



**Abordări
probabiliste în
Machine Learning**



PROBABILITAT

E?



Ce este probabilitatea?

Probabilitatea este un concept care estimează răspunsul la întrebarea:

- “Se va produce oare un anumit eveniment?”, unde răspunsul clasic "nu" sau "da" este înlocuit cu scara numerică de la 0 până la 1.
 - "0" arată că evenimentul nu va avea loc (echivalent răspunsului ferm "nu"),
 - "1" reprezintă faptul că evenimentul va avea loc (echivalent răspunsului ferm "da").

Definiții și notații

- Probabilitatea este adesea asociată cu cel puțin un eveniment
 - aruncarea zarului
 - bile negre/roșii din coș
- Rezultatul evenimentului este **random**
 - variabila care reprezintă rezultatul acestor evenimente se numește **variabilă aleatorie (VA)**

- De multe ori suntem interesați să cunoaștem valoarea acestei variabile
 - De exemplu, care este probabilitatea ca atunci când un zar „corect” să cadă 3?
 - corect ar însemna că probabilitatea de a cadea pe oricare dintre 6 fețe este **egală**
 - Intuitiv răspunsul este $\frac{1}{6}$ - **Corect!**
- Dar cum reprezentăm asta matematic?
- De obicei VA sunt notate prin majuscule
 - Vom nota variabila aleatorie prin **X**

- Prin urmare, vrem să știm care este probabilitatea că $X = 3$
- „Care este probabilitatea?” – notăm prin P
- Așadar, enunțul „care este probabilitatea ca atunci când arunc un zar corect să aterizeze pe un 3?”
 - Îl putem traduce matematic ca „ $P(X = 3)$ ”

Cele 3 tipuri de probabilitate

- Probabil că primul lucru care trebuie înțeles este că există diferite tipuri de probabilitate:
 - **Probabilitate marginală**
 - **Probabilitatea comună**
 - **Probabilitate condiționată**

Probabilitate marginală

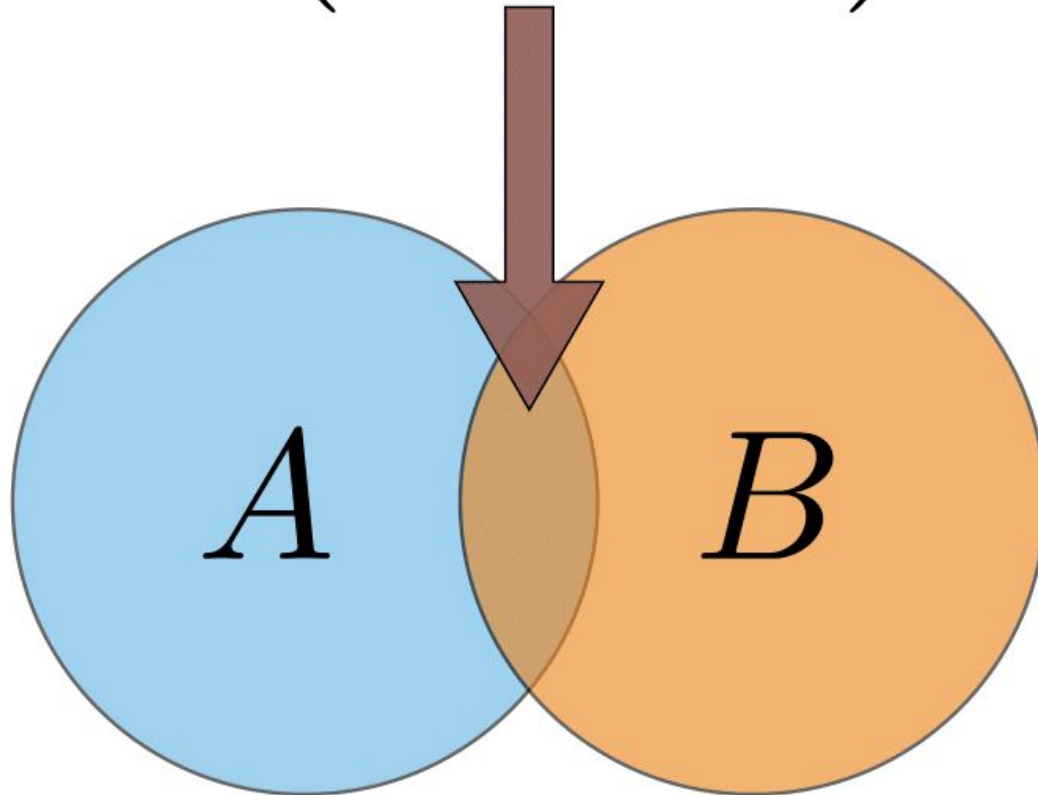
- Dacă **A** este un eveniment, atunci probabilitatea marginală este probabilitatea ca acel eveniment să se producă, $P(A)$.
- Presupunem că avem un pachet de cărți de joc
 - un exemplu de probabilitate marginală ar fi probabilitatea ca o carte extrasă dintr-un pachet să fie roșie: $P(\text{roșu}) = 0,5$.

