# Curs 8

Cristian Niculescu

# 1 Actualizare Bayesiană: şanse

## 1.1 Scopurile învăţării

1. Să poată converti şansele în probabilitate şi invers.
2. Să poată actualiza şansele a priori în şanse a posteriori folosind factori Bayes.
3. Să înţeleagă cum factorii Bayes măsoară gradul în care datele dau dovezi pentru sau împotriva ipotezelor.

## 1.2 Şanse

Când comparăm 2 evenimente, putem formula afirmaţiile probabilistice în termeni de şanse.

**Definiţie.** Şansele unui eveniment $E$ versus evenimentul $E'$ sunt raportul probabilităţilor lor $P(E)/P(E')$. Dacă nu se specifică evenimentul $E'$, al 2-lea eveniment este presupus a fi evenimentul complementar $E^c$. Astfel, şansele lui $E$ sunt:

$$O(E) = \frac{P(E)}{P(E^c)}.$$

De exemplu, $O(\text{ploaie}) = 2$ înseamnă că probabilitatea de a ploua este de 2 ori probabilitatea de a nu ploua (2/3 versus 1/3). Putem spune că "şansele de ploaie sunt 2 la 1".

**Exemplu.** Pentru o monedă corectă, $O(\text{avers}) = \frac{1/2}{1/2} = 1$. Putem spune că şansele aversului sunt 1 la 1 sau 50-50.

**Exemplu.** Pentru un zar standard, şansele de a da 4 sunt $\frac{1/6}{5/6} = \frac{1}{5}$. Putem spune că şansele sunt "1 la 5 pentru" sau "5 la 1 împotriva" a da un 4.

**Exemplu.** Probabilitatea unei perechi într-o mână de poker de 5 cărţi este 0.42257. Astfel, şansele unei perechi sunt $0.42257/(1 - 0.42257) = 0.73181$.

**Formule de conversie:** Dacă $P(E) = p$, atunci $O(E) = \frac{p}{1-p}$. Dacă $O(E) = q$, atunci $P(E) = \frac{q}{1+q}$.

Observaţii:
1. A 2-a formulă provine din rezolvarea ecuaţiei în $p$
$q = p/(1-p)$.
2. Probabilităţile sunt între 0 şi 1, în timp ce şansele sunt între 0 şi $\infty$.
3. Proprietatea $P(E^c) = 1 - P(E)$ devine $O(E^c) = 1/O(E)$.
**Exemplu.** Fie $F$ evenimentul că o mână de poker să fie ful (3 cărţi de un rang şi alte 2 cărţi de un alt rang). Atunci $P(F) = 0.0014521$ deci $O(F) = 0.0014521/(1 - 0.0014521) = 0.0014542$.
Şansele de a nu avea ful sunt
$O(F^c) = (1 - 0.0014521)/0.0014521 = 687.6578 = 1/O(F)$.
4. Dacă $P(E)$ sau $O(E)$ este mic, atunci $O(E) \approx P(E)$. Aceasta rezultă din formulele de conversie.
**Exemplu.** La exemplul din poker, unde $F =$ "ful", am văzut că $P(F)$ şi $O(F)$ diferă abia la a 6-a zecimală.

## 1.3 Actualizarea şanselor

### 1.3.1 Introducere

În actualizarea Bayesiană, am utilizat verosimilitatea datelor pentru a actualiza probabilităţile a priori ale ipotezelor la probabilităţi a posteriori. În limbaj de şanse, vom actualiza şansele a priori la şanse a posteriori. Datele dau dovezi pentru sau împotriva unei ipoteze dacă şansele ei a posteriori sunt mai mari sau mai mici decât şansele ei a priori.

### 1.3.2 Exemplu: sindromul Marfan

Sindromul Marfan este o boală genetică a ţesutului conjunctiv care apare la 1 din 15000 de oameni. Principalele caracteristici oculare ale sindromului Marfan includ ectopia de cristalin bilaterală, miopie şi detaşarea retinei. Aproximativ 70% din oamenii cu sindrom Marfan au cel puţin una din aceste caracteristici oculare, în timp ce doar 7% din oamenii fără sindrom Marfan au. (Nu garantăm acurateţea acestor numere.)
Dacă o persoană are cel puţin una din aceste caracteristici oculare, care sunt şansele să aibă sindromul Marfan?
**Răspuns:** Aceasta este o problemă standard de actualizare Bayesiană. Ipotezele noastre sunt:
$M =$ "persoana are sindromul Marfan",
$M^c =$ "persoana nu are sindromul Marfan".
Datele sunt:
$F =$ "persoana are cel puţin o caracteristică oculară".

Ştim probabilitatea a priori a lui $M$ şi verosimilităţile lui $F$ dat fiind $M$ sau $M^c$:

$$P(M) = 1/15000, \; P(F|M) = 0.7, \; P(F|M^c) = 0.07.$$

Putem calcula probabilităţile a posteriori folosind un tabel:

| hypothesis | prior | likelihood | Bayes numerator | posterior |
|---|---|---|---|---|
| $H$ | $P(H)$ | $P(F\|H)$ | $P(F\|H)P(H)$ | $P(H\|F)$ |
| $M$ | 0.000067 | 0.7 | 0.0000467 | 0.00066 |
| $M^c$ | 0.999933 | 0.07 | 0.069995 | 0.99933 |
| total | 1 | | 0.07004 | 1 |

Şansele a priori sunt:

$$O(M) = \frac{P(M)}{P(M^c)} = \frac{1/15000}{14999/15000} = \frac{1}{14999} \approx 0.000067.$$

Şansele a posteriori sunt date de raportul probabilităţilor a posteriori sau al numărătorilor Bayes, deoarece factorul normalizator este acelaşi la numărător şi la numitor.

$$O(M|F) = \frac{P(M|F)}{P(M^c|F)} = \frac{P(F|M)P(M)}{P(F|M^c)P(M^c)} = 0.000667.$$

Şansele a posteriori sunt de 10 ori mai mari decât şansele a priori. În acest sens, a avea o caracteristică oculară este o dovadă puternică în favoarea ipotezei $M$. Totuşi, deoarece şansele a priori sunt atât de mici, este încă foarte puţin probabil ca persoana să aibă sindromul Marfan.

## 1.4 Factorii Bayes şi puterea dovezii

Factorul 10 din exemplul precedent se numeşte factor Bayes.
**Definiţie:** Pentru ipoteza $H$ şi datele $D$, factorul Bayes este raportul verosimilităţilor:

$$\text{factorul Bayes} = \frac{P(D|H)}{P(D|H^c)}.$$

Să vedem exact unde factorul Bayes apare în actualizarea şanselor. Avem

$$\begin{aligned}
O(H|D) &= \frac{P(H|D)}{P(H^c|D)} \\
&= \frac{P(D|H)P(H)}{P(D|H^c)P(H^c)} \\
&= \frac{P(D|H)}{P(D|H^c)} \cdot \frac{P(H)}{P(H^c)} \\
&= \frac{P(D|H)}{P(D|H^c)} \cdot O(H).
\end{aligned}$$

şansele a posteriori = **factorul Bayes** $\times$ şansele a priori.

