Paraméteres intervallumbecslés

A pontbecslések, gyakorlati alkalmazhatóságuk ellenére, minden felmerülő problémára nem adhatnak teljes választ: végeredményben a folytonos esetben az az esemény, hogy a becslésfüggvény éppen a keresett paraméter igazi értékét veszi fel, nulla valószínűségű. A statisztikai következtetések másik fontos fajtája az **intervallumbecslés**: legyen $\xi_1, \dots, \xi_n$ véletlen minta a $f(x, \theta)$ sűrűségfüggvényű populációból, és legyen $t_1(\xi_1, \dots, \xi_n)$ és $t_2(\xi_1, \dots, \xi_n)$ statisztikai függvények, amelyekre teljesül, hogy $t_1 \leq t_2$, $P(t_1 < g(\theta) < t_2) = c$, ahol c független a θ paramétertől. A (t_1, t_2) intervallumot $100c$%**-os szintű konfidencia intervallum**nak, c-t **biztonsági szintnek**, $p = 1-c$-t **kockázati valószínűség**nek, a t_1 és t_2 valószínűségi változókat **alsó illetve felső konfidencia korlát**nak nevezzük.

Megjegyezzük, hogy ha egy θ valós paraméterre van (t_1, t_2) konfidencia intervallum adott szinten, és g szigorú monoton növekvő függvény, akkor $(g(t_1), g(t_2))$ a $g(\theta)$-ra konfidencia intervallum az adott szinten.

Legyen $\xi_1, \dots, \xi_n$ véletlen minta a $f(x, \theta)$ sűrűségfüggvényű populációból, és legyen $r(\xi_1, \dots, \xi_n, \theta)$ olyan valószínűségi változó, amelynek eloszlása az ismeretlen θ paramétertől nem függ, és létezik sűrűségfüggvénye. Ekkor bármely $0 < c < 1$ konfidencia szinthez található $r_1(c)$, $r_2(c)$ érték, hogy $P(r_1(c) < r < r_2(c)) = c$, továbbá kereshetünk $t_1(\xi_1, \dots, \xi_n)$ és $t_2(\xi_1, \dots, \xi_n)$ statisztikai függvényeket, melyek függetlenek a θ paramétertől, és $r_1(c) < r(\xi_1, \dots, \xi_n, \theta) < r_2(c)$ pontosan akkor, ha $t_1(\xi_1, \dots, \xi_n) < g(\theta) < t_2(\xi_1, \dots, \xi_n)$. Ekkor (t_1, t_2) $100c$%-os szintű konfidencia intervallum.

A statisztikus módszer. Legyen $\xi_1, \dots, \xi_n$ véletlen minta a $f(x, \theta_0)$ sűrűségfüggvényű populációból, ahol $\theta_0 \in \mathbb{R}$ és a Θ paramétertér intervallum. Legyen a $T(\xi_1, \dots, \xi_n)$ statisztika θ_0 pontbecslése, például a maximum likelihood becslés. Tételezzük fel, hogy $T(\xi_1, \dots, \xi_n)$ folytonos valószínűségi változó, ismeretes $f_T(t, \theta)$ sűrűségfüggvénye (a módszer, a megfelelő változtatásokkal, alkalmazható a diszkrét esetben is). Legyen p_1, p_2 rögzített pozitív számok, $p_1 + p_2 < 1$. Ha $t_0 = T(x_1, \dots, x_n)$ a statisztika megfigyelt értéke, legyen $h_1(\theta)$ a

$$p_1 = \int_{-\infty}^{h_1(\theta)} f_T(t, \theta)dt,$$

továbbá $h_2(\theta)$ a

$$p_2 = \int_{h_2(\theta)}^{\infty} f_T(t, \theta)dt,$$

egyenlőségek által meghatározott függvények, amelyekről feltesszük, hogy monoton növekvők. Tudjuk, hogy $h_1(\theta) < h_2(\theta)$. Legyen $r_1 = r_1(t_0)$, $r_2 = r_2(t_0)$

olyan, hogy $h_2(r_1) = h_1(r_2) = t_0$. Következésképp a $h_1(\theta) < t_0 < h_2(\theta)$ esemény pontosan akkor következik be, ha $r_1 < \theta_0 < r_2$ bármely megfigyelt $x_1, \ldots, x_n$ mintaérték esetén. Azonban $P(h_1(\theta) < t_0 < h_2(\theta)) = 1 - p_1 - p_2$, és innen $P(r_1 < \theta_0 < r_2) = 1 - p_1 - p_2$, azaz $(r_1(t_0), r_2(t_0))$ $100(1 - p_1 - p_2)$%-os szintű konfidencia intervallum θ-ra. Megjegyezzük, hogy elegendő a következő egyenleteket megoldani: $r_1(t_0)$ a

$$p_2 = \int_{t_0}^{\infty} f_T(t, \theta) dt,$$

továbbá $r_2(t_0)$ a

$$p_1 = \int_{-\infty}^{t_0} f_T(t, \theta) dt,$$

θ-ra vonatkozó egyenlet megoldása (a diszkrét esetben szummák vannak az integrálok helyett). Amennyiben a $h_1(\theta)$, $h_2(\theta)$ függvények nem monotonok, konfidencia intervallumrendszert kapunk.