Probabilitate comună

- Probabilitatea a două evenimente care sunt intersectate
- Dacă A și B sunt două evenimente, atunci probabilitatea comună a celor două evenimente este scrisă ca $P(A \cap B)$
- Presupunem că avem un pachet de cărți de joc:
 - probabilitatea ca o carte extrasă dintr-un pachet să fie roșie și să aibă valoarea 4 este $P(\text{roșu și } 4) = 2/52 = 1/26$

Probabilitate comuna

$$P(A \cap B)$$



Probabilitate condiționată

- Probabilitatea ca un eveniment să aibă loc știind că alte evenimente au avut deja loc
- Dacă A și B sunt două evenimente, atunci probabilitatea condiționată a apariției A având în vedere că B s-a produs se scrie ca $P(A | B)$
- Presupunem că avem un pachet de cărți de joc:
 - probabilitatea ca o carte să fie 4, dat fiind faptul că am extras o carte roșie este $P(4 | \text{roșu}) = 2/26 = 1/13$

Legarea celor 3 tipuri

- Regula generală de multiplicare
 - Regula generală de înmulțire este o ecuație frumoasă care leagă toate cele 3 tipuri de probabilitate:

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

Explicații

- Uneori, distincția dintre **probabilitatea comună** și **probabilitatea condiționată** poate fi destul de confuză, așa că folosind exemplul alegerii unei cărți dintr-un pachet de cărți de joc vom încerca să facem diferența

- În cazul în care dorim să găsim probabilitatea de a extrage o carte care este **roșie și 4** adică probabilitatea comună **P (roșu și 4)** imaginați-vă că toate cele 52 de cărți sunt cu fața în jos și alegeți una la întâmplare
- Dintre cele 52 de cărți, 2 dintre ele sunt „**roșu și 4**” (4 de dobă și 4 de inimă). Deci, probabilitatea comună este $\frac{2}{52} = \frac{1}{26}$

- În cazul în care dorim să găsim probabilitatea de a extrage o carte care este 4 având în vedere că știm că cardul este deja roșu, adică probabilitatea condiționată, $P(4 \mid \text{roșu})$, imaginați-vă că aveți toate cele 52 de cărți
- însă, înainte de a alege o carte la întâmplare, **sortați cărțile și selectați toate cele 26 de roșii**. Acum puneți cele 26 de cărți cu fața în jos și alegeți o carte la întâmplare. Din nou, 2 dintre acele cărți roșii sunt 4, astfel încât probabilitatea condiționată este $2/26 = 1/13$

- Putem folosi regula generală de înmulțire definită mai sus pentru a calcula probabilitatea comună
- Mai întâi ne reamenajăm pentru a face probabilitatea comună, $P(A \cap B)$, subiectul ecuației
- După rearanjament obținem

$$P(A \cap B) = P(A | B) \times P(B)$$

- Fie A în cazul în care cartea este un 4 și B este în cazul în care cartea este de roșu.

