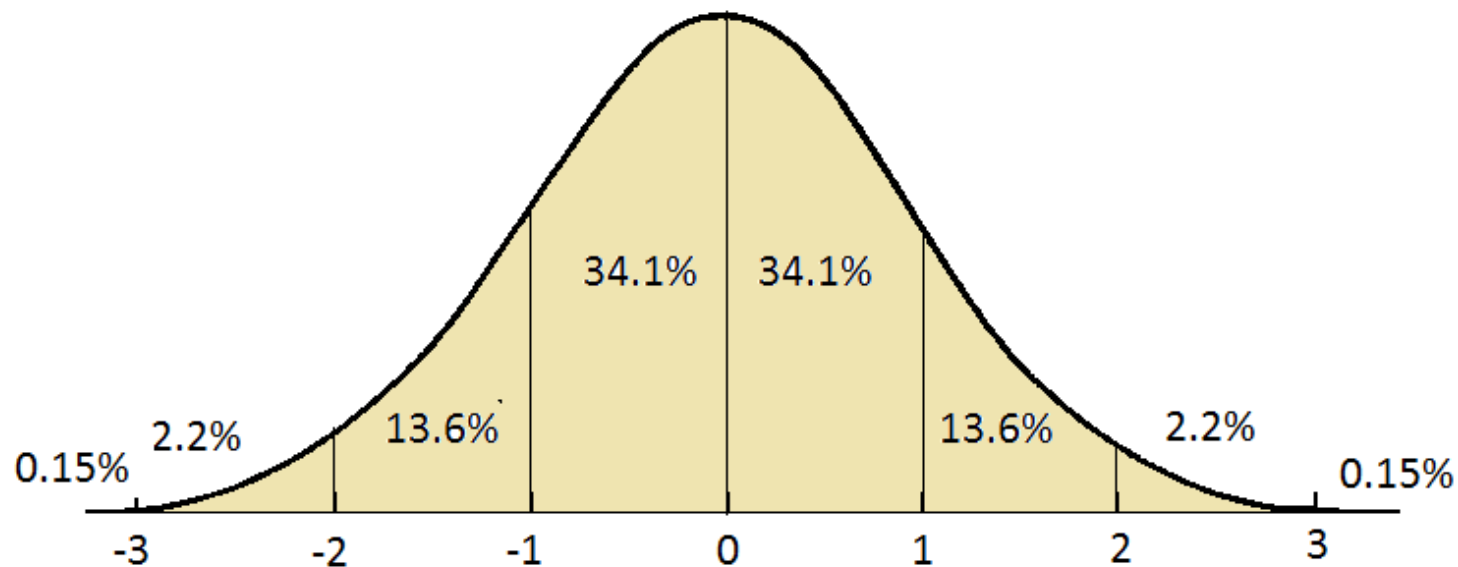
Distribuția normală **standard**



Este o distribuție normală cu $\mu=0$ și $\sigma=1$.

Formula:

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



- Mulțumesc!!!

