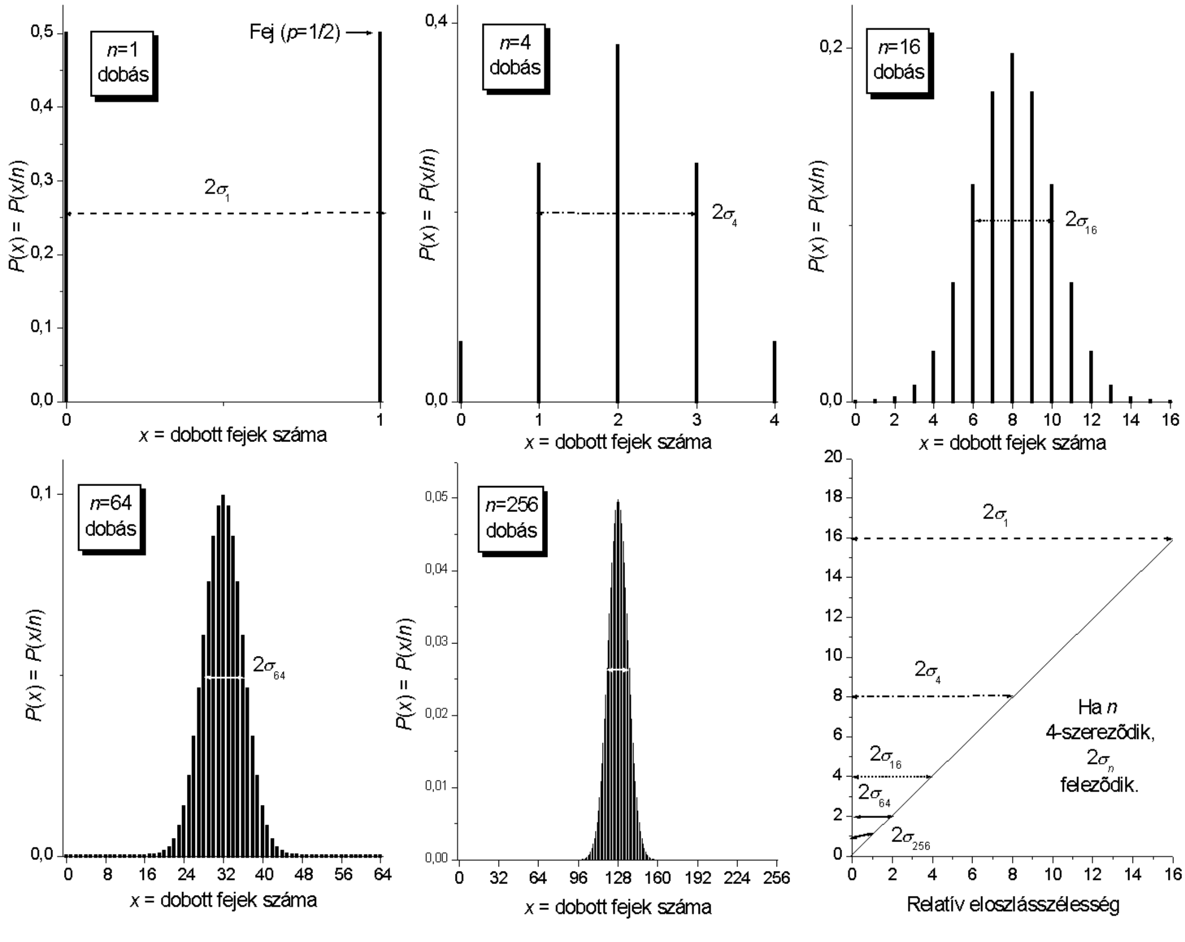
2.4 A centrális határeloszlás-tétel

Legyen X tetszőleges valószínűségi változó, melynek létezik várható értéke (μ) és szórása (σ). Legyenek az X_i-k az X egymástól független megfigyeléseivel kapcsolatos (nyilván független és egyforma eloszlású) valószínűségi változók (n db.). Ekkor ezek $\overline{X}$ átlaga (mintaátlag) aszimptotikusan $N(\mu, \sigma^2/n)$ **normális eloszlású** valószínűségi változó.

A centrális határeloszlás-tételnek az a **gyakorlati tartalma**, hogy ha egy mennyiség „valódi értékére” mért adatok átlagolása útján próbálunk következtetni, akkor **„normális dolog” az átlag szórását normális eloszlásúnak feltételezni**, akármilyen is a mért adatok tényleges eloszlása – feltéve, hogy az átlagolásnak egyáltalán értelme van, azaz feltéve, hogy létezik várható érték. (Amint a 2. ábra szövegéből kiderül, a centrális határeloszlás-tételt nemcsak átlagra, hanem összegre is ki lehet mondani, s ráadásul jóval gyengébb feltételekkel is, mint amit itt megadtam.)

