CAPITOLUL 2

ENTROPIE ŞI INFORMAŢIE

2.1. SURSE DE INFORMAŢIE

În această carte, ne ocupăm cu definiţia şi tratarea matematică a informaţiei. Aceasta ne permite să construim sisteme de comunucaţie cu ajutorul cărora luăm legătura, confortabil şi eficient, cu realităţi care se desfăşoară la sute şi poate mii de kilometri distanţă: să ne găndim la telefon, la radio, la Internet. Spunem că transmitem informaţie. Dar ce transmitem concret? Dacă analizăm tipurile de mesaje pe care simţim nevoia să le emitem şi să le recepţionăm, vom constata că ele se reduc la următoarele:

- voce, vorbire (mesaj verbal)
- muzică (sunet ca şi vocea, dar de bandă de frecvenţe mai largă)
- imagine, statică şi dinamică (mesaj vizual)
- semne şi simboluri, reprezentând tot un mesaj vizual, dar abstract.

Vedem că, din cele cinci simţuri, numai două, auzul şi văzul, sunt implicate în aceasta. Ca şi animalele, noi primim desigur informaţie şi prin alte simţuri (miros şi gust, la care putem avea acces în comun), dar dacă vrem să împărtăşim această informaţie cuiva, ne vom folosi de aparatul nostru fonator, vom gesticula, vom desena ideograme sau vom scrie. S-ar părea că simţul tactil este şi mai intim, dar să nu uităm că semenii noştri nevăzători citesc şi scriu (cu o maşină de dactilografiat specială) utilizând alfabetul Braille constituit din semne în relief.

În vederea înregistrării, prelucrării şi transmiterii la distanţă, sunetele, imaginile şi simbolurile zise alfanumerice (litere, numere, semne diacritice şi altele) sunt convertite în semnale electrice, electromagnetice sau optice. Atunci când vorbim la telefon, undele sonore pe care le produce organul nostru fonator sunt transformate de microfon într-un curent electric variabil care este transmis pe un circuit electric la centrala telefonică şi de acolo mai departe până la telefonul interlocutorului. Un dispozitiv care converteşte o mărime neelectrică într-una electrică sau invers se numeşte *traductor*. Microfonul este un exemplu de traductor care converteşte o undă sonoră într-un curent electric variabil; casca telefonică efectuează operația inversă. Din acest motiv, prin sursă de informație vom înțelege orice aparat care, incluzând dacă este cazul şi traductoare, poate genera mesaje în formă electrică – cel mai adesea, un curent sau o tensiune. Exemple de surse de informație: un microfon, doza unui pick-up, capul de citire al unui casetofon, o cameră de luat vederi, interfața serială a unui calculator electronic.

Prin semnal, înțelegem o mărime fizică a cărei variație în timp poate transporta mesaje. Dacă nu se precizează altfel, această mărime fizică este o tensiune electrică, măsurată în general între un punct de circuit numit „bornă caldă“ şi un punct de referință, numit „masă electrică“ sau „pământ“. Semnalele sunt *analogice* dacă sunt funcții continue de timp şi *discrete* sau *digitale* dacă pot lua numai un număr finit de valori, iar o anumită valoare este menținută constantă pe un interval elementar de timp. Denumirea de *analogic* are o explicație istorică: în epoca de pionierat a calculatoarelor electronice, cele zise analogice realizau diverse operații matematice cu ajutorul unor circuite electrice *R, L, C* a căror funcționare, după cum se ştie, este descrisă de ecuații integro-diferențiale şi astfel, toate mărimile implicate erau funcții continue de timp, iar funcționarea avea loc prin analogie cu funcționarea unor circuite electrice liniare. Între timp, calculatoarele digitale s-au impus, iar din cele analogice n-a supraviețuit decât denumirea de *analogic*, drept sinonim pentru continuu.

Sursele de informație se clasifică în *analogice* (care generează voce, muzică, imagini fixe şi în mişcare) şi *discrete* (care generează fişiere de calculator, şiruri de simboluri numerice sau alfanumerice). Conform teoremei eşantionării, care se studiază la cursul de „Semnale, circuite şi sisteme“, în general un semnal analogic poate fi transformat într-un semnal digital din care semnalul analogic originar poate fi recuperat, în limitele unei distorsiuni acceptabile. Având în vedere şi nivelul foarte ridicat la care a ajuns producția de circuite integrate digitale şi a celor de conversie (analogic-digitală şi digital-analogică), în prezent, indiferent de forma lor la sursa primară, fie ea analogică sau digitală, semnalele sunt tratate unificat, prin tehnici pur numerice. Este ceea ce se numeşte *multimedia*.

2.2. ENTROPIE

Entropia este o măsură a incertitudinii cu privire la o variabilă aleatoare. Termenul de *entropie* a fost împrumutat din termodinamică, o ramură a fizicii în care fenomenele studiate sunt preponderent probabiliste. Pentru un observator exterior, semnalul de ieşire generat de o sursă de informaţie la un moment dat este o variabilă aleatoare. O sursă de informaţie discretă emite un şir de litere dintr-un alfabet de L litere posibile, să spunem $\{x_1, x_2, \cdots, x_L\}$. Presupunem că fiecare literă din alfabetul $\{x_1, x_2, \cdots, x_L\}$ are o probabilitate de apariţie p_k dată:

$$p_k = P(X = x_k),\ 1 \le k \le L \tag{2.1}$$

unde

$$\sum_{k=1}^{L} p_k = 1 .$$

Entropia unei variabile aleatoare X este prin definiţie mărimea

$$H(X) = -\sum_{k=1}^{L} p_k \log_2 p_k \tag{2.2}$$

Se observă că baza logaritmului este 2, ceea ce face ca entropia să se exprime în biţi. Spre exemplu, în cazul aruncării unei monede nemăsluite, $p(cap) = p(ban) = \frac{1}{2}$ şi $H(X) = 1$ bit. Entropia se poate exprima şi într-o altă bază de logaritmi, spre exemplu numărul transcendent e sau numărul 10, în care caz unitatea de măsură este *nat* (de la natural) şi, respectiv, *hartley* (denumită astfel în onoarea lui R.V. Hartley, cel care, în 1928, a sugerat utilizarea logaritmului drept măsură a informaţiei). Mai observăm şi că entropia nu depinde de valorile luate de variabila aleatoare X, ci doar de probabilităţile cu care sunt luate aceste valori.

Ştim că, în teoria funcţiilor de o variabilă reală, funcţia logaritm se defineşte numai pe mulţimi de numere strict pozitive. Având însă în vedere că $x \log x \to 0$ pentru $x \to 0$, vom adopta convenţia că $0 \log 0 = 0$. Prin urmare, adăugarea unor termeni de probabilitate zero nu schimbă entropia.

Putem interpreta entropia lui X drept *expectaţia* (valoarea medie statistică, sau valoarea aşteptată) funcţiei $\log_2 \frac{1}{p(x)}$:

$$H(X) = E\left\{\log_2 \frac{1}{p(x)}\right\} \tag{2.3}$$

Entropia este întotdeauna pozitivă:

$$H(X) \geq 0 \tag{2.4}$$

Într-adevăr, întrucât $0 \leq p_k \leq 1$ pentru $1 \leq k \leq L$, urmează că $\log_2(1/p_k) \geq 0$.

