PROBABILITĂȚI

An universitar 2012-2013

Sorana D. Bolboacă

Conținut

- Concepte de bază
- Probabilități condiționate
- Proprietăți
- Reguli
- Aplicabilitate

Evenimentul aleatoriu binomial

- La aruncarea cu moneda avem două rezultate posibile (capul sau pajura) asociate cu o probabilitate specifică (ex. 0,5)
  - Testul: aplicarea unui experiment
  - Evenimentul: rezultatul testului
  - Evenimentul aleator: evenimentul care se obține la aplicarea unui singur test
  - Spațiu de evenimente:{cap, pajură}

Probabilitatea

- o măsură a şansei de realizare a unui eveniment
- **Pr(A) $\in$[0, 1]**

- Fie A un eveniment:
  - Pr(A) = probabilitatea evenimentului A
  - Dacă evenimentul este o certitudine: Pr(A) = 1
  - Dacă evenimentul este imposibil de realizat: Pr(A) = 0

Probabilitatea subiectivă vs obiectivă

Subiectivă:

- stabilită subiectiv sau empiric pe baza experienței sau prin studii populaționale foarte largi
- implică evenimente elementare care **nu sunt** echiprobabile (echiprobabil = are tot atâtea şanse de a se produce ca şi alte evenimente)

Formula de calcul:

- Dacă un eveniment A se poate realiza în S probe dintr-o serie de n încercări echiprobabile, atunci probabilitatea evenimentului A este dată de numărul de cazuri favorabile raportat la numărul de cazuri posibile
- Pr(A) = (nr cazuri favorabile)/(nr cazuri posibile)

Şanse şi rații

- Şansele sunt probabilități exprimate procentual
- Şansa ia valori între 0% şi 100%
  - Exemplu: o probabilitate de 0,75 este egală cu o şansă de 75%
- Rația unui eveniment este probabilitatea ca un eveniment să se întâmple împărțit la probabilitatea ca acel eveniment să nu se întâmple
  - Poate lua orice valoare pozitivă
  - Fie A evenimentul de interes. Rația de probabilitate = Pr(A)/[1-Pr(A)] (unde 1-Pr(A) = Pr(nonA))
  - Exemplu: dacă Pr(A) = 0,75 atunci rația de probabilitate este de 3 la 1 (0,75/(1-0,75)=0,75/0,25=3/1)

Spațiul unui eveniment

- Mulțimea tuturor rezultatelor posibile ale unui proces aleatoriu
  - La aruncarea cu zarul spațiul de evenimente este format din {1, 2, 3, 4, 5, 6}
  - La aruncarea unei monede spațiul de evenimente este {cap, pajură}.
- un eveniment este un membru al spațiului evenimentului
  - “cap” este un eveniment posibil la aruncarea unei monede
  - “un număr mai mic sau egal cu 3” este un eveniment posibil la aruncarea unui zar
- **Evenimentele au asociate probabilități de producere!**

Probabilități: proprietăți

- Iau valori între 0 şi 1:

$$\mathbf{0 \leq Pr(A) \leq 1}$$

- Pr(spațiu al unui eveniment) = 1
- Probabilitatea de a se întâmpla un eveniment este 1 minus probabilitatea de a nu se întâmpla acel eveniment

$$\mathbf{Pr(A) = 1 - Pr(nonA)}$$

Concepte de bază

- Evenimente compatibile: evenimente care se pot realiza simultan:
  - A = {TAS < 140 mmHg}
  - B = {TAD < 90 mmHg}
- Evenimente incompatibile: evenimente care nu se pot realiza simultan:
  - A = {TAS < 140 mmHg}
  - B = {140 ≤ TAS < 200 mmHg}

Concepte de bază

- Evenimentul A implică evenimentul B dacă evenimentul B se produce ori de câte ori se produce evenimentul A:
  - Simbol $A \subset B$
    - A = {TBC}
    - B = {testul la tuberculină pozitiv}