Din această formulă vedem că factorul Bayes ($BF$) ne spune dacă datele dau dovezi pentru sau împotriva ipotezei.
Dacă $BF > 1$, atunci şansele a posteriori sunt mai mari decât şansele a priori. Astfel, datele dau dovezi pentru ipoteză.
Dacă $BF < 1$, atunci şansele a posteriori sunt mai mici decât şansele a priori. Astfel, datele dau dovezi împotriva ipotezei.
Dacă $BF = 1$, atunci şansele a priori şi a posteriori sunt egale. Astfel, datele nu dau nicio dovadă pentru sau împotriva ipotezei.
Următorul exemplu este din cartea *Information Theory, Inference, and Learning Algorithms* de David J. C. Mackay, care spune cu privire la probele dintr-un proces:
”În opinia mea, sarcina unui juriu ar trebui în general să fie a înmulţi împreună rapoarte de verosimilitate evaluate cu grijă de la fiecare parte independentă de dovadă admisibilă cu o probabilitate a priori motivată la fel de grijuliu. Acest punct de vedere este împărtăşit de mulţi statisticieni, dar învăţaţii judecători de apel britanici nu au fost de acord recent şi în realitate au răsturnat verdictul unui proces deoarece juraţii fuseseră învăţaţi să utilizeze teorema lui Bayes pentru a trata dovezi ADN complicate.”
**Exemplul 1.** 2 oameni au lăsat urme de sânge la locul unei crime. Un suspect, Oliver, are grupa de sânge ”0”. Grupele de sânge ale celor 2 urme sunt ”0” (o grupă obişnuită în populaţia locală, având frecvenţa 60%) şi ”AB” (o grupă rară, cu frecvenţa 1%). Dau aceste date (grupele ”0” şi ”AB” găsite la locul crimei) dovezi în favoarea propoziţiei că ”Oliver a fost una dintre cele 2 persoane prezente la locul crimei”?
**Răspuns:** Sunt 2 ipoteze:
$S$ = ”Oliver şi o altă persoană necunoscută au fost la locul crimei”;
$S^c$ = ”2 persoane necunoscute au fost la locul crimei”.
Datele sunt:
$D$ = ”grupele de sânge ”0” şi ”AB” au fost găsite”.

Factorul Bayes pentru prezenţa lui Oliver este $BF_{\text{Oliver}} = \frac{P(D|S)}{P(D|S^c)}$. Calculăm numărătorul şi numitorul separat.
Datele spun că ambele grupe de sânge "0" şi "AB" au fost aflate la locul crimei. Dacă Oliver a fost la locul crimei, grupa "0" ar fi acolo. Deci, $P(D|S)$ este probabilitatea că cealaltă persoană avea grupa "AB". Ni s-a spus că aceasta este 0.01, deci $P(D|S) = 0.01$.
Dacă Oliver nu a fost la locul crimei, atunci au fost 2 persoane aleatoare, una cu grupa "0" şi una cu grupa "AB". Probabilitatea acestui fapt este $2 \cdot 0.6 \cdot 0.01$. Factorul 2 este deoarece sunt 2 moduri în care se poate întâmpla asta - prima persoană are grupa 0 şi a 2-a grupa AB sau viceversa. (Am presupus că grupele de sânge ale celor 2 persoane sunt independente. Aceasta nu este exact adevărat, dar pentru o populaţie suficient de mare este destul de aproape de adevăr. Fie $N_0$, numărul de oameni cu grupa 0, $N_{AB}$, numărul de oameni cu grupa AB şi $N$, mărimea populaţiei. Probabilitatea exactă este $\frac{(N_0-1)\cdot N_{AB}}{C^2_{N-1}} = \frac{2\cdot(N_0-1)\cdot N_{AB}}{(N-1)(N-2)} = 2 \cdot \frac{0.6N-1}{N-1} \cdot \frac{0.01N}{N-2} \xrightarrow[N\to\infty]{} 2 \cdot 0.6 \cdot 0.01$. Aceasta arată că pentru $N$ suficient de mare, probabilitatea este aproximativ $2 \cdot 0.6 \cdot 0.01$.)
Factorul Bayes pentru prezenţa lui Oliver la locul crimei este

$$BF_{\text{Oliver}} = \frac{P(D|S)}{P(D|S^c)} \approx \frac{0.01}{2 \cdot 0.6 \cdot 0.01} = 0.8(3).$$

Deoarece $BF_{\text{Oliver}} < 1$, datele dau dovezi (slabe) împotriva prezenţei lui Oliver la locul crimei.
**Exemplul 2.** Un alt suspect, Alberto, are grupa de sânge AB. Dau aceleaşi date dovezi în favoarea propoziţiei că Alberto a fost una din cele 2 persoane prezente la locul crimei?
**Răspuns:** Refolosind notaţiile de mai sus cu Alberto în locul lui Oliver, avem:

$$BF_{\text{Alberto}} = \frac{P(D|S)}{P(D|S^c)} \approx \frac{0.6}{2 \cdot 0.6 \cdot 0.01} = 50.$$

Deoarece $BF_{\text{Alberto}} \gg 1$, datele dau o dovadă puternică în favoarea prezenţei lui Alberto la locul crimei.
Observaţii:
1. În ambele exemple, am calculat doar factorul Bayes, nu şansele a posteriori. Pentru a le calcula pe ultimele, am avea nevoie să ştim şansele a priori pentru prezenţa lui Oliver (sau Alberto) la locul crimei, pe baza altor dovezi.
2. Dacă 50% din populaţie ar fi avut grupa 0 de sânge în loc de 60%, atunci factorul Bayes al lui Oliver ar fi fost 1 (nici pentru, nici împotrivă). Mai general, punctul de echilibru pentru dovezile cu grupe de sânge este când proporţia grupei de sânge a suspectului din populaţia generală egalează proporţia grupei de sânge a suspectului dintre cele aflate la locul crimei.

### 1.4.1 Reactualizare

Presupunem că adunăm datele în 2 etape, întâi $D_1$, apoi $D_2$. A posteriori finale pot fi calculate toate o dată sau în 2 etape, unde întâi actualizăm a priori folosind verosimilităţile pentru $D_1$ şi apoi actualizăm a posteriori rezultate după prima actualizare folosind verosimilităţile pentru $D_2$. Ultima abordare merge ori de câte ori verosimilităţile se înmulţesc:

$$P(D_1, D_2|H) = P(D_1|H)P(D_2|H).$$

Deoarece verosimilităţile sunt condiţionate de ipoteze, spunem că $D_1$ şi $D_2$ sunt condiţionat independente dacă relaţia de mai sus are loc pentru orice ipoteză $H$.

**Exemplu.** Sunt 5 zaruri într-un sertar, cu 4, 6, 8, 12 şi 20 de feţe (acestea sunt ipotezele). Luăm aleator un zar şi-l aruncăm de 2 ori, obţinând întâi 7, apoi 11. Sunt aceste rezultate condiţionat independente? Dar independente?