Fie o variabilă aleatoare *X* care nu ia decât două valori:

$$X = \begin{cases} 1 & \text{cu probabilitate } p \\ 0 & \text{cu probabilitate } 1-p \end{cases} \tag{2.5}$$

Conform relației de definiție (2.2), avem:

$$H(X) = -p\log_2 p - (1-p)\log_2(1-p) \tag{2.6}$$

Prin definiție, aceasta se mai notează şi *H*(*p*). Graficul funcției *H*(*p*) este arătat în figura 2.1.

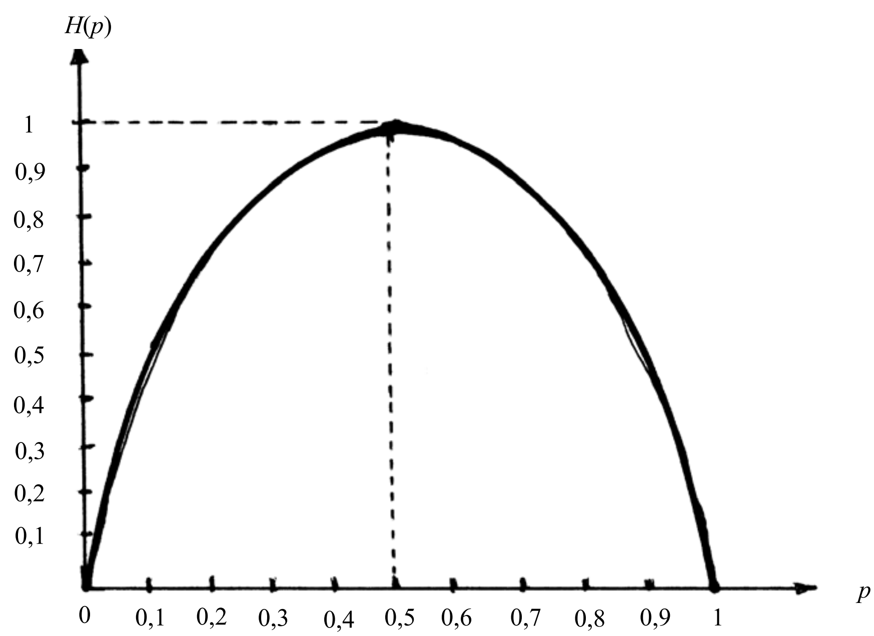

Fig. 2.1. Graficul funcției entropie $H(p)$.

Se observă că entropia este o funcţie concavă de p şi este egală cu 0 pentru p = 0 sau 1. Pentru aceste valori, variabila nu este aleatoare şi nu există nici o incertitudine, încât este normal ca entropia să fie zero. Incertitudinea este maximă pentru $p=\frac{1}{2}$, ceea ce corespunde valorii maxime a entropiei.

2.3. ENTROPIE COMUNĂ ŞI ENTROPIE CONDIŢIONATĂ

Fie două variabile aleatoare discrete, X cu rezultatele posibile x_i, $i=1, 2,\cdots,n$ şi Y cu rezultatele posibile y_j, $j=1, 2,\cdots,m$. Considerăm variabila aleatoare (X, Y) care ia valori vectoriale (x_i, y_j) cu probabilitatea $p(x_i, y_j)$, ceea ce se scrie simbolic $(X,Y)\sim p(x, y)$. Prin definiţie, entropia comună a variabilelor aleatoare discrete X şi Y este entropia variabilei aleatoare discrete (X,Y):

$$H(X,Y)=-\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(x_i,y_j) \tag{2.7}$$

Mai putem scrie aceasta şi astfel:

$$H(X,Y)=-E\log_2 p(X,Y) \tag{2.8}$$

Prin definiţie, entropia condiţionată $H(Y\,|\,X)$ este entropia variabilei aleatoare Y condiţionată de realizarea variabilei aleatoare X:

$$\begin{aligned} H(Y\,|\,X) &= \sum_{i=1}^{n}p(x_i)H(Y\,|\,X=x_i)\\ &= -\sum_{i=1}^{n}p(x_i)\sum_{j=1}^{m}p(y_j\,|\,x_i)\log_2 p(y_j\,|\,x_i)\\ &= -\sum_{i=1}^{n}\sum_{j=1}^{m}p(x_i,y_j)\log_2 p(y_j\,|\,x_i)\\ &= -E_{p(x,y)}\log_2 p(Y\,|\,X) \end{aligned} \tag{2.9}$$

Modul în care am definit entropia comună şi entropia condiţionată ne permite să scriem că entropia unei perechi de variabile aleatoare este egală cu entropia uneia dintre ele plus entropia condiționată a celeilalte, ceea ce constituie enunțul teoremei următoare:

TEOREMA 2.1 (*Regula lanțului*):

$$H(X,Y) = H(X) + H(Y \mid X) \tag{2.10}$$

DEMONSTRAȚIE

$$\begin{aligned}
H(X,Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i, y_j) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i)p(y_j \mid x_i) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(y_j \mid x_i) \qquad (2.11) \\
&= -\sum_{i=1}^{n} p(x_i)\log_2 p(x_i) - \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j)\log_2 p(y_j \mid x_i) \\
&= H(X) + H(Y \mid X)
\end{aligned}$$

O altă demonstrație a acestei teoreme, mai elegantă, este de a scrie

$$\log_2 p(X,Y) = \log_2 p(X) + \log_2 p(Y \mid X) \tag{2.12}$$

Aplicând în ambii membri ai ecuației (2.12) operatorul de expectație *E*, obținem (2.10).

EXEMPLUL 2.1: Fie perechea de variabile aleatoare (*X, Y*) având distribuția comună de probabilitate

$Y \backslash X$	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Să calculăm mai întâi distribuția marginală a lui *X*. Avem:

$$p(x_1) = \sum_{j=1}^{4} p(x_1, y_j) = \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{4} = \frac{1}{2}$$

$$p(x_2) = \sum_{j=1}^{4} p(x_2, y_j) = \frac{1}{16} + \frac{1}{8} + \frac{1}{16} + 0 = \frac{1}{4}$$

$$p(x_3) = \sum_{j=1}^{4} p(x_3, y_j) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + 0 = \frac{1}{8}$$

$$p(x_4) = \sum_{j=1}^{4} p(x_4, y_j) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + 0 = \frac{1}{8}$$

Deci, distribuția marginală a lui X este $\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right)$.

Calculăm şi distribuția marginală a lui Y.

$$p(y_1) = \sum_{i=1}^{4} p(x_i, y_1) = \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4}$$

$$p(y_2) = \sum_{i=1}^{4} p(x_i, y_2) = \frac{1}{16} + \frac{1}{8} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4}$$

$$p(y_3) = \sum_{i=1}^{4} p(x_i, y_3) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$$

$$p(y_4) = \sum_{i=1}^{4} p(x_i, y_4) = \frac{1}{4} + 0 + 0 + 0 = \frac{1}{4}$$

Distribuția marginală a lui Y este deci $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right)$.

