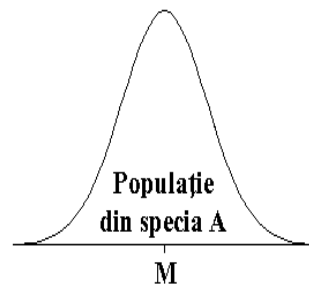


§ 3.8. Tratarea unei variabile calitative

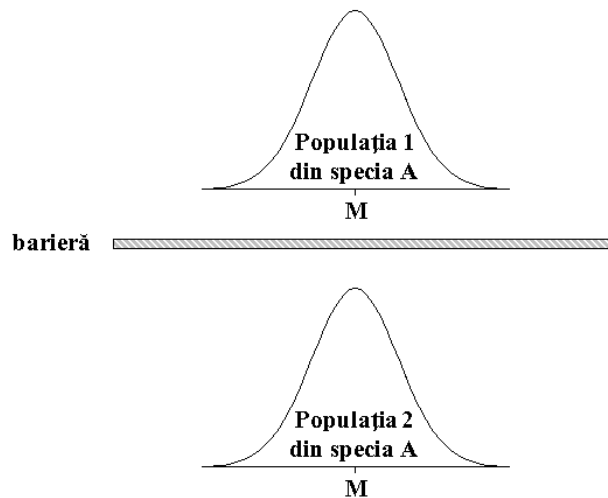
Paragraful cuprinde (a) tratarea cantitativă și (b) tratarea calitativă a unei variabile calitative. Mai întâi, însă, o observație originală foarte importantă.

Reamintim că dacă gândim cantitativ, variabilitatea apare ca o împrăștiere (de regulă) în jurul unei tendințe centrale, iar dacă gândim calitativ, variabilitatea înseamnă diversitate. Această distincție nu este suficient de clară în statistică. Ea este marcată mai mult de autorul acestei lucrări, care observă că un biolog poate înțelege cel mai bine această disociere inspirată de gândirea ecologică. Într-adevăr, să invocăm succint procesul de speciație geografică, urmărind variabilitatea unui caracter, de exemplu lungimea totală sau talia animalului.

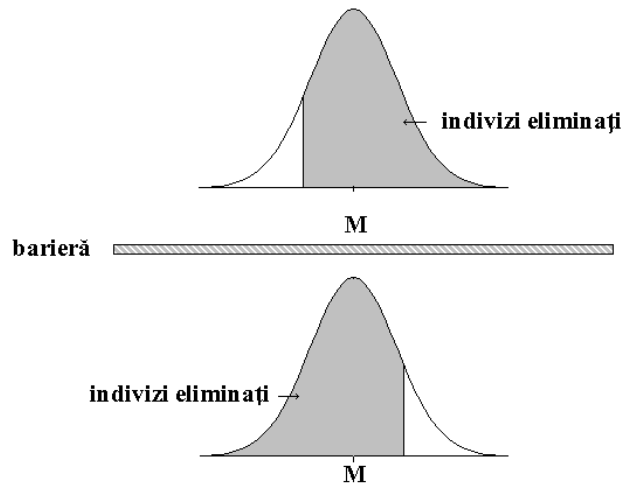
Să presupunem că într-un areal a existat o populație dintr-o specie A. (În acest caz **variabilitatea** taliei indivizilor poate fi explicată ca o **împrăștiere în jurul unei tendințe centrale M .**)



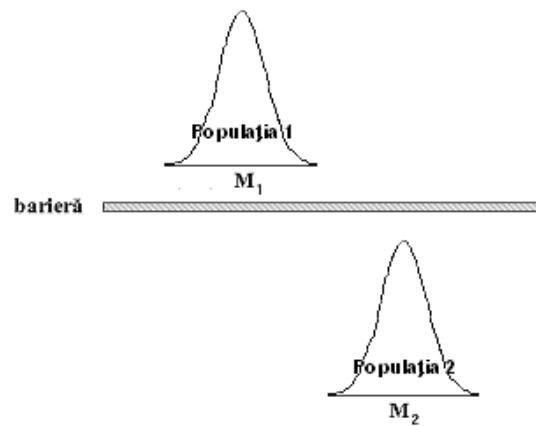
La un moment dat în areal a apărut o barieră care a despărțit populația în două populații cvasiidentice, prin urmare cu aproximativ aceeași medie M .



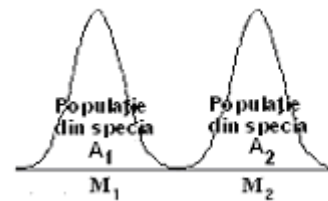
În cele două subareale separate de barieră condițiile de mediu au fost extrem de diferite producând, astfel, presiuni selectiv în sens contrar, conform modelului din observația de la 3.1.1. punctul 6^o.



