1. Súlyozott diszkrét valószínűségi változó

Legyen X egy diszkrét valószínűségi változó, amely egy d oldalú dobókocka kimenetelét reprezentálja. A dobókocka oldalaihoz tartozó kimeneteket az alábbi halmaz írja le:

$$\{x_1, x_2, \ldots, x_d\},$$

ahol tipikusan $x_i = i$ $(1 \leq i \leq d)$, de a levezetés általánosabb formában is érvényes.

Minden oldalhoz rendelünk egy nemnegatív súlyt:

$$w_1, w_2, \ldots, w_d \geq 0 \quad \text{úgy, hogy} \quad \sum_{i=1}^{d} w_i > 0.$$

Ezek a súlyok határozzák meg a kimenetelek relatív esélyeit.

A súlyokból kapott elméleti valószínűségek:

$$p_i = \frac{w_i}{\sum_{j=1}^{d} w_j}, \qquad i = 1, \ldots, d.$$

Ez biztosítja, hogy $0 \leq p_i \leq 1$ és $\sum_{i=1}^{d} p_i = 1$.

2. Valószínűségi tömegfüggvény

A súlyozott dobókocka eloszlását a relatív gyakoriságfüggvény (diszkrét „sűrűségfüggvény") írja le:

$$f_X(x_i) = P(X = x_i) = p_i, \qquad i = 1, \ldots, d.$$

Ez a projektben az elméleti (theoretical) PMF-et jelenti, míg a szimuláció során kapott relatív gyakoriságok az empirikus PMF-et adják:

$$\hat{p}_i = \frac{n_i}{n}, \qquad i = 1, \ldots, d,$$

ahol n a kísérletek száma, n_i pedig annak a kísérletekben előforduló számossága, hogy $X = x_i$.

3. Eloszlásfüggvény

A diszkrét eloszlásfüggvény definíciója:

$$F_X(k) = P(X \leq k) = \sum_{i:\, x_i \leq k} f_X(x_i) = \sum_{i:\, x_i \leq k} p_i.$$

Az empirikus eloszlásfüggvény (empirical CDF) definíciója:

$$\widehat{F}_n(k) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{X_j \leq k\}},$$

ahol $\mathbf{1}_A$ az A esemény indikátorfüggvénye, azaz

$$\mathbf{1}_A = \begin{cases} 1, & \text{ha } A \text{ bekövetkezik,} \\ 0, & \text{különben.} \end{cases}$$

4. Várható érték és második momentum

A valószínűségi változó várható értéke:

$$E(X) = \sum_{i=1}^{d} x_i \, p_i = \sum_{i=1}^{d} x_i \, f_X(x_i).$$

Az empirikus várható érték (mintaátlag) a relatív gyakoriságok segítségével:

$$\widehat{E}(X) = \sum_{i=1}^{d} x_i \, \hat{p}_i.$$

A második momentum definíciója:

$$E(X^2) = \sum_{i=1}^{d} x_i^2 \, p_i.$$

5. Variancia

A variancia (szórásnégyzet) definíciója:

$$D^2(X) = E(X^2) - \big(E(X)\big)^2.$$

A szimulációból származó empirikus szórásnégyzet:

$$\widehat{D^2}(X) = \sum_{i=1}^{d} x_i^2 \hat{p}_i - \left(\sum_{i=1}^{d} x_i \hat{p}_i \right)^2.$$

6. Hibamérés

Az elméleti és empirikus relatív gyakoriságfüggvény különbségét az alábbi hibamértékekkel jellemezzük.

Abszolút hiba:

$$\varepsilon_i^{(\mathrm{abs})} = \big| \hat{p}_i - p_i \big|.$$

Relatív hiba:

$$\varepsilon_i^{(\mathrm{rel})} = \frac{\big| \, \hat{p}_i - p_i \, \big|}{p_i}, \qquad p_i > 0.$$