Acum putem calcula entropiile lui X şi Y astfel:

$$\begin{aligned} H(X) &= \sum_{i=1}^{4} p(x_i) \log_2 \frac{1}{p(x_i)} \\ &= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{8}\log_2 8 = \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4} \text{ biti} \end{aligned}$$

$$H(Y) = \sum_{j=1}^{4} p(y_j)\log_2 \frac{1}{p(y_j)} = 4 \cdot \frac{1}{4}\log_2 4 = 2 \text{ biti}$$

Calculăm probabilitățile condiționate $p(y_j \mid x_i)$ astfel:

$$p(y_1 \mid x_1) = \frac{p(x_1, y_1)}{p(x_1)} = \frac{\frac{1}{8}}{\frac{1}{2}} = \frac{1}{4} \qquad p(y_1 \mid x_2) = \frac{p(x_2, y_1)}{p(x_2)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}$$

$$p(y_1 \mid x_3) = \frac{p(x_3, y_1)}{p(x_3)} = \frac{\frac{1}{32}}{\frac{1}{8}} = \frac{1}{4} \qquad p(y_1 \mid x_4) = \frac{p(x_4, y_1)}{p(x_4)} = \frac{\frac{1}{32}}{\frac{1}{8}} = \frac{1}{4}$$

$$p(y_2 \mid x_1) = \frac{p(x_1, y_2)}{p(x_1)} = \frac{\frac{1}{16}}{\frac{1}{2}} = \frac{1}{8} \qquad p(y_2 \mid x_2) = \frac{p(x_2, y_2)}{p(x_2)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}$$

$$p(y_2 \mid x_3) = \frac{p(x_3, y_2)}{p(x_3)} = \frac{\frac{1}{32}}{\frac{1}{8}} = \frac{1}{4} \qquad p(y_2 \mid x_4) = \frac{p(x_4, y_2)}{p(x_4)} = \frac{\frac{1}{32}}{\frac{1}{8}} = \frac{1}{4}$$

$$p(y_3 \mid x_1) = \frac{p(x_1, y_3)}{p(x_1)} = \frac{\frac{1}{16}}{\frac{1}{2}} = \frac{1}{8} \qquad p(y_3 \mid x_2) = \frac{p(x_2, y_3)}{p(x_2)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}$$

$$p(y_3 \mid x_3) = \frac{p(x_3, y_3)}{p(x_3)} = \frac{\frac{1}{16}}{\frac{1}{8}} = \frac{1}{2} \qquad p(y_3 \mid x_4) = \frac{p(x_4, y_3)}{p(x_4)} = \frac{\frac{1}{16}}{\frac{1}{8}} = \frac{1}{2}$$

$$p(y_4 \mid x_1) = \frac{p(x_1, y_4)}{p(x_1)} = \frac{\frac{1}{4}}{\frac{1}{2}} = \frac{1}{2} \qquad p(y_4 \mid x_2) = \frac{p(x_2, y_4)}{p(x_2)} = 0$$

$$p(y_4 \mid x_3) = \frac{p(x_3, y_4)}{p(x_3)} = 0 \qquad p(y_4 \mid x_4) = \frac{p(x_4, y_4)}{p(x_4)} = 0$$

Putem acum calcula entropia condiționată $H(Y \mid X)$ conform definiției:

$$
\begin{aligned}
H(Y \mid X) &= \sum_{i=1}^{4}\sum_{j=1}^{4} p(x_i, y_j)\log_2 \frac{1}{p(y_j \mid x_i)} \\
&= p(x_1, y_1)\log_2 \frac{1}{p(y_1 \mid x_1)} + p(x_1, y_2)\log_2 \frac{1}{p(y_2 \mid x_1)} + \\
&+ p(x_1, y_3)\log_2 \frac{1}{p(y_3 \mid x_1)} + p(x_1, y_4)\log_2 \frac{1}{p(y_4 \mid x_1)} + \\
&+ p(x_2, y_1)\log_2 \frac{1}{p(y_1 \mid x_2)} + p(x_2, y_2)\log_2 \frac{1}{p(y_2 \mid x_2)} + \\
&+ p(x_2, y_3)\log_2 \frac{1}{p(y_3 \mid x_2)} + p(x_2, y_4)\log_2 \frac{1}{p(y_4 \mid x_2)} + \\
&+ p(x_3, y_1)\log_2 \frac{1}{p(y_1 \mid x_3)} + p(x_3, y_2)\log_2 \frac{1}{p(y_2 \mid x_3)} + \\
&+ p(x_3, y_3)\log_2 \frac{1}{p(y_3 \mid x_3)} + p(x_3, y_4)\log_2 \frac{1}{p(y_4 \mid x_3)} + \\
&+ p(x_4, y_1)\log_2 \frac{1}{p(y_1 \mid x_4)} + p(x_4, y_2)\log_2 \frac{1}{p(y_2 \mid x_4)} + \\
&+ p(x_4, y_3)\log_2 \frac{1}{p(y_3 \mid x_4)} + p(x_4, y_4)\log_2 \frac{1}{p(y_4 \mid x_4)} \\
&= \frac{1}{8}\log_2 4 + \frac{1}{16}\log_2 8 + \frac{1}{16}\log_2 8 + \frac{1}{4}\log_2 2 + \frac{1}{16}\log_2 4 + \\
&+ \frac{1}{8}\log_2 2 + \frac{1}{16}\log_2 4 + 0 + \frac{1}{32}\log_2 4 + \frac{1}{32}\log_2 4 + \frac{1}{16}\log_2 2 \\
&+ 0 + \frac{1}{32}\log_2 4 + \frac{1}{32}\log_2 4 + \frac{1}{16}\log_2 2 + 0 \\
&= \frac{1}{8}\cdot 2 + \frac{1}{16}\cdot 3 + \frac{1}{16}\cdot 3 + \frac{1}{4} + \frac{1}{16}\cdot 2 + \frac{1}{8} + \frac{1}{16}\cdot 2 + \frac{1}{32}\cdot 2 + \\
&+ \frac{1}{32}\cdot 2 + \frac{1}{16} + \frac{1}{32}\cdot 2 + \frac{1}{32}\cdot 2 + \frac{1}{16} = \frac{13}{8} \text{ biti.}
\end{aligned}
$$

Să calculăm şi probabilitățile condiționate $p(x_i \mid y_j)$:

$$p(x_1 \mid y_1) = \frac{p(x_1, y_1)}{p(y_1)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2} \qquad p(x_1 \mid y_2) = \frac{p(x_1, y_2)}{p(y_2)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4}$$

$$p(x_1 \mid y_3) = \frac{p(x_1, y_3)}{p(y_3)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4} \qquad p(x_1 \mid y_4) = \frac{p(x_1, y_4)}{p(y_4)} = \frac{\frac{1}{4}}{\frac{1}{4}} = 1$$

$$p(x_2 \mid y_1) = \frac{p(x_2, y_1)}{p(y_1)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4} \qquad p(x_2 \mid y_2) = \frac{p(x_2, y_2)}{p(y_2)} = \frac{\frac{1}{8}}{\frac{1}{4}} = \frac{1}{2}$$

$$p(x_2 \mid y_3) = \frac{p(x_2, y_3)}{p(y_3)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4} \qquad p(x_2 \mid y_4) = \frac{p(x_2, y_4)}{p(y_4)} = 0$$

$$p(x_3 \mid y_1) = \frac{p(x_3, y_1)}{p(y_1)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{8} \qquad p(x_3 \mid y_2) = \frac{p(x_3, y_2)}{p(y_2)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{8}$$

$$p(x_3 \mid y_3) = \frac{p(x_3, y_3)}{p(y_3)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4} \qquad p(x_3 \mid y_4) = \frac{p(x_3, y_4)}{p(y_4)} = 0$$

$$p(x_4 \mid y_1) = \frac{p(x_4, y_1)}{p(y_1)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{8} \qquad p(x_4 \mid y_2) = \frac{p(x_4, y_2)}{p(y_2)} = \frac{\frac{1}{32}}{\frac{1}{4}} = \frac{1}{8}$$

$$p(x_4 \mid y_3) = \frac{p(x_4, y_3)}{p(y_3)} = \frac{\frac{1}{16}}{\frac{1}{4}} = \frac{1}{4} \qquad p(x_4 \mid y_4) = \frac{p(x_4, y_4)}{p(y_4)} = 0$$