**Răspuns.** Rezultatele sunt condiţionat independente. De exemplu, pentru ipoteza zarului cu 8 feţe avem:

$$\begin{aligned}
P(7 \text{ la aruncarea } 1|\text{zar cu 8 feţe}) &= 1/8,\\
P(11 \text{ la aruncarea } 2|\text{zar cu 8 feţe}) &= 0,\\
P(7 \text{ la aruncarea } 1, 11 \text{ la aruncarea } 2|\text{zar cu 8 feţe}) &= 0.
\end{aligned}$$

Pentru ipoteza zarului cu 20 de feţe avem:

$$\begin{aligned}
P(7 \text{ la aruncarea } 1|\text{zar cu 20 feţe}) &= 1/20,\\
P(11 \text{ la aruncarea } 2|\text{zar cu 20 feţe}) &= 1/20,\\
P(7 \text{ la aruncarea } 1, 11 \text{ la aruncarea } 2|\text{zar cu 20 feţe}) &= (1/20)^2.
\end{aligned}$$

Totuşi, rezultatele nu sunt independente. Adică:

$$P(7 \text{ la aruncarea } 1, 11 \text{ la aruncarea } 2) \neq P(7 \text{ la aruncarea } 1)P(11 \text{ la aruncarea } 2).$$

Intuitiv, aceasta este deoarece 7 la aruncarea 1 ne permite să eliminăm zarurile cu 4 şi 6 feţe, făcând un 11 la aruncarea 2 mai probabil. Ne verificăm intuiţia calcuând ambii membri. La membrul drept avem:

$$\begin{aligned}
P(7 \text{ la aruncarea } 1) &= \frac{1}{5}\cdot\frac{1}{8} + \frac{1}{5}\cdot\frac{1}{12} + \frac{1}{5}\cdot\frac{1}{20} = \frac{31}{600},\\
P(11 \text{ la aruncarea } 2) &= \frac{1}{5}\cdot\frac{1}{12} + \frac{1}{5}\cdot\frac{1}{20} = \frac{2}{75}.
\end{aligned}$$

Deci membrul drept este $\frac{31}{600} \cdot \frac{2}{75} = 0.0013(7)$.
Membrul stâng este:

$$\begin{aligned}
&P(7 \text{ la aruncarea } 1, 11 \text{ la aruncarea } 2) \\
&= P(11 \text{ la aruncarea } 2 | 7 \text{ la aruncarea } 1) \cdot P(7 \text{ la aruncarea } 1) \\
&= \left(\frac{10}{31} \cdot \frac{1}{12} + \frac{6}{31} \cdot \frac{1}{20}\right) \cdot \frac{31}{600} \\
&= \frac{17}{9000} = 0.001(8).
\end{aligned}$$

Aici $\frac{10}{31}$ şi $\frac{6}{31}$ sunt probabilităţile a posteriori ale zarurilor cu 12 şi cu 20 de feţe dat fiind un 7 la aruncarea 1 (se pot calcula cu un tabel de actualizare Bayesiană).
Concluzionăm că, fără condiţionarea de ipoteze, aruncările nu sunt independente.
Întorcându-ne la schema generală, dacă $D_1$ şi $D_2$ sunt condiţionat independente pentru $H$ şi $H^c$, atunci are sens să considerăm fiecare factor Bayes independent:

$$BF_i = \frac{P(D_i|H)}{P(D_i|H^c)}, i = 1, 2.$$

Şansele a priori ale lui $H$ sunt $O(H)$. Şansele a posteriori după $D_1$ sunt

$$O(H|D_1) = BF_1 \cdot O(H).$$

Şansele a posteriori după $D_1$ şi $D_2$ sunt

$$\begin{aligned}
O(H|D_1, D_2) &= BF_2 \cdot O(H|D_1) \\
&= BF_2 \cdot BF_1 \cdot O(H).
\end{aligned}$$

Actualizarea şanselor cu date noi, condiţionat independente de cele vechi, se face prin înmulţirea şanselor a posteriori curente cu factorul Bayes al noilor date.

**Exemplul 3. Alte simptome ale sindromului Marfan**

Reamintim din exemplul de mai devreme că factorul Bayes pentru cel puţin o caracteristică oculară ($F$) este

$$BF_F = \frac{P(F|M)}{P(F|M^c)} = \frac{0.7}{0.07} = 10.$$

Semnul încheieturii mâinii ($W$) este capacitatea de a înfăşura o mână în jurul celeilalte încheieturi a mâinii astfel încât unghia degetului mic să poată fi acoperită de degetul mare. Presupunem că 10% din populaţie are semnul încheiturii mâinii, în timp ce 90% din oamenii cu sindromul Marfan îl au. De aceea, factorul Bayes pentru semnul încheiturii mâinii este

$$BF_W = \frac{P(W|M)}{P(W|M^c)} = \frac{0.9}{0.1} = 9.$$

Presupunem că $F$ şi $W$ sunt simptome condiţionat independente. Adică, printre oamenii cu sindromul Marfan, caracteristicile oculare şi semnul încheieturii mâinii sunt independente şi printre oamenii fără sindromul Marfan, caracteristicile oculare şi semnul încheieturii mâinii sunt independente. Cu această presupunere, şansele a posteriori ale sindromului Marfan pentru cineva care are atât caracteristicile oculare cât şi semnul încheiturii mâinii sunt

$$O(M|F, W) = BF_W \cdot BF_F \cdot O(M) = 9 \cdot 10 \cdot \frac{1}{14999} \approx \frac{6}{1000}.$$

Putem converti şansele posterioare în probabilitate, dar, deoarece şansele sunt atât de mici, rezultatul este aproape acelaşi:

$$P(M|F, W) \approx \frac{6}{1000 + 6} \approx 0.596\%.$$

Aşadar, caracteristicile oculare şi semnul încheiturii sunt ambele dovezi puternice în favoarea ipotezei $M$ şi, luate împreună, sunt dovezi foarte puternice. Din nou, deoarece şansele a priori sunt atât de mici, este încă improbabil că persoana are sindromul Marfan, dar în această situaţie merită să fie făcute alte teste date fiind consecinţele potenţial fatale ale bolii (ca anevrismul aortic sau disecţia aortică).
Observăm de asemenea că, dacă o persoană are exact unul din cele 2 simptome, atunci produsul factorilor Bayes este aproape de 1 (9/10 sau 10/9). Astfel, cele 2 părţi ale datelor se anulează una pe alta în ce priveşte dovezile pe care le dau pentru sindromul Marfan.

## 1.5 Log şanse

În practică, este adesea convenabil a lucra cu logaritmii naturali ai şanselor în locul şanselor. Destul de firesc, aceştia sunt numiţi log şanse. Formula de actualizare Bayesiană

$$O(H|D_1, D_2) = BF_2 \cdot BF_1 \cdot O(H)$$

devine

$$\ln(O(H|D_1, D_2)) = \ln(BF_2) + \ln(BF_1) + \ln(O(H)).$$

Putem interpreta formula de mai sus pentru log şansele a posteriori ca suma log şanselor a priori cu toate dovezile $\ln(BF_i)$ furnizate de date. Luând logaritmii, dovezile pentru $(BF_i > 1)$ sunt pozitive şi dovezile împotrivă $(BF_i < 1)$ sunt negative. Log şansele joacă de asemenea un rol central în regresia logistică, un model statistic legat de regresia liniară.