Probabilități condiționate

- Probabilități condiționate:
  - Fie A şi B două evenimente
  - Prin probabilitatea condiționată a lui A de către B (simbol: Pr(A|B)) se înțelege probabilitatea de a se realiza evenimentul A dacă în prealabil s-a realizat evenimentul B
- Exemplu: Pr(Test pozitiv tuberculină|TBC) este probabilitatea de a obține un test pozitiv la tuberculină la un pacient care are TBC.
- **P(B|A)** nu este acelaşi lucru cu **P(A|B)**

Probabilități condiționate

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Fie:
  - A = {TBC+}
  - B = {Test+}

- Pr(A) = (15+25)/(15+12+25+18) = 0,57 (prevalența bolii)
- Pr(nonA) = (12+18) /(15+12+25+18) = 0,43
- Pr(B|A) = probabilitatea unui test pozitiv la un pacient cu TBC = 15/(15+25) = 0,38 = **SENSIBILITATE (Se)**

Probabilități condiționate

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Fie:
  - A = {TBC+}
  - B = {Test+}

- Pr(nonB|nonA) = probabilitatea de a obține un test negativ ştiind că testul se aplică unui pacient indemn de TBC = 18/(18+12) = 0,60 = **SPECIFICITATE (Sp)**
- Pr(A|B) = probabilitatea ca o persoană cu TBC să prezinte un test pozitiv = 15/(15+12) = 0,56 = **VALOAREA PREDICTIVĂ POZITIVĂ (VPP)**

Probabilități condiționate

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Fie:
  - A = {TBC+}
  - B = {Test+}

- Pr(nonA|nonB) = probabilitatea ca o persoană indemnă TBC să prezinte un test negativ = 18/(18+25) = 0,42 = **VALOAREA PREDICTIVĂ NEGATIVĂ (VPN)**

Probabilități condiționate

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Fie:
  - A = {TBC+}
  - B = {Test+}

- Rata falşilor pozitivi: RFP = Pr(B|nonA)
- Rata falşilor negativi: RFN = Pr(nonA|B)

Evenimente independente: probabilități condiționate

- Două evenimente A şi B se numesc <u>*independente*</u> dacă şi numai dacă

$$Pr(A \cap B) = Pr(A) \cdot Pr(B).$$

- În acest caz:
  - $Pr(B|A) = Pr(B|nonA) = Pr(B)$
  - $Pr(A|B) = Pr(A|nonB) = Pr(A)$

  - Exemplu: dacă aruncăm moneda de două ori probabilitatea ca la a doua aruncare să obțin cap este întotdeauna 0,5 indiferent ce am obținut la prima aruncare

Operații cu evenimente

- **REUNIUNEA (SAU):**
  - $A \cup B$ - se produce cel puțin unul dintre evenimentele A sau B
- **INTERSECȚIA (ŞI)**:
  - $A \cap B$ - evenimentele A şi B se produc simultan
- **NEGAREA:**
  - nonA

Reguli de probabilitate

- Probabilitatea de apariție a evenimentului A **sau** B: **ADUNARE**

$$\mathbf{Pr(A\ sau\ B) = Pr(A) + Pr(B)}$$

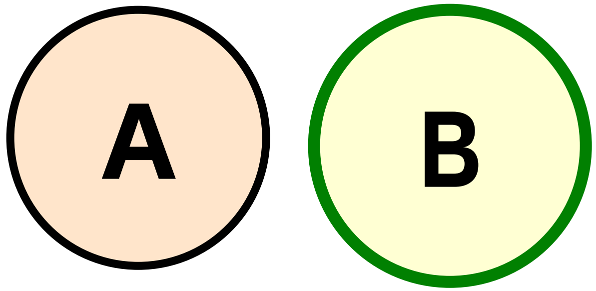
- evenimentele A şi B sunt mutual exclusive

