

2023  
InfoMed

...

# Aplicațiile medicale ale calculului probabilităților



Mădălina Văleanu

## ❖ Definiții:

- Experiment aleator
- Definiția clasică a probabilității
- Spațiul fundamental de evenimente
- Definiția axiomatică a probabilității

## ❖ Probabilități condiționate:

- ❖ Definiții
- ❖ Evenimente independente/dependente
  - Sensibilitate, specificitate, VPP, VPN
  - Teorema lui Bayes
  - Independența a două evenimente

- ❖ Teoria probabilităților folosește o serie de concepte fundamentale cum sunt:
  - *experimentul*
  - *proba*
  - *evenimentul*
  - *probabilitatea*
- ❖ Un **experiment** poate fi definit și ca un proces de colectare a datelor dintr-o populație.
- ❖ Prin **experiment aleator** se înțelege realizarea practică a unui complex de condiții, corespunzătoare unui criteriu de cercetare al diferitelor colectivități (populații), care aplicat unei entități a colectivității are un rezultat întâmplător.

- ❖ **Determinarea statusului de a fi seronegativ sau seropozitiv.**
- ❖ **Determinarea grupei sangvine.**
- ❖ **Determinarea prezenței sau absenței unui factor de risc (fumat, obezitate, stres, etc.).**



# Definiția clasică a probabilității



- ❖ **Presupunându-se că toate probele posibile ale unui experiment sunt la fel de verosimile (echiprobabile), probabilitatea de apariție a unei probe (sau a unui grup de probe) prin efectuarea experimentului este egală cu raportul dintre numărul de moduri în care această probă (sau grup de probe) se poate întâmpla la numărul total de probe pe care experimentul le poate genera.**

- ❖ dacă un eveniment  $A$  se poate realiza în  $s$  probe dintr-un total de  $n$  încercări (probe) echiprobabile pe care experimentul le poate produce, atunci probabilitatea evenimentului  $A$  se poate defini prin formula:

$$P(A) = \frac{s}{n} = \frac{\text{Numarul de cazuri favorabile}}{\text{Numarul de cazuri posibile}}$$

- ❖ Această definiție stă la baza calculării probabilității empirice a unui eveniment pe baza unor date statistice privind evenimentul respectiv

# Definiția clasică a probabilității



- ❖ Să presupunem că suntem interesați în cunoașterea probabilității nașterii unui băiat la o naștere oarecare.
- ❖ Care este această probabilitate?



# Definiția clasică a probabilității



Perioada	Nr. de născuți de sex masculin (a)	Nr. total de nașteri (b)	Probabilitatea empirică a nașterii unui copil de sex masculin (a/b)
1965	1927054	3760358	0.51247
1965-1969	9219202	17989361	0.51248
1965-1974	17857857	34832051	0.51268

## definiția clasică








- presupune probe (încercări) echiprobabile
- această cerință limitează utilizarea sa.

# Definiția clasică a probabilității



Dacă încercările constau în determinarea grupei sangvine, rezultatele posibile sunt: A, B, AB, O,

## The ABO Blood System

Blood Type (genotype)	Type A (AA, AO)	Type B (BB, BO)	Type AB (AB)	Type O (OO)
Red Blood Cell Surface Proteins (phenotype)	 A agglutinogens only	 B agglutinogens only	 A and B agglutinogens	 No agglutinogens
Plasma Antibodies (phenotype)	 b agglutinin only	 a agglutinin only	NONE. No agglutinin	 a and b agglutinin

Grupa Sangvina	Probabilitatea
O	?
A	?
B	?
AB	?

# Definiția clasică a probabilității



**Dacă încercările constau în determinarea grupei sangvine, rezultatele posibile fiind: A, B, AB, O, acestea nu sunt echiprobabile.**

Grupa Sangvina	Probabilitatea
O	0.42
A	0.43
B	0.11
AB	0.04

**Acesta este motivul pentru care este necesară o abordare mai generală a noțiunii de probabilitate printr-o introducere axiomatică a acestei noțiuni.**

# Definiția axiomatică a probabilității



# Spațiul fundamental de evenimente



- ❖ Fie  $H$  un experiment aleator dat pentru care  $E$  reprezintă mulțimea tuturor rezultatelor posibile.
- ❖  $E$  se numește ***mulțime fundamentală sau spațiu fundamental (spațiul evenimentelor elementare)***.
- ❖ Spațiul fundamental poate să fie finit sau infinit.
- ❖ Astfel că, o submulțime  $A$  a lui  $E$  se numește ***eveniment***, iar dacă  $A$  are un singur element din  $E$  el este un ***eveniment elementar***.
- ❖ Orice eveniment a cărui realizare depinde de cel puțin două evenimente elementare este un ***eveniment compus***.