# 2 Actualizare Bayesiană cu a priori continue

## 2.1 Scopurile învăţării

1. Să înţeleagă o familie parametrizată de repartiţii ca reprezentând un domeniu continuu de ipoteze pentru datele observate.
2. Să poată să enunţe teorema lui Bayes şi legea probabilităţii totale pentru densităţi continue.
3. Să poată să aplice teorema lui Bayes pentru a actualiza o funcţie densitate de probabilitate a priori la o pdf a posteriori cunoscând datele şi funcţia de verosimilitate.
4. Să poată interpreta şi calcula probabilităţi predictive a posteriori.

## 2.2 Introducere

Până acum am făcut actualizare Bayesiană când aveam un număr finit de ipoteze, de pildă exemplul nostru cu zaruri avea 5 ipoteze (4, 6, 8, 12 sau 20 de feţe). Acum vom studia actualizare Bayesiană când există un domeniu continuu de ipoteze. Procesul de actualizare Bayesiană va fi în esenţă acelaşi ca în cazul discret. Ca de obicei când ne mişcăm de la discret la continuu vom avea nevoie să înlocuim funcţia masă de probabilitate cu o funcţie densitate de probabilitate şi sumele cu integrale.

Primele câteva secţiuni ale acestui curs sunt dedicate lucrului cu pdf-uri. În particular, vom cuprinde legea probabilităţii totale şi teorema lui Bayes. Acestea sunt în esenţă identice cu versiunile discrete. Apoi, vom aplica teorema lui Bayes şi legea probabilităţii totale pentru actualizare Bayesiană.

## 2.3 Exemple cu domenii continue de ipoteze

Iată 3 exemple standard cu domenii continue de ipoteze.

**Exemplul 1.** Presupunem că avem un sistem care poate reuşi (sau eşua) cu probabilitatea $p$. Atunci putem face ipotezele că $p$ este oriunde în domeniul [0,1]. Adică, avem un domeniu continuu de ipoteze. Adesea vom modela

acest exemplu cu o monedă cu probabilitatea $p$ a aversului necunoscută.
**Exemplul 2.** Durata de viaţă a unui anumit izotop este modelată de o repartiţie exponenţială $\exp(\lambda)$. În principiu, media duratei de viaţă $1/\lambda$ poate fi orice număr în $(0, \infty)$.
**Exemplul 3.** Nu suntem restricţionaţi la un singur parametru. În principiu, parametrii $\mu$ şi $\sigma$ ai unei repartiţii normale pot fi orice numere reale în $(-\infty, \infty)$, respectiv în $(0, \infty)$. Dacă modelăm durata sarcinii fără gemeni printr-o repartiţie normală, atunci din milioane de date ştim că $\mu$ este aproximativ 40 de săptămâni şi $\sigma$ este aproximativ o săptămână.
În toate aceste exemple am modelat procesele aleatoare dând naştere la date printr-o repartiţie cu parametri - numită repartiţie parametrizată. Fiecare posibilă alegere a parametrului (parametrilor) este o ipoteză, de pildă putem face ipoteza că probabilitatea succesului în exemplul 1 este $p = 0.7313$. Avem o mulţime continuă de ipoteze deoarece am putea lua orice valoare între 0 şi 1.

## 2.4 Convenţii de notaţie

### 2.4.1 Modele parametrizate

Ca în exemplele de mai sus ipotezele noastre adesea iau forma un anumit parametru are valoarea $\theta$. Vom utiliza adesea litera $\theta$ pentru o ipoteză arbitrară. Aceasta va lăsa simboluri ca $p$, $f$ şi $x$ să ia sensurile lor uzuale de pmf, pdf şi date. De asemenea, decât să spunem "ipoteza că parametrul de interes are valoarea $\theta$" vom spune simplu ipoteza $\theta$.

### 2.4.2 Litere mari şi mici

Avem 2 notaţii paralele pentru rezultate şi probabilităţi:
1. (Litere mari) Evenimentul $A$, funcţia de probabilitate $P(A)$.
2. (Litere mici) Valoarea $x$, pmf $p(x)$, pdf $f(x)$.
Aceste notaţii sunt legate prin $P(X = x) = p(x)$, unde $x$ este o valoare a variabilei aleatoare discrete $X$ şi "$X = x$" este evenimentul corespunzător.
Ducem aceste notaţii la probabilităţile folosite în actualizarea Bayesiană.
1. (Litere mari) Din ipotezele $\mathcal{H}$ şi datele $\mathcal{D}$ calculăm câteva probabilităţi asociate:

$$P(\mathcal{H}),\ P(\mathcal{D}),\ P(\mathcal{H}|\mathcal{D}),\ P(\mathcal{D}|\mathcal{H}).$$

În exemplul cu moneda putem avea $\mathcal{H} =$ "moneda aleasă are probabilitatea aversului 0.6", $\mathcal{D} =$ "aruncarea a fost avers" şi $P(\mathcal{D}|\mathcal{H}) = 0.6$.
2. (Litere mici) Valorile ipotezei $\theta$ şi valorile datelor $x$ au ambele probabilităţi

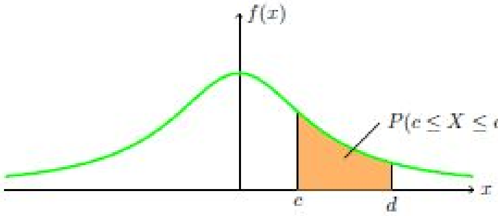
sau densităţi de probabilitate:

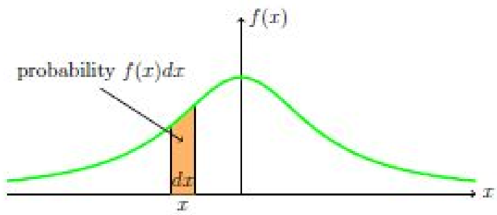
$$p(\theta),\ p(x),\ p(\theta|x),\ p(x|\theta)$$
$$f(\theta),\ f(x),\ f(\theta|x),\ f(x|\theta)$$

În exemplul cu moneda putem avea $\theta = 0.6$ şi $x = 1$, astfel $p(x|\theta) = 0.6$. Putem de asemenea scrie $p(x = 1|\theta = 0.6)$ pentru a evidenţia valorile lui $x$ şi $\theta$, dar niciodată nu vom scrie doar $p(1|0.6)$ deoarece este neclar care valoare este $x$ şi care este $\theta$.
Cu toate că încă vom folosi ambele tipuri de notaţii, de acum înainte vom folosi mai ales notaţia cu litere mici implicând pmf-uri şi pdf-uri. Ipotezele vor fi de obicei parametri reprezentaţi de litere greceşti $(\theta, \lambda, \mu, \sigma, ...)$ în timp ce valorile datelor vor fi de obicei reprezentate de litere româneşti $(x, x_i, y, ...)$.

## 2.5 Recapitulare rapidă a pdf şi probabilităţii

Presupunem că $X$ este o variabilă aleatoare cu pdf $f(x)$. Reamintim că $f(x)$ este o densitate; unităţile ei sunt probabilitate/(unităţile lui $x$).