Cu ajutorul acestor probabilităţi, calculăm entropia condiţionată $H(X \mid Y)$:

$$H(X|Y) = -\sum_{i=1}^{4}\sum_{j=1}^{4} p(x_i, y_j)\log_2 p(x_i, y_j) =$$

$$= p(x_1, y_1)\log_2 \frac{1}{p(x_1 | y_1)} + p(x_1, y_2)\log_2 \frac{1}{p(x_1 | y_2)} +$$

$$+ p(x_1, y_3)\log_2 \frac{1}{p(x_1 | y_3)} + p(x_1, y_4)\log_2 \frac{1}{p(x_1, y_4)} +$$

$$+ p(x_2, y_1)\log_2 \frac{1}{p(x_2 | y_1)} + p(x_2, y_2)\log_2 \frac{1}{p(x_2 | y_2)} +$$

$$+ p(x_2, y_3)\log_2 \frac{1}{p(x_2 | y_3)} + p(x_2, y_4)\log_2 \frac{1}{p(x_2 | y_4)} +$$

$$+ p(x_3, y_1)\log_2 \frac{1}{p(x_3 | y_1)} + p(x_3, y_2)\log_2 \frac{1}{p(x_3 | y_2)} +$$

$$+ p(x_3, y_3)\log_2 \frac{1}{p(x_3 | y_3)} + p(x_3, y_4)\log_2 \frac{1}{p(x_3 | y_4)} +$$

$$+ p(x_4, y_1)\log_2 \frac{1}{p(x_4 | y_1)} + p(x_4, y_2)\log_2 \frac{1}{p(x_4 | y_2)} +$$

$$+ p(x_4, y_3)\log_2 \frac{1}{p(x_4 | y_3)} + p(x_4, y_4)\log_2 \frac{1}{p(x_4 | y_4)} =$$

$$= \frac{1}{8}\cdot\log_2 2 + \frac{1}{16}\cdot\log_2 4 + \frac{1}{16}\cdot\log_2 4 + \frac{1}{4}\cdot\log_2 1 + \frac{1}{16}\cdot\log_2 4 +$$

$$+ \frac{1}{8}\cdot\log_2 2 + \frac{1}{16}\cdot\log_2 4 + 0 + \frac{1}{32}\cdot\log_2 8 + \frac{1}{32}\cdot\log_2 8 +$$

$$+ \frac{1}{16}\cdot\log_2 4 + 0 + \frac{1}{32}\cdot\log_2 8 + \frac{1}{32}\cdot\log_2 8 + \frac{1}{16}\cdot\log_2 4 + 0 =$$

$$= \frac{11}{8} \text{ biti.}$$

Să calculăm şi entropia comună pe baza definiţiei:

$$H(X, Y) = \sum_{i=1}^{4}\sum_{j=1}^{4} p(x_i, y_j)\log_2 \frac{1}{p(x_i, y_j)} =$$

$$= \frac{1}{8}\cdot\log_2 8 + \frac{1}{16}\cdot\log_2 16 + \frac{1}{32}\cdot\log_2 32 +$$

$$
\begin{aligned}
&+\frac{1}{32}\cdot\log_2 32+\frac{1}{16}\cdot\log_2 16+\frac{1}{8}\cdot\log_2 8+\\
&+\frac{1}{32}\cdot\log_2 32+\frac{1}{32}\cdot\log_2 32+\frac{1}{16}\cdot\log_2 16+\\
&+\frac{1}{16}\cdot\log_2 16+\frac{1}{16}\cdot\log_2 16+\frac{1}{16}\cdot\log_2 16+\\
&+\frac{1}{4}\cdot\log_2 4=\frac{27}{8}\text{ biti.}
\end{aligned}
$$

Constatăm că $H(Y\mid X)\neq H(X\mid Y)$. Observăm însă că

$$H(X)-H(X\mid Y)=H(Y)-H(Y\mid X)=\frac{7}{4}-\frac{11}{8}=2-\frac{13}{8}=\frac{3}{8}.$$

2.4. ENTROPIE RELATIVĂ ŞI INFORMAŢIE MUTUALĂ

Entropia unei variabile aleatoare X este o măsură a incertitudinii cu privire la X; ea este o măsură a cantităţii de informaţie necesare în medie pentru a descrie variabila aleatoare. Am văzut că entropia nu depinde de domeniul de existenţă al lui X, ci doar de distribuţia de probabilitate a lui X. În practică, s-ar putea să nu dispunem de adevărata distribuţie de probabilitate a unei variabile aleatoare, astfel încât suntem obligaţi să utilizăm în locul acesteia o distribuţie pe care o considerăm adecvată, dar care nu este decât o aproximaţie a celei reale. Entropia relativă este o măsură a distanţei dintre două distribuţii. Să presupunem că mulţimea valorilor pe care le poate lua variabila aleatoare X este $\{x_1, x_2, \cdots, x_L\}$. Conform adevăratei distribuţii, X ia valoarea x_k cu probabilitatea $p_k = P(X = x_k)$, dar noi considerăm că ea este $q_k = P(X = x_k), 1 \le k \le L$. Prin definiţie, *entropia relativă* sau *distanţa Kullback Leibler* dintre funcţiile masă de probabilitate $p(x)$ şi $q(x)$ este

$$D(p\,\|\,q)=\sum_{k=1}^{L} p_k \log_2 \frac{p_k}{q_k} = E_p \log_2 \frac{p(X)}{q(X)} \tag{2.13}$$

În definiţia de mai sus, utilizăm convenţia, bazată pe continuitatea funcţiilor, că $0\log\frac{0}{q}=0$ şi $p\log\frac{p}{0}=\infty.$

Observăm că entropia relativă astfel definită nu este o veritabilă distanță între distribuții căci nu este simetrică şi nu satisface inegalitatea triunghiului.