Cele două populații s-au refăcut în timp, dar cu diferențe semnificative ca tendințe centrale ($M_1 \neq M_2$).



La un moment dat bariera a dispărut și cele două populații au repopulat împreună vechiul areal. Având acum însă dimensiuni semnificativ diferite, ele nu mai sunt interfecunde. Au devenit, deci, specii diferite: A_1 și A_2 .



Acum putem spune că, sub presiune selectivă divergentă, **împrăștierea** dimensiunii a crescut atât de mult încât a produs ceva complet nou și anume a îmbogățit **diversitatea** caracterului calitativ "specie" (nu a caracterului cantitativ, a dimensiunii).

3.8.1. Tratarea cantitativă a unei variabile calitative

O variabilă calitativă se manifestă printr-o serie statistică univariată, calitativă $(x_j)_{j=1,2,\dots,N}$ unde x_j sunt *variante* distincte sau nu ale variabilei (la cele cantitative x_j se numeau *valorii*).

Exemplul 3.8.1.

{alb, roșu, alb, galben, alb, verde, roșu, galben, alb, alb}
poate fi o serie de culori ale unor flori.

Seria poate fi prezentată și grupat ca o distribuție de frecvențe absolute ale variantelor distincte x_j , notată

$$\begin{pmatrix} x_j \\ N_j \end{pmatrix}_{j=1,2,\dots,p} \quad \text{sau } (x_j, N_j)_{j=1,2,\dots,p}$$

Evident, $\sum_{j=1}^p N_j = N$.

Seria din exemplu devine:

$$\begin{pmatrix} \text{alb} & \text{roșu} & \text{galben} & \text{verde} \\ 5 & 2 & 2 & 1 \end{pmatrix}.$$

Dacă notăm $F_j = N_j / N$ se obține distribuția de frecvențe relative ale variantelor distincte x_j , notată:

$$\begin{pmatrix} x_j \\ F_j \end{pmatrix}_{j=1,2,\dots,p} \quad \text{sau } (x_j, F_j)_{j=1,2,\dots,p}$$

Evident, $\sum_{j=1}^p F_j = 1$.

Pentru exemplul de mai sus:

$$\begin{pmatrix} \text{alb} & \text{roșu} & \text{galben} & \text{verde} \\ \frac{5}{10} & \frac{2}{10} & \frac{2}{10} & \frac{1}{10} \end{pmatrix}.$$

1° Binarizarea unei variabile calitative

Tratarea cantitativă a unei variabile calitative presupune studierea unei singure variante în opoziție cu ceea ce rămâne în afara ei, adică binarizarea variabilei calitative.

De exemplu, dacă ne interesează doar culoarea alb în opoziție cu celelalte culori, sintetizăm datele în distribuția binară:

$$\begin{pmatrix} \text{alb} & \text{non - alb} \\ \frac{5}{10} & \frac{5}{10} \end{pmatrix}.$$

În general, pentru o distribuție de frecvențe relative a unei variabile calitative,

$$\begin{pmatrix} x_1 & x_2 & \dots & x_j & \dots & x_p \\ F_1 & F_2 & \dots & F_j & \dots & F_p \end{pmatrix}$$

dacă ne interesează varianta x_j în opoziție cu restul, sintetizăm distribuția binară:

$$\begin{pmatrix} x_j & \text{non } x_j \\ F_j & 1-F_j \end{pmatrix} \quad \text{notată mai simplu } \begin{pmatrix} x & \text{non } x \\ F & 1-F \end{pmatrix}$$

unde $F \in [0, 1]$ și este frecvența relativă a variantei x .

Cu aceste notații

tendința centrală e dată de proporția care ne interesează:

$$F \text{ ori } 1-F \text{ (după caz),}$$

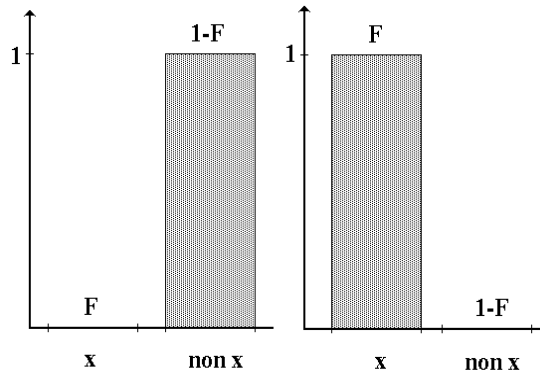
iar

împrăștierea în jurul tendinței centrale este dată în ambele situații de produsul:

$$F \cdot (1 - F), \text{ denumit } \underline{\text{dispersie}}.$$

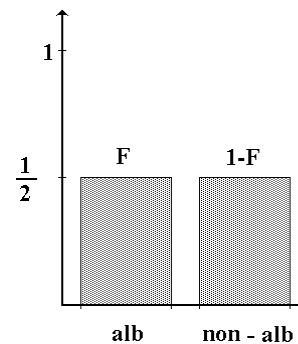
Proprietăți ale dispersiei

- Se observă că dispersia este minimă și anume zero dacă și numai dacă fie F , fie $1 - F$ este zero, adică atunci când într-adevăr nu există variabilitate, variabila realizând fie numai varianta *non x*, fie numai varianta *x*.



