AZ ENTRÓPIA FOGALMÁRÓL

BALATONI JÁNOS és RÉNYI ALFRÉD

Bevezetés

Az entrópia fogalma száz évvel ezelőtt a termodinamikában alakult ki és R. CLAUSIUS nevéhez fűződik [1]. Ugyancsak CLAUSIUS-tól származik a termodinamika második főtétele, amely szerint egy zárt rendszer entrópiája állandóan növekszik. Az entrópia fogalmát és a második főtételt egyaránt új megvilágításba helyezte L. BOLTZMANN [2], aki egyrészt kimutatta, hogy az entrópia az állapot valószínűségének logaritmusával arányos, és ennek alapján bebizonyította, hogy egy zárt rendszer entrópiája nem bizonyosan, hanem csak igen nagy valószínűséggel növekszik, feltéve, hogy a rendszer kicsiny valószínűségű állapotban tartózkodik. BOLTZMANN másik nagy érdeme az általa H-függvénynek nevezett fogalom bevezetése. BOLTZMANN eredményei a termodinamikai entrópia-fogalmat lényegében teljesen tisztázták. BOLTZMANN zseniális gondolatait azonban az ő korában kevesen értették meg, és elmélete körül hosszú ideig folyt a vita ; a felmerült ellenvetésekről, amelyek közül a legismertebb az úgynevezett LOSCHMIDT-féle „Umkehreinwand" (irreverzibilitási ellenvetés; erről az alábbiakban lesz még szó), kiderült, hogy félreértésen alapulnak. A kérdés tisztázásához nagyban hozzájárultak P. és T. EHRENFEST munkái [3]. Az 1. §-ban röviden kitérünk a kérdésnek a valószínűségszámítás alapján való megvilágítására. Jelen dolgozat főcélja azonban nem a statisztikus mechanika entrópia-fogalmának tisztázása — ezt lényegében BOLTZMANN elvégezte — hanem a valószínűségszámítási entrópia-fogalom tisztázása. Ez a kérdés a legutóbbi években az információelmélettel kapcsolatban került az érdeklődés homlokterébe.

Az információelméletbe az entrópia fogalmát elsőnek C. SHANNON [4], [5] vezette be és mutatta ki annak alapvető jelentőségét az információ továbbításának matematikai tárgyalásával kapcsolatban.

A BOLTZMANN-féle H-függvény és az információelméleti entrópia egy általánosabb fogalom, a valószínűségszámítási entrópia-fogalom speciális alkalmazásai. Dolgozatunkban az általános valószínűségszámítási entrópiafogalommal foglalkozunk. A statisztikus mechanika entrópia-fogalma, a BOLTZMANN-féle H-függvény és a valószínűségszámítási entrópia-fogalom összefüggésével az 1. §-ban foglalkozunk.

SHANNON meggondolásai matematikai szempontból nem voltak teljesen kidolgozottak. A kérdés első, matematikailag teljesen szabatos tárgyalását A. J. HINCSIN adta meg [6], a diszkrét valószínűség-eloszlásokra vonatkozólag. Tetszőleges valószínűség-eloszlások entrópiájával HINCSIN nem foglalkozik, bár SHANNON vizsgálataiban és általában az információelméletben folytonos eloszlások entrópiája szerepet játszik, és ezzel kapcsolatban a kérdés tüzetes matematikai tisztázásának hiánya még szembetűnőbb, mint a diszkrét valószínűség-eloszlások esetében.

Jelen dolgozat szerzői azt a célt tűzték ki maguk elé, hogy HINCSIN dolgozatából kiindulva az ő általa megadott tárgyalásmódot kiterjesszék tetszőleges valószínűség-eloszlásokra. A vizsgálatok során kitűnt, hogy ehhez a valószínűségszámítási entrópia-fogalom lényeges továbbfejlesztésére van szükség. Mielőtt a kapott eredményeket röviden összefoglalnánk, néhány szóval meg kívánjuk világítani a valószínűségszámítási entrópia-fogalom jelentését. Egy tetszőleges valószínűség-eloszlás úgy fogható fel, hogy megadja, hogy valamely, a véletlentől függő szituáció (kísérlet, megfigyelés, stb.) lehetséges kimenetelei (a lehetséges események) között a bizonyosság egységnyi valószínűsége hogyan oszlik meg. Minden valószínűségeloszlás felfogható mint egy mérték a lehetséges események terében. Mi itt csak olyan valószínűség-eloszlásokkal foglalkozunk, amelyeknél a bekövetkezendő esemény egy vagy több valós számmal, tehát az n-dimenziós euklidesi tér egy pontjával jellemezhető. Másszóval, kizárólag valószínűségi változók vagy általában vektor-változók valószínűség-eloszlásával, vagyis véges-dimenziós euklidesi terekben megadott valószínűség-eloszlásokkal foglalkozunk. A vizsgálatok kiterjesztése tetszőleges elemekből álló eseményterek valószínűségeloszlásaira elég kézenfekvő; erre a kérdésre egy további dolgozatban kívánunk visszatérni.

Vizsgáljuk először egy ξ valószínűségi változó eloszlását. Ha ez az eloszlás nem elfajult, vagyis ξ nem egyenlő bizonyosan (1 valószínűséggel) egy c állandóval, akkor ξ értékére vonatkozólag bizonyos *bizonytalanság* áll fenn. A kérdés, amelyre a valószínűségszámítási entrópia-fogalom választ ad, a következő: hogyan lehet a ξ értékére vonatkozó bizonytalanságot egy olyan mérőszámmal jellemezni, amely csak ξ valószínűség-eloszlásától függ, olymódon, hogy bármely két valószínűségi változó (illetve bármely két valószínűség-eloszlás) bizonytalanságát össze tudjuk hasonlítani és meg tudjuk mondani, hogy melyikre vonatkozólag nagyobb a bizonytalanság. Egy valószínűség-eloszlás entrópiáján az illető eloszlással bíró valószínűségi változó véletlentől függő értékére vonatkozó bizonytalanság mértékét értjük, vagyis röviden: *a valószínűségszámításban az entrópia a bizonytalanság mértékszáma.*

Ha ξ véges diszkrét eloszlású, mégpedig az $x_1, x_2, \ldots, x_n$ értékeket ($x_i \neq x_j$, ha $i \neq j$) rendre $p_1, p_2, \ldots, p_n$ valószínűséggel veszi fel (vagyis $\mathbf{P}(\xi = x_k) = p_k$, $k = 1,2, \ldots, n$, és így $p_k \geq 0$ és $\sum_{k=1}^{n} p_k = 1$), akkor, mint HINCSIN kimutatta, bizonyos egyszerű és természetes követelményekből kiindulva a ξ változó értékére vonatkozó „bizonytalanság" mértékének, amelyet $\mathbf{H}_0(\xi)$-vel jelölünk és ξ entrópiájának nevezünk, a következőnek kell lennie:

$$\mathbf{H}_0(\xi) = -\lambda \sum_{k=1}^{n} p_k \log p_k ,$$

ahol λ pozitív szám, amelynek megválasztása önkényes, konvenció kérdése. λ megválasztása tulajdonképpen az entrópia egységének megválasztásával ekvivalens, ami ugyanabban a mértékben konvenció kérdése, mint egy fizikai mennyiség mértékegységének megválasztása. Az információelméletben λ-nak az $1/\log 2$ értékét szokták adni, ami azt jelenti, hogy egy diszkrét valószínűségeloszlás entrópiáját a

$$\mathbf{H}_0(\xi) = -\sum_{k=1}^{n} p_k \log_2 p_k \tag{1}$$

SHANNON-féle képlettel definiálják, ahol $\log_2$ a 2 alapú logaritmust jelöli. Ez azt jelenti, hogy az entrópia (bizonytalanság) egységéül az egyszerű alternatívában rejlő bizonytalanságot[1] választják, tehát annak a ξ valószínűségi változónak tulajdonítanak egységnyi entrópiát, amely az a és b értéket ($a \neq b$) 1/2 valószínűséggel veszi fel.

Ilyen módon például egy pénzdarab feldobásának (a „fej vagy írás" kísérletnek) vagy egy a (0,1) intervallumban találomra (egyenletes eloszlással) választott szám diadikus kifejtése meghatározott sorszámú jegyének az entrópiája 1-gyel egyenlő.

Az (1)-gyel definiált entrópia a bizonytalansági mértékszámtól megkívánható összes tulajdonságokkal rendelkezik.

Ezek a tulajdonságok a következők:

1. Ha $\mathbf{P}(\xi = c) = 1$, vagyis ha ξ 1 valószínűséggel állandó, akkor $\mathbf{H}_0(\xi) = 0$, minden más esetben $\mathbf{H}_0(\xi) > 0$; ha ξ és η egyforma eloszlású valószínűségi változók, $\mathbf{H}_0(\xi) = \mathbf{H}_0(\eta)$, vagyis $\mathbf{H}_0(\xi)$ csak ξ eloszlásától függ.

2.1. Ha $\eta = f(\xi)$, ahol $f(x)$ egyrétű függvény, vagyis ha $x \neq x'$, akkor $f(x) \neq f(x')$, akkor $\mathbf{H}_0(\eta) = \mathbf{H}_0(\xi)$, vagyis ξ értékei közömbösek az entrópia szempontjából, csak az számít, hogy a különböző értékeknek mekkora a valószínűsége.

2.2. Ha $\eta = f(\xi)$, ahol $f(x)$ egy tetszőleges függvény, akkor $\mathbf{H}_0(\eta) \leq \mathbf{H}_0(\xi)$, mégpedig, ha legalább két érték, pl. x_k és x_l olyan, hogy $x_k \neq x_l$, $p_k > 0$ és $p_l > 0$, továbbá $f(x_k) = f(x_l)$, akkor $\mathbf{H}_0(\eta) < \mathbf{H}_0(\xi)$[2].

3.1. Legyen ξ és η két diszkrét, véges valószínűségi változó, $\mathbf{P}(\xi = x_j) = p_j$ $(j = 1, 2, \ldots, m)$ és $\mathbf{P}(\eta = y_k) = q_k$ $(k = 1, 2, \ldots, n)$, és az $\eta = y_k$ feltétel mellett ξ feltételes valószínűség-eloszlásának entrópiája legyen

$$\mathbf{H}_0(\xi \mid \eta = y_k) = -\sum_{j=1}^{m} p_{j|k} \log_2 p_{j|k}, \tag{2}$$

ahol $p_{j|k} = \mathbf{P}(\xi = x_j \mid \eta = y_k)$, és jelölje $\mathbf{H}_0(\xi, \eta)$ a ξ és η változók együttes eloszlásának, vagyis az

$$r_{j,k} = \mathbf{P}(\xi = x_j,\ \eta = y_k) \qquad (j = 1, 2, \ldots, m;\ k = 1, 2, \ldots, n)$$

---

[1] Az entrópia (bizonytalanság) egységét az angol nyelvű szakirodalomban „bit"-nek nevezik, ami a „binary digit" (diadikus számjegy) rövidítése.

[2] Ez abból következik, hogy ha $x > 0$, $y > 0$ és $x + y \leq 1$, akkor

$$x \log \frac{1}{x} + y \log \frac{1}{y} \geq (x + y) \log \frac{1}{x + y}.$$

valószínűség-eloszlásnak (másszóval a $\vec{\zeta} = (\xi, \eta)$ síkbeli valószínűségi vektorváltozó eloszlásának) az entrópiáját, vagyis legyen

$$(3) \qquad \mathbf{H}_0(\xi, \eta) = -\sum_{j=1}^{m}\sum_{k=1}^{n} r_{j,k} \log_2 r_{j,k}.$$

Legyen továbbá $\chi = \mathbf{H}_0(\xi|\eta)$ az a valószínűségi változó, amely a $\mathbf{H}_0(\xi|\eta = y_k)$ értéket q_k valószínűséggel veszi fel ($k = 1, 2, \ldots, n$), és legyen $\overline{\mathbf{H}_0(\xi|\eta)}$ a χ valószínűségi változó várható értéke, vagyis legyen

$$(4) \qquad \overline{\mathbf{H}_0(\xi|\eta)} = \sum_{k=1}^{n} q_k \mathbf{H}_0(\xi|\eta = y_k),$$

akkor

$$(5) \qquad \mathbf{H}_0(\xi, \eta) = \mathbf{H}_0(\eta) + \overline{\mathbf{H}_0(\xi|\eta)}.$$

Speciálisan, ha ξ és η függetlenek, akkor

$$\mathbf{H}_0(\xi|\eta = y_k) = \mathbf{H}_0(\xi), \ (k = 1, 2, \ldots, n) \text{ és így } \overline{\mathbf{H}_0(\xi|\eta)} = \mathbf{H}_0(\xi).$$

vagyis ez esetben

$$(6) \qquad \mathbf{H}_0(\xi, \eta) = \mathbf{H}_0(\xi) + \mathbf{H}_0(\eta).$$

Ebből következik, hogy ha $\xi_1, \xi_2, \ldots, \xi_n$ független valószínűségi változók és $\mathbf{H}_0(\xi_1, \xi_2, \ldots \xi_n)$ jelöli a $(\xi_1, \xi_2, \ldots, \xi_n)$ valószínűségi változók együttes eloszlásának entrópiáját, akkor

$$(7) \qquad \mathbf{H}_0(\xi_1, \xi_2, \ldots, \xi_n) = \sum_{k=1}^{n} \mathbf{H}_0(\xi_k),$$

vagyis *független valószínűségi változók együttes eloszlásának entrópiája egyenlő az egyes változók entrópiáinak összegével.*

3.2. Legyenek $\xi_1, \xi_2, \ldots, \xi_n$ tetszőleges véges diszkrét eloszlású valószínűségi változók, amelyek azzal a tulajdonsággal bírnak, hogy értékkészleteik idegenek, vagyis

$$\mathbf{P}(\xi_i = x)\, \mathbf{P}(\xi_j = x) = 0,$$

ha $i \neq j$, bármely x-re ($i, j = 1, 2, \ldots, n$), vagyis bármely x számot a ξ_k változók közül legfeljebb az egyik vehet fel pozitív valószínűséggel. Képezzük ezen eloszlások keverékét a $q_1, q_2, \ldots, q_n$ súlyokkal, vagyis az η valószínűségi változó legyen q_k valószínűséggel egyenlő ξ_k-val ($k = 1, 2, \ldots, n$); akkor

$$(8) \qquad \mathbf{H}_0(\eta) = \sum_{k=1}^{n} q_k \mathbf{H}_0(\xi_k) - \sum_{k=1}^{n} q_k \log_2 q_k,$$

vagyis *a keverékeloszlás entrópiája egyenlő a keverék komponensei entrópiájának a keverő eloszlással súlyozott középértékének és a keverő eloszlás entrópiájának összegével, feltéve, hogy a kevert eloszlások idegenek a fenti értelemben* (vagyis ha nincs olyan x szám, amelyet a ξ_k változók közül egynél több venne fel pozitív valószínűséggel).