Să considerăm acum că semnalul generat de o sursă de informație este descris de variabila aleatoare X care ia L valori discrete $x_k, k = 1, 2, \cdots, L.$ Acest semnal este aplicat la intrarea unui canal de comunicație. Din cauza zgomotului din canal, receptorul observă un semnal diferit de cel emis, acest semnal fiind descris de variabila aleatoare Y care ia M valori discrete $y_i, i = 1, 2, \cdots, M$. Nu este necesar ca $M = L$. La un moment dat, receptorul observă o valoare $Y = y_i$ şi trebuie să determine cantitatea de informație pe care apariția acestui eveniment o furnizează cu privire la evenimentul $X = x_k$. Dacă variabilele aleatoare X şi Y ar fi statistic independente, apariția lui $Y = y_i$ n-ar furniza nici o informație cu privire la apariția evenimentului $X = x_k$. Pe de altă parte, dacă X şi Y ar depinde total una de alta astfel încât apariția lui $Y = y_i$ să determine apariția lui $X = x_k$, conținutul de informație nu este altul decât cel furnizat de evenimentul $X = x_k$. Prin definiție, *informația mutuală* dintre x_k şi y_i este mărimea

$$I(x_k\,;\,y_i) = \log_2 \frac{P(x_k \mid y_i)}{P(x_k)} \tag{2.14}$$

Dacă variabilele aleatoare X şi Y sunt statistic independente, $P(x_k \mid y_i) = P(x_k)$ şi, deci, $I(x_k\,;\,y_i) = 0$. Pe de altă parte, dacă apariția evenimentului $Y = y_i$ determină univoc apariția evenimentului $X = x_k$, $P(x_k \mid y_i) = 1$ şi deci

$$I(x_k\,;\,y_i) = \log_2 \frac{1}{P(x_k)} = -\log_2 P(x_k) \tag{2.15}$$

Aceasta este chiar informația cu privire la evenimentul $X = x_k$. Din acest motiv, se numeşte *auto-informația* evenimentului $X = x_k$ şi se notează

$$I(x_k) = \log_2 \frac{1}{P(x_k)} = -\log_2 P(x_k) \tag{2.16}$$

Observăm că un eveniment cu probabilitate ridicată poartă mai puțină informație decât un eveniment cu probabilitate scăzută.

Conform definiţiei probabilităţii comune, avem că

$$\frac{P(x_k \mid y_i)}{P(x_k)} = \frac{P(x_k \mid y_i)P(y_i)}{P(x_k)P(y_i)} = \frac{P(x_k, y_i)}{P(x_k)P(y_i)} = \frac{P(y_i \mid x_k)}{P(y_i)} \tag{2.17}$$

Din (2.17), rezultă că

$$I(x_k; y_i) = I(y_i; x_k) \tag{2.18}$$

În cuvinte, informaţia furnizată de apariţia evenimentului $Y = y_i$ cu privire la evenimentul $X = x_k$ este egală cu informaţia furnizată de apariţia evenimentului $X = x_k$ cu privire la evenimentul $Y = y_i$. Acum, dacă cele două variabile aleatoare X şi Y au funcţia masă de probabilitate comună $p(x, y)$, funcţiile masă de probabilitate marginală sunt $p(x)$ şi $p(y)$, respectiv. Prin definiţie, *informaţia mutuală* $I(X;Y)$ este entropia relativă dintre distribuţia comună şi distribuţia produs $p(x)p(y)$:

$$\begin{aligned} I(X;Y) &= \sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 \frac{p(x_k, y_i)}{p(x_k)p(y_i)} \\ &= D(p(x,y)\,\|\,p(x)p(y)) = E_{p(x,y)} \log_2 \frac{p(X,Y)}{p(X)p(Y)} \end{aligned} \tag{2.19}$$

Putem rescrie informaţia mutuală astfel:

$$\begin{aligned} I(X;Y) &= \sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 \frac{p(x_k, y_i)}{p(x_k)p(y_i)} \\ &= \sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 \frac{p(x_k \mid y_i)}{p(x_k)} \\ &= -\sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 p(x_k) + \sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 p(x_k \mid y_i) \\ &= -\sum_{k=1}^{L} p(x_k)\log_2 p(x_k) - \left(-\sum_{k=1}^{L}\sum_{i=1}^{M} p(x_k, y_i)\log_2 p(x_k \mid y_i)\right) \\ &= H(X) - H(X \mid Y) \end{aligned} \tag{2.20}$$

Cu cuvinte, informaţia mutuală $I(X;Y)$ este reducerea incertitudinii cu privire la X datorită cunoaşterii lui Y. Prin simetrie, avem şi că

$$I(X;Y) = H(Y) - H(Y \mid X) \tag{2.21}$$

Conform regulii lanţului (2.10), avem:

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \tag{2.22}$$

Observăm că

$$I(X;X) = H(X) - H(X \mid X) = H(X) \tag{2.23}$$

În cuvinte, informaţia mutuală a unei variabile aleatoare cu ea însăşi este entropia variabilei aleatoare. Acesta este motivul pentru care entropia se mai numeşte şi *auto-informaţie.*

2.5. INEGALITATEA LUI JENSEN

O funcţie $f(x)$ se spune că este *convexă* pe un interval (a,b) dacă pentru orice $x_1, x_2 \in (a,b)$ şi $0 \le \lambda \le 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2) \tag{2.24}$$

O funcţie $f(x)$ se spune că este *strict convexă* dacă egalitatea are loc numai pentru $\lambda = 0$ sau $\lambda = 1$.

O funcţie $f(x)$ este *concavă* dacă $-f(x)$ este convexă.

O funcţie este convexă dacă se situează sub orice coardă. O funcţie este concavă dacă se situează deasupra oricărei corzi.

TEOREMA 2.2: Dacă funcţia $f(x)$ are o derivată a doua care este nenegativă (pozitivă) pe tot domeniul de definiţie, ea este convexă (strict convexă).

DEMONSTRAŢIE

Dezvoltăm funcţia în serie Taylor în jurul unui punct x_0:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x^*)}{2}(x - x_0)^2 \tag{2.25}$$

unde x^* este între x_0 şi x. Prin ipoteză, $f''(x^*) \ge 0$ astfel încât ultimul termen este nenegativ pentru orice x.

Fie $x_0 = \lambda x_1 + (1-\lambda)x_2$. Pentru $x = x_1$, obţinem:

$$f(x_1) \ge f(x_0) + f'(x_0)[(1-\lambda)(x_1 - x_2)]. \tag{2.26}$$

Pentru $x = x_2$, avem:

$$f(x_2) \ge f(x_0) + f'(x_0)[\lambda(x_2 - x_1)]. \tag{2.27}$$

Înmulţind (2.26) cu λ şi (2.27) cu $(1-\lambda)$ iar apoi adunând, obţinem (2.24).

TEOREMA 2.3 (*Inegalitatea lui Jensen*): Dacă $f(x)$ este o funcție convexă iar X este o variabilă aleatoare, avem

$$Ef(X) \geq f(EX) \tag{2.28}$$

Mai mult decât atât, dacă $f(x)$ este strict convexă, egalitatea din (2.28) implică faptul că $X = EX$ cu probabilitate 1, adică, X este o constantă.

DEMONSTRAȚIE

Vom demonstra teorema prin inducție după numărul de evenimente din spațiul de probabilitate.

