Parametrisk punktskattning

Vår statistiska modell:

1) Vi har en population av storlek N ur vilken vi drar ett slumpmässigt stickprov $X_1, \ldots, X_n$ (av storlek n).

2) Vi kommer (nästan) alltid låta $N = \infty$ så att $X_1, X_2, \ldots$ är i.i.d.

3) Vi antar att fördelningen av X_i är känd sånär som på värdet av en eller flera parametrar.

Ex: $X_i \sim \text{Exp}(\lambda)$ med λ okänt

Ex: $X_i \sim N(\mu, \sigma^2)$ med μ och/eller σ^2 okända.

Vi vill skatta (gissa) parametrarna, och låter θ beteckna dessa.

Def: En skattning är en funktion

$\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ av stickprovet $(X_1, \ldots, X_n) = \underline{X}$

Ex: 1) Om $\theta = \mu = \mathbb{E}[X_i]$ så använder vi skattningen $\hat{\theta}(X) = \overline{X} = \dfrac{X_1 + \dots + X_n}{n}$

2) Om $\theta = \sigma^2 = \mathrm{Var}(X_i)$ så använder vi ofta

$$\hat{\theta} = \widehat{(\sigma^2)} = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

om $\mu = \mathbb{E}[X_i]$ är okänd. Alternativt använder vi

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2$$

om μ är känd.

3) Om $X_i \sim \mathrm{Exp}(\lambda)$ så är $\mu = \mathbb{E}[X_i] = \frac{1}{\lambda}$

och därför är $\hat{\lambda} = \dfrac{1}{\overline{X}}$ naturligt.

Def: i) $\hat{\theta}$ är väntevärdesriktig (VVR)

om $\mathbb{E}[\hat{\theta}(X)] = \theta$. (unbiased)

ii) Effektiviteten (av VVR-skattningar)

bedöms genom variansen $\mathrm{Var}(\hat{\theta}(X))$

(liten = bra)

iii) $\hat{\Theta}$ är konsistent om $\hat{\Theta}(\bar{X}) \xrightarrow[n\to\infty]{P} \Theta$,

dvs om $\forall\ \varepsilon > 0$ $\quad \mathbb{P}(|\hat{\Theta}(\bar{X}) - \Theta| > \varepsilon) \to 0, \ n \to \infty.$

Anm: iii) kräver $N = \infty$. Fallet $N < \infty$ kräver mycket extra räkning med relativt liten vinst.

Observera att $\hat{\Theta}(\bar{X})$ är en s.v. Fördelningen för $\hat{\Theta}(\bar{X})$ är av central betydelse.

Ex1: Betrakta $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Fördelningen av $\bar{X}$ är i allmänhet svåråtkomlig men:

i) $$\mathbb{E}[\bar{X}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[X_i] = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu \quad \text{så VVR}$$

ii) $$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \underset{\text{ober.}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i) = \frac{\sigma^2}{n}$$

om $\sigma^2 = \mathrm{Var}(X_i)$.

iii) $$\mathbb{P}(|\bar{X} - \mu| > \varepsilon) \xrightarrow[n\to\infty]{} 0 \quad (\text{STL}) \quad \text{så } \bar{X} \text{ är kons.}$$

<u>Ex 2:</u> Betrakta $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$.

Lång räkning ger att $\mathbb{E}[s^2] = \sigma^2$ så VVR.

(se sid 211-212).

<u>Ex 3:</u> Vi vill skatta $p = \mathbb{P}(A)$ där $A \subset \Omega$

är någon händelse (utvald person har cancer etc.).

Vi vill skatta p och låter

$$X_i = \begin{cases} 1 & \text{om } A \text{ inträffar på försök } i \\ 0 & \text{annars.} \end{cases}$$

Vi tar $\hat{p} = \bar{X}$. Vi har att $X_1 + \dots + X_n \sim Bin(n,p)$

s.a. $\mathbb{E}[\hat{p}] = \frac{1}{n} \mathbb{E}[\sum_{i=1}^{n} X_i] = p$ och

$$Var(\hat{p}) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2} np(1-p) = \frac{p(1-p)}{n}$$

Vi är ofta intresserade av $Var(\hat{\theta}(X))$.

<u>Def:</u> <u>Standardfelet</u> är $\sqrt{Var(\hat{\theta}(X))}$

och betecknas ibland $D(\hat{\theta})$.

<u>Anm:</u> Ofta beror $Var(\hat{\theta}(X))$ på den

okända parametern θ. Därför måste $D(\hat{\theta})$ skattas. Vi skattar en egenskap hos en skattare!

Ex: Om $\bar{X}$ skattar μ har vi $Var(\bar{X}) = \frac{\sigma^2}{n}$.

Vi skattar $\sqrt{Var(\bar{X})}$ med $S_{\bar{X}} = \frac{S}{\sqrt{n}}$ där

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

Ex: Om $\hat{p}$ skattar p låter vi $S_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ skatta $\sqrt{Var(\hat{p})} = \sqrt{\frac{p(1-p)}{n}}$.

Hur skall vi hitta en bra skattning/skattare?

Momentmetoden

Om $X_1, X_2, \ldots$ är i.i.d. med parameter θ är det ofta så att $E[X_1] = f(\theta)$ för ngn funktion f. Gör följande:

1) Observera att STL säger att för n stort är

$$\bar{X} \approx E[X_1]$$

2) Lös ut $\hat{\theta}$ ur ekvationen $\bar{X} = f(\hat{\theta})$

Vi ser att $f(\theta) = \mathbb{E}[X_1] \approx \bar{X} = f(\hat{\theta})$ och därmed bör också $\theta \approx \hat{\theta}$ så att $\hat{\theta}$ är en bra skattare.

Detta kallas för en MME- (Method of Moment Estimator) skattare.

Tal: Väntetiden på en buss anses vara $U[0, \theta]$. a) Hitta MME-skattare för θ.

b) Om vi fick mätvärdena $x_1 = 2$, $x_2 = 7$, $x_3 = 4$, $x_4 = 15$, $x_5 = 11$ minuter, vilket blir det numeriska värdet på $\hat{\theta} (= \hat{\theta}(x_1, \dots, x_5))$?

L: a) Om $X_1 \sim U[0, \theta] \Rightarrow \mathbb{E}[X_1] = \frac{\theta}{2} \Rightarrow \bar{X} = \frac{\hat{\theta}}{2}$

$\Rightarrow \hat{\theta} = 2\bar{X}$ är min MME.

b) Vi får att $2\bar{x} = 2 \frac{x_1 + \dots + x_5}{5} = 15{,}6$ minuter.

Om vi har två okända parametrar θ_1, θ_2 beräknar vi

$$E[X] = f_1(\theta_1, \theta_2)$$

$$E[X^2] = f_2(\theta_1, \theta_2)$$

Vi ersätter: $\overline{X} = f_1(\hat{\theta}_1, \hat{\theta}_2)$

$\overline{X^2} = f_2(\hat{\theta}_1, \hat{\theta}_2)$

och löser ut $\hat{\theta}_1, \hat{\theta}_2$.

Här är $\overline{X^k} = \frac{1}{n} \sum_{i=1}^{n} X_i^k$.