Megjegyzendő, hogy a 3.2. tulajdonság a 3.1. tulajdonságból következik. Ugyanis, ha $\zeta = k$, feltéve, hogy $\eta = \xi_k$ $(k = 1, 2, \ldots, n)$, akkor η és ζ együttes eloszlása azonos η eloszlásával, és így (6) szerint

$$\mathbf{H}_0(\eta) = \mathbf{H}_0(\eta, \zeta) = \mathbf{H}_0(\zeta) + \overline{\mathbf{H}_0(\eta|\zeta)}\,; \tag{9}$$

továbbá

$$\overline{\mathbf{H}_0(\eta|\zeta)} = \sum_{k=1}^{n} q_k \mathbf{H}_0(\xi_k) \tag{10}$$

és

$$\mathbf{H}_0(\zeta) = -\sum_{k=1}^{n} q_k \log_2 q_k\,. \tag{11}$$

A (9), (10) és (11) összefüggésekből következik (8).

4. $\mathbf{H}_0(\xi)$ folytonosan függ ξ eloszlásától, vagyis bármely pozitív ε-hoz és bármely rögzített N természetes számhoz megadható olyan $\delta > 0$ szám, hogy ha

$$\mathbf{P}(\xi = x_k) = p_k\,, \qquad k = 1,2, \ldots, N\,; \qquad (x_i \neq x_k, \ \text{ha} \ i \neq k)$$

és

$$\mathbf{P}(\eta = y_k) = q_k\,, \qquad k = 1,2, \ldots, N\,; \qquad (y_i \neq y_k, \ \text{ha} \ i \neq k)$$

továbbá

$$|p_k - q_k| \leq \delta\,, \qquad k = 1,2, \ldots, N\,,$$

akkor

$$|\mathbf{H}_0(\xi) - \mathbf{H}_0(\eta)| \leq \varepsilon.$$

5. Ha ξ lehetséges értékeinek száma, n, adott, akkor ξ entrópiája akkor maximális, ha ξ az n különböző érték mindegyikét ugyanazzal a valószínűséggel (tehát $1/n$ valószínűséggel) veszi fel, vagyis ha ξ lehetséges értékei $x_1, x_2, \ldots, x_n$, akkor

$$\mathbf{H}_0(\xi) \leq \log_2 n\,. \tag{12}$$

Hincsin bebizonyította, hogy a 2.1., 3.1., 4. és 5. tulajdonságok az entrópiát egy pozitív konstans faktortól eltekintve meghatározzák. A 3. §-ban meg fogjuk mutatni, hogy az 1., 2.1., 3.2. és 4. tulajdonságok is meghatározzák az entrópiát egy konstans faktortól eltekintve.

Ha ξ végtelen diszkrét eloszlású, vagyis $\mathbf{P}(\xi = x_k) = p_k$ $(k = 1, 2, \ldots)$, definiálhatjuk ξ entrópiáját a

$$\mathbf{H}_0(\xi) = -\sum_{k=1}^{\infty} p_k \log_2 p_k \tag{13}$$

képlettel, ha a (13) jobboldalán álló sor konvergens; ebben az esetben azonban lehetséges, hogy a (13) jobboldalán álló sor divergens lesz; legyen ekkor definiciószerűen $\mathbf{H}_0(\xi) = +\infty$, vagyis ilyenkor azt mondjuk, hogy ξ entrópiája végtelen nagy.[3]

Megemlítjük még az entrópia következő fontos tulajdonságait, amelyek elvezetnek az entrópia-fogalom információelméleti jelentőségének megértéséhez.

[3] E dolgozatban nem foglalkozunk végtelen nagy entrópiájú valószínűségeloszlásokkal.

6.1. Ha ξ és η tetszőleges diszkrét eloszlású valószínűségi változók, akkor

$$\mathbf{H}_0(\xi, \eta) \leqq \mathbf{H}_0(\xi) + \mathbf{H}_0(\eta) ; \tag{14}$$

egyenlőség csak akkor áll fenn, ha ξ és η függetlenek.

Ez a tulajdonság a következő ekvivalens alakra is hozható :

6.2. Ha ξ és η tetszőleges diszkrét eloszlású valószínűségi változók

$$\overline{\mathbf{H}_0(\xi \mid \eta)} \leqq \mathbf{H}_0(\xi) . \tag{15}$$

A

$$\Delta(\xi, \eta) = \mathbf{H}_0(\xi) + \mathbf{H}_0(\eta) - \mathbf{H}_0(\xi, \eta) = \mathbf{H}_0(\xi) - \overline{\mathbf{H}_0(\xi \mid \eta)} \tag{16}$$

mennyiség információelméleti jelentősége a következő : Ha a ξ valószínűségi változó a leadott jel, és η a felvett jel (amely nem azonos ξ-vel zavaró körülmények, pl. zaj folytán), akkor $\Delta(\xi, \eta) = \mathbf{H}_0(\xi) - \overline{\mathbf{H}_0(\xi|\eta)}$ megadja, hogy mennyivel csökken a ξ értékére vonatkozó bizonytalanság átlagban azáltal, hogy η értékét megfigyeljük ; $\Delta(\xi, \eta)$ tehát a ξ-re vonatkozólag η-ból nyert információ mennyiségeként értelmezhető. A (15) egyenlőtlenség tehát úgy értelmezhető, hogy η megfigyelése mindig nyújt valami információt ξ-re nézve, a ξ értékére vonatkozó bizonytalanság η megfigyelésével átlagban csak csökkenhet. Ha $\eta = f(\xi)$, ahol $f(x) \neq f(x')$, ha $x \neq x'$, vagyis, ha különböző leadott jeleknek különböző felvett jelek felelnek meg (tehát η értékéből ξ értékére egyértelműen lehet következtetni), akkor $\mathbf{H}_0(\xi|\eta) = 0$, hiszen adott η mellett ξ 1 valószínűséggel állandó, tehát $\Delta(\xi, \eta) = \mathbf{H}_0(\xi)$, vagyis η megfigyelése *teljes* információt nyújt ξ-re nézve. A másik véglet, amikor ξ és η függetlenek ; ez esetben $\mathbf{H}_0(\xi|\eta = y_k) = \mathbf{H}_0(\xi)$, tehát $\overline{\mathbf{H}_0(\xi \mid \eta)} = \mathbf{H}_0(\xi)$ és így $\Delta(\xi, \eta) = 0$; ez esetben tehát η megfigyelése ξ-re vonatkozólag semmi információt nem ad ; ez szemléletesen is evidens, hiszen a függetlenség éppen azt jelenti, hogy semmilyen kapcsolat ξ és η értékei között nem áll fenn. Érdekes megjegyezni, hogy $\Delta(\xi, \eta) = \Delta(\eta, \xi)$, tehát η ugyanannyi információt nyújt ξ-re nézve, mint ξ η-ra nézve.

7. Ha $\xi_1, \xi_2, \ldots, \xi_n$ tetszőleges, diszkrét eloszlású valószínűségi változók, és $q_1, q_2, \ldots, q_n$ tetszőleges pozitív súlyok, $\sum_{k=1}^{n} q_k = 1$, továbbá ζ eloszlása a ξ_k változók eloszlásainak q_k súlyokkal vett keveréke, akkor

$$\mathbf{H}_0(\zeta) \leq \sum_{k=1}^{n} q_k \mathbf{H}_0(\xi_k) - \sum_{k=1}^{n} q_k \log_2 q_k \tag{17}$$

és egyenlőség csak akkor áll fenn, ha a ξ_k változók tényleges értékkészletei idegenek.

(17) úgy látható be legegyszerűbben, hogy először a ξ_k változók értékkészletét úgy módosítjuk, hogy azok idegenek legyenek ; ez esetben (17)-ben egyenlőség áll fenn ; ezután visszatérünk az eredeti ξ_k változókra ; eközben (17) jobboldala változatlan, baloldala azonban a 2.2. tulajdonság értelmében csökken.

8. Legyenek egy kísérlet lehetséges kimenetelei az $A_1, A_2, \ldots, A_n$ események és legyen A_k valószínűsége p_k $(k = 1, 2, \ldots, n)$. Ha egyszer végre-

hajtjuk a kísérletet és annak eredménye az A_k esemény, az eredmény valószínűségének reciprok értékének 2 alapú logaritmusa, $-\log_2 p_k$ valószínűségi változónak tekinthető. Ennek várható értéke, $-\sum_{k=1}^{n} p_k \log_2 p_k$, éppen a kísérlet entrópiája. Az entrópia ilyen értelmezésének jelentőségét az alábbi meggondolás mutatja meg: Ismételjük meg a kísérletet egymásután N-szer és jelölje N_k azoknak a kísérleteknek a számát, amelyeknél az A_k esemény következett be. Egy ilyen kísérletsoroztnak az eredmények sorrendjét is figyelembevéve a valószínűsége nyilván $P = p_1^{N_1} \cdot p_2^{N_2} \ldots, p_n^{N_n}$. Mivel az $N_1, N_2, \ldots N_n$ számok valószínűségi változók, maga P is annak tekinthető. Számítsuk ki $\log_2 P$ várható értékét:

$$\overline{\log_2 P} = \sum \overline{N_k} \log_2 p_k = N \sum p_k \log_2 p_k = -NH_0,$$

ahol H_0 jelenti a kísérlet entrópiáját. Mivel az N_k változók között negatív korreláció áll fenn, ha $\mathbf{D}^2(\xi)$ jelenti a ξ valószínűségi változó szorzásnégyzetét,

$$\mathbf{D}^2(\log_2 P) \leqq \sum \mathbf{D}^2(N_k) \log_2^2 p_k = N \sum_{k=1}^{n} p_k(1 - p_k) \log_2^2 p_k .$$

Ennélfogva $\log_2 P$ a $-NH_0$ középérték körül általában $\sqrt{N}$ nagyságrendű ingadozásokat végez, vagyis $(\log_2 P)/N$ általában igen közel lesz $-H_0$-hoz, vagyis nagy valószínűséggel

$$P \sim 2^{-NH_0}.$$

A Csebisev-egyenlőtlenség alkalmazásával könnyen belátható, hogy körülbelül 2^{NH_0}, egyenként közelítőleg 2^{-NH_0} valószínűségű lehetséges eredménysorozat van, a többi $n^N - 2^{NH_0}$ lehetséges eredménysorozat együttvéve is kicsiny valószínűségű. Röviden úgy lehetne jellemezni a helyzetet, hogy a kísérletet N-szer megismételve, ha N nagy szám, a helyzet közelítőleg olyan lesz, mintha minden egyes kísérletnek 2^{H_0} számú lehetséges, egyforma valószínűségű kimenetele volna, vagyis a szóbanforgó kísérlet eredménye körülbelül ugyanannyira bizonytalan, mint egy olyan kísérlet eredménye, amelynek 2^{H_0} számú egyformán valószínű eredménye lehetséges.

Megjegyzendő, hogy ha ahelyett, hogy $-\log_2 P$ várható értékét vesszük, tekintjük P várható értékének reciprokának logaritmusát, $H_0 = -\sum_{k=1}^{n} p_k \log_2 p_k$ helyett a $\varrho_0 = -\log_2 \sum_{k=1}^{n} p_k^2$ kifejezésre jutunk. Abban az esetben, ha $p_k = 1/n$ $(k = 1, 2, \ldots, n)$, $\varrho_0 = H_0$, egyébként (a logaritmusfüggvény konkávitása folytán) $\varrho_0 < H_0$. ϱ_0 is tekinthető a bizonytalanság mértékének, ez azonban kevésbé érzékeny, mint H_0, és ezért nem olyan alkalmas a bizonytalanság mérésére, mint H_0.