- ❖ **Mulțimea vidă  $\emptyset$  și mulțimea fundamentală  $E$  sunt și ele evenimente, și anume, evenimentul *imposibil* ( $\emptyset$ ) și respectiv evenimentul *cert* ( $E$ ).**
- ❖ **Evenimentul sigur se produce cu certitudine la orice efectuare a experimentului,**
- ❖ **Evenimentul imposibil este nerealizabil în urma efectuării experimentului.**

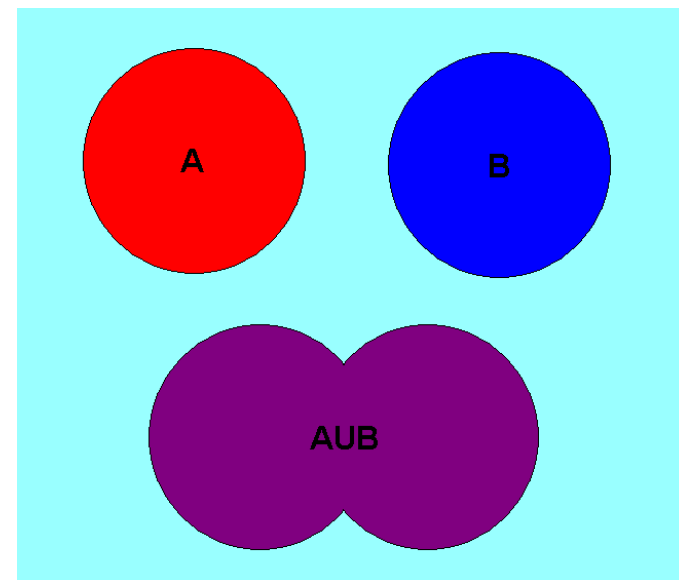
- ❖ In mod asemănător cu operațiile de reuniune și intersecție cu mulțimi se definesc operații similare cu evenimente.
- ❖ A și B sunt două evenimente oarecare
  - *reuniunea*  $C=A\cup B$  este un eveniment care are loc dacă cel puțin unul dintre evenimentele A sau B are loc
  - *intersecția*  $D=A\cap B$  este evenimentul care are loc numai când A și B au loc simultan.

- ❖ Prin **contrarul (complementarul)** unui eveniment  $A$  se înțelege un eveniment care se realizează ori de câte ori nu se realizează  $A$ . Acesta se notează prin  $\text{non } A$  (sau  $C(A)$ ).
- ❖ Dacă două evenimente  $A$  și  $B$  sunt disjuncte ( $A \cap B = \emptyset$ ) (nu se pot realiza simultan), se spune că ele sunt **incompatibile**.
- ❖ Două evenimente  $A$  și  $B$  care se pot realiza simultan se numesc **compatibile**.

# Definiția axiomatică



- ❖ Fie  $E$  un spațiu fundamental asociat unui experiment  $H$  și  $\Omega$  mulțimea tuturor evenimentelor, adică mulțimea părților lui  $E$ :
  - $\Omega = \Pr(E)$ .
- ❖ Se spune că funcția  $\Pr: \Omega \rightarrow \mathbb{R}$  este o funcție de probabilitate, iar prin  $\Pr(A)$  se notează probabilitatea evenimentului  $A$ , dacă satisface următoarele axiome:
  - **M1.**  $0 \leq \Pr(A) \leq 1, \quad \forall A \in \Omega$
  - **M2.**  $\Pr(E) = 1$
  - **M3.** Dacă  $A$  și  $B$  sunt incompatibile (adică nu pot avea loc simultan), atunci  $\Pr(A \cup B) = \Pr(A) + \Pr(B)$ .