A **centrális határeloszlás-tétel *n*-tagú összegekre** kimondva pl. a következő módon fogalmazható meg: A fenti feltételek mellett, a $\Sigma_n \equiv (X_1+X_2+\ldots+X_n)$ összeg aszimptotikusan $N(n\mu, n\sigma^2)$ normális valószínűségi változót állít elő.



2. ÁBRA. A fenti ábrasor öt *fej vagy írás* dobássor – öt Bernoulli-kísérletsorozat – kimeneteleinek eloszlását mutatja (mind az öt binomiális eloszlás – l. később). A fej valószínűségét p=0,5-nek vettem, s a dobássor hossza (n=1, 4, 16, 64, 256) mindig négyszereződik. Az ábra több dolgot demonstrál egyszerre. (1) **Centrális határeloszlás-tétel**: Elég sok független, egyforma eloszlású valószínűségi változó összege (ti. a dobott fejek x száma) normális eloszlást követ. (2) A binomiális eloszlás elég nagy n-re jól közelíthető normális eloszlással. (3) **Négyszer annyi adat átlaga kétszer olyan pontos** a jobb alsó grafikon szerint. Az átlag úgy jön ide, hogy a 0 és 1 közé eső x/n hányados n db. 0 vagy 1 értékű Bernoulli-változó (l. később) átlagaként is értelmezhető. (4) **Nagy számok törvénye**: Az n növekedésével a binomiális eloszlás egyre keskenyebb lesz, ami garantálja, hogy a relatív gyakoriságok $n\to\infty$ esetén tökéletesen reprezentálják a p=0,5 valószínűséget. (Vegyük észre, hogy az x/n hányados nemcsak a Bernoulli-változó átlagértékét, hanem a fej relatív gyakoriságát is jelenti egy-egy konkrét dobássorozatban.)

2.5 A hibaterjedés általános közelítő formulái

Tetszőleges valószínűségi változók esetén az ezekből számított $f(X_1, X_2,\ldots X_n)$ formula hibájára a következő **közelítő** összefüggést kapjuk:

$$D^2(f) \approx \sum_{i=1}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)^2_{\mathbf{x}=\boldsymbol{\mu}} D^2(x_i)\right] + 2\sum_{i<j}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)_{\mathbf{x}=\boldsymbol{\mu}}\left(\frac{\partial f}{\partial x_j}\right)_{\mathbf{x}=\boldsymbol{\mu}} Cov(x_i, x_j)\right] \qquad (57)$$

Ha a valószínűségi változók **függetlenek**, akkor a kovarianciákat tartalmazó második összeg kiesik, azaz:

$$D^2(f) \approx \sum_{i=1}^{n}\left[\left(\frac{\partial f}{\partial x_i}\right)^2_{\mathbf{x}=\boldsymbol{\mu}} D^2(x_i)\right] \qquad (58)$$

Az $\mathbf{x} = \boldsymbol{\mu}$ index azt jelenti, hogy a deriváltakat a változók várható, ill. „beállítani szándékozott" értékénél kell kiszámítani.

A fenti képletek az $f(X_1, X_2,\ldots X_n)$ függvény differenciáljából – tehát a függvény lineáris közelítéséből – származnak. Emiatt gyakorlatilag teljesen általánosan használhatók ugyan **folytonos változókra**, de csak **viszonylag kis hibák esetében adnak jó közelítést**. Ezzel szemben az összegre kapott szórási formulák egzaktak.

2.6 Független valószínűségi változók szorzatára/hányadosára vonatkozó képletek

A **várható értékre egzakt** módon teljesül, hogy:

$$E\left(\frac{X\,Y}{Z}\right) = E(X)E(Y)E\left(\frac{1}{Z}\right) \qquad (59)$$

Ha $D^2(Z)/E^3(Z)$ elég kicsi, akkor az $1/Z$ kifejezés $E(Z)$ körüli sorfejtéséből **közelítőleg** azt kapjuk, hogy:

$$E\left(\frac{X\,Y}{Z}\right) \approx \frac{E(X)E(Y)}{E(Z)} \qquad (60)$$

A **szórásra** a következő **közelítés** nyerhető a megfelelő hibaterjedési formulából:

$$\left[\frac{D\left(\frac{XY}{Z}\right)}{E\left(\frac{XY}{Z}\right)}\right]^2 \approx \left[\frac{D(X)}{E(X)}\right]^2 + \left[\frac{D(Y)}{E(Y)}\right]^2 + \left[\frac{D(Z)}{E(Z)}\right]^2 \qquad (61)$$

Szorzásnál és osztásnál tehát a **relatív szórások (hibák) négyzetei adódnak össze** a varianciák helyett. Vegyük észre, hogy az **osztásnál is összeadás szerepel** a hibaformulában.