Ha ξ abszolút folytonos eloszlású valószínűségi változó és ξ sűrűségfüggvénye $f(x)$, akkor SHANNON ξ entrópiáját a

$$(18) \qquad -\int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx$$

kifejezéssel definiálja. A (18) kifejezés bizonyos hasonlóságot mutat az (1) kifejezéssel, azonban attól sok szempontból különbözik. Így például a (18) definíció szerint a $(0, a)$ intervallumban egyenletes valószínűség-eloszlás entrópiája $\log_2 a$, tehát egy folytonos eloszlás entrópiája *negatív* értékeket is felvehet, hiszen ha $0 < a < 1$, akkor $\log_2 a < 0$. Ebből a megjegyzésből is látszik, hogy a diszkrét és folytonos eloszlás entrópiái nem tekinthetők ugyanazon fajta mennyiségeknek és nem hasonlíthatók egyszerűen számértékük alapján össze, hiszen lehetetlen, hogy a $(0, a)$ intervallumban egyenletes eloszlás bizonytalanságának mértéke, ha $a < 1$, kisebb legyen, mint egy konstans entrópiája. Már ebből a megjegyzésből sejteni lehet, hogy egy abszolút folytonos eloszlás entrópiája *magasabb* „*dimenziójú*", mint egy diszkrét eloszlás entrópiája: bármely abszolút folytonos eloszlás entrópiája, tekintet nélkül a (18) képlet által megadott számértékre, „nagyobb", mint bármely diszkrét eloszlás entrópiája, tekintet nélkül utóbbi számértékére.[4] Erre való tekintettel a (18) kifejezés által definiált entrópiát egydimenziójú entrópiának nevezzük — szemben a diszkrét eloszlások entrópiájával, amelyet 0-dimenziójú entrópiának nevezünk, és a ξ abszolút folytonos eloszlású valószínűségű változó (18) által definiált entrópiáját $\mathbf{H}_1(\xi)$-vel jelöljük; vagyis ha ξ sűrűségfüggvénye $f(x)$,

$$\mathbf{H}_1(\xi) = -\int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx . \tag{19}$$

Ezen összehasonlításra való tekintettel jelöltük kezdettől fogva a diszkrét eloszlású valószínűségi változók entrópiáját $\mathbf{H}_0$-lal.

A tüzetesebb vizsgálat kimutatta, hogy a 0 dimenziójú és 1 dimenziójú entrópián kívül még végtelen sok másfajta entrópia is van. Bármely ξ valószínűségi változóhoz, amelynek van entrópiája, tartozik egy meghatározott d szám $(0 \leq d \leq 1)$, amelyet ξ eloszlása dimenziójának nevezünk, továbbá tartozik egy $\mathbf{H}_d(\xi)$ számérték (ha $d = 0$, $0 \leq \mathbf{H}_0(\xi) \leq +\infty$, ha $0 < d \leq 1$, akkor $-\infty \leq \mathbf{H}_d(\xi) \leq +\infty$), amelyet d-dimenziójú entrópiának nevezünk. Ha ξ eloszlása d-dimenziójú és $0 \leq c < d$, akkor $\mathbf{H}_c(\xi) = +\infty$ ha pedig $d < c \leq 1$, akkor $\mathbf{H}_c(\xi) = -\infty$. Minden 0 és 1 közé eső d számhoz tartoznak olyan valószínűségi változók, amelyek eloszlása pontosan d dimenziójú, rögzített d érték mellett a $\mathbf{H}_d(\xi)$ dimenziójú entrópia számértéke még $-\infty$-től $+\infty$-ig változik; ha $0 \leq d_1 < d_2 \leq 1$ és ξ és η két olyan valószínűségi változó, hogy ξ eloszlása d_1 dimenziójú, η eloszlása pedig d_2 dimenziójú, akkor ξ entrópiája — számértékre való tekintet nélkül — kisebbnek tekintendő, mint η entrópiája, tekintet nélkül utóbbi számértékére.

Ha ξ egy diszkrét eloszlású, η pedig abszolút folytonos eloszlású valószínűségi változó, és ζ eloszlása ξ és η eloszlásának b és $c = 1 - b$ súlyokkal való keverésével származtatható, akkor ζ eloszlása c-dimenziójú.

---

[4] Ezt megvilágítja például a következő megjegyzés is: a (0,1) intervallum egy x pontjának meghatározásához végtelen sok diadikus jegy megadása szükséges; ha ezek mindegyike 1/2 valószínűséggel a 0 és 1/2 valószínűséggel az 1 értéket veszi fel, egymástól függetlenül, akkor, mint jól ismeretes, maga az x szám egyenletes eloszlású lesz a (0, 1) intervallumban. Az egyes jegyek entrópiájának össze kell adódnia, tehát egy a (0,1)-ben egyenletes eloszlású ξ valószínűségi változóra $\mathbf{H}_0(\xi) = 1 + 1 + \ldots + 1 + \ldots = +\infty$. Ezzel szemben $\mathbf{H}_1(\xi) = 0$.

Ilyenmódon tehát a különböző valószínűség-eloszlások entrópiái ábrázolására nem elegendő egyetlen számskála, hanem kontinuum számosságú különböző számskálákra van szükség.

A 2. §-ban azzal foglalkozunk, hogyan lehet megállapítani, hogy egy valószínűségeloszlásnak mi a dimenziója, s ha ez d, akkor d-dimenziójú entrópiájának mi az értéke.

A 3. §-ban azzal foglalkozunk, hogy kimutatjuk, hogyan lehet a dimenzió és az entrópia bizonyos plauzibilis tulajdonságainak posztulálásával az entrópia kifejezését levezetni, vagyis az entrópia általános definícióját ugyanolyan módon megalapozni, ahogyan ezt Hincsin a 0-dimenziójú entrópiára vonatkozólag megtette.[5)]

Az entrópiafogalommal az elmúlt években Szilárd L. [7], Gábor D. [8] és D. Brillouin [9] — [12] más vonatkozásban is foglalkoztak: a fizikai kísérlet bizonytalanságának és a megfigyelt rendszer entrópiájának összefüggését vizsgálták; új szempontokból vizsgálták a kérdést J. Ville [13] és A. Blanc-Lapierre [14] is; túl messze vezetne, ha mindezekre a kérdésekre itt részletesen kitérnénk.

1. §. A statisztikus mechanikai entrópia, a Boltzmann-féle H-függvény és a valószínűségszámítás entrópia-fogalma közötti összefüggés[6)]

E § célja annak tisztázása, hogy a valószínűségszámítási entrópiafogalom hogyan függ össze a statisztikus mechanika entrópia-fogalmával. A kérdést az egyszerűség kedvéért az ideális gáz példáján igyekszünk megvilágítani. Ha N számú, egyenként m tömegű atomból álló egyatomos gáz hőmérséklete T, és a gáz egy V köbtartalmú edénybe van bezárva, akkor — ha a gázt elegendő ideig magára hagyjuk — az atomok ütközése következtében egy olyan állapot fog kialakulni, amelyben az atomok eloszlása az edényben közel egyenletes lesz, és az atomok sebessége közelítőleg a Maxwell-eloszlást követi, vagyis az atomok sebességének komponensei egymástól függetlenek és külön-külön közelítőleg normális eloszlásúak, 0 várható értékkel és $\sigma = \sqrt{kT/m}$ szórással, ahol k a Boltzmann-féle állandó. A gáz egy találomra „kiválasztott" atomjának helye tehát az egész edényben egyenletes eloszlású, 3-dimenziós valószínűségi vektorváltozónak, az atom sebességének komponensei egymástól és az atom helyétől független normális eloszlású valószínűségi változóknak tekinthetők.

A 3. §-ban meg fogjuk mutatni, hogy érvényes a következő két tétel (lásd: a (101) és (105) egyenleteket):

A) Ha a $\vec{\zeta}$ n-dimenziós valószínűségi vektorváltozó sűrűségfüggvénye $f(x)$, ahol $x = (x_1, x_2, \ldots, x_n)$, akkor n-dimenziós entrópiája

---

5) Jelen dolgozat nyomdába adása után jelent meg D. K. Faggyejev egy dolgozata [Д. К. Фаддеев: „К понятии энтропии конечной вероятностной схемы." *Успехи Математических Наук* **11**(1956) 227—231], melyben a 0-dimenziójú entrópiának néhány egyszerű tulajdonságával való jellemzését adja, valamivel kevesebb feltevéssel mint A. J. Hincsin.

6) E §-ban mindenütt természetes logaritmust használunk 2 alapú logaritmus helyett, mert a statisztikus mechanikában ez a szokásos.

$$\mathbf{H}_n(\vec{\zeta}) = -\int f(\boldsymbol{x}) \log f(\boldsymbol{x})\, d\boldsymbol{x} \tag{20}$$

ahol $d\boldsymbol{x} = dx_1\, dx_2 \ldots dx_n$ és az integrál az egész n-dimenziós térre terjesztendő ki. Speciálisan, ha $\vec{\zeta}$ eloszlása egyenletes a V köbtartalmú n-dimenziós tartományban, (20)-ból következik, hogy

$$\mathbf{H}_n(\vec{\zeta}) = \log V. \tag{21}$$

B) Ha $\xi_1, \xi_2, \ldots, \xi_n$ független, abszolút folytonos eloszlású változók, együttes eloszlásuk n-dimenziós entrópiája egyenlő az egyes változók 1-dimenziós entrópiáinak összegével.

Az *A)* és *B)* tételeken kívül szükségünk lesz továbbá a normális eloszlás entrópiájának kifejezésére. Ha ξ normális eloszlású, m várható értékkel és σ szórással, akkor sűrűségfüggvénye $\frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(x-m)^2}{2\sigma^2}}$, tehát

$$\mathbf{H}_1(\xi) = \int\limits_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{x^2}{2\sigma^2}} \left(\log \sqrt{2\pi}\,\sigma + \frac{x^2}{2\sigma^2}\right) dx = \log \sigma \sqrt{2\pi e}\,. \tag{22}$$

Ennélfogva, ha ξ_1, ξ_2, ξ_3 jelölik egy atom sebességének koordinátáit és $\vec{\zeta}$ a helyzetét jellemző vektort, akkor az atomok, illetve az azt jellemző $\vec{\alpha} = (\xi_1, \xi_2, \xi_3, \vec{\zeta})$ 6-dimenziós valószínűségi vektorváltozónak az entrópiája, (21) és (22), valamint *B)* szerint:

$$\mathbf{H}_6(\vec{\alpha}) = \mathbf{H}_1(\xi_1) + \mathbf{H}_1(\xi_2) + \mathbf{H}_1(\xi_3) + \mathbf{H}_1(\vec{\zeta}) = 3 \log\left(\sigma\sqrt{2\pi e}\right) + \log V. \tag{23}$$

Figyelembe véve, hogy $\sigma = \sqrt{\frac{kT}{m}}$, következik, hogy

$$\mathbf{H}_6(\alpha) = \frac{3}{2} \log T + \log V + \frac{3}{2} \log \frac{2\pi e k}{m}. \tag{24}$$

Mivel minden egyes atomnak ennyi az entrópiája, ha összesen N atom van jelen, az egész Γ gáz entrópiája

$$\mathbf{H}_{6N}(\Gamma) = N \left(\frac{3}{2} \log T + \log V + \frac{3}{2} \log \frac{2\pi e k}{m}\right). \tag{24}$$

Az entrópia termodinamikai definiciója, mint ismeretes, a logaritmus alapjának választásától és a k faktortól eltekintve, ugyanerre az eredményre vezet, vagyis, ha S jelöli a termodinamikai entrópiát,

$$S = kN \left(\frac{3}{2} \log T + \log V + c\right) \tag{26}$$

ahol a c konstans értéke nem függ sem a köbtartalomtól, sem a hőmérséklettől.[7]

Boltzmann [2] az úgynevezett H-függvényt[8] a

$$H = \int f(x) \log f(x)\, dx \tag{27}$$

képlettel definiálta, vagyis (a logaritmus alapjának megválasztásától eltekintve a Boltzmann-féle H-függvény azonos a valószínűségszámítási entrópia (-1)-szeresével.

Az elmondottak a legvalószínűbb állapotra vonatkoztak; azt kapjuk, hogy a Boltzmann-féle H-függvény értékének $(-1/\log 2)$-szerese azonos a rendszer valószínűségszámítási entrópiájával. Most vizsgáljuk a statisztikus mechanikai entrópiát, amelyre a Boltzmann-féle $S = k \log W = - H + C$ képlet érvényes, ahol W a pillanatnyi állapot valószínűsége, H meg az állapothoz tartozó H-függvény. Mivel maga a pillanatnyi állapot valószínűségi változó, az S entrópia és a H-függvény is az. A legvalószínűbb állapot entrópiája tehát az entrópiának, mint valószínűségi változónak a maximális értéke.