- Dispersia este maximă dacă și numai dacă $F = 1 - F = 1/2$, ceea ce corespunde, de asemenea, intuiției. Este ceea ce se întâmplă în exemplul nostru, dacă ne interesează culoarea *alb*.

- tendința centrală este proporția $F = 5 / 10 = 1 / 2$, iar
- dispersia este $F \cdot (1 - F) = 1/2 \cdot (1 - 1/2) = 1 / 4$. Dacă ne interesează culoarea *non-alb*,
- tendința centrală este $1 - F = 1 - 5 / 10 = 5 / 10 = 1 / 2$,



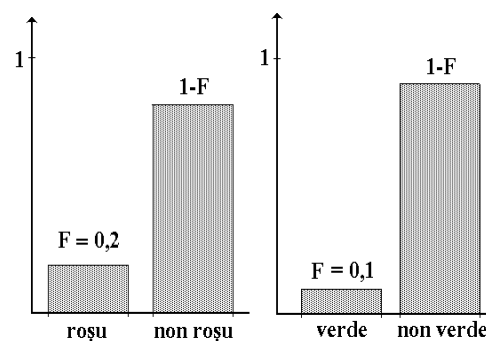
iar

- dispersia este aceeași, și anume $1 / 4$.

- Dispersia scade pe măsură ce crește decalajul (diferența) între cele două frecvențe, F și $1 - F$.

- De pildă în cazul culorii *roșu* dispersia este $F \cdot (1 - F) = 2 / 10 \cdot 8 / 10 = 16 / 100 = 0,16$ ($< 0,25$);
- iar în cazul celei *verzi* dispersia este $F \cdot (1 - F) = 1 / 10 \cdot 9 / 10 = 9 / 100 = 0,09$ ($< 0,16 < 0,25$).

Se observă cum cele două grafice se apropie din ce în ce mai mult de primul grafic (cel cu $F = 0$ și $1 - F = 1$), cu dispersia nulă.



+2^o Justificarea acestui mod de tratare

În cele ce urmează vom arăta “ce se ascunde în spatele” acestui mod de tratare a unei variabile calitative. În final se va putea înțelege cum o proporție poate fi un indicator de tendință centrală (o medie) iar produsul a două proporții complementare poate juca rolul unei dispersii.

Pentru acesta este suficient să observăm că, în cazul în care ne interesează doar o variantă, șirul de variante $(x_i)_{i=1, 2, \dots, N}$ poate fi înlocuit cu un șir de valori $(y_i)_{i=1, 2, \dots, N}$, care se impun în mod natural. Este vorba de valoarea 1, în cazul variantei de interes și valoarea 0, în rest.

Astfel, în cazul exemplului 3.8.1., dacă ne interesează varianta “alb”, se va obține șirul de valori următor:

Exemplul 3.8.1’.

$$\{1, 0, 1, 0, 1, 0, 0, 0, 1, 1\}$$

Fiind un șir numeric care exprimă cu sens o variabilă cantitativă - prezența sau absența unei anumite variante – îl putem trata cantitativ. Pentru conciziunea explicației vom considera distribuția de frecvențe relative corespunzătoare:

$$\mathbf{Y}_1 = \begin{pmatrix} 1 & 0 \\ \frac{5}{10} & \frac{5}{10} \end{pmatrix}.$$

În general, tratarea cantitativă a variabilei calitative binarizate $\mathbf{X} = \begin{pmatrix} x & \text{non} - x \\ F & 1 - F \end{pmatrix}$

înseamnă, de fapt, tratarea variabilei cantitative binare $\mathbf{Y} = \begin{pmatrix} 1 & 0 \\ F & 1 - F \end{pmatrix}$.

Într-adevăr, să calculăm media și dispersia acestei noi variabile:

$$M = (\text{cf. 3.3.4. pct. 2}^{\text{o}}\text{b}) = F \cdot 1 + (1 - F) \cdot 0 = F, \text{ iar}$$

$$S^2 = (\text{cf. 3.5.4. pct. 2}^{\text{o}}\text{b}) = F \cdot (1 - F)^2 + (1 - F) \cdot (0 - F)^2 = F \cdot (1 - F)^2 + (1 - F) \cdot F^2 = F \cdot (1 - F) \cdot [(1 - F) + F] = F \cdot (1 - F).$$

Se observă că media M este tocmai frecvența relativă F , iar dispersia S^2 este produsul $F \cdot (1 - F)$.