Pentru o distribuție de probabilitate cu două evenimente elementare, inegalitatea se scrie

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2). \tag{2.29}$$

Având în vedere că $p_1 + p_2 = 1$, aceasta urmează direct din definiția funcțiilor convexe. Să presupunem că teorema este adevărată pentru distribuții cu $k - 1$ evenimente elementare. Pentru o distribuție cu k evenimente elementare, avem

$$p_1 + p_2 + \cdots + p_{k-1} + p_k = 1 \tag{2.30}$$

Din (2.30), rezultă imediat că

$$\frac{p_1 + p_2 + \cdots + p_{k-1}}{1 - p_k} = 1 \tag{2.31}$$

Notăm $p_i^{'} = p_i / (1 - p_k)$ pentru $i = 1, 2, \cdots, k-1$. Avem atunci

$$\begin{aligned}
\sum_{i=1}^{k} p_i f(x_i) &= p_k f(x_k) + \sum_{i=1}^{k-1} p_i f(x_i) \\
&= p_k f(x_k) + (1 - p_k) \sum_{i=1}^{k-1} p_i^{'} f(x_i) \\
&\geq p_k f(x_k) + (1 - p_k) f\left(\sum_{i=1}^{k-1} p_i^{'} x_i \right) \qquad (2.32) \\
&\geq f\left(p_k x_k + (1 - p_k) \sum_{i=1}^{k-1} p_i^{'} x_i \right) \\
&= f\left(\sum_{i=1}^{k} p_i x_i \right)
\end{aligned}$$

Rămâne să demonstrăm partea a doua a teoremei. Dacă $f(x)$ este strict convexă, egalitatea are loc în (2.24) numai pentru $\lambda = 0$ sau $\lambda = 1$. Aceasta înseamnă că, în (2.29), $p_1 = 0$ şi $p_2 = 1$ sau $p_1 = 1$ şi $p_2 = 0$. Dar aceasta înseamnă că mulţimea evenimentelor elementare se reduce la un singur punct şi că X este o constantă cu probabilitate 1.

TEOREMA 2.4: Fie $p(x_k), q(x_k), k = 1, 2, \cdots, L$ două funcţii masă de probabilitate. Entropia relativă $D(p \| q)$ este nenegativă

$$D(p \| q) \geq 0 \tag{2.33}$$

cu egalitate dacă şi numai dacă

$$p(x_k) = q(x_k) \text{ pentru toţi } k. \tag{2.34}$$

DEMONSTRAŢIE

Fie $A = \{x_i : p(x_i) > 0\}$ mulţimea suport a lui $p(x)$. Reamintim că mulţimea suport a unei funcţii este submulţimea domeniului de definiţie pe care funcţia este diferită de zero. Putem scrie

$$\begin{aligned} -D(p \| q) &= -\sum_{x_k \in A} p(x_k) \log_2 \frac{p(x_k)}{q(x_k)} \\ &= \sum_{x_k \in A} p(x_k) \log_2 \frac{q(x_k)}{p(x_k)} \end{aligned} \tag{2.35}$$

Fie funcţia $f(t) = \log_2 t$. Avem $f'(t) = \dfrac{\log_2 e}{t}$ şi $f''(t) = -\dfrac{\log_2 e}{t^2}$.

Conform Teoremei 2.1, funcţia $\log_2 t$ este strict concavă. Aplicăm în (2.35) inegalitatea lui Jensen:

$$\begin{aligned} -D(p \| q) &\leq \log_2 \sum_{x_k \in A} p(x) \frac{q(x)}{p(x)} \\ &= \log_2 \sum_{x_k \in A} q(x) \leq \log_2 \sum_{k=1}^{L} q(x_k) = \log_2 1 = 0 \end{aligned} \tag{2.36}$$

COROLAR: Pentru oricare două variabile aleatoare X şi Y, informaţia mutuală $I(X;Y)$ este nenegativă

$$I(X;Y) \geq 0 \tag{2.37}$$

cu egalitate dacă şi numai dacă X şi Y sunt independente.

DEMONSTRAŢIE

Am arătat că $I(X;Y) = D(p(x,y) \| p(x)p(y))$, iar conform Teoremei 2.3, aceasta este nenegativă. Egalitatea are loc dacă şi numai dacă $p(x,y) = p(x)p(y)$, adică, dacă X şi Y sunt independente.

TEOREMA 2.5: Fie o variabilă aleatoare X al cărei domeniu de definiție are L evenimente elementare. Avem $H(X) \le \log_2 L,$ cu egalitate dacă şi numai dacă X are o distribuție de probabilitate uniformă (toate evenimentele elementare au aceeaşi probabilitate egală cu $\frac{1}{L}$).

DEMONSTRAȚIE

Fie $u(x_k) = \frac{1}{L}, k = 1, 2, \cdots, L$ funcția masă de probabilitate uniformă şi fie $p(x_k)$ funcția masă de probabilitate pentru X. Avem atunci

$$D(p \parallel u) = \sum_{k=1}^{L} p(x_k) \log_2 \frac{p(x_k)}{u(x_k)} = \log_2 L - H(X) \tag{2.38}$$

Având în vedere că entropia relativă este nenegativă, avem

$$0 \le D(p \parallel u) = \log_2 L - H(X) \tag{2.39}$$

2.6. ENTROPIE DIFERENȚIALĂ

Entropia diferențială este entropia unei variabile aleatoare de tip continuu. Ea este similară în multe privințe cu entropia unei variabile aleatoare de tip discret, dar sunt şi unele diferențe importante.

Fie o variabilă aleatoare de tip continuu X cu funcția densitate de probabilitate $p(x)$. Din capitolul 1, ştim că

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1.$$

Mulțimea unde $p(x) > 0$ se numeşte *mulțimea suport* a lui X şi o notăm cu S. Prin definiție, *entropia diferențială* $h(X)$ a variabilei aleatoare de tip continuu X cu densitate $p(x)$ este integrala

$$h(X) = -\int_{\mathrm{S}} p(x) \log_2 p(x)\mathrm{d}x \tag{2.40}$$

Definiția de mai sus este valabilă cu condiția ca integrala respectivă să existe.

Distribuție uniformă

Considerăm o variabilă aleatoare distribuită uniform de la 0 la a, astfel încât densitatea sa de probabilitate este $1/a$ de la 0 la a şi 0 în rest. Entropia ei diferențială este

$$h(X) = -\int_0^a \frac{1}{a}\log_2 \frac{1}{a}\,\mathrm{d}x = \log_2 a \tag{2.41}$$

Pentru $a < 1$, $\log_2 a < 0$, astfel încât entropia diferențială este negativă. Deci, spre deosebire de entropia discretă, entropia diferențială poate fi şi negativă.

Distribuție normală (Gaussiană)

Fie

$$X \sim \phi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \tag{2.42}$$

Calculăm entropia diferențială astfel:

$$\begin{aligned} h(\phi) &= -\int \phi(x)\log_2\phi(x)\mathrm{d}x = -\log_2 e\int \phi(x)\ln\phi(x)\mathrm{d}x \\ &= -\log_2 e\int \phi(x)\left[-\frac{x^2}{2\sigma^2} - \ln\left(\sqrt{2\pi}\sigma\right)\right]\mathrm{d}x \\ &= \frac{\log_2 e}{2\sigma^2}\int x^2\phi(x)\mathrm{d}x + \log_2\left(\sqrt{2\pi}\sigma\right)\int \phi(x)\mathrm{d}x \end{aligned} \tag{2.43}$$

Din Capitolul 1 ştim că, pentru distribuția normală cu medie zero, avem:

$$EX^2 = \int x^2\phi(x)\mathrm{d}x = \sigma^2$$

iar ca pentru orice funcție densitate de probabilitate, $\int \phi(x)\mathrm{d}x = 1$. De aceea, (2.43) se scrie:

$$\begin{aligned} h(\phi) &= \frac{1}{2}\log_2 e + \frac{1}{2}\log_2\left(2\pi\sigma^2\right) \\ &= \frac{1}{2}\log_2\left(2\pi e\sigma^2\right)\text{biti.} \end{aligned} \tag{2.44}$$

2.7. RELAȚIA DINTRE ENTROPIA DIFERENȚIALĂ ȘI ENTROPIA DISCRETĂ

Considerăm o variabilă aleatoare X cu densitate de probabilitate $p(x)$ reprezentată grafic în figura 2.2.