Probabilitatea că valoarea lui $X$ este în $[c, d]$ este dată de

$$\int_c^d f(x)dx.$$

Probabilitatea că $X$ este într-un interval infinitezimal $dx$ în jurul lui $x$ este $f(x)dx$. De fapt, formula integrală este doar "suma" acestor probabilităţi infinitezimale. Putem vizualiza aceste probabilităţi văzând integrala ca aria de sub graficul lui $f(x)$. Pentru a manipula probabilităţi în loc de densităţi în cele ce urmează, vom utiliza frecvent noţiunea că $f(x)dx$ este probabilitatea că $X$ este într-un interval infinitezimal în jurul lui $x$ de lungime $dx$.

## 2.6 A priori continue, verosimilităţi discrete

În cadrul Bayesian avem probabilităţi ale ipotezelor - numite probabilităţi a priori şi a posteriori - şi probabilităţi ale datelor cunoscând o ipoteză - numite verosimilităţi. În cursurile precedente atât ipotezele cât şi datele aveau domenii discrete de valori. Am văzut în introducere că putem avea un domeniu continuu de ipoteze. Acelaşi lucru este valabil şi pentru date, dar pentru acum vom presupune că datele noastre pot lua doar o mulţime discretă de valori. În acest caz, verosimilitatea datelor $x$ cunoscând ipoteza $\theta$ este scrisă folosind o pmf: $p(x|\theta)$.

Vom utiliza următorul exemplu cu monedă pentru a explica aceste noţiuni.

**Exemplul 4.** Presupunem că avem o monedă cu probabilitatea aversurilor necunoscută $\theta$. Valoarea lui $\theta$ este aleatoare şi poate fi oriunde între 0 şi 1. Pentru acest exemplu şi pentru următoarele vom presupune că valoarea lui $\theta$ are o repartiţie cu densitatea de probabilitate a priori continuă $f(\theta) = 2\theta$. Avem o verosimilitate discretă deoarece aruncarea monedei are doar 2 rezultate, $x = 1$ pentru avers şi $x = 0$ pentru revers.

$$p(x = 1|\theta) = \theta, \; p(x = 0|\theta) = 1 - \theta.$$

**Gândiţi:** Aceasta poate fi dificil de înţeles. Avem o monedă cu o probabilitate necunoscută $\theta$ a aversului. Valoarea parametrului $\theta$ este ea însăşi aleatoare şi are o pdf a priori $f(\theta)$. Poate ajuta să vedem că exemplele discrete din cursurile anterioare sunt similare. De exemplu, am avut o monedă care putea avea probabilitatea aversului 0.5, 0.6 sau 0.9. Deci, puteam numi ipotezele noastre $H_{0.5}, H_{0.6}, H_{0.9}$ şi acestea aveau probabilităţile a priori $P(H_{0.5})$, etc. Cu alte cuvinte, aveam o monedă cu o probabilitate necunoscută a aversurilor, aveam ipoteze despre acea probabilitate şi fiecare dintre acele ipoteze aveau o probabilitate a priori.

## 2.7 Legea probabilităţii totale

Legea probabilităţii totale pentru repartiţii de probabilitate continue este în esenţă aceeaşi ca pentru repartiţii discrete. Înlocuim pmf a priori cu o pdf a priori şi suma cu o integrală. Începem prin a recapitula legea pentru cazul discret.

Reamintim că pentru o mulţime discretă de ipoteze $\{\mathcal{H}_1, \mathcal{H}_2, ..., \mathcal{H}_n\}$ legea probabilităţii totale spune

$$P(\mathcal{D}) = \sum_{i=1}^{n} P(\mathcal{D}|\mathcal{H}_i)P(\mathcal{H}_i). \tag{1}$$

Aceasta este probabilitatea a priori totală a lui $\mathcal{D}$ deoarece am folosit probabilităţile a priori $P(\mathcal{H}_i)$.

În notaţiile cu litere mici cu $\theta_1, \theta_2, ..., \theta_n$ pentru ipoteze şi $x$ pentru date, legea probabilităţii totale este scrisă

$$p(x) = \sum_{i=1}^{n} p(x|\theta_i)p(\theta_i). \tag{2}$$

De asemenea am numit aceasta probabilitatea predictivă a priori a rezultatului $x$ pentru a o distinge de probabilitatea a priori a ipotezei $\theta$.
Analog, este o lege a probabilităţii totale pentru pdf contiunue. O formulăm ca o teoremă folosind notaţia cu litere mici.
**Teoremă.** Legea probabilităţii totale. Presupunem că avem un parametru continuu $\theta$ în domeniul $[a, b]$ şi datele discrete aleatoare $x$. Presupunem că $\theta$ este el însuşi aleator cu densitatea $f(\theta)$ şi că $x$ şi $\theta$ au verosimiliatea $p(x|\theta)$. În acest caz, probabilitatea totală a lui $x$ este dată de formula

$$p(x) = \int_a^b p(x|\theta) f(\theta) d\theta. \tag{3}$$

**Demonstraţie.** Demonstraţia noastră va fi analoagă versiunii discrete: termenul de probabilitate $p(x|\theta)f(\theta)d\theta$ este perfect analog termenului $p(x|\theta_i)p(\theta_i)$ din relaţia 2 (sau termenului $P(\mathcal{D}|\mathcal{H}_i)P(\mathcal{H}_i)$ din relaţia 1). Continuând analogia: suma din relaţia 2 devine integrala din relaţia 3.
Ca în cazul discret, când gândim $\theta$ ca o ipoteză care explică probabilitatea datelor, numim $p(x)$ probabilitatea predictivă a priori pentru $x$.
**Exemplul 5.** Continuarea exemplului 4. Avem o monedă cu probabilitatea $\theta$ a aversului. Valoarea lui $\theta$ este aleatoare cu pdf a priori $f(\theta) = 2\theta$ pe [0,1]. Presupunem că aruncăm moneda o dată. Care este probabilitatea totală a aversului?
**Răspuns:** În exemplul 4 am observat că verosimilităţile sunt $p(x = 1|\theta) = \theta$ şi $p(x = 0|\theta) = 1 - \theta$. Probabilitatea totală a lui $x = 1$ este

$$p(x = 1) = \int_0^1 p(x = 1|\theta) f(\theta) d\theta = \int_0^1 \theta \cdot 2\theta d\theta = \int_0^1 2\theta^2 d\theta = \left. \frac{2\theta^3}{3} \right|_0^1 = \frac{2}{3}.$$

## 2.8 Teorema lui Bayes pentru densităţi de probabilitate continue

Enunţul teoremei lui Bayes pentru pdf-uri continue este în esenţă identic cu cel pentru pmf-uri. O enunţăm incluzând $d\theta$ pentru a avea probabilităţi autentice:
**Teoremă.** Teorema lui Bayes. Folosim aceleaşi presupuneri ca în legea probabilităţii totale, i.e., $\theta$ este un parametru continuu cu pdf $f(\theta)$ şi domeniul $[a, b]$; $x$ sunt datele discrete aleatoare; împreună au verosimilitatea $p(x|\theta)$. Cu aceste presupuneri:

$$f(\theta|x) d\theta = \frac{p(x|\theta) f(\theta) d\theta}{p(x)} = \frac{p(x|\theta) f(\theta) d\theta}{\int_a^b p(x|\theta) f(\theta) d\theta}. \tag{4}$$