Nagy számú részecskékből álló rendszer esetében S általában (az idő legnagyobb részében) közel lesz a maximális értékhez. Ez az oka annak, hogy az entrópiát, mint állapothatározót, és ennek maximumát gyakran összecserélik; azonban a kérdés teljes tisztázása érdekében a két fogalom, az entrópia mint állapothatározó (amely tehát *valószínűségi változó*) és a maximális (vagy valószínűségszámítási) entrópia, mint a rendszert (az eloszlást) jellemző *konstans* adat között helyes különbséget tenni.

Nem segíti persze elő a fogalmak tisztázását, hogy mind a két mennyiséget entrópiának nevezik. Mivel nem akarunk a kialakult terminológiától eltérni, csak azt szögezzük le mégegyszer, félreértések elkerülése végett, hogy a valószínűség-eloszlás entrópiájának a statisztikus mechanikában nem az entrópia, hanem egy konstans faktortól eltekintve annak maximuma: az egyensúlyi állapot entrópiája felel meg.

---

[7] Ezt, mint ismeretes, a következőképpen bizonyítják be: Ha S jelöli a termodinamikai entrópiát, akkor

$$dS = \frac{dQ}{T}$$

ahol dQ a gázzal (reverzibilisen) közölt hőmennyiséget jelenti; $dQ = dE + p dV$, ahol E a gáz kinetikus energiáját, p a nyomását és V a térfogatát jelenti. Ha a kinetikus gázelméletből jól ismert $E = 3/2\; kTN$ képletet, továbbá a $pV = NkT$ gáztörvényt felhasználjuk, következik, hogy

$$dQ = \frac{3}{2} NkT + NkT \frac{dV}{V},$$

tehát

$$dS = \frac{3}{2} Nk \frac{dT}{T} + Nk \frac{dV}{V} \text{ és így } S = Nk\left(\frac{3}{2} \log T + \log V + c\right)$$

vagyis, mivel az additív állandó a termodinamikában önkényesen választható, tehát a termodinamikai entrópia egy konstans faktortól eltekintve egyenlő a gáz valószínűségszámítási entrópiájával.

[8] A *H-függvény* elnevezés nem szerencsés, hiszen valójában egy eloszláshoz, illetve sűrűségfüggvényhez rendelt funkcionálról van szó, helyesebb volna a *H-funkcionál* elnevezés.

A helyzetet még bonyolítja, hogy míg a H-függvény kiszámításánál a folytonos eloszlással számolnak (egyenletes térbeli eloszlás, a sebesség Maxwell-eloszlása), ezzel szemben a W állapot-valószínűség kiszámításánál a fázisteret véges sok cellára bontják, és a rendszer (pl. egy ideális gáz) állapotát azzal jellemzik, hogy a rendszert alkotó részecskék a fázistér celláiban hogyan oszlanak meg. Ez az oka bizonyos additív tagok felléptének. Ha a fázistér celláinak valószínűségei egyenlők, és a k-adik cellában N_k részecske van $\sum_{k=1}^{n} N_k = N$, akkor ennek az eloszlásnak a valószínűsége a polinomiális eloszlás képlete szerint

$$\frac{N!}{N_1!\,N_2!\ldots N_n!}\,\frac{1}{n^N} \tag{28}$$

volna, ha az N_k számok nem volnának alávetve a $\sum_{k=1}^{n} N_k E_k = E$ feltételnek, ahol E_k a k-adik cellához tartozó energiaérték, E pedig az egész rendszer összenergiája; ezért csak azok az $(N_1, N_2, \ldots, N_n)$ eloszlások lehetségesek, amelyekre a $\sum_{k=1}^{n} N_k = N$ feltétel mellett a $\sum_{k=1}^{n} N_k E_k = E$ feltétel is teljesül; a megengedett eloszláshoz tartozó valószínűség ezáltal csak egy állandó faktorral változik meg; minden tagot el kell osztani az összes megengedett eloszlások valószínűségeinek összegével, tehát

$$W = c\,\frac{N!}{N_1!\,N_2!\ldots N_n!}, \tag{29}$$

ahol c nem függ az N_k számoktól. Számítsuk ki W közelítő értékét, abban az esetben, ha az N_k számok mind igen nagyok. Legyen $\frac{N_k}{N} = f_k$, akkor a Stirling-formula szerint

$$\log W \sim -N \sum_{k=1}^{n} f_k \log f_k + \log c, \tag{30}$$

vagyis W logaritmusa egy additív állandótól eltekintve arányos az empirikus eloszlás entrópiájával, vagyis az állapotvalószínűség logaritmusa (az állapot entrópiája) közelítőleg arányos az empirikus eloszlás entrópiájával. Ennélfogva a legvalószínűbb állapot az, amelynél $-\sum_{k=1}^{n} f_k \log f_k$ maximális; itt természetesen a $\sum_{k=1}^{n} f_k = 1$ és $\sum_{k=1}^{n} f_k E_k = \frac{E}{N}$ mellékfeltételeket tekintetbe kell venni. Egyszerű számítással (az úgynevezett Lagrange-féle multiplikátormódszerrel) adódik, hogy ez akkor következik be, ha

$$f_k = Ae^{-\beta E_k}, \tag{31}$$

ahol a A és β állandók a $\sum_{k=1}^{n} f_k = 1$ és $\sum_{k=1}^{n} f_k E_k = \frac{E}{N}$ egyenletekből határozhatók meg.

Ha most a vizsgált rendszer egy V köbtartalmú edénybe zárt, m tömegű atomokból álló ideális egyatomos gáz, felhasználva az $E/N = {}^3/_2\, kT$ összefüggést (amelyet ebben a vonatkozásban a hőmérséklet *definíciójának* lehet tekinteni) és feltéve, hogy a fázistér egyenlő valószínűségű cellái egyenlő (egységnyi) köbtartalmú térrészeknek és a sebesség-komponensek terének egyenlő köbtartalmú térrészeinek felelnek meg, az összegeket közelítőleg integrálokkal helyettesítve adódik, hogy az A és β állandók a következő feltételeknek kell, hogy eleget tegyenek :

$$VA \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-\frac{1}{2} m\beta(x^2+y^2+z^2)}\, dx\, dy\, dz = 1 \tag{32}$$

$$\frac{1}{2} m V A \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x^2+y^2+z^2)\, e^{-\frac{1}{2} m\beta(x^2+y^2+z^2)}\, dx\, dy\, dz = \frac{3kT}{2} \tag{33}$$

A (32) egyenletből adódik, hogy

$$VA\left(\sqrt{\frac{2\pi}{m\beta}}\right)^3 = 1 \tag{34}$$

A (33) egyenletből pedig

$$\frac{1}{2} m V A\left(\sqrt{\frac{2\pi}{m\beta}}\right)^3 \frac{3}{m\beta} = \frac{3kT}{2}. \tag{35}$$

Így nyerjük, hogy

$$\beta = \frac{1}{kT} \quad \text{és} \quad A = \frac{(2\pi m/k)^{3/2}}{VT^{3/2}}. \tag{36}$$

Ezen összefüggések alapján az entrópia maximumára a

$$H = -N \sum_{k=1}^{n} f_k \log f_k + \log c = -N \sum_{k=1}^{n} f_k (\log A - \beta E_k) + \log c =$$

$$= -N \log A + \beta E + \log c$$

tehát

$$H = N\left(\frac{3}{2} \log T + \log V\right) + C \tag{37}$$

eredményt kapjuk, ahol C-be foglaltuk össze az összes T-től és V-től nem függő tagokat. Ilymódon ezen az úton is eljutottunk az ideális gáz (maximális) entrópiájának jólismert kifejezéséhez.

A kérdés további megvilágítására foglalkozunk most a hőátadás úgynevezett EHRENFEST-féle modelljével.[9] Ez a modell, mint ismeretes, a következő :

Két urnában összesen N számú golyó van, amelyek 1-től N-ig meg vannak számozva. Ugyanakkor egy dobozba $N+1$ cédulát helyezünk,

---

[9] Lásd például M. KAC [15].

melyek 0-tól N-ig vannak megszámozva.[10] Kihúzunk egy cédulát a dobozból. Ha a nullát húzzuk, nem változtatunk semmit, ha pozitív számot húztunk, megkeressük a kihúzott számmal megegyező számozású golyót és azt áttesszük a másik urnába. Ezután a kihúzott cédulát a dobozba visszahelyezzük, a cédulákat összekeverjük, egy újabb cédulát húzunk, és így tovább. A stacionér eloszlás, mint ismeretes, a következő: ha W_k jelenti annak a valószínűségét, hogy az első urnában pontosan k golyó legyen, akkor

$$W_k = \binom{N}{k} \frac{1}{2^N}, \tag{38}$$

vagyis a stacioner eloszlás N-edrendű binomiális eloszlás $N/2$ várható értékkel.

Könnyen be lehet látni, hogy ez a sztochasztikus folyamat *reverzibilis* Markov-lánc, tehát, ha $A_k^{(n)}$ azt az eseményt jelöli, hogy az n-edik húzás után az első urnában k golyó van, akkor

$$\mathbf{P}(A_k^{(n)} A_l^{(n+m)}) = \mathbf{P}(A_l^{(n)} A_k^{(n+m)}), \tag{39}$$

vagyis $\mathbf{P}(A_l^{(n+m)} \mid A_k^{(n)})$-val jelölve annak a feltételes valószínűségét, hogy az $(n+m)$-edik húzás után az első urnában l golyó legyen, ha az n-edik húzás után ugyanott k golyó volt,

$$\mathbf{P}(A_l^{(n)})\, \mathbf{P}(A_k^{(n+m)} \mid A_l^{(n)}) = \mathbf{P}(A_k^{(n)})\, \mathbf{P}(A_l^{(n+m)} \mid A_k^{(n)}); \tag{40}$$

mivel a folyamat stacioner, s így $\mathbf{P}(A_k^{(n)}) = \binom{N}{k} \frac{1}{2^N}$ $(k = 1, 2, \dots, N)$

és $\mathbf{P}(A_k^{(n+m)} \mid A_l^{(n)}) = \mathbf{P}(A_k^{(m)} \mid A_l^{(0)})$, tehát

$$\frac{\mathbf{P}(A_k^{(m)} \mid A_l^{(0)})}{\mathbf{P}(A_l^{(m)} \mid A_k^{(0)})} = \frac{\binom{N}{k}}{\binom{N}{l}} \qquad (k, l = 0, 1, \dots, N;\ m = 1, 2, \dots). \tag{41}$$

(41)-et a következőképpen lehet belátni: legyen a rövidség kedvéért $\mathbf{P}(A_l^{(m)} \mid A_k^{(0)}) = p_m(l \mid k)$, akkor fennáll a következő két rekurziós formula:

$$p_m(l \mid k) = \frac{1}{N+1} \big((N - l + 1)\, p_{m-1}(l - 1 \mid k) + p_{m-1}(l \mid k) + {} \\ + (l + 1)\, p_{m-1}(l + 1 \mid k)\big) \tag{42}$$

és

$$p_m(k \mid l) = \frac{1}{N+1} \big(l\, p_{m-1}(k \mid l - 1) + p_{m-1}(k \mid l) + {} \\ + (N - l)\, p_{m-1}(k \mid l + 1)\big). \tag{43}$$

---

[10] A modellnek az a módosítása, hogy egy 0-val számozott cédula is van a dobozban, VINCZE ISTVÁNtól származik. Ennek az az előnye, hogy a stacioner határeloszlás páros és páratlan számú húzás esetére ugyanaz.

Bevezetve a

$$Q_m(l|k) = \frac{p_m(l|k)}{\binom{N}{l}} \tag{44}$$

jelölést, a bizonyítandó (41) összefüggés nyilván

$$Q_m(l|k) = Q_m(k|l) \tag{45}$$

alakra hozható. Könnyen ellenőrizhető, hogy (45) $m = 1$-re fennáll (csak az $l = k + 1$ és $l = k - 1$ eseteket kell megvizsgálni). (45)-öt teljes indukcióval fogjuk bebizonyítani. Tegyük fel, hogy (45), m helyett $m - 1$-et írva, minden l és k értékre ($l, k = 0, 1, \ldots, N$) fennáll. Akkor, (42) szerint

$$\begin{aligned} Q_m(l|k) &= \frac{p_m(l|k)}{\binom{N}{l}} = \\ &= \frac{(N - l + 1)p_{m-1}(l-1|k) + p_{m-1}(l|k) + (l+1)p_{m-1}(l+1|k)}{(N+1)\binom{N}{l}} \end{aligned} \tag{46}$$

és így

$$\begin{aligned} Q_m(l|k) = \frac{1}{N+1} \big(l\, Q_{m-1}(l-1|k) + Q_{m-1}(l|k) + \\ + (N-l)\, Q_{m-1}(l+1|k)\big). \end{aligned} \tag{47}$$

Az indukciós feltevés szerint tehát

$$Q_m(l|k) = \frac{lQ_{m-1}(k|l-1) + Q_{m-1}(k|l) + (N-l)\, Q_{m-1}(k|l+1)}{N+1} \tag{48}$$

és így

$$Q_m(l|k) = \frac{l\, p_{m-1}(k|l-1) + p_{m-1}(k|l) + (N-l)\, p_{m-1}(k|l+1)}{\binom{N}{k}(N+1)} \tag{49}$$

és így (43)-at felhasználva következik, hogy

$$Q_m(l|k) = \frac{p_m(k|l)}{\binom{N}{k}} = Q_m(k|l), \tag{50}$$

amivel kimutattuk, hogy (45) $m - 1$-ről m-re következik.