- Probabilitate de A **şi** B: **ÎNMULȚIRE**

$$\mathbf{P(A\ şi\ B) = P(A) \cdot P(B)}$$

- evenimentele A şi B sunt independente

Reguli de adunare a probabilităților



- Fie A şi B două evenimente:

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

- Evenimente mutual exclusive: $Pr(A \cap B) = 0$

Reguli de adunare a probabilităților

- A = {TAS mamă > 140 mmHg}, $Pr(A) = 0,25$
- B = {TAS tată > 140 mmHg}, $Pr(B) = 0,15$
- Care este probabilitatea ca într-o familie să avem un părinte hipertensiv?

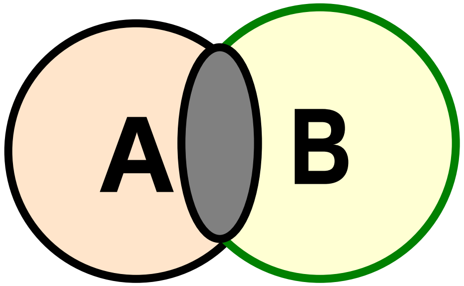
$$\mathbf{Pr(A \cup B) = 0,25 + 0,15 - 0 = 0,40}$$

Reguli de adunare a probabilităților

- Într-o cafenea există 20 de persoane; la 10 le place ceaiul, la alți 10 cafeaua şi la 2 le place şi ceaiul şi cafeaua.
- Care este probabilitatea de a extrage la întâmplare din populație o persoană căreia să-i placă ceaiul sau cafeaua?

$$\mathbf{Pr(ceai \cup cafea) = Pr(ceai) + Pr(cafea) - Pr(ceai \cap cafea) = 0{,}50 + 0{,}50 - 0{,}10 = 0{,}90}$$

Reguli de înmulțire a probabilităților



- Fie A şi B două evenimente:

$$\Pr(A \cap B) = \Pr(A) \cdot \Pr(B|A)$$

- Evenimente independente $\Pr(B|A) = \Pr(B)$

Reguli de înmulțire a probabilităților

- A = {TAS mamă > 140 mmHg}, $Pr(A) = 0,10$
- B = {TAS tată > 140 mmHg}, $Pr(B) = 0,20$
- $Pr(A \cap B) = 0,05$
- Evenimentele A Şi B sunt dependente sau independente?

$Pr(A \cap B) = Pr(A) \cdot Pr(B)$ – evenimente independente

$0,05 \neq 0,10 \cdot 0,20 \rightarrow$ evenimente dependente

Probabilități în tabelul de contingentă

	Disease+	Disease -	Total
Test +	AP	FP	= AP+FP
Test -	FN	AN	= FN+AN
Total	= AP+FN	=FP+AN	= n

Denumire parametru	**Formula**
Rata falşilor pozitivi	=FP/(FP+AN)
Rata falşilor negativi	=FN/(FN+AP)
Sensibilitatea	=AP/(AP+FN)
Specificitatea	=AN/(AN+FP)
Acuratețea	=(AP+AN)/n
Valoarea predictivă pozitivă	=AP/(AP+FP)
Valoarea predictivă pozitivă	=AN/(AN+FN)
Riscul relativ	=AP(FP+AN)/FN(AP+FP)
Rata şansei	=(AP·AN)/(FN·FP)
Riscul atribuabil	=AP/(AP+FP)-FN/(FN+AN)

Probabilități: exemplu

- http://www.biomedcentral.com/1746-6148/8/68
- *BMC Veterinary Research* 2012, **8**:68 doi:10.1186/1746-6148-8-68

Background

Brucella ovis causes an infectious disease responsible for infertility and subsequent economic losses in sheep production. The standard serological test to detect *B. ovis* infection in rams is the complement fixation test (CFT), which has imperfect sensitivity and specificity in addition to technical drawbacks. Other available tests include the indirect enzyme-linked immunosorbent assays (I-ELISA) but no I-ELISA kit has been fully evaluated.

