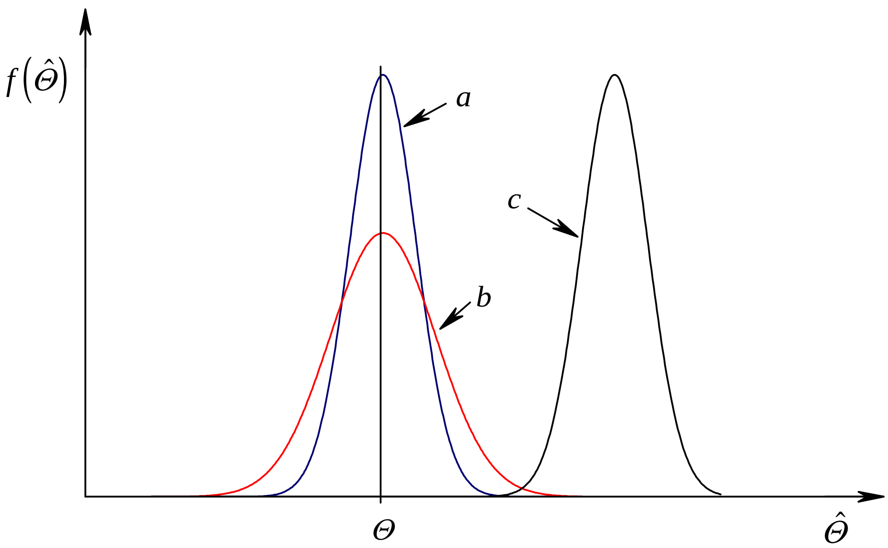
2.3. Paraméterbecslés

A becslés művelete: a sokaság tulajdonságára (paraméterére) következtetünk a minta adatai (jellemzői) alapján. A becslés: a mintából kiszámított statisztika; pl. a várható érték egyik lehetséges becslése a mintaelemek számtani középértéke ($\bar{x}$). A becslés is valószínűségi változó, eloszlása van. A becsült érték, amelyet jelölésben a becsléstől nem különböztetünk meg, egy számérték, amely nem valószínűségi változó, nincs eloszlása. Általában: ha Θ jelenti a paramétert, a becslés jele $\hat{\Theta}$; szokás még a paraméterek valódi értékére görög betűket (μ, σ, α), a becslésre pedig a megfelelő latin betűket (m, s, a) használni.

A 2-20. ábra szerint *a* jobb becslés, mint *b*, mert kisebb a valódi érték körüli ingadozás; *c*-re a várható érték nem Θ.



2-20. ábra. Becslések összehasonlítása

2.3.1. A becslések tulajdonságai

Jelölje $\hat{\Theta}_n$ az n elemű mintából meghatározható becslést. A $\hat{\Theta}_n$ valószínűségi változó, amely mintáról mintára más-más értéket vehet fel.

Torzítatlan egy becslés, ha $\hat{\Theta}_n$ mint valószínűségi változó Θ körül ingadozik, várható értéke Θ:

$$E\left(\hat{\Theta}_n\right) = \Theta. \qquad (2.41)$$

A torzítás: $E\left(\hat{\Theta}_n\right) - \Theta$;

a korrekció pedig: $\Theta - E\left(\hat{\Theta}_n\right)$.

Vannak paraméterek, amelyekre nem létezik torzítatlan becslés. Ilyenkor olyan becslést keresünk, amelynek torzítása n növelésével csökken. A becslés *aszimptotikusan torzítatlan*, ha

$$\lim_{n\to\infty} E\left(\hat{\Theta}_n\right) = \Theta. \tag{2.42}$$

Kívánatos, hogy a $\hat{\Theta}_n$ becslés ingadozásai ne legyenek túlságosan nagyok, vagyis a $\hat{\Theta}_n$ aktuális becslés kis valószínűséggel térjen el a Θ paraméter értékétől, azaz a becslésnek mint valószínűségi változónak a varianciája lehetőleg kicsi legyen. A becslés varianciája a *becslés hatásosságának* mértéke. Két torzítatlan becslés közül az a becslés hatásosabb (*efficiensebb*), amelyiknek kisebb a varianciája.