Ha tehát az l-edik állapot λ-szor valószínűbb, mint a k-adik állapot, akkor a k-adik állapotból az l-edik állapotba m lépésben való átmenet valószínűsége λ-szor akkora, mint az l-edik állapotból a k-adik állapotba m lépésben való átmenet valószínűsége. Ebből világos, hogy miért látszik a valójában reverzibilis folyamat irreverzibilisnek. Ez az úgynevezett LOSCHMIDT-féle ellenvetés legegyszerűbb cáfolata.

Vizsgáljuk meg az entrópia kérdését is az EHRENFEST-modell esetében. Ha a rendszer a k-adik állapotban van, statisztikus mechanikai entrópiája

$$S = \log W_k = \log \frac{1}{2^N}\binom{N}{k}. \tag{51}$$

Mivel a Moivre—Laplace-tétel szerint, ha k közel van $N/2$-hez

$$W_k \sim \sqrt{\frac{2}{\pi N}}\, e^{-2N\left(\frac{k}{N}-\frac{1}{2}\right)^2}, \tag{52}$$

tehát (51)-ből és (52)-ből

$$S \sim -2N\left(\frac{k}{N}-\frac{1}{2}\right)^2 + \frac{1}{2}\log\frac{2}{\pi N}. \tag{53}$$

Természetesen itt is fennáll, hogy a statisztikus mechanikai entrópia egy additív állandótól eltekintve arányos az empirikus eloszlás valószínűségszámítási entrópiájával, ugyanis, ha x közel van $^1/_2$-hez Taylor-sorfejtéssel nyerjük, hogy

$$-x\log x-(1-x)\log(1-x) \sim \log 2 - 2\left(x-\frac{1}{2}\right)^2. \tag{54}$$

Ennélfogva

$$-2N\left(\frac{k}{N}-\frac{1}{2}\right)^2 \sim -N\log 2 - N\left[\frac{k}{N}\log\frac{k}{N}+\frac{N-k}{N}\log\frac{N-k}{N}\right] \tag{55}$$

és így (53) és (55) összevetéséből

$$S \sim -N\left(\frac{k}{N}\log\frac{k}{N}+\frac{N-k}{N}\log\frac{N-k}{N}\right) - N\log 2 + \frac{1}{2}\log\frac{2}{\pi N}, \tag{56}$$

tehát

$$S \sim NH_e + C, \tag{57}$$

ahol H_e a pillanatnyi empirikus eloszlás entrópiáját jelöli. Mivel az entrópia csak egy additív állandótól eltekintve van meghatározva, feltehetjük, hogy $C = 0$, vagyis, hogy

$$S = NH_e. \tag{58}$$

Az empirikus eloszlás entrópiája természetesen akkor maximális, ha $k = N/2$ illetve páratlan N esetében, ha $k = (N \pm 1)/2$, ez esetben

$$H_e = \log 2$$

és

$$S_{\max} = N\log 2.$$

Megjegyzendő, hogy $S_{\max}$ nem azonos a $\{W_k\}$ valószínűségeloszlás entrópiájával. $S_{\max}$ ugyanis kizárólag a legvalószínűbb állapottól függ és azt fejezi ki, hogy a legvalószínűbb állapotban az egyes részecskék elhelyezkedésére vonatkozólag mekkora bizonytalanság áll fenn. Mivel a legvalószínűbb állapotban (páros N esetében) minden golyó egyforma valószínűséggel helyezkedhet el mindkét urnában, egy golyóra vonatkozólag a bizonytalanság $\log 2$, N golyóra $N \log 2$.

A statisztikus mechanika entrópia-fogalma mindig egyetlen eloszlásra vonatkozik, és az egyes részecskéknek a fázistér celláiban való eloszlására vonatkozó bizonytalanság mértéke.

Az az összefüggés, hogy az eloszlás entrópiája egy additív állandótól eltekintve megegyezik az eloszlás valószínűségének logaritmusának maximumával és ugyanakkor megegyezik az empirikus eloszlás entrópiájának maximumával is, nemcsak a statisztikus mechanikában áll fenn, hanem megadható egy általános valószínűségszámítási tétel, amely ezt az összefüggést kifejezi.

Az egyszerűség kedvéért ezt az összefüggést nem teljes általánosságban vezetjük le.

Legyenek $\xi_1, \xi_2, \ldots, \xi_N$ független valószínűségi változók, amelyek n különböző értéket, pl. az $y_1, y_2, \ldots, y_n$ értékeket egyforma valószínűséggel veszik fel, vagyis

$$\mathbf{P}(\xi_k = y_j) = \frac{1}{n} \qquad (k = 1, 2, \ldots, N\,;\; j = 1, 2, \ldots, n)\,.$$

Akkor együttes eloszlásuk entrópiája:

$$H_0 = \mathbf{H}_0(\xi_1, \xi_2, \ldots, \xi_N) = N \log n\,. \tag{59}$$

Figyeljük most meg a $\xi_1, \xi_2, \ldots, \xi_N$ változók értékeit egy kísérletnél: legyenek ezek $x_1, x_2, \ldots, x_N$. Jelölje N_k az $x_1, x_2, \ldots, x_N$ számok közül az y_k-val egyenlők számát és legyen $f_k = N_k/N$ az y_k érték relatív gyakorisága ($k = 1, \ldots, n$). A kapott $(f_1, f_2, \ldots, f_n)$ empirikus eloszlás létrejöttének valószínűsége nyilván

$$W = \frac{N!}{N_1!\,N_2!\ldots N_n!}\,\frac{1}{n^N}\,. \tag{60}$$

Ha N nagy n-hez képest, az N_k számok nagy valószínűséggel közel lesznek N/n-hez, tehát mind nagy számok lesznek és így alkalmazható $N_k!$ közelítő kifejezésére a Stirling-formula. Így kapjuk, hogy

$$\log W \sim -N \sum_{k=1}^{n} f_k \log f_k + C_1\,, \tag{61}$$

ahol

$$C_1 = -\log n^{\frac{n}{N^2}} (2\pi N)^{\frac{n-1}{2}}\,.$$

Másrészt figyelembevéve, hogy ha $x - \frac{1}{n}$ igen kicsiny,

$$-x \log x \sim \frac{1}{n} \log n + \left(x - \frac{1}{n}\right)(\log n - 1) - \frac{\left(x - \frac{1}{n}\right)^2}{2} n$$

és így

$$(62) \qquad -N \sum_{k=1}^{n} f_k \log f_k \sim N \log n - \frac{N}{2} \frac{\sum_{k=1}^{n}\left(f_k - \frac{1}{n}\right)^2}{\frac{1}{n}} = H_0 - \frac{\chi^2}{2},$$

ahol χ^2 a jólismert PEARSON-féle χ^2-eltérés az $(f_1, \ldots, f_n)$ empirikus eloszlás és a megfelelő $(1/n, \ldots, 1/n)$ elméleti eloszlás között. Ennélfogva a következő összefüggésre jutottunk:

$$(63) \qquad \log W \sim -N \sum_{k=1}^{n} f_k \log f_k + C_1 = H_0 - \frac{\chi^2}{2} + C_1,$$

vagyis *a minta valószínűségének logaritmusa egy additív és a mintától nem függő konstanstól eltekintve közelítőleg egyenlő a minta empirikus eloszlása entrópiájának N-szeresével (ahol N a megfigyelt értékek száma), továbbá ugyanazon additív állandótól eltekintve közelítőleg $\chi^2/2$-vel kisebb, mint a kísérlet entrópiája.* Ezt az összefüggést a következőkben BOLTZMANN tételének fogjuk nevezni.

A valószínűségszámítás szempontjából kiemeljük (62) alábbi következményét:

Ha egy ξ valószínűségi változó n különböző értéket egyforma valószínűséggel vesz fel, és ξ értékére N számú megfigyelést végzünk, akkor, ha N elég nagy, 1-hez közeli valószínűséggel

$$\mathbf{H}_0(\xi) - \mathbf{H}_0(x) \sim \frac{\chi^2}{2N},$$

ahol $\mathbf{H}_0(\xi)$ jelöli ξ entrópiáját, $\mathbf{H}_0(x)$ az n megfigyelt érték empirikus eloszlásának entrópiáját és χ^2 a megfigyelt értékek empirikus eloszlása és ξ eloszlása közötti Pearson-féle χ^2-eltérés. A nyert összefüggésből következik, hogy az empirikus eloszlás entrópiájának a megfigyelt változó entrópiájától való eltérése közelítőleg χ^2-eloszlású valószínűségi változó, ha maga a ξ változó összes értékeit egyenlő valószínűséggel veszi fel.

2. § Valószínűség-eloszlások dimenziójának és entrópiájának értelmezése

Legyen ξ egy tetszőleges korlátos valószínűségi változó; legyen $\xi^{(n)} = [n\xi]/n$, másszóval $\xi^{(n)} = k/n$, ha $k/n \leqq \xi < (k+1)/n$ $(k = 0, \pm 1, \pm 2, \pm \ldots)$, vagyis $\xi^{(n)}$-et úgy kapjuk hogy ξ értékét lekerekítjük $1/n$ legnagyobb ξ-nél még nem nagyobb többszörösére; ha például $n = 10^s$, akkor $\xi^{(n)}$ értékét ξ értékéből úgy nyerjük, hogy ξ értékét végtelen tizedestörtben fejtjük és csak s tizedesjegyet tartunk meg.

Ha $|\xi| \leqq K$, akkor $\xi^{(n)}$ lehetséges értékeinek száma $2Kn$ és így $\mathbf{H}_0(\xi^{(n)}) \leqq \log_2 2Kn = \log_2 n + \log_2 2K$. Ennélfogva

$$(64) \qquad \limsup_{n\to\infty} \frac{\mathbf{H}_0(\xi^{(n)})}{\log_2 n} \leqq 1 .$$

Ha létezik a

$$(65) \qquad \mathbf{d}(\xi) = \lim_{n\to\infty} \frac{\mathbf{H}_0(\xi^{(n)})}{\log_2 n},$$

határérték, akkor *a* $\mathbf{d}(\xi)$ *számot* ($0 \leqq \mathbf{d}(\xi) \leqq 1$) ξ *eloszlása dimenziójának nevezzük*. Legyen $\mathbf{d}(\xi) = d$; ha létezik a (véges)

$$(66) \qquad \lim_{n\to\infty} [\mathbf{H}_0(\xi^{(n)}) - d \log_2 n] = \mathbf{H}_d(\xi)$$

határérték is, akkor *a* $\mathbf{H}_d(\xi)$ *számot* ξ *eloszlása d-dimenziós entrópiájának nevezzük.*

Könnyen be lehet látni, hogy ha ξ eloszlása abszolút folytonos, és ξ sűrűségfüggvénye, $f(x)$, szakaszonkint folytonos, akkor $\mathbf{d}(\xi) = 1$, és

$$(67) \qquad \mathbf{H}_1(\xi) = -\int_{-\infty}^{+\infty} f(x) \log f(x)\, dx ;$$

ugyanis

$$\mathbf{H}_0(\xi^{(n)}) = \sum_k \Delta_k F \log_2 \frac{1}{\Delta_k F} = -\sum_k \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)\,dx \log_2 \int_{\frac{k}{n}}^{\frac{k+1}{n}} f(x)\,dx =$$

$$= \sum_k \left[f(z_k) \log_2 f(z_k) \frac{1}{n} \right] + \log_2 n .$$

(Itt $k/n \leqq z_k < (k+1)/n$ és $\Delta_k F = F\big((k+1)/n\big) - F(k/n)$.) Ha ξ diszkrét eloszlású, $\mathbf{P}(\xi = x_k) = p_k$ $(k = 1, 2, \ldots)$, ahol $x_j \neq x_k$, ha $j \neq k$, akkor $\mathbf{H}_0(\xi) = \sum_{k=1}^{\infty} p_k \log_2 p_k$, feltéve, hogy utóbbi sor konvergens.