The study aimed to compare an I-ELISA kit and the standard CFT. Our study was carried out on serum samples from 4599 rams from the South of France where the disease is enzootic. A Bayesian approach was used to estimate tests characteristics (diagnostic sensitivity, Se and diagnostic specificity, Sp). The tests were then studied together in order to optimise testing strategies to detect *B. ovis*.

Probabilități: exemplu

- *http://www.biomedcentral.com/1746-6148/8/68*
- *BMC Veterinary Research* 2012, **8**:68 doi:10.1186/1746-6148-8-68

Results

After optimising the cut-off values in order to avoid doubtful results without deteriorating the concordance between the results of the two tests, the I-ELISA appeared to be slightly more sensitive than CFT ($Se_{I\text{-}ELISA}=0.917$ [0.822; 0.992], 95% Credibility Interval (CrI) compared to $Se_{CFT}=0.860$ [0.740; 0.967], 95% CrI). However, CFT was slightly more specific than I-ELISA ($Sp_{CFT}=0.988$ [0.947; 1.0], 95% CrI) compared to $Sp_{I\text{-}ELISA}=0.952$ [0.901; 1.0], 95% CrI).

The tests were then associated with two different interpretation schemes. The series association increased the specificity of screening and could be used for pre-movement testing in rams from uninfected flocks. The parallel association increased sequence sensitivity, thus appearing more suitable for eradicating the disease in infected flocks.

Probabilități: exemplu

- *http://www.biomedcentral.com/1746-6148/8/184/abstract*
- *BMC Veterinary Research* 2012, **8**:184 doi:10.1186/1746-6148-8-184

Background

This study aimed to identify risk factors for active porcine reproductive and respiratory syndrome virus (PRRSV) infection at farm level and to assess the probability of an infected farm being detected through passive disease surveillance in England. Data were obtained from a cross-sectional study on 147 farrow-to-finish farms conducted from April 2008 -- April 2009. The risk factors for active PRRSV infection were identified using multivariable logistic regression analysis. The surveillance system was evaluated using a stochastic scenario tree model.

Results

Evidence of PRRSV circulation was confirmed on 35.1% (95%CI: 26.8-43.4) of farms in the cross sectional study, with a higher proportion of infected farms in areas with high pig density (more than 15000 pigs within 10 km radius from the farm). Farms were more likely to have active PRRSV infection if they used the live virus vaccine-Porcilis PRRS (OR=7.5, 95%CI: 2.5-22.8), were located in high pig density areas (OR=2.9, 95%CI: 1.0-8.3) or had dead pigs collected (OR=5.6, 95%CI: 1.7-18.3). Farms that weaned pigs at 28 days of age or later had lower odds of being PRRSV positive compared to those weaning at 21-27 days (OR=0.2, 95%CI: 0.1-0.7). The probability of detecting an infected farm through passive surveillance for disease was low (mode=0.074, 5th and 95th percentiles: 0.067; 0.083 respectively). In particular farms which used live virus vaccine had lower probabilities for detection compared to those which did not.

De reținut! Operații cu probabilități

- **Adunare:**
- **$Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$**
- **$Pr(A \cup B) = Pr(A) + Pr(B)$: evenimente mutual exclusive**
- **Înmulțire:**
- **$Pr(A \cap B) = Pr(A) \cdot Pr(B|A)$**
- **$Pr(A \cap B) = Pr(A) \cdot Pr(B)$: evenimente independente**

Variabile Aleatoare

Variabile aleatoare... despre

- Definiție
- Introducere: distribuții de probabilitate
  - Discrete
  - Continue

Definiție

- Fie X o variabilă cantitativă măsurată sau observată rezultată dintr-un experiment
- Valoarea pe care o ia variabila X în urma experimentului este o variabilă aleatoare