# Definiția axiomatică - proprietăți

❖ **T1.** Dacă  $A_1, A_2, \dots, A_n$  sunt evenimente incompatibile două câte două atunci:

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(A_i)$$

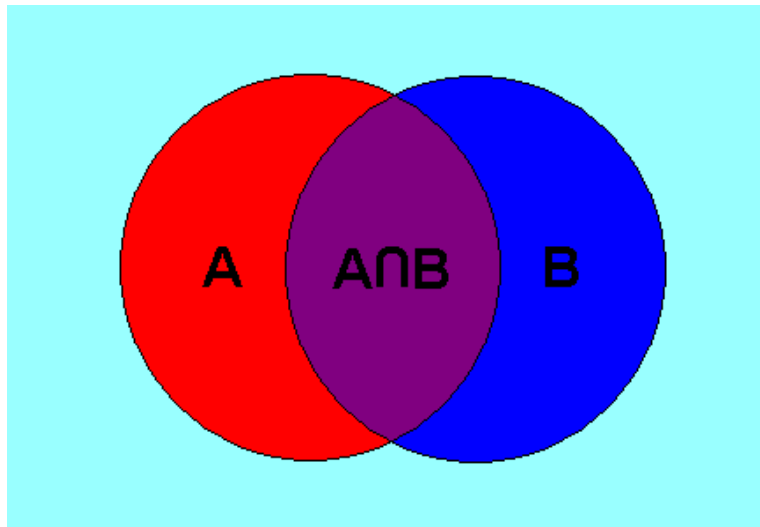
❖ **T2.**  $\Pr(\emptyset) = 0$ .

❖ **T3.**  $\Pr(\text{non } A) = 1 - \Pr(A)$ .

# Definiția axiomatică - proprietăți

❖ **T4.** Pentru orice evenimente A și B are loc egalitatea:

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$  .



# Independența a două evenimente



- ❖ Două evenimente A și B se numesc independente dacă și numai dacă
- ❖  $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$ .
- ❖ Această proprietate se mai numește și **legea de înmulțire a probabilităților**.
- ❖ Două evenimente A și B sunt **dependente** dacă  $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$ .

# Exemplul 1



- Două evenimente sunt independente dacă și numai dacă:  $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
- Două evenimente sunt dependente dacă:  
 $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$
- Studiu:
  - Pentru studiul agregării familiale a HTA s-a determinat probabilitatea HTA la mama  $P(A)=0,1$ , la primul copil  $P(B)=0,2$  și frecvența apariției HTA la copii și la mama în același timp  $\Pr(A \cap B) = 0,05$
  - Există o relație de cauzalitate între HTA la mama și cea de la copil?



# Exemplul 2



- Două evenimente sunt independente dacă și numai dacă:
  - $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$
- Doua evenimente sunt dependente dacă :
  - $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$
- Studiu:
  - Pentru studiul agregarii familiale a HTA s-a determinat probabilitatea HTA la barbati  $P(A)=0,2$  si la femei  $P(B)=0,1$
  - Care este probabilitatea de a avea o familie de hipertensivi?  $\Pr(A \cup B)$ 
    - $\Pr(A \cap B) = ?$

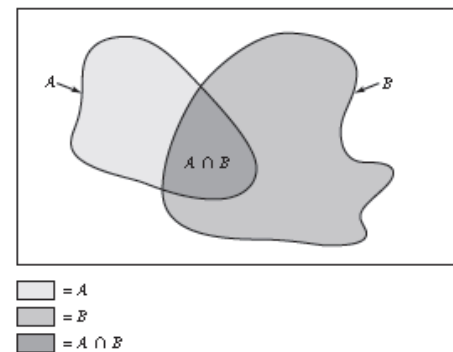


# Probabilitate condiționată



- ❖ Dacă A și B sunt două evenimente arbitrare, prin **probabilitatea condiționată a lui A de către B**, notată prin  **$\Pr(A|B)$** , se înțelege probabilitatea de a se realiza evenimentul A dacă în **prealabil** s-a realizat evenimentul B.
- ❖ Prin definiție:

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



- ❖ sau raportul dintre numărul elementelor din B care sunt și în A la numărul elementelor lui B.