Revenind la exemplul numeric, cititorul poate verifica singur că media și dispersia distribuției \mathbf{Y}_1 (respectiv ale seriei 3.8.1’) sunt tocmai frecvența F , respectiv produsul

$$F \cdot (1 - F) \text{ calculate pentru distribuția calitativă binarizată } \mathbf{X}_1 = \begin{pmatrix} \text{alb} & \text{non} - \text{alb} \\ \frac{5}{10} & \frac{5}{10} \end{pmatrix}.$$

3.8.2. Tratarea calitativă a unei variabile calitative

1^o Entropia informațională - măsură a diversității unei biocenoze

În general, în statistică, variabilele calitative sunt tratate cantitativ, ca mai sus. Ecologiei însă, abordează problema biodiversității prin mai multe categorii de indicatori de tratare calitativă. Acești indicatori pot fi însă aplicați, după părerea noastră, oricăror variabile calitative.

În acest cadru vom aborda doar doi indicatori dintr-o categorie denumită “indicatori entropici”. Denumirea provine din faptul că la baza construcției acestor indicatori se află *entropia informațională* definită și studiată în cadrul “teoriei matematice a informației”.

O prezentare mai puțin riguroasă, dar mai intuitivă și specific adaptată ecologilor, a conceptului de entropie informațională folosit pentru măsurarea diversității specifice (adică în specii) a unei biocenoze este realizată de Daget în [9], aproximativ în felul următor.

Să considerăm, pentru început, o biocenoză formată din p specii distribuite uniform.

$$\begin{pmatrix} S_1 & S_2 & \dots & S_j & \dots & S_p \\ \frac{1}{p} & \frac{1}{p} & \dots & \frac{1}{p} & \dots & \frac{1}{p} \end{pmatrix}$$

Vom construi un indicator pentru diversitatea acestei biocenoze bazat pe conceptul de nedeterminare asociat experimentului de capturare la întâmplare a unor specii din această biocenoză, respectiv pe conceptul de informație obținută după efectuarea capturilor.

Nedeterminarea asupra rezultatului trebuie să îndeplinească condițiile:

1. Să fie o funcție f crescătoare în raport cu p , căci „cu cât numărul de specii p este mai mare, cu atât va fi mai dificilă predicția rezultatului unei capturi”.
2. Dacă $p = 1$ atunci $f = 0$. Altfel scris, $f(1) = 0$.
3. Dacă presupunem independența a două capturi succesive va trebui ca $f(p^2) = 2 \cdot f(p)$.

Cerința de la punctul 3 rezultă astfel: prima captură are p posibilități egale ca șansă. A doua captură va avea, de asemenea, p posibilități cu aceeași șansă dacă, prin efectuarea primei capturi, nu a fost schimbată în nici un fel compoziția biocenozei. Acest lucru se întâmplă, de exemplu, prin repunerea în biocenoză a exemplarului capturat fără ca acesta să fie afectat de captură în vreun fel, deci fără a i se modifica "șansa" de fi capturat. Prin urmare, două capturi independente vor avea $p \cdot p$ posibilități cu șanse egale, iar informația obținută din acestea trebuie să fie egală cu suma informațiilor după fiecare captură, adică:

$$f(p \cdot p) = f(p) + f(p)$$

ceea ce altfel scris este tocmai condiția de la punctul 3.

O funcție care satisface cele trei condiții este funcția $f(p) = \log_a p$ cu $a > 1$. Într-adevăr:

1. $\log_a p$ este crescătoare, dacă $a > 1$
2. $\log_a 1 = 0$
3. $\log_a (p^2) = 2 \cdot \log_a p$

În cazul biocenozei de mai sus, nedeterminarea sau informația I este, deci, $\log_a p$.

Observăm în continuare că putem considera informația I a întregii biocenoze ca fiind formată din suma informațiilor aduse de fiecare specie, informații care în acest caz sunt egale și au valoarea $1/p \cdot \log_a p$. Astfel:

$$I = \log_a p = \sum 1/p \cdot \log_a p = - \sum 1/p \cdot \log_a (1/p)$$

Ultimul mod de scriere are avantajul că oferă o formulă în care apar numai frecvențele relative ale speciilor ceea ce ne conduce la generalizarea finală. Într-adevăr, să considerăm o biocenoză în care proporțiile speciilor F_j nu sunt neapărat egale:

$$\begin{pmatrix} S_1 & S_2 & \dots & S_j & \dots & S_p \\ F_1 & F_2 & \dots & F_j & \dots & F_p \end{pmatrix}$$

unde $F_j \leq 1$ și $\sum F_j = 1$.

Evident, în acest caz informația I va fi egală cu:

$$I = - \sum F_j \cdot \log_a F_j$$

Această formulă de măsurare a informației este tocmai **entropia informațională** notată tradițional cu H . *Entropia informațională* este, în concluzie, o măsură a diversității unei biocenoze.