- **Exemple**:
  - Numărul de globule roşii dintr-un frotiu
  - Numărul de bacterii de pe mâinile studenților
  - Scorul de depresie mediu obținut la aplicarea unui test pe un eşantion de pacienți cu patologie terminală

Variabile aleatoare

- Media aritmetică a eşantionului
- Deviația standard
- Proporția
- Frecvența
  - Toate sunt variabile aleatoare

Tipuri de variabile aleatoare

Discrete:

- Poate lua un număr finit măsurabil de valori
  - Numărul de persoane cu RH- dintr-un eşantion
  - Numărul de copii cu gripă dintr-o colectivitate
  - Numărul de studenți anorexici din universitate
  - Pulsul

Continue:

- Poate lua orice valoare din nenumăratele valori posibile într-un interval definit
- Variază în mod continuu în intervalul dat
  - Temperatura corporală
  - Concentrația zahărului în sânge
  - Tensiunea arterială

Tipuri de variabile aleatoare

- În general, mediile sunt variabile aleatoare continue iar frecvențele sunt discrete
  - Media capacității pulmonare a unei persoane care munceşte în domeniul minier
  - Numărul de pacienți cu parvoviroză internați în Clinică.

Distribuția de probabilitate: variabila aleatoare discretă

Spațiul unui eveniment

- Fie X numărul de fețe “cap” obținute la aruncarea de 3 ori a unei monede
- X este o variabilă aleatoare care poate lua una din următoarele valori {0,1,2,3}

Spațiul unui eveniment

- Dintr-un sac care conține bile albe şi negre sunt extrase 2 bile. La extragerea unei bile albe se câştigă 1 Ron iar la extragerea unei bile negre se pierde 1 Ron.
- X este o variabilă aleatoare care poate lua una din valorile {-2,0,2}

Distribuția de probabilitate: variabila aleatoare discretă

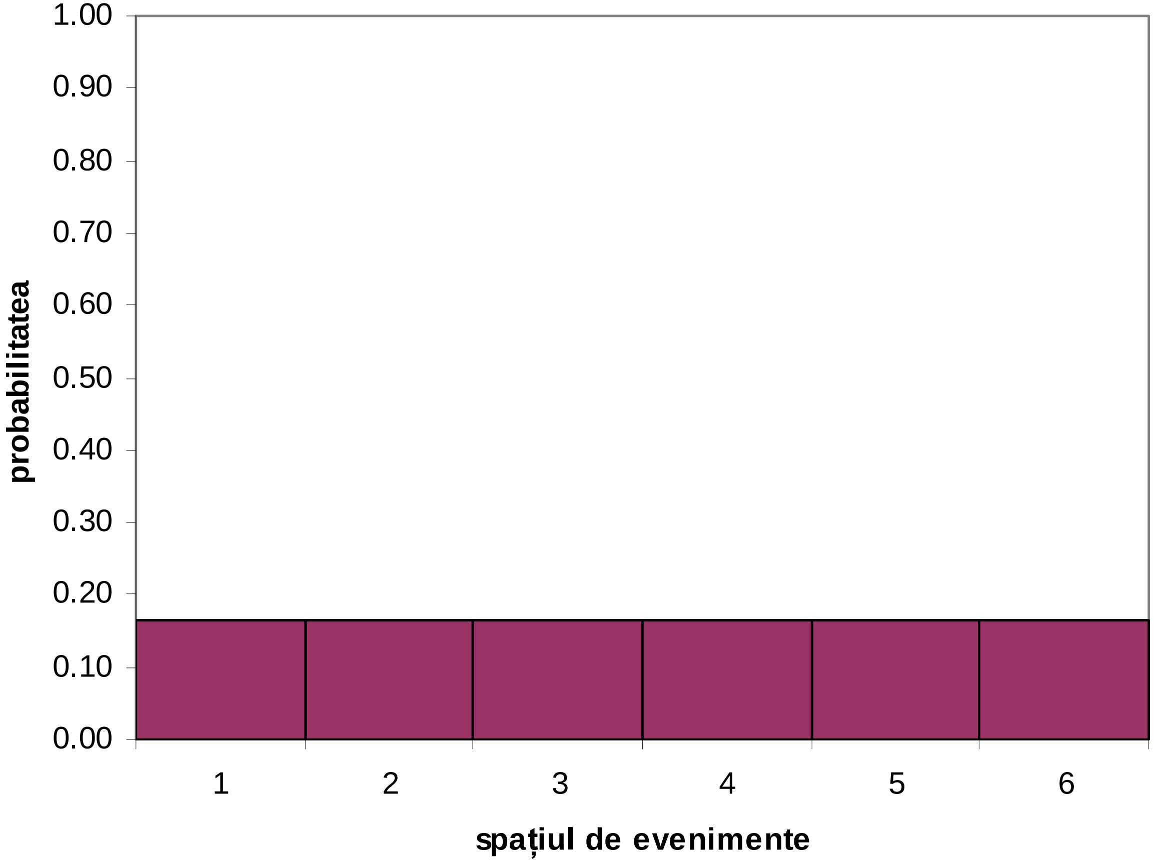
- Probabilitatea distribuției lui X: listă de valori ale spațiului de evenimente şi probabilitățile asociate acestora

