Kort om konfidensintervall

Andreas Rasmusson

November 14, 2024

Definition

Låt

1. X vara en slumpvariabel med (okänt) väntevärde $\mu < \infty$ och (okänd) standardavvikelse $0 < \sigma < \infty$
2. $\{X_i\}_{i=1}^n$ vara ett i.i.d stickprov för X.
3. $\bar{X}_n$ vara slumpvariabeln definierad genom $\bar{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$
4. S vara slumpvariabeln definierad genom $S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2}$
5. $\delta \in [0, 1]$
6. z_δ vara det reella tal för vilket gäller att en standardiserad normalfördelad slumpvariabel Z uppfyller $\mathbb{P}(-z_\delta \leq Z \leq z_\delta) = \delta$
7. $Y_{1,n}$ vara slumpvariabeln $\bar{X}_n - z_\delta \cdot \frac{S_n}{\sqrt{n}}$
8. $Y_{2,n}$ vara slumpvariabeln $\bar{X}_n + z_\delta \cdot \frac{S_n}{\sqrt{n}}$

Vi kallar *slumpvektorn* $(Y_{1,n}, Y_{2,n})$ för ett konfidensinterval av ordning n med konfidensgrad δ för μ. Ett *observerat* konfidensintervall av ordning n med konfidensgrad δ är ett talpar $(Y_{1,n}(\omega), Y_{2,n}(\omega))$ för något $\omega \in \Omega$ där $(\Omega, \mathcal{F}, \mathbb{P})$ är det underliggande sannolikhetsrummet.

Sats

Låt X vara en slumpvariabel med väntevärde $-\infty < \mu < \infty$ och standardavvikelse $0 < \sigma < \infty$. Låt vidare, för varje $n \geq 1$, $(Y_{1,n}, Y_{2,n})$ vara ett konfidensintervall av ordning n med konfidensgrad $0 < \delta < 1$ för μ. Då gäller att

$$\lim_{n\to\infty} \mathbb{P}(Y_{1,n} \leq \mu \leq Y_{2,n}) = \delta$$

Anmärkning

Observera att teoremet ovan är ett uttalande om kvantiteter som varierar slumpmässigt i förhållande till en fix kvantitet μ. Teoremet säger *inte* att i efterhand, när man fått sitt *observerade* konfidensintervall I, gäller att "μ tillhör I med sannolikhet nära δ". Faktum är att det senare uttalandet är falskt eftersom $0 < \delta < 1$ och såväl μ som I är fixa. Det finns ingen icketrivial sannolikhet involverad i huruvida μ tillhör I. Det vore som att säga "Sannolikheten att $2 \leq 3 \leq 4$ är 0.95" eller "Sannolikheten att $2 < 8 < 4$ är 0.95". På samma sätt är det inte korrekt att säga att "I täcker μ med sannolikhet nära δ". Båda dessa påståenden är alltså falska. Däremot är det, under ovanstående definition av konfidensintervall, korrekt att säga att "För stora stickprov gäller att ett konfidensintervall med konfidensgrad δ täcker μ med sannolikhet nära δ" eller "För stora stickprov tillhör μ ett konfidensintervall med konfidensgrad δ med sannolikhet nära δ" eftersom det är uttalanden om en slumpvektor och inte om fixa kvantiteter.

Bevis av satsen

Vi har att

$$\mathbb{P}(Y_{1,n} \leq \mu \leq Y_{2,n}) = \mathbb{P}\left(\bar{X}_n - z_\delta \cdot \frac{S_n}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_\delta \cdot \frac{S_n}{\sqrt{n}}\right)$$

$$=$$

$$\mathbb{P}\left(\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z_\delta \frac{S_n}{\sigma} \& \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \geq -z_\delta \frac{S_n}{\sigma}\right)$$

$$=$$

$$\mathbb{P}\left(-z_\delta \frac{S_n}{\sigma} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z_\delta \frac{S_n}{\sigma}\right)$$

Eftersom S_n är en konsistent skattning av σ, vet vi att S_n konvergerar mot σ i sannolikhet och därmed också i fördelning. Vidare vet vi från centrala gränsvärdessatsen att $\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$ konvergerar i fördelning mot Z. Således måste det gälla att

$$\lim_{n\to\infty} \mathbb{P}(Y_{1,n} \leq \mu \leq Y_{2,n}) = \lim_{n\to\infty} \mathbb{P}\left(-z_\delta \frac{S_n}{\sigma} \leq \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq z_\delta \frac{S_n}{\sigma}\right) = \mathbb{P}(-z_\delta \leq Z \leq z_\delta) = \delta$$

, precis som utlovat. Beviset är nu färdigt. $\Diamond$