February 6, 2018

1 Föreläsning VIII

1.1 Punktskattning

Punktskattning av μ

- Vi låter $\{\xi_1, \xi_2, \ldots, \xi_n\}$ vara oberoende likafördelade stokastiska variabler (med ett gemensamt μ).
- $\overline{\xi} =: \mu^*$ är en *punktskattning* av μ, en stokastisk variabel.
- Vi antar att man har observerat värden $x_1, x_2, \ldots, x_n$ av dessa. Då är $\overline{x} =: \mu^*_{\text{obs}}$ en *observerad* punktskattning av μ, ett tal.

Ex: Med (observerade) värden

$$\{x_2, x_2, x_3, x_4\} = \{48.0, 49.0, 50.5, 50.5\}$$

är

$$\overline{x} = \mu^*_{\text{obs}} = 49.5\,.$$

Kommentarer

Man kan tänka sig andra typer av punktskattningar av μ, ex.vis genom att skriva dem i storleksordning (som stor. var.), kan man ta

$$\mu^* = \frac{\xi_{\min} + \xi_{\max}}{2}$$

eller

$$\mu^* = \text{det mittersta av } \xi_k.$$

Punktskattning av σ^2 och σ

- Med $\{\xi_1, \xi_2, \ldots, \xi_n\}$ som ovan använder man punktskattningen av σ^2, som ser ut så här, om μ känd:

$$\sigma^{2^*} := \frac{1}{n}\sum_{j=1}^{n}(\xi_j - \mu)^2\,. \tag{1}$$

eller då μ okänd:

$$\sigma^{2^*} := \frac{1}{n-1}\sum_{j=1}^{n}(\xi_j - \overline{\xi})^2\,. \tag{2}$$

(Man kan visa att dessa är v.v.r. , d.v.s. $E(\sigma^{2^*}) = \sigma^2$.)

- Med samma exempel som ovan är motsvarande observerade pumktskattning av σ^2

$$\sigma^{2^*}_{\text{obs}} = 1.5\,.$$

- Punktskattningen av σ får genom att dra roten ur (2) och p.s.s. med observerad punktskattning, som alltså är $\sigma^*_{\text{obs}} = 1.22474\ldots$

Obs! σ^*_{obs} finns på miniräknaren.

1.2 Intervallskattning

Vi erinrar oss om frekvens- och fördelningsfunktion för standardnrmalfördelningen är

$$\begin{cases} \varphi(x) = & \dfrac{1}{\sqrt{2\pi}}\, e^{-x^2/2} \text{ och} \\ \\ \Phi(x) = & \displaystyle\int_{-\infty}^{x} \varphi(t)dt, \end{cases}$$

där $\Phi(x)$ finns tabellerad.

Ibland vill man skatta en parameter med ett intervall, såsom μ för en normalfördelning. Detta innebär att man med stor säkerhet, ex.vis 95%, vill påstå att $\mu \in [a, b]$, ett intervall som inte är alltför stort. Vi skall nedan intervallskatta μ och σ (bara) för en normalfördelning. För detta inför vi, för $\mathrm{N}(0, 1)$, *kvantiler*. Ex,vis är $x = \lambda_{0.05} = 1.6455$, det x, sådant att $0.95 = 95\%$ av arean under $\varphi(x)$ är t.v. om detta x.

1.2.1 Intervallskattning av μ i normalfördelning

Vi har sedan tidigare att för ξ_k, $k = 1, 2, \ldots n$ likafördelade och oberoende och med v.v. μ och standardavvikelse σ, att

$$\begin{array}{l} \zeta := \sum_{k=1}^{n} \xi_k \text{ har } E(\zeta) = n \cdot \mu \text{ och } V(\zeta) = n\sigma^2 \\ \text{och} \\ \eta := \overline{z} \text{ har } E(\eta) = n \cdot \mu \text{ och } V(\eta) = \dfrac{\sigma^2}{n}. \end{array} \tag{3}$$