$$P(A | B) = 1/13 \text{ și } P(B) = 1/2$$

Inferența Bayesiană

- Înainte de a introduce inferența bayesiană, este necesar să înțelegem teorema lui Bayes

Cunoștințe apriorii - **Prior**

- De exemplu, dacă dorim să găsim probabilitatea de a vinde înghețată într-o zi caldă și însorită, teorema lui Bayes ne oferă instrumentele pentru a utiliza cunoștințe prealabile despre probabilitatea de a vinde înghețată în orice alt tip de zi

Definiția matematică

- Teorema lui Bayes:

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

- unde A și B sunt evenimente, $P(B|A)$ este probabilitatea condiționată de a se produce evenimentul A, având în vedere că evenimentul B a avut deja loc și $P(A)$ și $P(B)$ sunt probabilitățile marginale ale evenimentului A și B

Exemplu

- În pachet sunt 52 de cărți, 26 dintre ele sunt roșii și 26 negre. Care este probabilitatea ca cartea aleasă să fie un 4 având în vedere că știm că cartea este de roșu?
- Traducerea în matematică:
 - evenimentul A este evenimentul în care cartea aleasă este un 4, iar evenimentul B este că cartea este de roșu
- $P(A | B)$ în ecuația de mai sus este $P(4 | \text{roșu})$ în exemplul nostru, și asta este ceea ce vrem să calculăm, dar după teorema lui Bayes

- Înclocuim în Teorema lui Bayes:

$$P(4|roșu) = \frac{P(roșu|4) \times P(4)}{P(roșu)}$$

- Trebuie să găsim probabilitățile pentru termenii din partea dreaptă. Respectiv:

1. $P(B|A) = P(roșu|4) = 1/2$

2. $P(A) = P(4) = 4/52 = 1/13$

Cum ne permite teorema lui
Bayes să încorporăm
observările anterioare?

- Fie A să reprezinte evenimentul în care vindem înghețată și B să fie evenimentul vremii
- Apoi am putea întreba care este probabilitatea de a vinde înghețată într-o zi dată, având în vedere tipul de vreme?
- Din punct de vedere matematic, aceasta este scrisă ca $P(A = \text{vânzare de înghețată} \mid B = \text{tip de vreme})$, echivalent cu partea stângă a ecuației

- **P (A)** pe partea dreaptă este expresia care este cunoscută drept anterioară. În exemplul nostru, acesta este $P (A = \text{vânzare de înghețată})$, adică probabilitatea (marginală) de a vinde înghețată indiferent de tipul de vreme afară
 - **P (A)** este cunoscut ca prior (anterior), deoarece am putea cunoaște deja probabilitatea marginală a vânzării înghețatei

- De exemplu, aş putea privi datele care spuneau că 30 de persoane din 100 au cumpărat de fapt îngheţată la un magazin undeva. Deci,