Ha tekintünk egy $\{p_k\}$ diszkrét eloszlást, amelynek létezik a (0-dimenziójú entrópiája és egy $f(x)$ szakaszonkint folytonos sűrűségfüggvényű korlátos és abszolút folytonos eloszlást, és képezzük ezeknek az eloszlásoknak a keverékét p és q súlyokkal, a kapott eloszlás entrópiája q-dimenziós lesz és értéke a következő képlettel fejezhető ki :

$$(68) \qquad H_q = -q \int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx - p \sum_{k=1}^{\infty} p_k \log_2 p_k - p \log_2 p - q \log_2 q .$$

Igen általános feltételek mellett érvényes a következő összefüggés: ha a ξ_k ($k = 1, 2, \ldots, n$) korlátos valószínűségi változók értékkészletei idegenek[11] és ξ_k eloszlásának dimenziója $d_k = \mathbf{d}(\xi_k)$ és ξ_k d_k-dimenziójú entrópiája, $\mathbf{H}_{d_k}(\xi_k)$, létezik, és vesszük a ξ_k változók eloszlásainak q_k súlyokkal a keverékét, továbbá ζ egy ilyen keverékeloszlású változó, akkor

$$d = \mathbf{d}(\zeta) = \sum_{k=1}^{n} q_k \, \mathbf{d}(\xi_k) \tag{69}$$

és

$$\mathbf{H}_d(\zeta) = \sum_{k=1}^{n} p_k \, \mathbf{H}_{d_k}(\xi_k) - \sum_{k=1}^{n} q_k \log_2 q_k \, , \tag{70}$$

vagyis *a keverékeloszlás dimenziója az egyes eloszlások dimenzióinak a keverősúlyokkal súlyozott középértéke és a keverékeloszlás entrópiája egyenlő az egyes komponensek entrópiáinak a keverősúlyokkal súlyozott középértékének és a keverőeloszlás entrópiájának összegével.*

A mondottakat megvilágítja a következő megjegyzés: Ha a ξ valószínűségi változó eloszlása $d = \mathbf{d}(\xi)$ dimenziójú és d-dimenziójú entrópiája $\mathbf{H}_d(\xi) = H$, akkor, ha $n \to \infty$

$$\mathbf{H}_0\left(\frac{[n\xi]}{n}\right) = d \log_2 n + H + o(1) \, . \tag{71}$$

Megfordítva, ha egy ξ valószínűségi változóra

$$\mathbf{H}_0\left(\frac{[n\xi]}{n}\right) = A \log_2 n + B + o(1) \, , \tag{72}$$

ha $n \to \infty$, akkor $A = \mathbf{d}(\xi)$ és $B = \mathbf{H}_{\mathbf{d}(\xi)}(\xi)$. Ha a ξ (korlátos) változóra nem érvényes (72) alakú aszimptotikus reláció, akkor ξ entrópiája nincsen értelmezve.

Eddig csak korlátos valószínűségi változók entrópiáját értelmeztük. A dimenzió és a entrópia értelmezését bizonyos, nem korlátos változókra is kiterjeszthetjük; e kérdéssel azonban itt nem foglalkozunk.

3. §. A dimenzió és az entrópia jellemző tulajdonságai

Nevezzük eleminek azokat a valószínűség-eloszlásokat, amelyek előállíthatók, mint egy véges diszkrét valószínűség-eloszlás és egy olyan abszolút folytonos valószínűség-eloszlás keveréke, amelynek sűrűségfüggvénye egy véges intervallumban szakaszonként folytonos és azon kívül eltűnik. Tegyük

[11] Azon, hogy a ξ és η valószínűségi változók idegen értékkészletűek, azt értjük, hogy a számegyenes bármely E Borel-halmaza felbontható olyan idegen E_1 és E_2 halmazokra, hogy $\mathbf{P}(\xi \in E_1) = 0$ és $\mathbf{P}(\eta \in E_2) = 0$ (ezt úgy szokták kifejezni, hogy a két változó által származtatott mérték *ortogonális*). Vegyük észre, hogy egy tetszőleges diszkrét eloszlású változó és egy tetszőleges abszolút folytonos eloszlású változó értékkészlete idegen.

fel, hogy minden ξ elemi valószínűség-eloszlású valószínűségi változóhoz hozzá van rendelve egy $\mathbf{d}(\xi)=d$ és egy $\mathbf{H}_d(\xi)$ szám, amelyek eleget tesznek az alábbi feltételeknek[12])

1.a. $0 \leq \mathbf{d}(\xi) \leq 1$; $\mathbf{d}(\xi)$ csak ξ eloszlásától függ. Ha $\mathbf{P}(\xi = a) = 1$, akkor $\mathbf{d}(\xi) = 0$.

1.b. $\mathbf{H}_d(\xi)$ csak ξ eloszlásától függ; $\mathbf{H}_0(\xi) \geq 0$ és ha $\mathbf{P}(\xi = a) = 1$, akkor $\mathbf{H}_0(\xi) = 0$,

2.a. Ha $y = f(x)$ olyan függvény, hogy ha $x \neq x'$, akkor, $f(x) \neq f(x')$, továbbá ξ olyan valószínűségi változó, amelyre $\mathbf{d}(\xi) = 0$ és $\eta = f(\xi)$, akkor $\mathbf{d}(\eta) = 0$ és $\mathbf{H}_0(\eta) = \mathbf{H}_0(\xi)$.

2.b. Ha ξ és η tetszőleges elemi valószínűségi változók és $\eta = \xi + C$, ahol C állandó, akkor $\mathbf{d}(\eta) = \mathbf{d}(\xi) = d$ és $\mathbf{H}_d(\eta) = \mathbf{H}_d(\xi)$.

3.a. Ha ξ eloszlása a $\xi_1, \xi_2, \ldots, \xi_n$ idegen értékkészletű változók eloszlásának $q_1, q_2, \ldots, q_n$ súlyokkal vett keveréke, akkor

$$\mathbf{d}(\xi) = \sum_{k=1}^{n} q_k \mathbf{d}(\xi_k).$$

3.b. Ha ζ eloszlása a $\xi_1, \xi_2, \ldots, \xi_n$ idegen értékkészletű valószínűségi változók eloszlásának $q_1, q_2, \ldots, q_n$ súlyokkal vett keveréke és η egy olyan valószínűségi változó, amely az $1, 2, \ldots, n$ értékeket rendre $q_1, q_2, \ldots, q_n$ valószínűséggel veszi fel, akkor bevezetve a $d_k = \mathbf{d}(\xi_k)$ és $d = \sum_{k=1}^{n} q_k d_k$ jelöléseket

$$\mathbf{H}_d(\zeta) = \sum_{k=1}^{n} q_k \mathbf{H}_{d_k}(\xi_k) + \mathbf{H}_0(\eta).$$

Megjegyzés: Az 1.a., 2.a. és 3.a. feltevésekből következik, hogy bármely ξ diszkrét eloszlású (elemi) valószínűségi változóra $\mathbf{d}(\xi) = 0$, ugyanis minden diszkrét valószínűség-eloszlás felfogható, mint elfajult valószínűség-eloszlások (konstansok eloszlásainak) keveréke. Ezért írható 3.b.-ben $\mathbf{H}_{\mathbf{d}(\eta)}(\eta)$ helyett $\mathbf{H}_0(\eta)$.

4.a. Ha a ξ_n diszkrét eloszlású valószínűségi változó eloszlását a p_{nk} $(k = 1, 2, \ldots, N)$ számok alkotják és $\lim_{n \to \infty} p_{nk} = p_k$ $(k = 1, 2, \ldots, N)$, továbbá ξ egy olyan valószínűségi változó, amelynek eloszlását a p_k számok alkotják, akkor $\lim_{n \to \infty} \mathbf{H}_0(\xi_n) = \mathbf{H}_0(\xi)$.

4.b. Ha ξ_n elemi abszolút folytonos eloszlású valószínűségi változó, amelynek sűrűségfüggvénye $f_n(x)$ és x-ben egyenletesen $\lim_{n \to \infty} f_n(x) = f(x)$, továbbá ξ olyan valószínűségi változó, amelynek sűrűségfüggvénye $f(x)$, akkor

$$\lim_{n \to \infty} \mathbf{d}(\xi_n) = \mathbf{d}(\xi)$$

és

$$\lim_{n \to \infty} \mathbf{H}_{\mathbf{d}(\xi_n)}(\xi_n) = \mathbf{H}_{\mathbf{d}(\xi)}(\xi).$$

12) E feltételek lényegesen enyhíthetők volnának, erre azonban itt nem törekszünk.

5.a. Ha $\mathbf{P}(\xi = \pm 1) = 1/2$, akkor $\mathbf{H}_0(\xi) = 1$.

5.b. Ha ξ egyenletes eloszlású a (0,1) intervallumban, akkor $\mathbf{d}(\xi) = 1$ és $\mathbf{H}_1(\xi) = 0$.

Az 1. — 5. feltevésekből következik, hogy *ha ξ valószínűség-eloszlását a $p_1, p_2, \ldots, p_n$ számok alkotják, akkor* $\mathbf{H}_0(\xi) = -\sum_{k=1}^{n} p_k \log_2 p_k$, *továbbá, ha ξ tetszőleges elemi eloszlású valószínűségi változó, akkor a* $\mathbf{d}(\xi) = d$ és $\mathbf{H}_d(\xi)$ *számok a*

$$\mathbf{H}_0\left(\frac{[n\xi]}{n}\right) = d \log_2 n + \mathbf{H}_d(\xi) + o(1)$$

aszimptotikus relációnak tesznek eleget, vagyis $\mathbf{d}(\xi)$ és $\mathbf{H}_d(\xi)$ *azonosak a 2. §-ban definiált dimenzióval, illetve entrópiával.*

Bizonyítás: Először bebizonyítjuk, hogy az 1.b., 2.a., 3.b. és 4.a. feltevésekből következik, hogy ha $\mathbf{P}(\xi = x_k) = p_k$ $(k = 1, 2, \ldots, n)$, akkor

$$\mathbf{H}_0(\xi) = -\lambda \sum_{k=1}^{n} p_k \log_2 p_k .$$

Legyenek ugyanis n, m és k tetszőleges, a $kn \leqq m < (k+1)n$ egyenlőtlenséget kielégítő pozitív egész számok. Legyenek $\xi_1, \xi_2, \ldots, \xi_{k+1}$ véges, diszkrét eloszlású és idegen értékkészletű valószínűségi változók; ξ_j lehetséges értékeinek száma legyen n és vegye fel értékeit rendre $1/n$ valószínűséggel $(j = 1, 2, \ldots, k)$. ξ_{k+1} értékeinek száma legyen $m - kn$ és vegye fel ezeket rendre $1/(m - kn)$ valószínűséggel. Legyen η olyan valószínűségi változó, amelynek eloszlása a ξ_j változók eloszlásának q_j súlyokkal vett keveréke, ahol $q_j = n/m$, ha $j = 1, 2, \ldots, k$ és $q_{k+1} = 1 - kn/m$.

Nyilván η-nak m különböző értéke van és ezeket mind $1/m$ valószínűséggel veszi fel.

A 3.b. feltevést alkalmazva és egy olyan valószínűségi változó 0-dimenziójú entrópiáját, amely r különböző érték mindegyikét $1/r$ valószínűséggel veszi fel, a rövidség kedvéért $H(r)$-rel jelölve kapjuk, hogy

$$H(m) = \frac{kn}{m} H(n) + \left(1 - \frac{kn}{m}\right) H(m - kn) + \mathbf{H}_0(\zeta),$$

ahol ζ egy olyan valószínűségi változó, amely az $1, 2, \ldots, k$ értékeket n/m valószínűséggel, a $k+1$ értéket pedig $1 - kn/m$ valószínűséggel veszi fel. Felhasználva 1.b.-t és a $kn \leqq m < (k+1)n$, vagy másképpen írva az $1 - n/m \leqq kn/m < 1$ egyenlőtlenséget, adódik

$$H(m) \geqq \left(1 - \frac{n}{m}\right) H(n) \qquad (m \geqq n). \tag{73}$$

Most legyenek r és s tetszőleges prímszámok; adott N-hez tartozik egy és csakis egy M, amelyre fennállnak az

$$r^M \leqq s^N < r^{M+1}$$

egyenlőtlenségek. Alkalmazva (73)-at az $n = r^M$ és $m = s^N$, illetve $n = s^N$ és $m = r^{M+1}$ számokra, kapjuk, hogy

$$H(s^N) \geqq \left(1 - \frac{r^M}{s^N}\right) H(r^M) \tag{74}$$

és

$$H(s^N) \leqq \frac{1}{1 - \dfrac{s^N}{r^{M+1}}} H(r^{M+1}). \tag{75}$$

Most bebizonyítjuk, hogy ha A és B tetszőleges pozitív egész számok,

$$H(AB) = H(A) + H(B) \tag{76}$$

(76)-ot a következőképpen láthatjuk be: Legyenek $\xi_1, \xi_2, \ldots, \xi_A$ idegen értékkészletű valószínűségi változók, melyek mindegyike B számú különböző értéket rendre $1/B$ valószínűséggel vesz fel és legyen η eloszlása $\xi_1, \xi_2, \ldots, \xi_A$ eloszlásának $1/A$ súlyokkal vett keveréke. Akkor η nyilván AB számú különböző érték mindegyikét $1/AB$ valószínűséggel veszi fel, és így, mivel 3.b. szerint

$$\mathbf{H}_0(\eta) = \frac{1}{A} \sum_{k=1}^{A} \mathbf{H}_0(\xi_k) + H(A),$$

tehát

$$H(AB) = H(A) + H(B).$$

Ezzel (76)-ot bebizonyítottuk. (74)-ből, (75)-ből és (76)-ból következik, hogy

$$\frac{M}{N}\left(1 - \frac{r^M}{s^N}\right) \leqq \frac{H(s)}{H(r)} \leqq \frac{M+1}{N} \frac{1}{1 - \dfrac{s^N}{r^{M+1}}}. \tag{77}$$

Mivel s és r különböző prímszámok, $\dfrac{\log s}{\log r}$ irracionális.