# Probabilitate condiționată



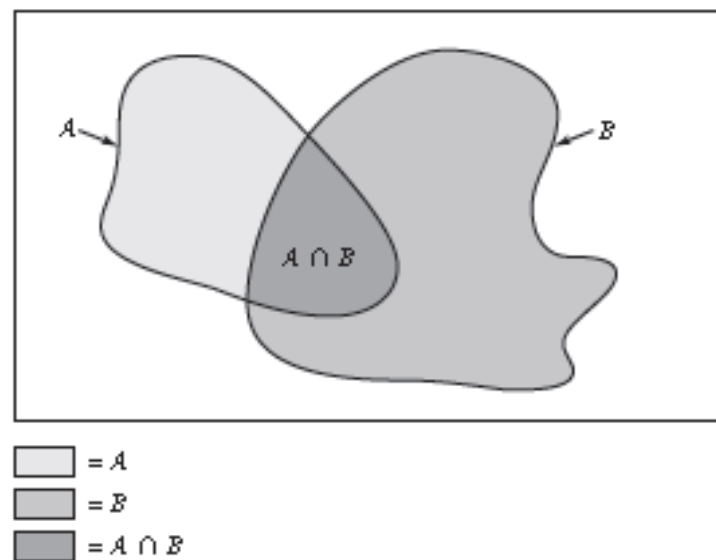
## ❖ Au loc următoarele proprietăți privind probabilitățile condiționate:

- Dacă A și B sunt evenimente independente, atunci  $\Pr(B|A) = \Pr(B)$
- Dacă A și B sunt evenimente dependente, atunci  $\Pr(B|A) \neq \Pr(B)$  și  $\Pr(A \cap B) \neq \Pr(A) \times \Pr(B)$ .

# Exemplul 3



- Doi medici A si B dintr-un spital testează toți pacienții pentru sifilis: primul medic, A, diagnostichează 10% din pacienți ca pozitivi, iar medicul B 17%; 8% din pacienți sunt diagnosticați pozitiv de ambii medici



# Exemplul 3



- Date
  - $P(A) = 0,1$
  - $P(B) = 0,17$
  - $P(A \cap B) = 0,08$
- Calculați probabilitatea condiționată ca doctorul B să pună un diagnostic pozitiv dacă doctorul A a pus deja un diagnostic pozitiv la același pacient
- $P(B | A) = P(A \cap B) / P(A) = 0,08 / 0,1 = 0,8$
- Deci, doctorul B confirmă 80% din pacienții doctorului A



# Măsurarea riscului: RR



# Riscul relativ (RR)



= raportul dintre probabilitatea condiționată de B de a avea evenimentul A și probabilitatea condiționată de B de a nu avea evenimentul A

$$RR = \frac{P(B|A)}{P(B|\bar{A})}$$

$$RR = \frac{P(Boala|Factor\ de\ risc)}{P(Boala|\overline{Factor\ de\ risc})}$$

Pentru două evenimente independente

$$P(B|A) = P(B) = P(B|\bar{A})$$

Pentru două evenimente dependente

$$P(B|A) \neq P(B) \neq P(B|\bar{A})$$

$$RR=1$$

$$RR \neq 1$$

# Riscul relativ (RR)



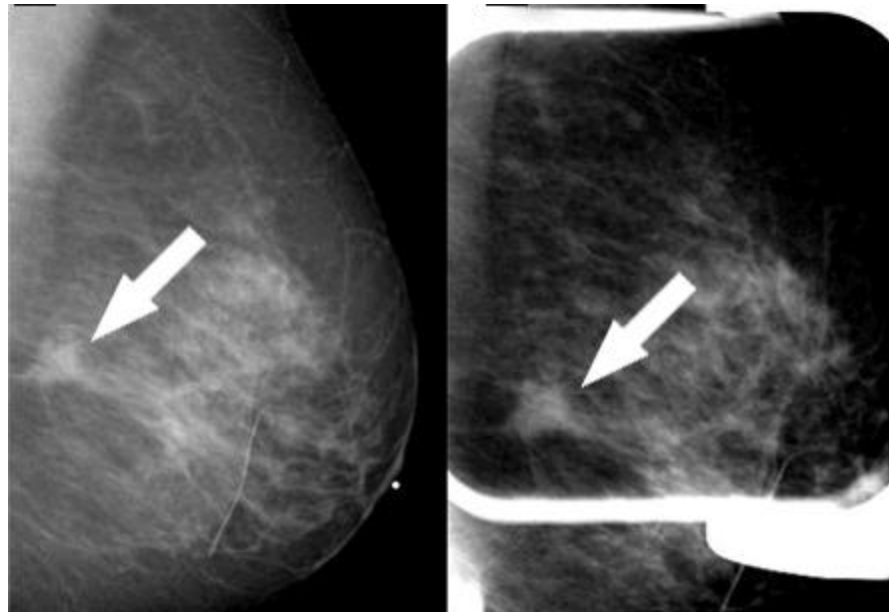
$$RR=1$$



# EXEMPLU



- ❖ Se recomandă femeilor peste 50 de ani să își facă o mamografie odată la 1-2 ani; testul de “aur” pentru cancerul de san este biopsia, dar este prea invaziv