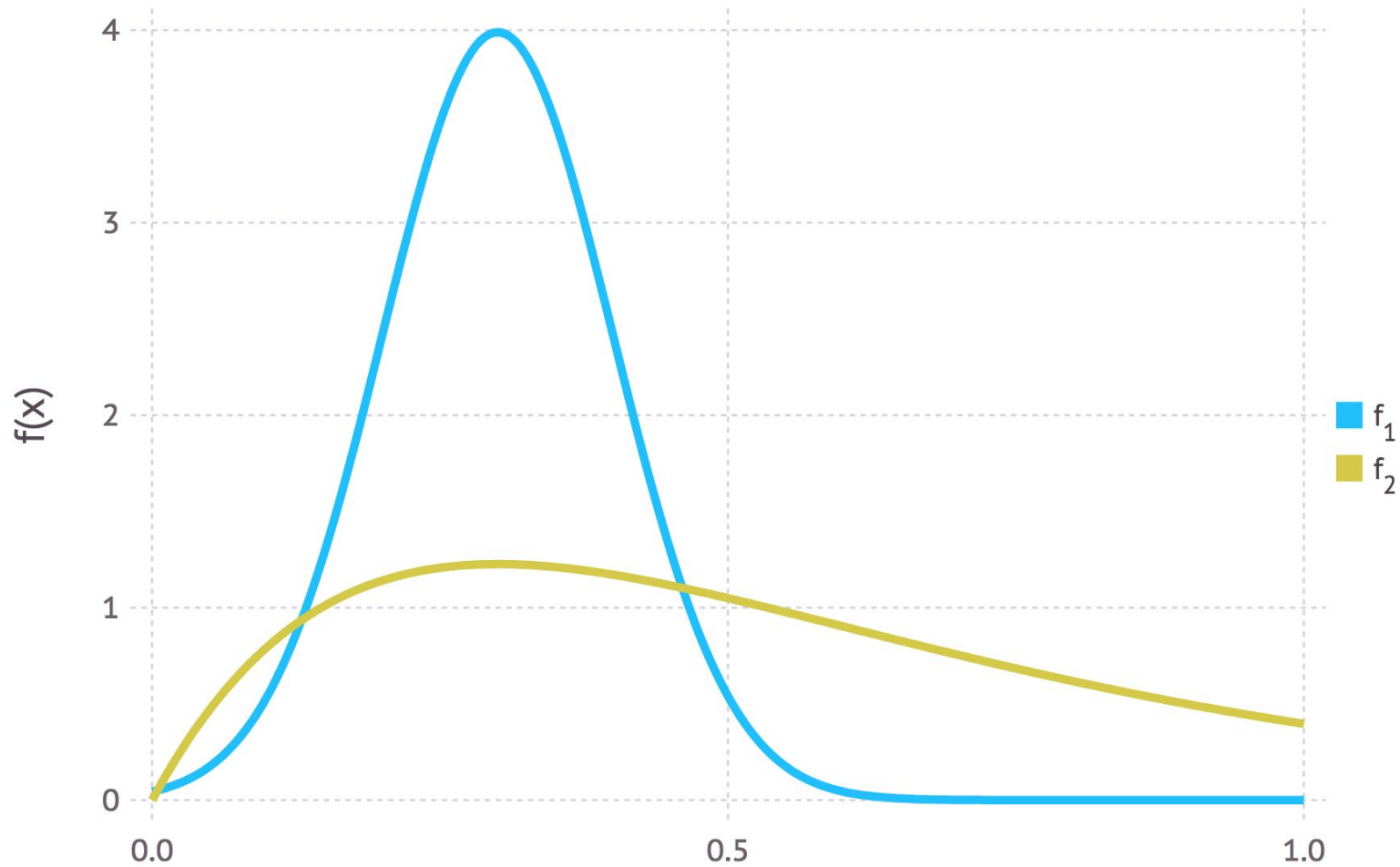
$$P(A = \text{vânzare de îngheţată}) = 30/100 = 0,3$$

- înainte să ştiu ceva despre vreme. Acesta este modul în care teorema lui Bayes ne permite să includem informaţii anterioare

Acum știm care este teorema
lui Bayes și cum să o
folosim, putem începe să
răspundem la întrebarea:
Ce este inferența bayesiană?

- **Inferența** este procesul de distribuire a probabilității din date
- **Inferența bayesiană** este, așadar, doar procesul de distribuire a probabilităților din date folosind teorema lui Bayes

- În exemplul de înghețată de mai sus, am văzut că probabilitatea anterioară de a vinde înghețată era de 0,3
- Cu toate acestea, dacă 0.3 ar fi doar cea mai bună presupunere a mea, dar eram un pic nesigur cu privire la această valoare, probabilitatea ar putea fi, de asemenea, 0,25 sau 0,4.
- Această distribuție este cunoscută sub numele de **distribuție anterioară**.