- Fie X rezultatul aruncării unui zar
- X este o variabilă aleatoare care ia una din următoarele valori 1, 2, 3, 4, 5, 6

X_i	Pr_i
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Distribuția de probabilitate: variabila aleatoare discretă

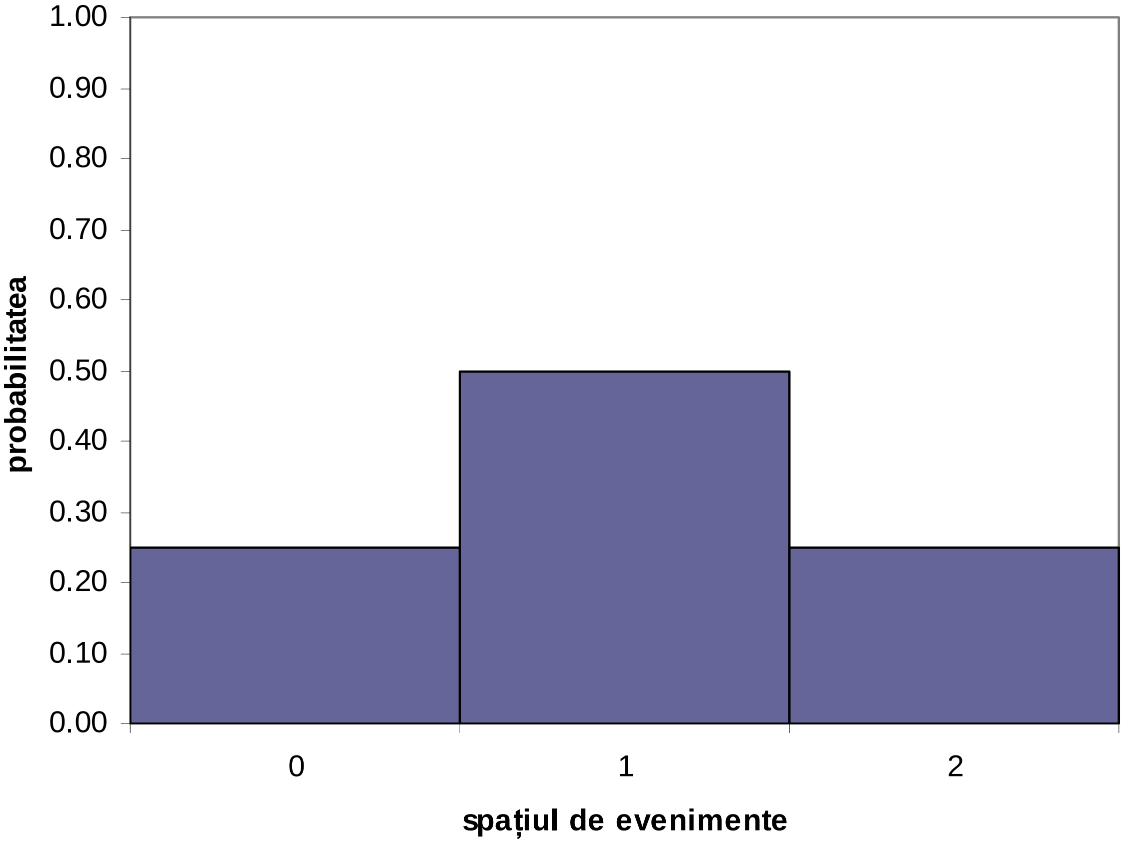
- Probabilitatea distribuției lui X listează valorile spațiului de evenimente şi probabilitățile asociate