- ✓ O observație foarte importantă pentru aplicații este aceea că dacă speciile foarte rare nu apar în distribuția de frecvențe (ceea ce se întâmplă de regulă în practică), entropia nu se modifică foarte mult, frecvențele relative F_j foarte mici fiind nesemnificative în sumă. Cu alte cuvinte acest indicator de diversitate specifică, ca de altfel toți indicatorii entropici (adică bazați pe entropie), nu este sensibil la speciile rare.

2° Entropia informațională. Definiție, proprietăți

Shannon, fondând în 1948 (în lucrarea intitulată "A mathematical theory of communications") *teoria matematică a informației*, a introdus conceptul de entropie ca măsură a cantității de informație transmisă de o sursă. Vom prezenta, în continuare, această definiție și unele proprietăți, pornind de la exemplarele note de curs ale colegului Marcel Țena, la cursul de referință de "Teoria Informației" al Prof. S. Guiașu (1974-1975, Facultatea de Matematică - Mecanică). Am adăugat și unele observații necesare acestui cadru.

Fie o sursă de informație E despre care știm doar că emite p semnale S_j ($j = 1, 2, \dots, p$), fiecare semnal S_j având frecvența relativă de apariție F_j , necunoscând nimic despre regula (ordinea) de

apariție a semnalelor. Evident $0 \leq F_j \leq 1$ și $\sum_{j=1}^p F_j = 1$. Notăm sursa E astfel:

$$E = \begin{pmatrix} S_1 & S_2 & \dots & S_j & \dots & S_p \\ F_1 & F_2 & \dots & F_j & \dots & F_p \end{pmatrix}$$

Definiție: Se numește **entropia informațională** a sursei E valoarea

$$H_p(E) = -\sum_{j=1}^p F_j \cdot \log_a F_j$$

în care $a > 1$.

Deoarece valoarea expresiei $H_p(E)$ depinde de baza de logaritmare, pentru cele trei baze a , folosite în mod curent, 2, e și 10, s-au introdus "unități de măsură" ale entropiei corespunzătoare. Acestea se numesc:

bit (prescurtare de la englezescul "binary digit" - număr binar),

nit și respectiv

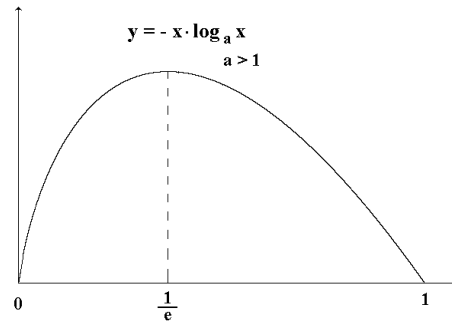
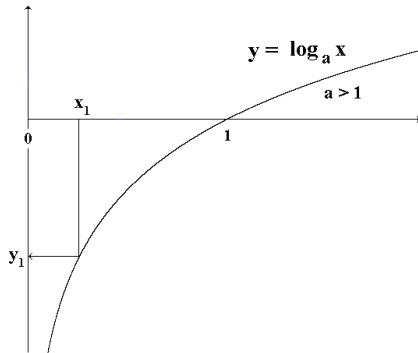
dit.

Litera **n** din *nit* provine de la inițiala matematicianului Neper, descoperitorul numărului irațional e , iar **d** din *dit* de la englezescul "decimal" - zecimal.

- ✓ H_p a fost denumit *entropie informațională* și nu simplu, *entropie*, pentru a se diferenția de *entropia termodinamică* care, din punct de vedere formal, are aceeași expresie. *Entropia termodinamică* a fost definită de Boltzman, în 1898, în fizica statistică, pentru măsurarea gradului de dezorganizare a unui sistem închis și exprimarea, cu ajutorul acesteia, a principiului doi al termodinamicii. Mulți autori fac diverse speculații mai mult sau mai puțin valide bazate pe coincidența celor două formule.
- ✓ Un experiment oarecare E cu p rezultate posibile pentru care cunoaștem doar frecvențele relative de apariție ale rezultatelor, F_j , poate fi considerat o sursă de informație. Putem spune că *informația* pe care ne-o oferă efectuarea experimentului este tot una cu *nedeterminarea* pe care o aveam înainte de efectuare. Guiașu spune că "*informație înseamnă nedeterminare înlăturată*" și că acestea sunt egale pentru un experiment dat. Vorbim despre *nedeterminare* dacă ne situăm înaintea experimentului și despre *informație după efectuarea acestuia*.
- ✓ Reciproc, o sursă de informație poate fi considerată un experiment cu această sursă. Ca atare, cele două "limbaje" (sursă de informație cu semnale etc. și experiment cu rezultate posibile etc.) sunt echivalente. În continuare vom prefera limbajul bazat pe experiment deoarece pare de o mai mare generalitate.