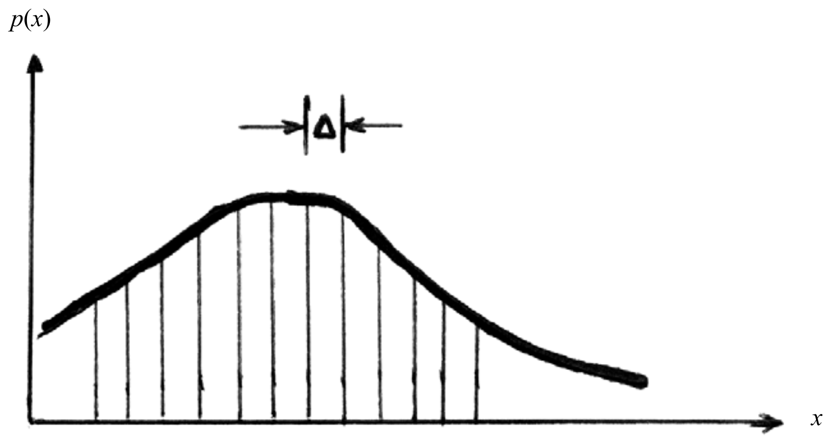

Fig. 2.2. Cuantizarea unei variabile aleatoare de tip continuu.

Împărțim domeniul de definiție al lui X în intervale elementare de lungime Δ. În interiorul fiecărui interval elementar există un punct x_i astfel încât $p(x_i)$ este media lui $p(x)$ în acel interval elementar:

$$p(x_i) = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} p(x)\mathrm{d}x \tag{2.45}$$

Definim variabila aleatoare cuantizată X_Δ astfel:

$$X_\Delta = x_i \text{, dacă } i\Delta \le X < (i+1)\Delta \tag{2.46}$$

Probabilitatea ca $X_\Delta = x_i$ este

$$p_i = \int_{i\Delta}^{(i+1)\Delta} p(x)\mathrm{d}x = p(x_i)\Delta \tag{2.47}$$

Entropia lui X_Δ este

$$\begin{aligned} H(X_\Delta) &= -\sum_{i=-\infty}^{\infty} p_i \log_2 p_i \\ &= -\sum_{i=-\infty}^{\infty} p(x_i)\Delta \log_2 (p(x_i)\Delta) \\ &= -\sum_{i=-\infty}^{\infty} \Delta p(x_i) \log_2 p(x_i) - \sum_{i=-\infty}^{\infty} p(x_i)\Delta \log_2 \Delta \end{aligned} \tag{2.48}$$

Având în vedere că $\sum p(x_i)\Delta = \int p(x)\mathrm{d}x = 1$, aceasta se poate scrie:

$$H(X_\Delta) = -\sum_{i=-\infty}^{\infty} \Delta p(x_i) \log_2 p(x_i) - \log_2 \Delta \tag{2.49}$$

Dacă funcția $p(x)\log_2 p(x)$ este integrabilă în sens Riemann, condiție necesară pentru a asigura că limita este bine definită, primul termen din (2.49) tinde spre integrala lui $-p(x)\log_2 p(x)$. Am demonstrat astfel următoarea teoremă.

TEOREMA 2.6: Dacă densitatea de probabilitate $p(x)$ a unei variabile aleatoare *X* este integrabilă în sens Riemann, avem:

$$H(X_\Delta) + \log_2 \Delta \to h(X) \text{ pentru } \Delta \to 0 \tag{2.50}$$

Să presupunem că variabila aleatoare X variază într-un domeniu finit, să spunem $x_1 \le X \le x_2$. Pentru o cuantă Δ, avem $(x_2 - x_1)/\Delta$ nivele de cuantizare. Dacă exprimăm fiecare eşantion printr-un cuvânt binar de *n* biți, avem că

$$\frac{x_2 - x_1}{\Delta} = 2^n \ . \tag{2.51}$$

Întrucât $x_2 - x_1$ este o constantă pe care o putem norma astfel încât să fie egală cu 1, avem din (2.51) că

$$\Delta = 2^{-n}. \tag{2.52}$$

Combinând (2.52) cu (2.50), obținem următorul rezultat: entropia versiunii cuantizate pe n biți a unei variabile aleatoare de tip continuu X este aproximativ egală cu $h(X) + n$.

2.8. ENTROPIE DIFERENȚIALĂ COMUNĂ ŞI CONDIȚIONATĂ

Fie două variabile aleatoare de tip continuu X şi Y având funcția densitate de probabilitate comună $p(x, y)$. Definim entropia diferențială comună a lui X şi Y astfel:

$$h(X, Y) = -\int p(x, y)\log_2 p(x, y)\mathrm{d}x\mathrm{d}y \tag{2.53}$$

De asemenea, definim entropia diferențială condiționată $h(X \mid Y)$ astfel:

$$h(X \mid Y) = -\int p(x, y)\log_2 p(x \mid y)\mathrm{d}x\mathrm{d}y \tag{2.54}$$

Întrucât în general $p(x \mid y) = p(x, y) / p(y)$, putem scrie şi că

$$h(X \mid Y) = h(X, Y) - h(Y) \tag{2.55}$$

2.9. ENTROPIE RELATIVĂ ŞI INFORMAȚIE MUTUALĂ PENTRU VARIABILE ALEATOARE DE TIP CONTINUU

Prin definiție, *entropia relativă*, sau *distanța Kullback Leibler*, $D(f \,\|\, g)$ a două densități de probabilitate $f(x)$ şi $g(x)$ este

$$D(f \,\|\, g) = \int f(x)\log_2 \frac{f(x)}{g(x)}\mathrm{d}x \tag{2.56}$$

Aceasta este finită numai dacă mulțimea suport a lui $f(x)$ este conținută în mulțimea suport a lui $g(x)$.

Prin definiție, informația mutuală $I(X; Y)$ dintre două variabile aleatoare de tip continuu cu densitate de probabilitate comună $p(x, y)$ este:

$$I(X; Y) = \int p(x, y)\log_2 \frac{p(x, y)}{p(x)p(y)}\mathrm{d}x\mathrm{d}y \tag{2.57}$$

Din definiția (2.57), rezultă imediat că

$$I(X; Y) = D(p(x, y) \,\|\, p(x)p(y)) \tag{2.58}$$

şi că

$$I(X; Y) = h(X) - h(X \mid Y) = h(Y) - h(Y \mid X) \tag{2.59}$$

Proprietățile lui $D(f \,\|\, g)$ şi ale lui $I(X; Y)$ sunt aceleaşi ca şi în cazul variabilelor aleatoare de tip discret. În particular, informația mutuală dintre două variabile aleatoare de tip continuu X şi Y este limita informației mutuale dintre versiunile lor cuantizate, căci

$$\begin{aligned} I(X_\Delta; Y_\Delta) &= H(X_\Delta) - H(X_\Delta \mid Y_\Delta) \\ &\approx h(X) - \log_2 \Delta - (h(X \mid Y) - \log_2 \Delta) \\ &= I(X; Y) \end{aligned} \tag{2.60}$$

Proprietățile entropiei diferențiale, ale entropiei relative şi ale informației mutuale sunt similare celor din cazul variabilelor aleatoare de tip discret.