X_i	Pr_i
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Distribuția de probabilitate: variabila aleatoare discretă

- Fie X numărul de fețe ‘cap’ rezultate la aruncarea a două monezi de două ori. Care este distribuția de



X_i	Pr_i
0	1/4
1	2/4
2	1/4

Distribuția de probabilitate: variabila aleatoare discretă

- Legea de probabilitate: simbolistică

$$X : \begin{pmatrix} X_1 & X_2 & \dots & X_n \\ \Pr(x_1) & \Pr(x_2) & \dots & \Pr(x_n) \end{pmatrix}$$

- **Proprietate:** probabilitățile care apar în distribuția unei variabile aleatoare finite X verifică

$$\sum_{i=1}^{n} \Pr(X_i) = 1$$

Distribuția de probabilitate: variabila aleatoare discretă

- **Media** distribuției de probabilitate discretă (denumită şi **valoare expectată** sau **speranța matematică**) este dată de formula

$$M(X) = \sum_{i=1}^{n} X_i \cdot \Pr(X_i)$$

- Este media ponderată a valorilor posibile, fiecare valoare fiind ponderată cu probabilitatea ei de apariție

Distribuția de probabilitate: variabila aleatoare discretă

Exemplu:

- Fie X o variabilă aleatoare reprezentând numărul de episoade de otită în primii doi ani de viață într-o colectivitate. Această variabilă aleatoare are distribuția:

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- Care este numărul aşteptat (mediu) de episoade de otită în primii doi ani de viață?

Distribuția de probabilitate: variabila aleatoare discretă

- Care este numărul aşteptat (mediu) de episoade de otită în primii doi ani de viață?

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- M(X) = 0·0.129 + 1·0.264 + 2·0.271 + 3·0.185 + 4·0.095 + 5·0.039 + 6·0.017
- M(X) = 0 + 0.264 + 0.542 + 0.555 + 0.38 + 0.195 + 0.102
- M(X) = 2.038

Distribuția de probabilitate: variabila aleatoare discretă

- Variația: media ponderată a pătratului deviației lui X

$$V(X) = \sum_{i=1}^{n} (X_i - M(X))^2 \cdot \Pr(X_i)$$

- Abaterea standard sau ecartul tip:

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^{n} (X_i - M(X))^2 \cdot \Pr(X_i)}$$

Variabila aleatoare discretă: V(X),σ(X)

X_i	$Pr(X_i)$	$X_i*Pr(X_i)$	X_i-M(X)	$(X_i-M(X))^2$	$(X_i-M(X))^2*Pr(X_i)$
0	0.129	0	-2.038	4.153	0.536
1	0.264	0.264	-1.038	1.077	0.284
2	0.271	0.542	-0.038	0.001	0.000
3	0.185	0.555	0.962	0.925	0.171
4	0.095	0.38	1.962	3.849	0.366
5	0.039	0.195	2.962	8.773	0.342
6	0.017	0.102	3.962	15.697	0.267
		M(X)=**2.038**			V(X)=**1.967**
					σ(X)=**1.402**