**Demonstraţie.** Fie $\Theta$ variabila aleatoare care produce valoarea $\theta$. Considerăm evenimentele

$$H = \text{"}\Theta \text{ este într-un interval de lungime } d\theta \text{ în jurul valorii } \theta\text{"}$$

şi

$$D = \text{" valoarea datelor este } x\text{"}.$$

Atunci $P(H) = f(\theta)d\theta,\ P(D) = p(x)$ şi $P(D|H) = p(x|\theta)$. Din forma uzuală a teoremei lui Bayes avem

$$f(\theta|x)d\theta = P(H|D) = \frac{P(D|H)P(H)}{P(D)} = \frac{p(x|\theta)f(\theta)d\theta}{p(x)}, \text{q.e.d.}$$

Păstrarea factorului $d\theta$ în enunţul teoremei lui Bayes este mai propice pentru a gândi în termeni de probabilităţi. Dar deoarece apare în toţi membrii relaţiei 4 se poate simplifica şi putem scrie teorema lui Bayes în termeni de densităţi astfel:

$$f(\theta|x) = \frac{p(x|\theta)f(\theta)}{p(x)} = \frac{p(x|\theta)f(\theta)}{\int_a^b p(x|\theta)f(\theta)d\theta}.$$

## 2.9 Actualizare Bayesiană cu a priori continue

Caracteristici ale tabelului de actualizare Bayesiană pentru a priori continue:
1. Tabelul pentru a priori continue este foarte simplu: deoarece nu putem avea o linie pentru fiecare dintr-un număr infinit de ipoteze, vom avea doar o linie care foloseşte o variabilă pentru toate ipotezele $\theta$.
2. Incluzând $d\theta$, toate apariţiile din tabel sunt probabilităţi şi se aplică toate regulile de probabilităţi obişnuite.
**Exemplul 6.** (Actualizare Bayesiană.) Continuarea exemplelor 4 şi 5. Avem o monedă cu probabilitate necunoscută $\theta$ a aversului. Valoarea lui $\theta$ este aleatoare cu pdf a priori $f(\theta) = 2\theta$. Presupunem că aruncăm moneda o dată şi obţinem avers. Calculaţi pdf a posteriori pentru $\theta$.
**Răspuns:** Facem un tabel de actualizare cu coloanele obişnuite. Deoarece acesta este primul nostru exemplu, prima linie este versiunea abstractă a actualizării Bayesiene în general şi a 2-a linie este actualizarea Bayesiană pentru acest exemplu particular. "hypothesis" = "ipoteză", "prior" = "a priori", "likelihood" = "verosimilitate", "Bayes numerator" = "numărător Bayes", "posterior" = "a posteriori".

| hypothesis | prior | likelihood | Bayes numerator | posterior |
|---|---|---|---|---|
| $\theta$ | $f(\theta)\,d\theta$ | $p(x=1\|\theta)$ | $p(x=1\|\theta)f(\theta)\,d\theta$ | $f(\theta\|x=1)\,d\theta$ |
| $\theta$ | $2\theta\,d\theta$ | $\theta$ | $2\theta^2\,d\theta$ | $3\theta^2\,d\theta$ |
| total | $\int_a^b f(\theta)\,d\theta = 1$ | | $p(x=1)=\int_0^1 2\theta^2\,d\theta = 2/3$ | 1 |

De aceea pdf a posteriori (după vederea unui avers) este $f(\theta|x) = 3\theta^2$.
Comentarii:
1. Deoarece am folosit probabilitatea a priori $f(\theta)d\theta$, ipoteza ar fi trebuit să fie:
"parametrul necunoscut este într-un interval de lungime $d\theta$ în jurul lui $\theta$".
Chiar dacă este prea mult de scris pentru noi, asta trebuie să gândim de fiecare dată când scriem că ipoteza este $\theta$.
2. Pdf a posteriori pentru $\theta$ se află înlăturând $d\theta$ din probabilitatea a posteriori din tabel.

$$f(\theta|x) = 3\theta^2.$$

3. i) $p(x)$ este probabilitatea totală. Deoarece avem o repartiţie continuă, în loc de o sumă calculăm o integrală.
ii) Incluzând $d\theta$ în tabel, este clar ce integrală avem nevoie să calculăm pentru a afla probabilitatea totală $p(x)$.
4. Tabelul organizează versiunea continuă a teoremei lui Bayes. Anume, pdf a posteriori este legată de pdf a priori şi verosimilitate prin

$$f(\theta|x)d\theta = \frac{p(x|\theta)f(\theta)d\theta}{\int_a^b p(x|\theta)f(\theta)d\theta} = \frac{p(x|\theta)f(\theta)d\theta}{p(x)}.$$

Împărţind la $d\theta$ avem enunţul în termeni de densităţi.
5. Putem exprima teorema lui Bayes în forma:

$$f(\theta|x) \propto p(x|\theta)\cdot f(\theta)$$

$$\text{a posteriori} \propto \text{verosimilitatea} \times \text{a priori}.$$

### 2.9.1 A priori plate

O a priori plată sau uniformă presupune că fiecare ipoteză este egal probabilă. De exemplu, dacă $\theta$ are domeniul [0,1], atunci $f(\theta) = 1$ este o a priori plată.
**Exemplul 7.** (A priori plate.) Avem o monedă cu probabilitatea necunoscută $\theta$ a aversului. Presupunem că o aruncăm o dată şi obţinem revers. Presupunând o a priori plată, aflaţi probabilitatea a posteriori pentru $\theta$.

**Răspuns:** Procedăm ca la exemplul 6, cu a priori şi verosimilitatea modificate.

| hypothesis $\theta$ | prior $f(\theta)\,d\theta$ | likelihood $p(x=0\|\theta)$ | Bayes numerator | posterior $f(\theta\|x=0)\,d\theta$ |
|---|---|---|---|---|
| $\theta$ | $1\cdot d\theta$ | $1-\theta$ | $(1-\theta)\,d\theta$ | $2(1-\theta)\,d\theta$ |
| total | $\int_a^b f(\theta)\,d\theta = 1$ | | $p(x=0)=\int_0^1 (1-\theta)\,d\theta = 1/2$ | 1 |

### 2.9.2 Folosirea pdf a posteriori

**Exemplul 8.** În exemplul anterior probabilitatea a priori a fost plată. Întâi arătaţi că aceasta înseamnă că a priori moneda este egal părtinitoare spre avers sau revers. Apoi, după observarea unui avers, care este probabilitatea (a posteriori) ca moneda să fie părtinitoare spre avers?
**Răspuns.** Deoarece parametrul $\theta$ este probabilitatea aversului, întâi ni se cere să arătăm că $P(\theta > 0.5) = 0.5$, apoi ni se cere să calculăm
$P(\theta > 0.5|x = 1)$. Acestea se calculează din pdf-urile a priori, respectiv a posteriori.
Probabilitatea a priori ca moneda să fie părtinitoare spre avers este

$$P(\theta > 0.5) = \int_{0.5}^{1} f(\theta)d\theta = \int_{0.5}^{1} 1d\theta = \theta|_{0.5}^{1} = \frac{1}{2}.$$

Probabilitatea de 1/2 înseamnă că moneda este egal părtinitoare spre avers sau revers.
Ca în exemplul 7 se calculează că $f(\theta|x = 1) = 2\theta$.
Probabilitatea a posteriori că moneda este părtinitoare spre avers este

$$P(\theta > 0.5|x = 1) = \int_{0.5}^{1} f(\theta|x = 1)d\theta = \int_{0.5}^{1} 2\theta d\theta = \theta^2|_{0.5}^{1} = \frac{3}{4}.$$

Vedem că observarea unui avers a crescut probabilitatea ca moneda să fie părtinitoare spre avers de la 1/2 la 3/4.