+*Proprietăți:*

1. $H_p(E) = H_p(F_1, F_2, \dots, F_p) \geq 0$ deoarece este o sumă cu semn schimbat de produse negative, frecvențele relative F_j fiind pozitive, iar logaritmiile negativi [baza a fiind supraunitară, iar frecvențele relative subunitare. Reamintim în desenul din stânga al figurii următoare, graficul funcției logaritmice cu bază supraunitară în care se observă că pentru valori subunitare (x_1) logaritmiile sunt negativi (y_1). În dreapta figurii prezentăm graficul funcției $-x \log_a x$ pentru $a > 1$. Se observă că pentru valori subunitare ale lui x , așa cum sunt frecvențele relative F_j , funcția ia valori pozitive și este maximă în valoarea $1/e$].



2. Dacă există un indice j_0 astfel încât $F_{j_0} = 1$, respectiv $F_j = 0$ oricare ar fi j diferit de j_0 , atunci $H_p(E) = 0$ cu convenția $0 \cdot \infty = 0$ (deci prelungind prin continuitate funcția $-x \cdot \log_a x$, în origine).

Interpretarea acestei proprietăți este următoarea: "Un experiment în care se produce un singur rezultat nu conține nici o nedeterminare, respectiv o sursă care emite continuu un singur semnal nu comunică nici o informație."

3. $H_{p+1}(F_1, F_2, \dots, F_p, 0) = H_p(F_1, F_2, \dots, F_p)$

Interpretare: "Dacă la un experiment se adaugă un rezultat care nu se produce niciodată, atunci nedeterminarea nu se schimbă". Demonstrația este evidentă.

Interpretare în cazul diversității specifice: "Dacă la o biocenoză cu p specii apare în plus o specie cu frecvența nulă sau cvasinulă, entropia rămâne aceeași, respectiv aproape aceeași. Entropia nu este, deci, sensibilă la speciile rare."

4. $H_p(F_1, F_2, \dots, F_p) \leq H_p(1/p, 1/p, \dots, 1/p)$ [^{not 1} $= H_{p, \max}$]

Interpretare: "Entropiile asociate distribuțiilor uniforme (echitabile, regulate) sunt maxime față de celelalte entropii cu același număr de variante p ." Demonstrația depășește cadrul prezent.

- 4'. $H_{p, \max} = \log_a p$

Într-adevăr, $H_{p, \max} = H_p(1/p, 1/p, \dots, 1/p) = -\sum 1/p \cdot \log_a(1/p) = -1/p \cdot \log_a(1/p) \cdot \sum 1 = -p \cdot 1/p \cdot \log_a(1/p) = -\log_a(1/p) = \log_a p$.

- 4". Minimul de nedeterminare, respectiv de informație, dintre toate experimentele distribuite uniform este conținut, respectiv este dat de experimentul cu 2 rezultate posibile ($p=2$):

$$\begin{pmatrix} S_1 & S_2 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

și are valoarea $\log_a 2$.

Dacă baza $a=2$ atunci $H_2 = \log_2 2 = 1$ bit. Sublinierea din enunț pune în evidență faptul că există și experimente cu o cantitate de informație mai mică de 1 bit, de exemplu orice experiment binar distribuit neuniform.

5. $H_{p, \max} < H_{p+1, \max}$. Demonstrația rezultă imediat din proprietatea 4'.

¹ "not" = se notează.

Interpretare: “Entropia unui experiment cu p rezultate cu frecvențe relative egale crește o dată cu numărul de rezultate posibile p .”

3° Entropia relativă

Din proprietatea 5 de mai sus, rezultă că pe măsură ce crește numărul de rezultate posibile ale unui experiment entropia poate crește. Ca atare, pentru a compara cantitățile de informație a două sau mai multe experimente cu numere diferite de rezultate posibile, este nevoie de relativizarea cantității de informație față de entropia maximă a experimentului cu același număr de rezultate. Pentru aceasta s-a introdus noțiunea de **entropie relativă**, definită prin:

$$H_{rel} = \frac{H_p}{H_{p,max}} = \frac{H_p}{\log_a p}$$

Proprietăți:

1. $0 \leq H_{rel} \leq 1$.
2. $H_{rel} = 0 \Leftrightarrow p = 1$ (căci doar atunci $H_p = 0$).
3. $H_{rel} = 1 \Leftrightarrow$ distribuția este uniformă (căci numai atunci $H_p = H_{p,max}$).
4. H_{rel} este adimensional.
5. H_p este omologul dispersiei σ^2 sau al abaterii standard σ , de la o variabilă cantitativă, iar H_{rel} este omologul coeficientului de variabilitate $CV\%$ de la aceeași variabilă cantitativă. Spre deosebire de $CV\%$, care este exprimat procentual, H_{rel} este exprimat ca raport.
6. H_{rel} este cu atât mai mare cu cât distribuția este mai apropiată de distribuția uniformă.

