

2.3.2. Az entrópia

Eddig olyan példák kerültek szóba, amelyekben egyetlen esemény bekövetkezési valószínűségét, illetve az ebből eredő információtartalmat vizsgáltuk. Ha több esemény bekövetkezését vizsgáljuk, akkor *eseményrendszer*ről beszélünk. Ha egy eseményrendszert n darab esemény alkot, és ebből kiválasztunk k darabot, ezek közül vagy egyik, vagy másik, vagy valamelyik (bármelyik) bekövetkezésének valószínűsége a k darab részesemény bekövetkezési valószínűségeinek összegével lesz egyenlő:

$$p_{(1,2,\dots,k)} = p_1 + p_2 + \dots + p_k. \quad (7)$$

Tulajdonképpen a 4. sz. példa (páros számú pettyek dobása) kérdését is átfogalmazhatjuk úgy, hogy eseményrendszert alkosson a lehetséges kimenetek száma: Mennyi a valószínűsége annak, hogy 2-est, vagy 4-est, vagy 6-ost dobunk? A megoldás magyarázata során már utaltunk erre a tényre, de ott még egy kalap alá vettük ezeket az eseményeket, és csak mint páros szám dobását fogalmaztuk meg a feltételt, nem volt jelentősége a részeseményeknek.

8. *példa.* Van 10 darab golyó egy dobozban: 1 fehér, 2 sárga, 4 piros, a többi kék. Mennyi a valószínűsége annak, hogy találmomra kiválasztva egyet közülük, fehéret, vagy sárgát, vagy pirosat húzunk ki? Megoldás: a golyók száma: $n=10$, ezek között fehér: $k(f)=1$, sárga: $k(s)=2$, piros: $k(p)=4$ darab található. A három szín valamelyike közül történő húzás valószínűsége (5) és (7) képletekből adódóan:

$$p(f, s, p) = p(f) + p(s) + p(p) = \frac{k(f)}{n} + \frac{k(s)}{n} + \frac{k(p)}{n} = \frac{1}{10} + \frac{2}{10} + \frac{4}{10} = \frac{7}{10} = 0,7.$$

Ha a 3 db. kék golyó [$k(k)=3$] húzása is megengedett lenne, nyilván a biztos valószínűséget kapnánk, így alakulna a számítás:

$$\begin{aligned} p(f, s, p, k) &= p(f) + p(s) + p(p) + p(k) = \\ &= \frac{k(f)}{n} + \frac{k(s)}{n} + \frac{k(p)}{n} + \frac{k(k)}{n} = \frac{1}{10} + \frac{2}{10} + \frac{4}{10} + \frac{3}{10} = \frac{10}{10} = 1. \end{aligned}$$

Általános esetben az n db. eseményt tartalmazó eseményrendszerrel elmondható a megállapítás: *A rendszer eseményeire vonatkozó összes valószínűség értéke 1, azaz, a biztos esemény.*

$$\sum_{i=1}^n p(i) = p(1) + p(2) + \dots + p(n) = 1. \quad (8)$$

Ha feltételezzük, hogy a lehetséges esetek között nincs nulla valószínűségű esemény, és az összes esemény lefedi a teljes valószínűséget, azaz a (8) képlet érvényes a rendszerre, akkor számítható a rendszer információátlaga.

A rendszer átlagos információtartalmát, vagy más szóval várható értékét az események előfordulási valószínűségének és azok információtartalmának szorzatösszege adja. Ezt nevezzük entrópiának:

$$H(x) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + \dots + p_n \log_2 \frac{1}{p_n} = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i} = \sum_{i=1}^n p_i I(x_i). \quad (9)$$

A (9) képletben a $H(x)$ az entrópiát jelöli. A képlet segítségével számolható érték entrópia elnevezése a fizikából ismert entrópia képletéhez való – pusztán formai – hasonlósága miatt adódott.

9. *példa.* Mennyi az entrópiája annak a rendszernek, amelyben $p_1=0,25$; $p_2=0,25$; és $p_3=0,5$? A megoldás a (9) képletet felhasználva:

$$\begin{aligned} H(x) &= p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + p_3 \log_2 \frac{1}{p_3} = 0,25 \log_2 \frac{1}{0,25} + 0,25 \log_2 \frac{1}{0,25} + 0,5 \log_2 \frac{1}{0,5} = \\ &= \frac{1}{4} \log_2 \frac{1}{\frac{1}{4}} + \frac{1}{4} \log_2 \frac{1}{\frac{1}{4}} + \frac{1}{2} \log_2 \frac{1}{\frac{1}{2}} = \frac{1}{4} \log_2 4 + \frac{1}{4} \log_2 4 + \frac{1}{2} \log_2 2 = \frac{1}{4} 2 + \frac{1}{4} 2 + \frac{1}{2} 1 = \frac{3}{2} = 1,5. \end{aligned}$$