## 2.10 Probabilităţi predictive

La fel ca în cazul discret, suntem de asemenea interesaţi să folosim probabilităţile a posteriori ale ipotezelor pentru a face predicţii pentru ce se va întâmpla mai departe.
**Exemplul 9.** (Predicţie a priori şi a posteriori.) Continuarea exemplelor 4, 5, 6: avem o monedă cu probabilitate necunoscută $\theta$ a aversului şi valoarea

lui $\theta$ are pdf a priori $f(\theta) = 2\theta$. Aflaţi probabilitatea predictivă a priori a aversului. Apoi presupunând că prima aruncare a fost avers, aflaţi probabilităţile predictive a posteriori atât pentru avers cât şi pentru revers la a 2-a aruncare.
**Răspuns:** Fie $x_1$ rezultatul primei aruncări şi $x_2$ al celei de-a 2-a. Probabilitatea predictivă a priori este exact probabilitatea totală calculată în exemplele 5 şi 6.

$$p(x=1) = \int_0^1 p(x=1|\theta)f(\theta)d\theta = \int_0^1 \theta \cdot 2\theta d\theta = \int_0^1 2\theta^2 d\theta = \left.\frac{2\theta^3}{3}\right|_0^1 = \frac{2}{3}.$$

Probabilităţile predictive a posteriori sunt probabilităţile totale calculate folosind pdf a posteriori. Din exemplul 6 ştim că pdf a posteriori este $f(\theta|x_1 = 1) = 3\theta^2$. Probabilităţile predictive a posteriori sunt

$$p(x_2 = 1|x_1 = 1) = \int_0^1 p(x_2 = 1|\theta, x_1 = 1)f(\theta|x_1 = 1)d\theta = \int_0^1 \theta \cdot 3\theta^2 d\theta$$
$$= \left.\frac{3\theta^4}{4}\right|_0^1 = \frac{3}{4},$$
$$p(x_2 = 0|x_1 = 1) = \int_0^1 p(x_2 = 0|\theta, x_1 = 1)f(\theta|x_1 = 1)d\theta = \int_0^1 (1-\theta) \cdot 3\theta^2 d\theta$$
$$= \left.\left(\theta^3 - \frac{3\theta^4}{4}\right)\right|_0^1 = \frac{1}{4}.$$

(Mai simplu, puteam calcula
$p(x_2 = 0|x_1 = 1) = 1 - p(x_2 = 1|x_1 = 1) = 1 - 3/4 = 1/4$.)

## 2.11 De la actualizarea Bayesiană discretă la actualizarea Bayesiană continuă

Pentru a dezvolta intuiţia pentru tranziţia de la actualizarea Bayesiană discretă la actualizarea Bayesiană continuă, vom:
i) aproxima domeniul continuu de ipoteze printr-un număr finit de ipoteze;
ii) crea tabelul de actualizare discretă pentru numărul finit de ipoteze;
iii) considera cum se schimbă tabelul când numărul de ipoteze tinde la infinit.
În acest fel, vom vedea că pmf-urile a priori şi a posteriori converg la pdf-urile a priori şi a posteriori.
**Exemplul 10.** Pentru a păstra lucrurile concrete, vom lucra cu moneda cu o a priori plată $f(\theta) = 1$ din exemplul 7. Scopul nostru este să mergem de la discret la continuu crescând numărul de ipoteze.

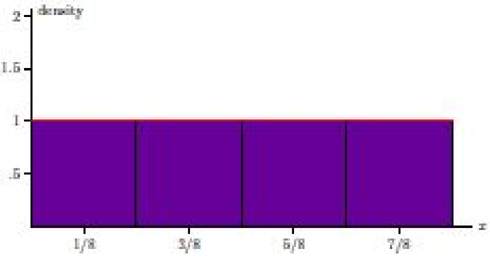
**4 ipoteze.** Împărţim [0,1] în 4 intervale egale: [0,1/4], [1/4,1/2] [1/2,3/4], [3/4,1]. Fiecare interval are lungimea $\Delta\theta = 1/4$. Punem cele 4 ipoteze ale noastre $\theta_i$ în centrele celor 4 intervale:

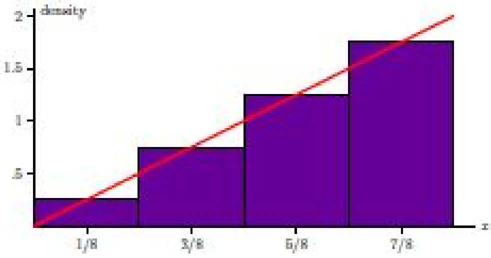
$$\theta_1 : "\theta = 1/8", \ \theta_2 : "\theta = 3/8, \ \theta_3 : "\theta = 5/8", \ \theta_4 : "\theta = 7/8".$$

A priori plată dă fiecărei ipoteze o probabilitate de $1/4 = 1 \cdot \Delta\theta$. Avem tabelul:

| hypothesis | prior | likelihood | Bayes num. | posterior |
|---|---|---|---|---|
| $\theta = 1/8$ | $1/4$ | $1/8$ | $(1/4) \times (1/8)$ | $1/16$ |
| $\theta = 3/8$ | $1/4$ | $3/8$ | $(1/4) \times (3/8)$ | $3/16$ |
| $\theta = 5/8$ | $1/4$ | $5/8$ | $(1/4) \times (5/8)$ | $5/16$ |
| $\theta = 7/8$ | $1/4$ | $7/8$ | $(1/4) \times (7/8)$ | $7/16$ |
| Total | 1 | – | $\sum_{i=1}^{n} \theta_i \Delta\theta$ | 1 |

Iată histogramele de densitate ale pmf-urilor a priori şi a posteriori. Pdf-urile a priori şi a posteriori din exemplul 7 sunt trasate pe histograme cu roşu.