Az $r^M \leqq s^N < r^{M+1}$ egyenlőtlenségből adódik, hogy $M \leqq N \dfrac{\log s}{\log r} < M + 1$ és így, hogy

$$M = \left[N \frac{\log s}{\log r}\right], \tag{78}$$

továbbá, hogy

$$\lim_{N \to \infty} \frac{M}{N} = \frac{\log s}{\log r}. \tag{79}$$

Felhasználva azt a tételt, hogy ha α irracionális szám, akkor az $n\alpha - [n\alpha]$ számsorozat ($n = 1, 2, \ldots$) mindenütt sűrűn helyezkedik el a $(0, 1)$ intervallumban, bármely pozitív ε-hoz meg lehet választani N értékét úgy, hogy $0 \leqq N \frac{\log s}{\log r} - M < \varepsilon$ legyen, vagyis, hogy teljesüljön az

$$\frac{s^N}{r^{M+1}} < \frac{1}{r^{1-\varepsilon}} \tag{80}$$

egyenlőtlenség. Másrészt választható N úgy is, hogy teljesüljön a

$$1 - \varepsilon \leqq N \frac{\log s}{\log r} - M \leqq 1$$

egyenlőtlenség, vagyis, hogy teljesüljön a

$$\frac{r^M}{s^N} \leqq \frac{1}{r^{1-\varepsilon}} \tag{81}$$

egyenlőtlenség. A (77), (79), (80) és (81) összefüggésekből következik, hogy

$$\frac{\log s}{\log r}\left(1 - \frac{1}{r}\right) \leqq \frac{H(s)}{H(r)} \leqq \frac{\log s}{\log r} \frac{1}{1 - \frac{1}{r}}. \tag{82}$$

Legyen

$$\liminf_{r\to\infty} \frac{H(r)}{\log r} = \lambda \quad \text{és} \quad \limsup_{r\to\infty} \frac{H(r)}{\log r} = \mu . \tag{83}$$

Nyilván (82) írható a

$$\frac{H(r)}{\log r}\left(1 - \frac{1}{r}\right) \leqq \frac{H(s)}{\log s} \leqq \frac{H(r)}{\log r} \frac{1}{1 - \frac{1}{r}} \tag{84}$$

alakba is. (84)-ből következik, hogy

$$\mu \leqq \frac{H(s)}{\log s} \leqq \lambda , \tag{85}$$

de mivel $\lambda \leqq \mu$, (85)-ből következik, hogy $\lambda = \mu$ és hogy

$$\frac{H(s)}{\log s} = \lambda , \tag{86}$$

ha s törzsszám. Azonban (77)-ből következik, hogy ha az n szám törzstényezős felbontása $n = p_1^{\alpha_1} p_2^{\alpha_2} \dots p_j^{\alpha_j}$ ($p_1, p_2, \dots, p_j$ törzsszámok) akkor

$$H(n) = \sum_{i=1}^{j} \alpha_i H(p_i) \tag{87}$$

(86)-ból és (87)-ből következik, hogy

$$H(n) = \lambda \log n . \tag{88}$$

Ez az állítás azonban nem más, mint bizonyítandó tételünk speciális esete. Vizsgáljuk most az általánosabb esetet, amikor $p_1, p_2, \ldots, p_n$ tetszőleges racionális számok. Legyen $p_k = g_k/g$ $(k = 1, 2, \ldots, n)$, ahol $g_1, g_2, \ldots, g_n$ természetes számok és $\sum_{=1}^{n} g_n = g$. Legyen η_k olyan valószínűségi változó, amely g_k különböző értéket $1/g_k$ valószínűséggel vesz fel, és tegyük fel, hogy az η_k $(k = 1, 2, \ldots, n)$ valószínűségi változók értékkészletei idegenek. Képezzük az η_k változók eloszlásainak $p_k = g_k/g$ súlyokkal vett keverékét, és legyen ζ egy olyan valószínűségi változó, amelynek eloszlása az így nyert keverékeloszlás.

Nyilvánvaló, hogy a ζ változó $\sum_{k=1}^{n} g_k = g$ számú különböző érték mindegyikét $1/g$ valószínűséggel veszi fel.

3.b. szerint

$$\mathbf{H}_0(\zeta) = \sum_{k=1}^{n} p_k \mathbf{H}_0(\eta_k) + \mathbf{H}_0(\xi), \tag{89}$$

ahol ξ egy olyan valószínűségi változó, amely az $1, 2, \ldots, n$ értékeket $p_1, p_2, \ldots, p_n$ valószínűséggel veszi fel. Mivel $\mathbf{H}_0(\zeta) = \lambda \log g$ és $\mathbf{H}_0(\eta_k) = \lambda \log g_k$, tehát (89)-ből

$$\mathbf{H}_0(\xi) = -\lambda \sum_{k=1}^{n} p_k \log p_k. \tag{90}$$

A (90) összefüggést eddig racionális p_k számokra igazoltuk; azonban 4.a.-ból következik, hogy érvényes kell, hogy legyen az argumentumok irracionális értékeire is. Így a (90) összefüggést tetszőleges véges diszkrét eloszlású változóra bebizonyítottuk.

A λ szám értéke 5.a. miatt $1/\log 2$ kell, hogy legyen. Ilyenmódon

$$\mathbf{H}_0(\xi) = -\sum_{k=1}^{n} p_k \log_2 p_k, \tag{91}$$

tehát állításunk diszkrét eloszlású változókra vonatkozó részét igazoltuk.

Megjegyezzük, hogy a diszkrét eloszlás entrópiájának fentebb adott jellemzése a HINCSIN-féle tárgyalásból elsősorban abban tér el, hogy nem használtuk fel azt, hogy egy n tagú $(p_1, p_2, \ldots, p_n)$ valószínűség-eloszlás entrópiája akkor maximális, ha $p_k = 1/n$ $(k = 1, 2, \ldots, n)$. Ennek következtében nem tudtuk előre, hogy $H(n)$ n-nek monoton növekvő függvénye, ami megnehezítette annak kimutatását, hogy a $H(n)$ függvényre fennálló

$$H(mn) = H(m) + H(n)$$

függvényegyenletből következik, hogy $H(n) = \lambda \log n$. Az ugyanis régóta ismeretes, hogy e függvényegyenletnek nincs más monoton növekvő megoldása, mint a $\lambda \log n$ függvények $(\lambda > 0)$. A fenti bizonyítás lényege tulajdonképpen az, hogy ha a $H(n)$ számelméleti függvény eleget tesz a $H(mn) = H(m) + H(n)$ függvényegyenletnek és „majdnem monoton”, abban az értelemben, hogy fennáll a $H(m) \geqq (1 - n/m)\, H(n)$ egyenlőtlenség, ha $m \geqq n$, akkor $H(n) = \lambda \log n$.

Itt jegyezzük meg, hogy az entrópiára vonatkozólag még számos más szélsőértéktulajdonság ismeretes. Így például az összes olyan abszolút folytonos eloszlások közül, amelyeknek sűrűségfüggvénye az I intervallumon kívül eltűnik, az I-ben egyenletes eloszlásnak maximális az entrópiája, az összes abszolút folytonos eloszlások közül, amelyek szórása, σ, adott, a σ szórású normális eloszlásnak maximális az entrópiája. Ez a normális eloszlás egy érdekes jellemzése.

Most vizsgáljuk a $(0, a)$ intervallumban egyenletes eloszlású ξ valószínűségi változót, legyen ennek dimenziója D_a és D_a-dimenziós entrópiája $h(a)$. Ha a pozitív egész szám, ez az eloszlás előállítható, mint a $(0, 1), (1, 2), \ldots, (a - 1, a)$ intervallumokban egyenletes eloszlások $1/a$ súlyokkal vett keveréke. Ennélfogva 3.a.-ból és 5.b.-ből következik, hogy $D_a = 1$, ha a pozitív egész szám, és 3.b.-ből és 5.b.ből következik, hogy

$$h(a) = \log_2 a\,. \tag{92}$$

Ha a racionális szám, $a = b/c$, ahol b és c természetes számok, akkor a $(0, b)$ intervallumban egyenletes eloszlás előállítható, mint a $(0, a), (a, 2a), \ldots, ((c - 1)a, b)$ intervallumokban egyenletes eloszlások $1/c$ súlyokkal vett keveréke. Ennélfogva $D_a = 1$ és

$$h(b) = h(c) + h(a) \tag{93}$$

(92)-ből és (93)-ból következik, hogy ha $a = b/c$, akkor

$$h(a) = \log_2 \frac{b}{c} = \log_2 a \tag{94}$$

vagyis, hogy (92) tetszőleges racionális a-ra érvényes.

4.b. felhasználásával következik, hogy ha a tetszőleges pozitív szám, $D_a = 1$ és $h(a) = \log_2 a$. Mármost legyen $f(x)$ egy tetszőleges sűrűségfüggvény, amely egy véges $(-K, +K)$ intervallumon kívül eltűnik és a $(-K, +K)$ zárt intervallumban folytonos.[13] Akkor ugyanott $f(x) \log f(x)$ is folytonos, tehát a $\int\limits_{-\infty}^{+\infty} f(x) \log f(x)\, dx$ Riemann-integrál létezik; legyen ξ valószínűségsűrűségfüggvénye $f(x)$ és legyen ξ_n sűrűségfüggvénye $g_n(x) = n \int\limits_{\frac{k}{n}}^{\frac{k+1}{n}} f(t)\, dt$, ha $\frac{k}{n} \leqq x < \frac{k+1}{n}$ $(k = 0, \pm 1, \pm 2, \ldots)$. Nyilvánvaló, hogy ξ_n eloszlása a

[13] Az intervallum bal-, illetve jobboldali végpontjában csak jobbról, illetve balról való folytonosságot kötünk ki.

$\left(\frac{k}{n}, \frac{k+1}{n}\right)$ intervallumban egyenletes eloszlások $q_k = \int\limits_{\frac{k}{n}}^{\frac{k+1}{n}} f(t)\,dt$ súlyokkal vett keveréke, és így $\mathbf{d}(\xi_n) = 1$, és

$$\mathbf{H}_1(\xi_n) = \sum_{k=-\infty}^{+\infty} q_k \left(\log_2 \frac{1}{n} - \log_2 q_k\right),$$

tehát

$$\mathbf{H}_1(\xi_n) = -\sum_{k=-\infty}^{+\infty} q_k \log_2 n\, q_k\,. \tag{95}$$

Mivel az integrálszámítás középértéktétele szerint

$$n q_k = f(x_k) \qquad \frac{k}{n} \leqq x_k < \frac{k+1}{n},$$

tehát

$$\mathbf{H}_1(\xi_n) = -\sum_{k=-\infty}^{+\infty} \left[f(x_k) \log_2 f(x_k)\, \frac{1}{n} \right] \tag{96}$$

és így 4.b. figyelembevételével $\lim\limits_{k \to \infty} \mathbf{d}(\xi_n) = 1 = \mathbf{d}(\xi)$ és

$$\mathbf{H}_1(\xi) = \lim_{n \to \infty} \mathbf{H}_1(\xi_n) = -\int\limits_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx\,. \tag{97}$$

Az áttérés folytonos sűrűségfüggvényű eloszlásról szakaszonként folytonos sűrűségfüggvényű eloszlásra a 3.b. feltétel újbóli alkalmazásával történhet.

Ha mármost ξ eloszlása egy $f(x)$ szakaszonként folytonos sűrűségfüggvényű elemi eloszlás és egy $\{p_k\}$ diszkrét eloszlás q és p súlyokkal vett keveréke, a már bebizonyított (91) és (97) összefüggésekből, továbbá a 3.a. és 3.b. feltevésekből következik, hogy $\mathbf{d}(\xi) = q$ és

$$\mathbf{H}_q(\xi) = -q \int\limits_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx - p \sum_{k=1}^{\infty} p_k \log_2 p_k - p \log_2 p - q \log_2 q. \tag{98}$$

Ezzel állításunkat teljes egészében bebizonyítottuk.

4. § Többdimenziós valószínűség-eloszlások entrópiája

Legyenek $\xi_1, \xi_2, \ldots, \xi_r$ valószínűségi változók; értelmezzük $\xi_1, \xi_2, \ldots, \xi_r$ együttes eloszlásának entrópiáját a következőképpen: Ha az $\frac{[n\xi_1]}{n}, \frac{[n\xi_2]}{n}, \ldots, \frac{[n\xi_r]}{n}$ diszkrét változók együttes eloszlásának entrópiáját $H_0^{(n)}$-nel jelöljük és ha $n \to \infty$, esetén fennáll egy

$$H_0^{(n)} = D \log_2 n + H + o(1) \tag{99}$$

alakú reláció, akkor legyen

$$\mathbf{H}_D(\xi_1, \xi_2, \ldots, \xi_r) = H \tag{100}$$

és $\mathbf{d}(\xi_1, \xi_2, \ldots, \xi_r) = D$ és ez esetben nevezzük a D számot a $(\xi_1, \xi_2, \ldots, \xi_r)$ változók együttes eloszlása dimenziójának és a H számot a $\xi_1, \xi_2, \ldots, \xi_r$ változók együttes eloszlása D-dimenziós entrópiájának. Ebből a definícióból azonnal adódik, hogy ha pl. a $\vec{\zeta}$ r-dimenziós valószínűségi vektorváltozó sűrűségfüggvénye a folytonos és egy $|\boldsymbol{x}| \leqq R$ gömbön kívül eltűnő $f(\boldsymbol{x})$ függvény, ahol

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_r), \quad |\boldsymbol{x}| = \sqrt{x_1^2 + x_2^2 + \ldots + x_r^2},$$

akkor $\mathbf{d}(\vec{\zeta}) = r$ és

$$\mathbf{H}_r(\vec{\zeta}) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty} f(\boldsymbol{x}) \log_2 f(\boldsymbol{x})\, d\boldsymbol{x}, \tag{101}$$

ahol $d\boldsymbol{x} = dx_1\, dx_2 \ldots dx_r$.