Principalele distribuții de probabilitate: variabile aleatoare discrete

- **Bernoulli** (cap versus pajură): două rezultate posibile
- **Binomială** (numărul de 'cap' în n aruncări): variabile aleatoare finite
- **Poisson** (numărul de pacienți care sunt consultați în serviciul de urgență într-o zi): variabile aleatoare discrete infinite

Distribuția Binomială

- Un experiment e alcătuit din repetarea unei încercări elementare de *n* ori (n = un număr natural dat)
- Rezultatele posibile ale fiecărei încercări elementare sunt două evenimente numite *succes* şi *eşec*
- Probabilitatea de succes este notată cu *p* iar probabilitatea de eşec este notată cu *q* ($q = 1-p$)
- Cele *n* încercări repetate sunt independente

Distribuția Binomială

- Numărul X de succese obținute în cele n încercări este o variabilă aleatoare de tip binomial care depinde de parametrii n şi p şi se notează cu Bi(n,p)
- Variabila aleatoare X poate să ia valorile 0,1,2,...n
- Probabilitatea ca X să fie egal cu o valoare k este dată de formula:

$$\Pr(X = k) = C_n^k p^k q^{n-k} \qquad C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

Distribuția Binomială

- Media sau speranța matematică a distribuției binomiale:

$$M(X) = n \cdot p$$

- Variația:

$$V(X) = n \cdot p \cdot q$$

- Abaterea standard:

$$\sigma(X) = \sqrt{(n \cdot p \cdot q)}$$

Distribuția Binomială

- Care este probabilitatea de ca din 5 copii 2 să fie băieți dacă probabilitatea de a naşte un băiat este de 0,47 pentru fiecare naştere şi sexul copiilor născuți succesiv în familie este considerat o variabilă aleatoare independentă?

- $p=0.47$
- $q=1-0.47=0.53$
- $n=5$
- $k=2$
- **$Pr(X=2)=10 \cdot 0.47^2 \cdot 0.53^3$**
- **$Pr(X=2) = 0.33$**

$$\Pr(X = k) = C_n^k p^k q^{n-k}$$

$$C_5^2 = \frac{5!}{2! \cdot (5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (3 \cdot 2 \cdot 1)} = \frac{120}{12} = 10$$

Distribuția POISSON

- Variabila aleatoare POISSON ia o infinitate numărabilă de valori: 0,1,2,...,k,... , care reprezintă numărul de realizări într-un interval dat de timp sau spațiu ale unui eveniment:
  - numărul de intrări pe an într-un spital
  - numărul de globule albe de pe frotiu
  - numărul de dezintegrări ale unei substanțe radioactive într-un interval de timp T dat

Distribuția POISSON

- Variabila aleatoare POISSON
- Este caracterizată de parametrul teoretic θ (numărul mediu aşteptat de realizări ale evenimentului în intervalul considerat)
- Simbol: Po(θ)
- Legea de distribuție:

$$X : \begin{pmatrix} k \\ e^{-\theta} \cdot \frac{\theta^k}{k!} \end{pmatrix} \qquad \Pr(X = k) = \frac{e^{-\theta} \cdot \theta^k}{k!}$$

Distribuția POISSON

- Speranța matematică: $M(X) = \theta$
- Variația: $V(X) = \theta$

Distribuția POISSON

- Rata de mortalitate pentru a anumită patologie virală este de 7 la 1000 de cazuri. Care este probabilitatea ca într-un grup de 400 entități această patologie să determine 5 decese?
- n=400
- p=7/1000=0,007

- $\theta = n \cdot p = 400 \cdot 0{,}007 = 2{,}8$
- e=2,718281828=2,72

Pr(X=5)

$$= (2{,}72^{-2{,}8} \cdot 2{,}8^{5}) \,/\, (5 \cdot 4 \cdot 3 \cdot 2 \cdot 1)$$

=10,45/120

=0,09