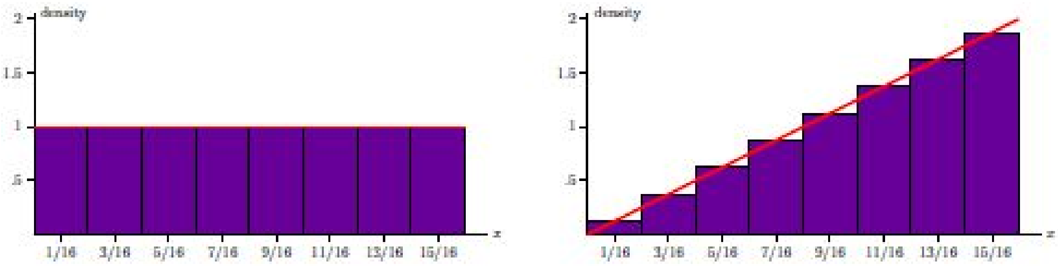




**8 ipoteze.** Acum împărţim [0,1] în 8 intervale, fiecare cu lungimea $\Delta\theta = 1/8$ şi punem cele 8 ipoteze ale noastre $\theta_i$ în centrele celor 4 intervale:

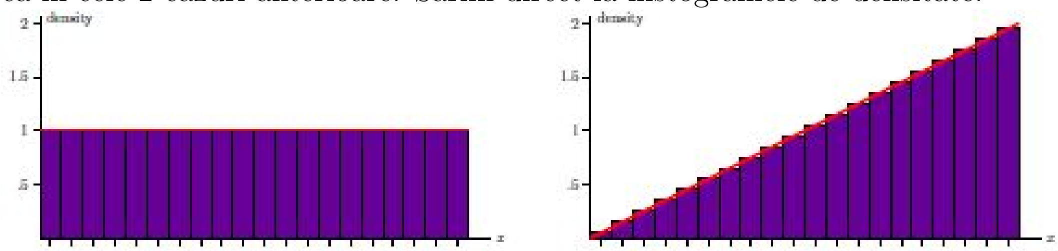
$$\theta_1 : "\theta = 1/16", \ \theta_2 : "\theta = 3/16, \ \theta_3 : "\theta = 5/16", \ \theta_4 : "\theta = 7/16",$$
$$\theta_5 : "\theta = 9/16", \ \theta_6 : "\theta = 11/16, \ \theta_7 : "\theta = 13/16", \ \theta_8 : "\theta = 15/16".$$

A priori plată dă fiecărei ipoteze o probabilitate de $1/8 = 1 \cdot \Delta\theta$. Iată tabelul şi histogramele de densitate:

| hypothesis | prior | likelihood | Bayes num. | posterior |
|---|---|---|---|---|
| $\theta = 1/16$ | $1/8$ | $1/16$ | $(1/8) \times (1/16)$ | $1/64$ |
| $\theta = 3/16$ | $1/8$ | $3/16$ | $(1/8) \times (3/16)$ | $3/64$ |
| $\theta = 5/16$ | $1/8$ | $5/16$ | $(1/8) \times (5/16)$ | $5/64$ |
| $\theta = 7/16$ | $1/8$ | $7/16$ | $(1/8) \times (7/16)$ | $7/64$ |
| $\theta = 9/16$ | $1/8$ | $9/16$ | $(1/8) \times (9/16)$ | $9/64$ |
| $\theta = 11/16$ | $1/8$ | $11/16$ | $(1/8) \times (11/16)$ | $11/64$ |
| $\theta = 13/16$ | $1/8$ | $13/16$ | $(1/8) \times (13/16)$ | $13/64$ |
| $\theta = 15/16$ | $1/8$ | $15/16$ | $(1/8) \times (15/16)$ | $15/64$ |
| Total | 1 | – | $\sum_{i=1}^{n} \theta_i \, \Delta\theta$ | 1 |




**20 ipoteze.** Acum împărţim [0,1] în 20 de intervale. Tabelul se face analog ca în cele 2 cazuri anterioare. Sărim direct la histogramele de densitate.




Privind la reprezentări, vedem cum histogramele de densitate a priori şi a posteriori converg la funcţiile densitate de probabilitate a priori şi a posteriori.

# 3 Convenţii de notaţie

## 3.1 Scopul învăţării

Să poată lucra cu notaţiile şi termenii pe care îi folosim pentru a descrie probabilităţile şi verosimilitatea.

## 3.2 Introducere

Am introdus un număr de notaţii diferite pentru probabilitate, ipoteze şi date. Le adunăm aici, pentru a le avea într-un singur loc.

## 3.3 Notaţii şi terminologie pentru date şi ipoteze

Când am început cursul am vorbit despre rezultate (sau cazuri), de exemplu avers sau revers. Apoi, când am introdus variabilele aleatoare am dat rezultatelor valori numerice, de exemplu 1 pentru avers şi 0 pentru revers. Aceasta ne-a permis să facem lucruri precum calculul mediilor şi dispersiilor.
Reamintim convenţiile noastre de notaţie:
Evenimentele sunt notate cu majuscule, de exemplu $A, B, C$.
O variabilă aleatoare este majuscula $X$ şi ia valorile $x$.
Legătura dintre valori şi evenimente: "$X = x$" este evenimentul că $X$ ia valoarea $x$.
Probabilitatea unui eveniment este majuscula $P(A)$.
O variabilă aleatoare discretă are funcţia masă de probabilitate $p(x)$. Legătura dintre $P$ şi $p$ este că $P(X = x) = p(x)$.
O variabilă aleatoare continuă are funcţia densitate de probabilitate $f(x)$. Legătura dintre $P$ şi $f$ este că $P(a \le X \le b) = \int_a^b f(x)dx$.
Pentru o variabilă aleatoare continuă $X$, probabilitatea că $X$ este într-un interval infinitezimal de lungime $dx$ în jurul lui $x$ este $f(x)dx$.

În contextul actualizării Bayesiene avem convenţii similare.
Folosim litere mari caligrafice (de mână), în special $\mathcal{H}$, pentru a indica o ipoteză, de exemplu $\mathcal{H} =$ "moneda este corectă".
Folosim litere mici greceşti, în special $\theta$, pentru a indica valoarea ipotetică a unui parametru al modelului, de exemplu probabilitatea ca moneda să aterizeze avers este $\theta = 0.5$.
Folosim litere mari caligrafice (de mână), în special $\mathcal{D}$, când vorbim despre date ca evenimente. De exemplu, $\mathcal{D} =$ "secvenţa de aruncări a fost HTH".
Folosim litere mici, în special $x$, când vorbim despre date ca valori. De exemplu, secvenţa de date a fost $x_1, x_2, x_3 = 1, 0, 1$.

Când mulţimea de ipoteze este discretă, putem folosi probabilitatea ipotezelor individuale, de exemplu $p(\theta)$. Când mulţimea este continuă, trebuie să folosim probabilitatea pentru un domeniu infinitezimal de ipoteze, de exemplu $f(\theta)d\theta$.
Următorul tabel rezumă aceasta pentru $\theta$ discretă şi $\theta$ continuă. În ambele cazuri presupunem o mulţime discretă de rezultate posibile (date) $x$.

| | hypothesis | prior | likelihood | Bayes numerator | posterior |
|---|---|---|---|---|---|
| | $\mathcal{H}$ | $P(\mathcal{H})$ | $P(\mathcal{D}\|\mathcal{H})$ | $P(\mathcal{D}\|\mathcal{H})P(\mathcal{H})$ | $P(\mathcal{H}\|\mathcal{D})$ |
| Discrete $\theta$: | $\theta$ | $p(\theta)$ | $p(x\|\theta)$ | $p(x\|\theta)p(\theta)$ | $p(\theta\|x)$ |
| Continuous $\theta$: | $\theta$ | $f(\theta)\,d\theta$ | $p(x\|\theta)$ | $p(x\|\theta)f(\theta)\,d\theta$ | $f(\theta\|x)\,d\theta$ |

Reamintim că ipoteza continuă $\theta$ este în realitate o prescurtare pentru "parametrul $\theta$ este într-un interval de lungime $d\theta$ în jurul lui $\theta$".