Probabilitatea vânzării de înghețata

Modelul teoremei lui Bayes

- În loc de evenimentul A , vom vedea de obicei Θ , acest simbol se numește Theta.
- Deci, dacă încercăm să estimăm valorile parametrilor unei distribuții gaussiene, atunci Θ reprezintă atât media, μ cât și abaterea de la standard, σ , scrisă matematic ca $\Theta = \{\mu, \sigma\}$.
- În loc de evenimentul B , vom vedea date sau

$$\mathbf{x} = \{x_1, x_2, \dots, x_n\}$$

Acestea reprezintă datele, adică setul de observații pe care le avem.

$$P(\Theta | data) = \frac{P(data | \Theta) \times P(\Theta)}{P(data)}.$$

- Am văzut că $P(\Theta)$ este **distribuția anterioară**. Reprezintă observările noastre despre adevărata valoare a parametrilor, la fel cum am avut distribuții care reprezintă observarea noastră despre probabilitatea de a vinde înghețată.

- $P(\Theta \mid \text{data})$ din partea stângă este cunoscută sub numele de **distribuție posterioară**. Aceasta este distribuția care reprezintă observarea noastră despre valorile parametrilor după ce am calculat tot pe partea dreaptă, luând în considerare datele observate.



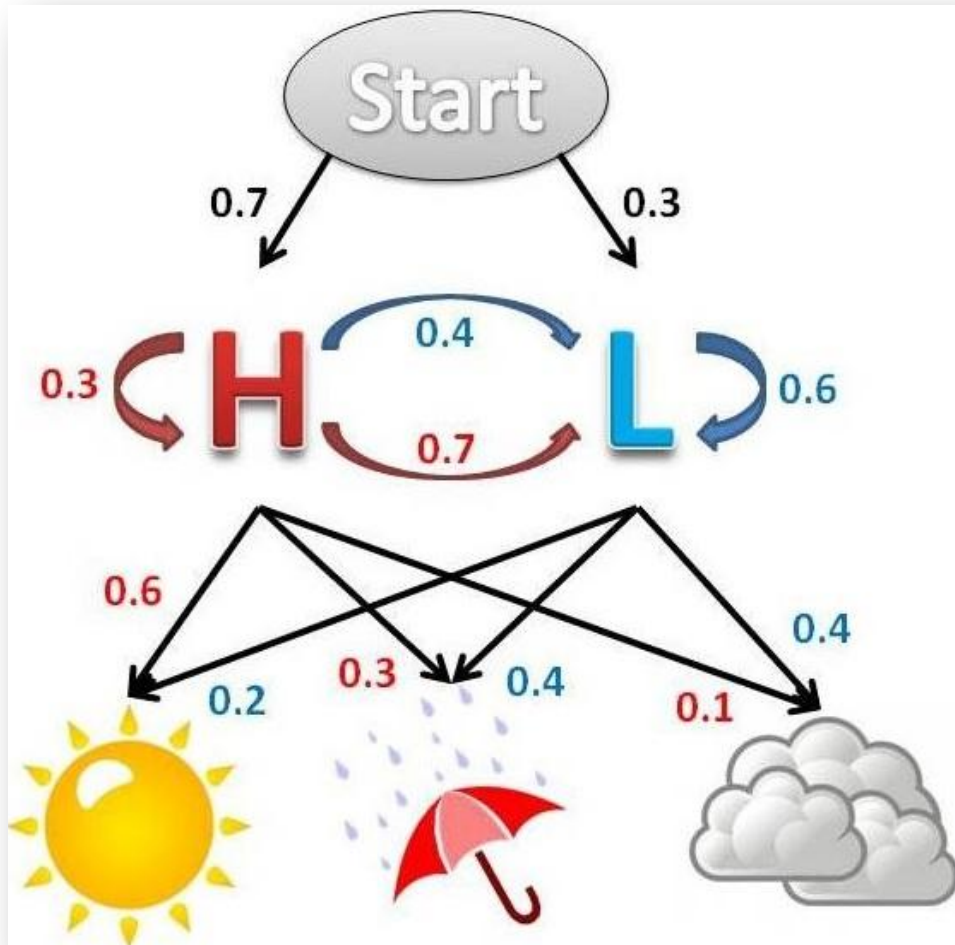
TEORIA PROBABILITĂȚII & ML

- În general vorbind, teoria probabilității este studiul matematic al incertitudinii.
- Acesta joacă un rol central în procesul de învățare a mașinilor,
 - întrucât proiectarea algoritmilor de învățare se bazează adesea pe asumarea probabilistică a datelor.