# EXEMPLU



- ❖ Din 100000 de femei cu mamografii negative, 20 vor dezvolta cancer de sân în următorii 2 ani
- ❖ Din 10 femei cu mamografii pozitive, 1 va dezvolta cancer de san în următorii 2 ani
- ❖ Care este RR de a dezvolta cancer de sân în următorii 2 ani după o mamografie pozitivă?

$$\begin{aligned}\text{❖ } \mathbf{RR} &= \Pr(B|A) / \Pr(B|\text{non } A) \\ &= 0,1 / 0,0002 \\ &= 500\end{aligned}$$



# Semne diagnostice

**Legătura dintre semnele sau testele diagnostice și diagnosticul pacientului este una probabilistică**



## ❖ Să considerăm următoarele evenimente în legătură cu aplicarea unui test diagnostic:

- B - evenimentul ca o persoană luată la întâmplare dintr-o populație să aibă o anumită afecțiune B (TBC , HIV etc.),
- T - evenimentul de obținere a unui test pozitiv în cazul aplicării unui test diagnostic T pentru detectarea afecțiunii B la o persoană.
- $\text{non}(B)$  = persoană fără afecțiunea B
- $\text{non}(T)$  = test negativ.

- ❖ Să presupunem că populația căreia i s-a aplicat testul are  $n$  persoane și s-au obținut următoarele rezultate:

Afecțiunea / Testul	B Bolnavi	non(B) Sanatoși	Total
T Test pozitiv	$a$ (AP)	$b$ (FP)	$a+b$
non (T) Test negativ	$c$ (FN)	$d$ (AN)	$c+d$
Total	$a+c$	$b+d$	$n$

# Valoarea pozitivă predictivă VPP



- ❖ este probabilitatea ca un test pozitiv să indice o persoană cu afecțiunea B:

$$VPP = \Pr(B / T) = \frac{\Pr(T \cap B)}{\Pr(T)} = \frac{AP}{AP + FP} = \frac{a}{a + b}$$

Afecțiunea / Testul	B bolnavi	non(B) sanatosi	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) Test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

# Valoarea predictivă negativă VPN

- ❖ este probabilitatea ca un test negativ să indice o persoană fără afecțiune:

$$VPN = \Pr(nonB / nonT) = \frac{\Pr(nonB \cap nonT)}{\Pr(nonT)} = \frac{AN}{FN + AN} = \frac{d}{c + d}$$

Afecțiunea / Testul	B bolnavi	non(B) sanatosi	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) Test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

# Sensibilitatea testului



- ❖ Se = probabilitatea de a obține un test pozitiv, știind că testul este aplicat unei persoane care posedă afecțiunea

$$Se = \Pr(T / B) = \frac{AP}{AP + FN} = \frac{a}{a + c} = \frac{\frac{a}{n}}{\frac{a + c}{n}} = \frac{\Pr(T \cap B)}{\Pr(B)}$$

Afecțiunea / Testul	B bolnavi	non(B) sanatosi	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) Test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

# Specificitatea testului



- ❖ Pentru caracterizarea unui test diagnostic se utilizează și specificitatea testului care se definește prin probabilitatea de a obține un test negativ pentru o persoană sănătoasă:

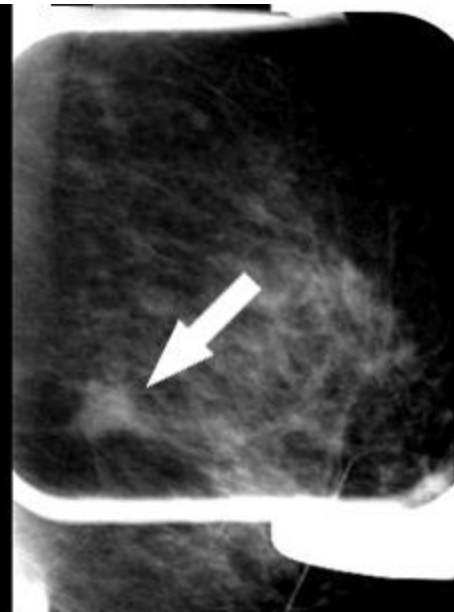
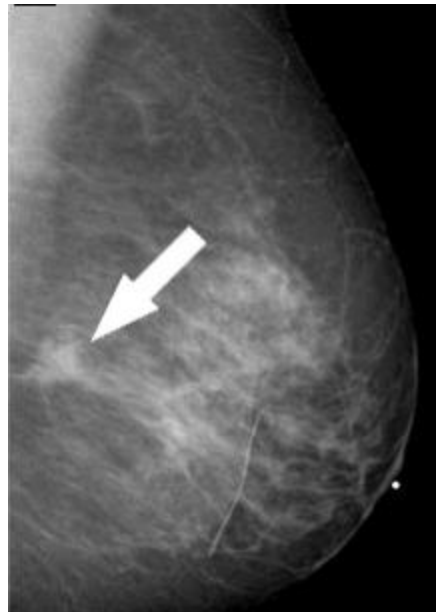
$$Sp = \Pr(\text{non}(T) / \text{non}(B)) = \frac{AN}{FP + AN} = \frac{d}{b + d} = \frac{\frac{d}{n}}{\frac{b + d}{n}} = \frac{\Pr(\text{non}T \cap \text{non}B)}{\Pr(\text{non}B)}$$

Afecțiunea / Testul	B bolnavi	non(B) sanatosi	Total
T Test pozitiv	a (AP)	b (FP)	a+b
non (T) Test negativ	c (FN)	d (AN)	c+d
Total	a+c	b+d	n

# EXEMPLU



- ❖ Se recomandă femeilor peste 50 de ani să își facă o mamografie odată la 1-2 ani; testul de "aur" pentru cancerul de san este biopsia dar este prea invaziv



# EXEMPLU



- ❖ Din 100000 de femei cu mamografii negative, 20 vor dezvolta cancer de sân în următorii 2 ani
- ❖ Din 10 femei cu mamografii pozitive 1 va dezvolta cancer de sân în următorii 2 ani
- ❖ **Calculați Se, Sp, VPP, VPN**

# Exemplu: Se, Sp, VPP, VPN



$$Se = 1/21 = 0,047$$

$$P = 21/100.010 = 0,21 \text{ la mie}$$

Afecțiunea / Testul	Subiecți care au dezvoltat afecțiunea	Subiecți sănătoși	Total
Mamografie pozitivă	1	9	10
Mamografie negativă	20	99.980	100.000
Total	21	99.989	100.010

$$VPP = 1/10 = 0,1$$

$$VPN = \frac{99980}{100000} = 0,99$$

$$Sp = 99980/99989 = 0,99$$

Semn sau  
test (T)

Se, Sp

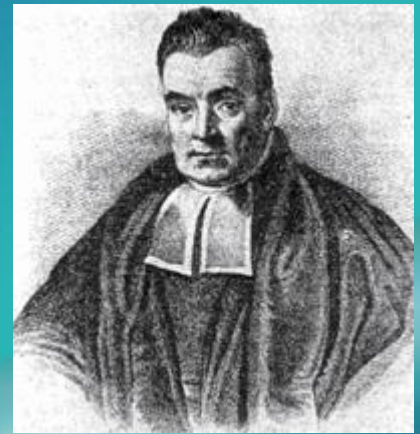
VPP, VPN

Afectiune  
(B)

Clinică

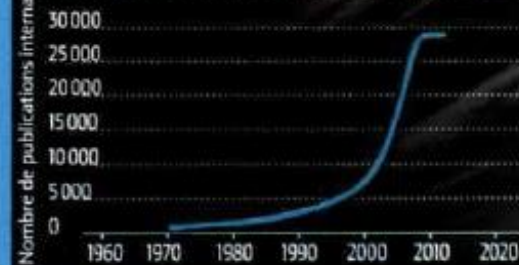
Semiologie

# Formula lui Bayes



> Oubliée pendant 250 ans, reléguée aux manuels de statistiques, la petite formule du révérend est aujourd'hui utilisée par les chercheurs du monde entier. En 2010, près de 30 000 publications y ont eu recours.