-
- ✓ În ecologie distribuția uniformă se numește distribuție regulată sau echitabilă. De aceea H_{rel} este o bună măsură a conceptului de **Regularitate** sau **echitabilitate** utilizat în ecologie și notat cu **R**.
 - ✓ Prin H_{rel} putem compara două biocenoze cu număr egal sau diferit de specii, în timp ce prin H_p , nu putem compara decât biocenoze cu același număr de specii.
-

4° Calcul manual entropie în biți și entropie relativă

Exemplul 3.8.2.

Fie distribuția de abundențe absolute:

$$\begin{pmatrix} x_j \\ N_j \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & S_3 \\ 5 & 3 & 2 \end{pmatrix}$$

Calcul entropie în biți folosind tablele cu valori $-F \cdot \log_2 F$ (vezi anexa 2):

Calculăm $N = \sum N_j = 5 + 3 + 2 = 10$

și distribuția de frecvențe relative $\begin{pmatrix} x_j \\ F_j \end{pmatrix}$ unde $F_j = N_j / N$:

$$\begin{pmatrix} S_1 & S_2 & S_3 \\ \frac{5}{10} & \frac{3}{10} & \frac{2}{10} \end{pmatrix} = \begin{pmatrix} S_1 & S_2 & S_3 \\ 0,5 & 0,3 & 0,2 \end{pmatrix}$$

Entropia în biți va fi:

$H_3 = - \sum F_j \cdot \log_2 F_j = -0,5 \cdot \log_2 0,5 - 0,3 \cdot \log_2 0,3 - 0,2 \cdot \log_2 0,2 =$ (vezi în tabel valorile funcției $-F \cdot \log_2 F$ pentru $F = 0,5; 0,3$ și $0,2$) $= 0,5 + 0,5211 + 0,4644 = 1,4855$ biți.

[Pentru eventuale transformări în niți și diți utilizăm formulele:

✓ 1 bit = 1 / 1,442695 niți,

✓ 1 bit = 1 / 3,321928 diți.]

Deci $H_3 = 1,4855$ biți = 1,02967 niți = 0,4471801 diți.

Calcul entropie relativă

$H_{rel} = H_3 / H_{3, max}$ unde $H_{3, max} = \log_2 p$ dacă H_3 e măsurat în biți și p este numărul de specii (aici $p = 3$).

$H_{max} = \log_2 3 = 1,584962$ biți.

[În general $\log_2 p$ se poate calcula, dacă avem tabele sau calculator cu funcțiile **ln** sau **lg**, prin relațiile $\log_2 p = 1,442695 \cdot \ln p = 3,321928 \cdot \lg p$]

$H_{rel} = H_3 / H_{3, max} = 1,4855$ biți / 1,584962 biți $\approx 0,94$.

5° Interpretarea ecologică a diversității specifice

Conceptul ecologic complex de biodiversitate cuprinde printre altele și componenta "diversitate specifică a unei biocenoze". Aceasta se exprimă pornind de la distribuția de frecvențe ale speciilor care alcătuiesc biocenoza respectivă. Una dintre ideile cele mai răspândite este aceea că diversitatea specifică nu poate fi exprimată, sintetizată într-un singur număr, ci trebuie avute în vedere două entități distincte:

- *numărul de specii* și
- *regularitatea* distribuției lor de frecvențe.

Regularitatea poate fi măsurată prin diverși indicatori, cum ar fi, de exemplu, entropia relativă.

Interpretarea diversității specifice (Legendre L. & Legendre P.) [22]

1. Numărul de specii p depinde în mod direct de stabilitatea mediului și de capacitatea de suport a sa, am adăuga noi².

Se consideră că stabilitatea antrenează un grad mai ridicat de organizare sau de complexitate a nivelelor trofice, altfel spus că un mediu mai stabil va constitui mai multe nișe, deci mai multe specii. Numărul p este evident proporțional cu numărul de nișe, căci prin definiție **nișa** ocupată de o specie este "*ansamblul condițiilor mediului pe care acea specie nu le împărtășește cu nici o altă specie simpatrică*" [22]. Această definiție prezintă un interes special deoarece permite conexarea conceptului de diversitate specifică de cel al diversității mediului.

2. Regularitatea distribuției de specii depinde în mod invers de activitatea biologică din mediu.

Adică, cu cât este mai crescută activitatea biologică măsurată în producție, cicluri vitale, transfer de energie de la un nivel trofic la altul etc., cu atât va fi mai scăzută regularitatea.