Exemplul

1

- Avem următoarea figură:



Există o anumită stare - de exemplu, temperatură ridicată sau scăzută a aerului.

- Ce fel de stare - nu știm, dar cunoaștem semnele: soarele strălucea, ploaia, erau nori.
- Avem un set de valori colectate în timpul perioadei statistice: soarele - ploaia, soarele - ploaia - norii.

Exemplul

1

- **Numerele de deasupra săgeților sunt probabilități, adică :**
 - ❖ **dacă este însorit, atunci cu o probabilitate de căldură de 60% și cu o probabilitate de 20% este rece.**
 - ❖ **dacă există nori, atunci căldura este de 10% și este rece cu o probabilitate de 40%.**

- **Din exemplul anterior concluzionăm că ML întipărește informația statistică pe o perioadă studiată, fie zile, luni, ani,
– luând în considerare probabilitatea producerii evenimentului anumit.**
- **Pe măsură ce acumulează informația ML poate face o prognoză a vremii pe viitor.**

Scurt istoric

- Pentru prima dată abordări probabiliste în Machine Learning au avut loc în anii '90 și 2000 până când a început revoluția profundă declanșată de apariția rețelelor neuronale.

Abordări probabiliste: Bayes

□ **Abordarea Bayes este cea mai academică viziune a învățării mașinilor.**

Acest lucru este bun și rău.

Abordări probabiliste:

Bayes

De ce e

bun?

Pentru că oferă o descriere matematică clară a formării și a estimărilor numerice ale fiabilității ipotezelor.

De ce e

rău?

- Pentru că în viața reală nu este întotdeauna posibilă implementarea unui model matematic impecabil.

Avem notațiile:

- $P(h)$ - probabilitatea că se petrece evenimentul h (h , în acest caz nu este număr, ci înseamnă o abstractizare, care, desigur, poate fi și un număr; vom desemna prin h - o anumită ipoteză)
- $P(A \wedge B)$ - este probabilitatea că a avut loc atât evenimentul A cât și evenimentul B
- $P(A \setminus B)$ - probabilitatea că s-a întâmplat evenimentul A , cu condiția că avut loc și B (adică, în cazul în care A - posesia mașinii roșii și B - posesia mașinii, atunci $P(A \setminus B)$ - probabilitatea de a deține o mașină roșie, calculată numai pentru proprietarii de autoturisme)

Teorema Bayes

- Teorema lui Bayes este o consecință a afirmației evidente:

$$P(A \cap B) = P(A \setminus B)P(B) = P(B \setminus A)P(A)$$

- Adică:

$$P(A \setminus B) = \frac{P(B \setminus A)P(A)}{P(B)}$$

- Bineînțeles, $P(B)$ trebuie să fie mai mare decât 0, dar acest lucru este de înțeles, dacă B este un eveniment improbabil, atunci $P(A \setminus B)$ este de asemenea improbabil.

- Fie că notăm datele noastre – D , iar ipotezele noastre – h . Atunci trebuie să găsim probabilitatea ipotezei pentru datele noastre $P(h \setminus D)$, care prin teorema lui Bayes este egală cu:

$$P(h \setminus D) = \frac{P(D \setminus h)P(h)}{P(D)}$$

- Suntem interesați doar de relația de probabilitate, deci putem elimina $P(D)$ din această expresie (D nu depinde de h) și $P(h)$ (presupunem că toate ipotezele sunt la fel de probabile, strict vorbind, acest lucru nu este întotdeauna așa, dar în multe cazuri acest lucru este adevărat).
- Reiese că trebuie să găsim o ipoteză h pentru care $P(D|h)$ este maximă.

- **Până ce nu avem nici date concrete, nici mulțimea ipotezelor.**
- **Hai să le inventăm.**

Exemplul

2

- Să presupunem că am extras $N=3$ mere: $R=1$ roșii și $G=2$ verzi.
- Să prezentăm două ipoteze:
 - ◆ h_1 - în coș sunt 10 mere dintre care 3 sunt roșii și 7 verzi
 - ◆ h_2 - în coș sunt 10 mere dintre acestea 7 sunt verzi și 3 roșii
- Acum, trebuie doar să calculăm probabilitatea obținerii datelor noastre în fiecare dintre aceste ipoteze.