Ha így definiáljuk a többdimenziós eloszlások entrópiáját, akkor érvényes a következő tétel: Ha ξ és η tetszőleges valószínűségi változók, melyek együttes eloszlásának entrópiája értelmezve van és $\mathbf{d}(\xi|\eta)$ jelöli ξ feltételes eloszlásának dimenzióját azon feltevés mellett, hogy η értéke rögzítve van, továbbá $\mathbf{H}_{\mathbf{d}(\xi|\eta)}(\xi|\eta)$ jelöli η rögzített értéke mellett ξ feltételes eloszlásának $\mathbf{d}(\xi|\eta)$-dimenziójú entrópiáját, akkor (feltéve, hogy az összes bevezetett mennyiségek értelmezhetők),

$$\mathbf{d}(\xi, \eta) = \mathbf{d}(\eta) + \overline{\mathbf{d}(\xi|\eta)} \tag{102}$$

és

$$\mathbf{H}_{\mathbf{d}(\xi,\eta)}(\xi, \eta) = \mathbf{H}_{\mathbf{d}(\eta)}(\eta) + \overline{\mathbf{H}_{\mathbf{d}(\xi|\eta)}(\xi|\eta)}. \tag{103}$$

Abban az esetben, ha ξ és η függetlenek, $\mathbf{d}(\xi|\eta) = \mathbf{d}(\xi)$ és $\mathbf{H}_{\mathbf{d}(\xi)}(\xi|\eta) = \mathbf{H}_{\mathbf{d}(\xi)}(\xi)$, tehát

$$\mathbf{d}(\xi, \eta) = \mathbf{d}(\xi) + \mathbf{d}(\eta) \tag{104}$$

és

$$\mathbf{H}_{\mathbf{d}(\xi,\eta)}(\xi, \eta) = \mathbf{H}_{\mathbf{d}(\xi)}(\xi) + \mathbf{H}_{\mathbf{d}(\eta)}(\eta). \tag{105}$$

Ez a tétel érthetővé teszi, hogy egy r-dimenziós térben megadott abszolút folytonos eloszlásnak a fenti értelemben is r-dimenziós az eloszlása és egyben fényt vet az eloszlások dimenziója és a geometriai dimenzió fogalmának összefüggésére és érthetővé teszi, miért neveztük a $\mathbf{d}(\xi)$ számot ξ eloszlása dimenziójának.

Arra a kérdésre, hogy az e dolgozatban tárgyaltakon kívül mely valószínűség-eloszlásokra definiálható a dimenzió és az entrópia, és egyéb nyitvahagyott kérdésekre egy további dolgozatban fogunk visszatérni.

IRODALOM

[1] R. Clausius: *Die mechanische Wärmetheorie.* 1876.
[2] L. Boltzmann: *Vorlesungen über Gastheorie.* 1896—1898.
[3] P. Ehrenfest—T. Ehrenfest: „Über zwei bekannte Einwände gegen das Boltzmannsche H-theorem." *Physikalische Zeitschrift* **8** (1907) 311—314.
[4] C. E. Shannon: „A mathematical theory of communication." *Bell System Technical Journal* **27** (1948) 399—429, 623—656.
[5] C. E. Shannon—W. Weaver: *The mathematical theory of communication.* The University of Illinois Press, Urbana, 1949.
[6] А. Я. Хинчин: „Понятие энтропии в теорий вероятностей." *Успехи Математических Наук* **8** (1953) 3—51.
[7] L. Szilárd: „Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen." Berlin, 1928. 840—856.
[8] D. Gábor: *La théorie des communications et la physique.* „*La Cybernétique.*" Editions Revue d'Optique, Paris, 1951. 115 —
[9] D. Brillouin: „Maxwell's Demon cannot operate: Information and entropy, I." *Journal of Applied Physics* **22** (1951) 334—337.
[10] D. Brillouin: „Physical entropy and information II." *Journal of Applied Physics* **22** (1951) 338—344.
[11] D. Brillouin: „The negentropy principle of information." *Journal of Applied Physics* **24** (1953) 1152—1163.
[12] D. Brillouin: „Information theory and uncertainity principle." *Journal of Applied Physics* **25** (1954) 887—893.
[13] J. A. Ville: „Lecons sur quelques aspects nouveaux de la théorie des probabilités." *Annales de l'Institut H. Poincaré* **14** (1954) 61—143.
[14] A. Blanc-Lapierre: „Considérations sur la théorie de la transmission de l'information (etc)." *Annales de l'Institut H. Poincaré* **13** (1953) 245—296.
[15] M. Kac: „Random walk and the theory of Brownian motion." *American Mathematical Monthly* **54** (1947) 369—391.

(Beérkezett: 1956. II. 7.)

ЗАМЕЧАНИЯ ОБ ЭНТРОПИИ

Я. БАЛАТОНЪИ и А. РЕНЬИ

Резюме

Цель настоящей работы — обоснование понятия энтропии в теории вероятностей. А. Я. Хинчин [6] изучал лишь энтропию конечного дискретного распределения. В теории информации встречается и энтропия некоторых непрерывных распределений вероятностей. Однако, изучение энтропии для этих распределений в том смысле в каком это сделал Хинчин для случая дискретных распределений, не было проведено.

В ходе исследования выяснилось, что для характеристики энтропии (меры неопределенности распределения) не достаточно одного числа, а требуется два числа. Мы введем для них следующие названия: размерность и относящаяся к данной размерности энтропия.

В § 1 изучается связь между H-функцией Болцманна, понятием энтропии статистической механики и понятием энтропии теории вероятностей.

На примере идеального газа показывается, что физическая энтропия есть случайная величина, которая совпадает с теоретико-вероятностной энтропией моментального распределения в фазовом пространства (если не обращать внимания на выбор единицы измерения). В этом же §-е исследуется модель теплопередачи Эренфеста.

В § 2 определяется размерность и энтропия распределений вероятностей.

Если ξ — случайная величина с дискретным распределением, $\mathbf{P}(\xi = x_k) = p_k$ $(k = 1,2,\ldots;\ \Sigma p_k = 1)$ размерность её распределения считается равной нулю, а её 0-мерная энтропия для которой вводится обозначение $\mathbf{H}_0(\xi)$ определяется формулой Шэннона

$$\mathbf{H}_0(\xi) = -\sum_{k=1}^{\infty} p_k \log_2 p_k$$

если ряд (1) сходится. Пусть ξ любая ограниченная случайная величина, и $[x]$ обозначает целую часть числа x. Если имеет место соотношение

$$\mathbf{H}_0\left(\frac{[n\,\xi]}{n}\right) = d \log_2 n + H + o(1)$$

где d и n постоянные, то распределение ξ называется d-мерным, а число n называется d-мерной энтропией распределения ξ и обозначается через $\mathbf{H}_d(\xi)$ Доказывается что если функция плотности $f(x)$ случайной величины ξ кусочно непрерывна на некотором конечном замкнутом отрезке и равна нулю вне его, то размерность распределения ξ равна 1, а одномерная энтропия распределения ξ даётся формулой Шэннона

$$\mathbf{H}_1(\xi) = -\int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx$$

Авторы подчеркивают, что энтропии разных размерностей не должны сравниваться по их численным значениям; энтропия с большей размерностью всегда «превосходит» энтропию с меньшей размерностью.

Если распределение ξ есть смесь, с весами q и p $(p + q = 1)$ абсолютно непрерывного распределения с кусочно непрерывной функцией плотности $f(x)$ и дискретного распределения $\{p_k\}$ то распределение ξ есть q-мерное распределение, а его q-мерная энтропия

$$\mathbf{H}_q(\xi) = -q \int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx - p \sum_{k=1}^{\infty} p_k \log_2 p_k - p \log_2 p - q \log_2 q.$$

В § 3 перечисляются характеристические свойства размерности и энтропии. Исходя из этих свойств, доказывается единственность для распределений, представимых в вида смеси распределения с кусочнонепрерывной функцией плотности и конечного дискретного распределения.

В § 4 результаты § 2 обобщаются на случай многомерных распределений.

REMARKS ON ENTROPY

J. BALATONI and A. RÉNYI

Summary

The paper deals with the notion of the entropy of a probability distribution as a measure of the uncertainity implied by the distribution. In §. 1. the relation of this notion to the entropy in statistical mechanics resp. to Boltzmann's H-function, further to the entropy in information theory is discussed. It is shown, by discussing the example

of an ideal gas contained in a vessel, further the model of heat exchange given by P. and T. EHRENFEST, that the quantity which is called entropy in statistical mechanics is a random variable, characterising the momentary state of the physical system considered and the maximal value of this quantity is up to an additive constant and (up to the choice of the unity of uncertainity) identical with the entropy in the sense of probability theory of the equilibrium distribution (= most probable distribution). In §. 2. two notions are introduced : the *dimension* of a probability distribution, and the entropy of dimension d of a probability distribution which has the dimension d. If ξ is random variable, which has a discrete distribution, i. e. $\mathbf{P}(\xi = x_k) = p_k$ $(k = 1, 2, \ldots)$ and $\sum_{k=1}^{\infty} p_k = 1$, then the distribution of ξ is said to have dimension 0, and the entropy of dimension 0 of its distribution, which is denoted by $\mathbf{H}_0(\xi)$ is defined by the well known formula of SHANNON :

$$\mathbf{H}_0(\xi) = -\sum_{k=1}^{\infty} p_k \log_2 p_k \tag{1}$$

provided that the series (1) converges. If ξ is an arbitrary (bounded) random variable, we consider the random variables $[n\xi]/n$ where $[x]$ denotes the integer part of x. If

$$\mathbf{H}_0\left(\frac{[n\xi]}{n}\right) = d \log_2 n + H + o(1) \tag{2}$$

is satisfied, we say that the distribution of ξ has the dimension d and its d-dimensional entropy $\mathbf{H}_d(\xi)$ is equal to H. The extension of these notions to probability distributions for which the relation (2) does not hold, is left to a forthcoming paper. If the probability density function $f(x)$ of the distribution of ξ exists, and is continuons up to a finite number of points in a finite interval and equal to 0 outside this interval, it follows that the probability distribution of ξ has the dimension 1 and its one-dimensional entropy $\mathbf{H}_1(\xi)$ is given by

$$\mathbf{H}_1(\xi) = -\int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx \tag{3}$$

This formula is identical with that used hitherto in information theory; what is essentially new in the present paper is that due emphasisis is laid on the fact, that the one-dimensional entropy of an absolutely continuous distribution defined by (3) is qualitatively different from the 0-dimensional entropy of a discrete distribution, defined by (1) ; a one-dimensional entropy, whatever its numerical value be, is always considered to be „greater" than a 0 dimensional entropy, whatever the latters numerical value be : the scale of 1-dimensional entropy is considered to be on a higher level. If a (bounded) probability distribution is obtained as a mixture of an absolutely continuous distribution, with the density function $f(x)$ continuous except for a finite number of points, and of a discrete distribution $\{p_k\}$ the two distributions being mixed with the weights q and p, then according to (2) this distribution has the dimension q and its q-dimensional entropy is equal to

$$\mathbf{H}_q(\xi) = -q\int_{-\infty}^{+\infty} f(x) \log_2 f(x)\, dx - p\sum_{k=1}^{\infty} p_k \log_2 p_k - p \log_2 p - q \log_2 q.$$

Thus there exists probability distributions with dimension d for any d in $(0, 1)$. In §. 3. some characteristic properties of the dimension and the entropy of a probability distribution are discussed. It is shown that the above definition of these notions is the consequence of a set of plausible properties. The dimension and the d-dimensional entropy of a probability distribution is characterized by properties closely related (though not identical even for $d=0$) to the properties used recently by A. J. KHINTCHIN [6] to characterize the 0-dimensional entropy.

In §. 4. the dimension and the entropy of the joint probability distribution of any finite number of random variables is defined. If the joint distribution of r random variables $\xi_1, \xi_2, \ldots, \xi_r$ has a Riemann-integrable density function $f(\boldsymbol{x})$ $(\boldsymbol{x} = (x_1, x_2, \ldots, x\ )$ which vanishes outside a finite r-dimensional sphere, the dimension of the joint distribution of $\xi_1, \xi_2, \ldots, \xi_r$ is equal to r, and the r-dimensional entropy of this distributon is defined by

$$\mathbf{H}_r(\xi_1, \xi_2, \ldots, \xi_r) = -\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{\infty} f(\boldsymbol{x}) \log_2 f(\boldsymbol{x})\, d\boldsymbol{x}$$

where $d\boldsymbol{x} = dx_1\, dx_2 \ldots dx_r$. The investigations started in this paper will be continued iin a forthcoming paper.