2.10. LANȚURI MARKOV

În Capitolul 1, am văzut că un proces stochastic este un şir indexat de variabile aleatoare. În general, aceste variabile aleatoare nu sunt independente. Un exemplu simplu de proces stochastic cu dependență este unul în care fiecare variabilă aleatoare depinde de cea precedentă şi nu este condiționată de nici una din celelalte variabile aleatoare precedente.

Prin definiție, un *lanț Markov* sau un *proces Markov* este un proces stochastic discret $X_1, X_2, \cdots$ dacă, pentru $n = 1, 2, \cdots$,

$$\begin{aligned} &\Pr(X_{n+1} = x_{n+1} \mid X_n = x_n, X_{n-1} = x_{n-1}, \cdots, X_1 = x_1) \\ &\qquad = \Pr(X_{n+1} = x_{n+1} \mid X_n = x_n) \end{aligned} \tag{2.61}$$

pentru toate valorile $x_1, x_2, \cdots, x_n, x_{n+1}$ pe care le poate lua procesul.

În acest caz, funcția masă de probabilitate comună a variabilelor aleatoare se poate scrie

$$p(x_1, x_2, \cdots, x_n) = p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_2) \cdots p(x_n \mid x_{n-1}) \tag{2.62}$$

Prin definiție, se spune că lanțul Markov este *invariant în timp* dacă probabilitatea condiționată $p(x_{n+1} \mid x_n)$ nu depinde de *n*, adică, pentru $n = 1, 2, \cdots$

$$\Pr\{X_{n+1} = b \mid X_n = a\} = \Pr\{X_2 = b \mid X_1 = a\} \text{ pentru orice } a, b \tag{2.63}$$

Dacă $\{X_i\}$ este un lanț Markov, X_n se numeşte *starea* la timpul *n*. Un lanț Markov invariant în timp se caracterizează prin starea sa inițială şi o *matrice a tranziției de probabilitate*

$$P = [P_{ij}], i, j \in \{1, 2, \cdots, m\} \text{ unde } P_{ij} = \Pr\{X_{n+1} = j \mid X_n = i\}.$$

Dacă este posibil să se meargă cu probabilitate pozitivă din orice stare a lanțului Markov în oricare altă stare într-un număr finit de paşi, lanțul Markov se spune că este *ireductibil.*

Dacă funcția masă de probabilitate a unei variabile aleatoare la timpul *n* este $p(x_n)$, funcția masă de probabilitate a unei variabile aleatoare la timpul *n*+1 este

$$p(x_{n+1}) = \sum_{x_n} p(x_n) P_{x_n x_{x+1}} \tag{2.64}$$

O distribuție de probabilitate a stărilor astfel încât distribuția la timpul *n*+1 este aceeaşi cu distribuția la timpul *n* se numeşte *distribuție staționară.*

Dacă un lanț Markov cu un număr finit de stări posibile este ireductibil şi aperiodic, distribuția staționară este unică, astfel încât din orice distribuție inițială, distribuția lui X_n tinde către distribuția staționară când $n \to \infty$.

EXEMPLUL 2.2: Considerăm un lanț Markov cu două stări având matricea tranziției de probabilitate

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{2.65}$$

Ilustrăm aceasta printr-un graf orientat în care cele două vârfuri reprezintă starea 1 şi starea 2, arcele reprezintă tranzițiile iar etichetele arcelor sunt probabilitățile de tranziție corespunzătoare, aşa cum se arată în figura 2.3.

Reprezentăm distribuția staționară prin vectorul $\mu = [\mu_1, \mu_2]$ ale cărui componente sunt probabilitățile staționare în starea 1 şi în starea 2, respectiv. Putem găsi probabilitatea staționară rezolvând ecuația matriceală

$$\mu P = \mu \tag{2.66}$$

ceea ce ne dă că

$$\mu_1 \alpha = \mu_2 \beta \tag{2.67}$$

Având în vedere că $\mu_1 + \mu_2 = 1$, distribuția staționară este

$$\mu_1 = \frac{\beta}{\alpha+\beta}, \ \mu_2 = \frac{\alpha}{\alpha+\beta} \tag{2.68}$$

Dacă lanțul Markov are o distribuție inițială egală cu distribuția staționară, procesul care rezultă va fi staționar. Entropia stării X_n la timpul *n* este

$$H(X_n) = H\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}\right) \tag{2.69}$$

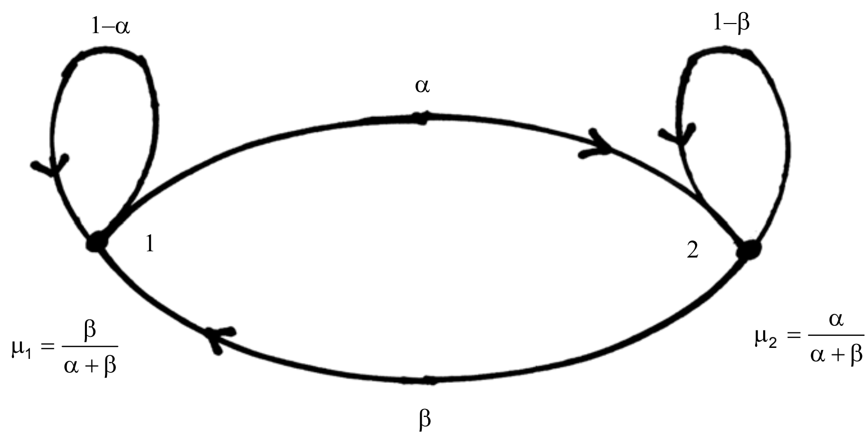

Fig. 2.3. Lanț Markov cu două stări.

Dacă însă în starea inițială distribuția de probabilitate diferă de cea staționară, la aceasta din urmă se ajunge după un proces tranzitoriu.

Prin definiție, *rata de entropie* a unui proces stochastic $\{X_i\}$ este

$$H(X) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \cdots, X_n) \tag{2.70}$$

dacă limita există.

Pentru un lanț Markov staționar, rata de entropie este dată de

$$\begin{aligned} H(X) &= \lim_{n\to\infty} H(X_n \mid X_{n-1}, \cdots, X_1) \\ &= \lim_{n\to\infty} H(X_n \mid X_{n-1}) = H(X_2 \mid X_1) \end{aligned} \tag{2.71}$$

Dacă lanțul Markov este ireductibil şi aperiodic, el are o distribuție staționară unică a stărilor, iar orice distribuție inițială tinde către distribuția staționară când $n \to \infty$. În acest caz, rata de entropie este dată de (2.71).

În capitolul următor, vom utiliza noțiunea de entropie a unei variabile aleatoare introdusă mai sus pentru a optimiza codarea surselor de informație.