- Notăm prin R_0 și G_0 numărul de mere roșii și verzi din coș.
- În câte moduri putem extrage R mere din R_0 ?

$$C(R_0, R) = \frac{R_0!}{(R_0 - R)!R!}$$

- Aceasta este formula coeficientului binomial. Adică dacă avem 3 mere, atunci putem să extragem 2 mere din ele în trei moduri ● ● ○, ● ○ ●, ○ ● ●:

$$C(3, 2) = \frac{3!}{(1! \cdot 2!)} = \frac{1 \cdot 2 \cdot 3}{(1 \cdot (1 \cdot 2))} = 3$$

- Formula dată este necesară doar p/u rezolvarea problemei respective .

- **Atunci probabilitatea căutată este egală cu:**

$$P(R_0, R, G_0, G) = \frac{C(R_0, R) \cdot C(G_0, G)}{C(R_0 + G_0, R + G)}$$

- **A rămas să aplicăm aceasta (probabilitatea) la ipotezele noastre:**

$$h_1 : P(3, 7, 1, 2) = 0.525$$

◆ **ipoteza** $h_2 : P(7, 3, 1, 2) = 0.175$

◆ **ipoteza**

- **Se observă că ipoteza I este în câștig.**

MĂRIM COLECȚIA DE IPOTEZE

- Hai să nu ne limităm la 2 ipoteze. Luăm în considerare oricare combinație R_0 și G_0 (R și G rămân aceleași). Obținem tabelul :

$R_0 \setminus G_0$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	0	1.0000	0.7500	0.6000	0.5000	0.4286	0.3750
2	0	0	0.5000	0.6000	0.6000	0.5714	0.5357	0.5000
3	0	0	0.3000	0.4500	0.5143	0.5357	0.5357	0.5250
4	0	0	0.2000	0.3429	0.4286	0.4762	0.5000	0.5091
5	0	0	0.1429	0.2679	0.3571	0.4167	0.4545	0.4773
6	0	0	0.1071	0.2143	0.3000	0.3636	0.4091	0.4406
7	0	0	0.0833	0.1750	0.2545	0.3182	0.3671	0.4038

Se observă 2 probleme:

- Problema mai mică: multe ipoteze au aceeași probabilitate (De ex. $(R_0 = 1, G_0 = 4)$ și $(R_0 = 2) G_0 = 4$)
- Problema mai mare: am obținut că soluția cea mai bună este $(R_0 = 1, G_0 = 2)$. Acest rezultat este atât exact, cât și fără sens. Adică ML ne-a spus:
 - *”dacă ați extras din coș 1 măr roșu și 2 mere verzi, înseamnă că cel mai probabil acolo și era 1 măr roșu și 2 mere verzi.”*
- Așa și este. ML are dreptate, dar avem nevoie de așa un răspuns?
- Evident că acesta nu e rezultatul la care ne așteptam.
- Aceasta se numește supraînvățare a mașinii.

- **Inteligența artificială a noastră a învățat pe de rost toate datele de învățare. Cel mai probabil ML nu e în stare să rezolve problemele calitativ, dar la toate întrebările de învățare ea dă un răspuns de învățare exact.**

- **Să presupunem că știm de undeva că în coș se află doar 6, 7 sau 8 mere. Probabilitatea altor ipoteze este 0. Atunci tabelul nostru ia forma:**

$R_0 \setminus G_0$	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0.5000	0.4286	0.3750
2	0	0	0	0	0.6000	0.5714	0.5357	0
3	0	0	0	0.4500	0.5143	0.5357	0	0
4	0	0	0.2000	0.3429	0.4286	0	0	0
5	0	0	0.1429	0.2679	0	0	0	0
6	0	0	0.1071	0	0	0	0	0
7	0	0	0	0	0	0	0	0

MULȚUMESC

PENTRU

ATENȚIE !!!