O situație interesantă este următoarea. Fie un mediu în care competiția interspecifică este scăzută (deci regularitatea este mare). Se produce o diminuare a resurselor, de exemplu din motive sezoniere - schimbare de climă. Aceasta implică o creștere a competiției și favorizarea anumitor specii în detrimentul altora, ceea ce va diminua regularitatea. Legendre L & Legendre P. conchid că acesta este și cazul poluării.

² De exemplu într-un mediu polar, orcât de stabil ar fi, nu vom pute găsi un număr mai mare de specii decât într-un mediu tropical. Diferența este dată de capacitățile de suport ale celor două medii.

- ✓ După părerea noastră, numărul de specii, p , exprimă capacitatea de suport a biotopului corespunzător biocenozei respective, evident în condiții de stabilitate suficient de îndelungate pentru ca acea capacitate de suport să se poată exprima.

Problema 3.8.2.

O biocenoză a fost studiată la 3 momente de timp diferite. Numerele de indivizi din fiecare specie prezentă au fost următoarele:

| specia | <i>a</i> | <i>b</i> | <i>c</i> | <i>d</i> | <i>e</i> | <i>f</i> |
|-------------------|----------|----------|----------|----------|----------|----------|
| momentul 1 | 40 | 120 | 10 | 10 | 10 | 10 |
| momentul 2 | 25 | 70 | 1 | 2 | 1 | 1 |
| momentul 3 | 27 | 258 | 0 | 15 | 0 | 0 |

- a. Să se calculeze, pentru **momentul 3**, prin formula lui Shannon, diversitatea specifică măsurată în *biți*, precum și entropia relativă ca măsură a echitabilității (regularității, R). Calculul se va face manual, folosind tabelele cu valorile $-F \cdot \log_2 F$. Se dă $\log_2 3 = 1,58$.
- b. Știind că entropia, în *biți* (H), respectiv regularitatea (R) pentru primele două momente au valorile:

| | H | R |
|-------------------|------|------|
| momentul 1 | 1,77 | 0,59 |
| momentul 2 | 1,17 | 0,45 |

să se compare biocenoza la **momentul 1** cu cea de la **momentul 2**, respectiv **momentul 2** cu **momentul 3**, interpretându-se ecologic, după concepția lui fraților Legendre și în acord cu observația anterioară, componentele diversității și oferindu-se astfel o descriere a dinamicii biocenozei și a biotopului respectiv.

Rezolvare:

a.

- Frecvențele absolute ale distribuției de abundențe sunt $N_1 = 27$, $N_2 = 258$, $N_3 = 15$.
- Total indivizi $N = N_1 + N_2 + N_3 = 27 + 258 + 15 = 300$.
- Frecvențele relative sunt $F_1 = 27 / 300 = 9 / 100 = 0,09$; $F_2 = 258 / 300 = 86 / 100 = 0,86$; $F_3 = 15 / 300 = 5 / 100 = 0,05$ \Rightarrow
 $F_2 = 1 - (0,09 + 0,05) = 1 - 0,14 = 0,86$.
- $H = \sum -F_j \cdot \log_2 F_j = -F_1 \cdot \log_2 F_1 - F_2 \cdot \log_2 F_2 - F_3 \cdot \log_2 F_3 =$
 $= -0,09 \cdot \log_2 0,09 - 0,86 \cdot \log_2 0,86 - 0,05 \cdot \log_2 0,05 =$
 $= (\text{vezi anexa 2}) \approx 0,31 + 0,19 + 0,22 = 0,72$.
- $H_{3, \max} = \log_2 3 = 1,58 \Rightarrow R = H / H_{3, \max} = 0,72 / 1,58 \approx 0,46$.

b.

- Comparație **momentul 1** cu **momentul 2**:
 $p_1 = p_2 (= 6) \Rightarrow$ Aceeași stabilitate a mediului în condiții de capacități de suport asemănătoare.
 $R_1 > R_2 \Rightarrow$ Probabil o diminuare a resurselor, căci a crescut competiția cu favorizarea unor specii în detrimentul altora. Alarmă, eventual poluare.
- Comparație **momentul 2** cu **momentul 3**:
 $p_2 > p_3 \Rightarrow$ Scăderea capacității de suport a mediului. Degradare, șoc major în biotop: nișe mai puține,
 $R_2 = R_3 \Rightarrow$ dar noul biotop - deși mai sărac - este ocupat în aceeași măsură.

- ✓ În încheierea paragrafului 3.8., dedicat unei variabile calitative observăm că aceasta nu poate fi tratată decât în cele două moduri prezentate mai sus. Astfel, de exemplu, pentru o serie de grupe sanguine putem calcula doar (1) proporția unei grupe, respectiv complementul față de 1 - ca indicatori de tendință centrală - și produsul acestora - ca indicator de împrăștiere - sau (2) entropia și entropia relativă, ca măsuri ale diversității. În nici-un caz nu are sens să vorbim despre medie, modă sau mediană pentru o variabilă calitativă.