## Le nombre de publications utilisant la formule de Bayes explose depuis 2000



# Formula lui BAYES



- ❖ Să considerăm două evenimente A și B care nu sunt independente (A = afecțiunea, B = semnul). Atunci din formulele:

$$\Pr(A/B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

$$\Pr(B/A) = \frac{\Pr(A \cap B)}{\Pr(A)}$$

- ❖ și se deduce formula lui BAYES:

$$\Pr(A/B) = \frac{\Pr(B/A) \cdot \Pr(A)}{\Pr(B)}$$

# Formula lui BAYES



❖ Fie **A** o maladie și **B** un semn:

$$\Pr(A/B) = \frac{\Pr(B/A) \cdot \Pr(A)}{\Pr(B)}$$

❖ Dar fiindcă

- $\Pr(B) = \Pr((B \cap \text{non}A) \cup (B \cap A)) = \Pr(B \cap \text{non}A) + \Pr(B \cap A),$

❖ aplicând formula probabilităților condiționate se obține:

- $\Pr(B) = \Pr(B|A) \Pr(A) + \Pr(B|\text{non}A) \Pr(\text{non}A).$

❖ De aici rezultă următoarea formă a formulei lui Bayes:

$$\Pr(A|B) = \frac{\Pr(B|A) \cdot \Pr(A)}{\Pr(B|A) \cdot \Pr(A) + \Pr(B|\text{non}A) \cdot \Pr(\text{non}A)}$$

# Formula lui BAYES



$$Pr(A | B) = \frac{Pr(B | A) \cdot Pr(A)}{Pr(B | A) \cdot Pr(A) + Pr(B | nonA) \cdot Pr(nonA)}$$

$$VPP = Pr(A|B)$$

$$VPN = Pr(\bar{A}|\bar{B})$$

$$Sensibilitate = Pr(B|A)$$

$$Specificitate = Pr(\bar{B}|\bar{A})$$

$$VPP = \frac{sensibilitate \times prevalenta}{sensibilitate \times prevalenta + (1 - specificitate) \times (1 - prevalenta)}$$
$$VPN = \frac{specificitate \times (1 - prevalenta)}{specificitate \times (1 - prevalenta) + (1 - sensibilitate) \times prevalenta}$$

# EXEMPLU



$$VPP = \frac{\text{sensibilitate} \times \text{prevalenta}}{\text{sensibilitate} \times \text{prevalenta} + (1 - \text{specificitate}) \times (1 - \text{prevalenta})}$$
$$VPN = \frac{\text{specificitate} \times (1 - \text{prevalenta})}{\text{specificitate} \times (1 - \text{prevalenta}) + (1 - \text{sensibilitate}) \times \text{prevalenta}}$$

- ❖ In farmacii s-au introdus sisteme automate de determinare a TA. Un astfel de sistem clasifică 84% din hipertensivi și 23% din normotensivi ca având HTA.
- ❖ **Dacă 20% din populatia adulta are HTA, care sunt VPP si VPN ale acestui aparat?**
- ❖ Sensibilitatea = 0,84
- ❖ Specificitatea =  $1 - 0,23 = 0,77$
- ❖  $VPP = 0,84 \times 0,2 / (0,84 \times 0,2 + 0,23 \times 0,8) = 0,168 / 0,352 = 0,48$
- ❖  $VPN = 0,77 \times 0,8 / (0,77 \times 0,8 + 0,16 \times 0,2) = 0,616 / 0,648 = 0,95$

# EXEMPLU



$$VPP = \frac{\text{sensibilitate} \times \text{prevalenta}}{\text{sensibilitate} \times \text{prevalenta} + (1 - \text{specificitate}) \times (1 - \text{prevalenta})}$$
$$VPN = \frac{\text{specificitate} \times (1 - \text{prevalenta})}{\text{specificitate} \times (1 - \text{prevalenta}) + (1 - \text{sensibilitate}) \times \text{prevalenta}}$$

- ❖ **90% din persoanele care au facut cancer pulmonar au fumat iar 30% din populația sănătoasă fumează**
- ❖ **Dacă 0,1% din populatia adultă dezvoltă această afecțiune ce putem spune despre fumat ca și semn diagnostic?**
- ❖ **Sensibilitatea = 0,90** – un test sensibil ne indică sănătoșii
- ❖ **Specificitatea = 0,70** – nu este suficient de specific
- ❖ **VPP= 0,003** – prezenta semnului nu indica boala
- ❖ **VPN= 0,999** – absenta semnului indica un sănătos

Nu putem spune cuiva care fumează că este afectat, dar putem fi aproape siguri că cineva care nu fumează nu are afecțiunea