Hasznos tulajdonsága egy becslésnek, ha a minta elemszámának növelésével a becslés a paraméter igazi értékéhez tart, pontosabban n növelésével egyre csökken annak valószínűsége, hogy Θ -tól jelentősen eltérjen:

$$\lim_{n\to\infty} P\left(\left|\hat{\Theta}_n - \Theta\right| > \varepsilon\right) = 0. \tag{2.43}$$

Ez a *konzisztens becslés*. A konzisztens becslés nem feltétlenül torzítatlan, de legalább aszimptotikusan torzítatlan (Vincze, 1968).

2-19. példa

Vizsgáljuk meg, hogy

a) a számtani átlag $(\bar{x})$

b) az n-edik mért érték (x_n)

milyen becslése a várható értéknek! (Ha nincs rendszeres hiba, akkor a várható érték azonos a valódi értékkel; ilyen pl. a tömegmérés jó mérlegen, parallaxismentes skálaleolvasással.)

a) $$\hat{\mu} = \bar{x} = \frac{\sum_i x_i}{n}. \tag{2.44}$$

A becslés várható értéke:

$$E(\bar{x}) = \frac{1}{n} E\left(\sum_i x_i\right) = \frac{1}{n} n \sum_i E(x_i) = E(x_i) = \mu, \tag{2.45}$$

vagyis torzítatlan, és belátható, hogy konzisztens is. A becslés varianciája:

$$Var(\bar{x}) = \frac{1}{n^2} Var\left(\sum_i x_i\right) = \frac{nVar(x_i)}{n^2} = \frac{\sigma^2}{n}. \tag{2.46}$$

b) $E(\hat{\mu}) = E(x_n) = \mu$,

torzítatlan, de kevésbé efficiens az előzőnél, mivel $Var(x_n) = \sigma^2 > \sigma^2/n$, és nem konzisztens.

2-20. példa

Vizsgáljuk meg, hogy a korrigálatlan tapasztalati szórásnégyzet milyen becslése a varianciának!

$$\widehat{Var(x)} = s^{2*} = \frac{\sum_i (x_i - \bar{x})^2}{n}. \tag{2.47}$$

A (2.10) összefüggéssel bizonyítottuk, hogy a $\sum (x_i - \bar{x})^2$ kifejezés $\chi^2 \sigma^2$ eloszlású, várható értéke $(n-1)\sigma^2$, mivel $\nu = n-1$, tehát

$$E(s^{2*}) = \frac{n-1}{n}\sigma^2. \tag{2.48}$$

Az s^{2*} tehát torzított, bár aszimptotikusan torzítatlan, konzisztens becslés. Ennek $n/(n-1)$-szeresét célszerű használni, hogy torzítatlan legyen. Ez az ún. *korrigált tapasztalati szórásnégyzet*:

$$s^2 = \frac{1}{n-1}\sum (x_i - \bar{x})^2 . \tag{2.49}$$

Varianciája: $$Var(s^2) = \frac{2\sigma^4}{\nu}. \tag{2.50}$$

Tehát a (2.49) szerinti korrigált tapasztalati szórásnégyzet konzisztens becslése a varianciának, mert a minta elemszámának növelésével a torzítatlan becslés varianciája csökken.

2-21. példa

a) Vizsgáljuk meg, hogy a súlyozott átlag $(\bar{\bar{x}})$ a várható értéknek milyen becslése!

$$\hat{\mu} = \bar{\bar{x}} = w_1 x_1 + w_2 x_2 + \ldots + w_n x_n = \sum_i w_i x_i , \tag{2.51}$$

ahol a w_i értékek tetszőleges számok, az ún. súlyok. A súlyozott átlag várható értéke:

$$E(\bar{\bar{x}}) = E\left(\sum_i w_i x_i\right) = \sum_i w_i E(x_i) = \mu \sum_i w_i .$$

Látható, hogy amennyiben

$$\sum_i w_i = 1, \tag{2.52}$$

a becslés torzítatlan.

b) A továbbiakban vizsgáljuk meg, hogy milyen súlyozás szükséges ahhoz, hogy az előbbi torzítatlan becslés efficiens (a legkisebb varianciájú) legyen!