μ då σ känd Givet ξ_k och η som i (3) där vi dessutom antar att $\xi_k \in \mathrm{N}(\mu, \sigma)$. Vi skall då utnyttja punktskattningen $\overline{z}$ för att göra en intervallskattning av μ. Vi vet att $\overline{z} \in \mathrm{N}(\mu, \sigma/\sqrt{n})$. Vi börjar med att bestämma a, så att

$$-a < \overline{\xi} - \mu < a$$

med, säg $1 - \alpha = 95\%$:s säkerhet, d.v.s. $\alpha = 5\% = 0.05$. Vi gör om mittenledet till en $\mathrm{N}(0, 1)$ stok. var. Vi skriver detta ekvivalent som

$$-\frac{a}{\sigma/\sqrt{n}} < \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} < \frac{a}{\sigma/\sqrt{n}} \text{ och } \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} \in \mathrm{N}(0, 1). \tag{4}$$

P.g.a. symmetri får vi vänster och höger gräns som $x = -\lambda_{\alpha/2}$ och $x = \lambda_{\alpha/2}$. Vi sätter vänsterled till $-\lambda_{\alpha/2}$ och HL till $\lambda_{\alpha/2}$ i (4) och får en tvåsidig *intervallskattning* av μ som

$$-\lambda_{\alpha/2} < \frac{\overline{\xi} - \mu}{\sigma/\sqrt{n}} < \lambda_{\alpha/2} \Longleftrightarrow \overline{\xi} - \lambda_{\alpha/2}\sigma/\sqrt{n} < \mu < \overline{\xi} + \lambda_{\alpha/2}\sigma/\sqrt{n} \tag{5}$$

eller som intervall

$$\left[\overline{\xi} - \lambda_{\alpha/2}\sigma/\sqrt{n}, \overline{\xi} + \lambda_{\alpha/2}\sigma/\sqrt{n}\right].$$

Motsvarande *observerade intervallskattning* ges av

$$\overline{x} - \sigma/\sqrt{n} < \mu < \overline{x} + \sigma/\sqrt{n}. \tag{6}$$

eller som intervall

$$\left[\overline{x} - \lambda_{\alpha/2}\sigma/\sqrt{n},\, \overline{x} + \lambda_{\alpha/2}\sigma/\sqrt{n}\right].$$

Ex 1 Ett *observerat* stickprov på μ är $\{48.0, 49.0, 50.5, 50.5\}$ av storlek $n = 4$ av en normalförd. $\mathrm{N}(\mu, 0.5)$. Bestäm

(a) ett tvåsidigt 95% konfidensintervall för μ.

(b) ett ensidigt nedåt begränsat 95% konfidensintervall för μ.

Lösning

(a) ett tvåsidigt (symmetriskt) 95% konfidensintervall för μ: $\lambda_{\alpha/2} = \lambda_{0.025} = \{\text{tabell}\} = 1.96$Nu är $\overline{x} = 49.5$. Konfidensintervallet är

$$[49.01,\ 49.99].$$

(b) Vi behöver kvantilen $-\lambda_\alpha = -\lambda_{0.05} = -1.6455\ldots$, som har 95% t.h., se figur.

Ett ensidigt nedåt begränsat 95% konfidensintervall för μ ser ut så här

$$[\overline{x} - \lambda_\alpha \cdot \frac{\sigma}{\sqrt{n}},\, \infty) = [49.1, \infty).$$

■

μ då σ okänd I detta fall betraktar vi först variansen som en stokastisk variabel, en punktskattning:

$$\sigma^{2*} = \frac{1}{n-1}\sum_{k=1}^{n}(\xi_k - \overline{\xi})^2. \tag{7}$$

- Obs, man borde dividera med n och inte $n-1$ men divisionen med $n-1$ gör att σ^{2*} blir v.v.r., d.v.s.

  $$E(\sigma^{2*}) = \sigma^2.$$

  Den är dessutom den mest effektiva.

- Motsvarande observerade punktskattning är

  $$\sigma^{2*}\mathrm{obs} = s^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \overline{x})^2. \tag{8}$$

  Vidare måste vi veta fördelningen för detta σ.

  Det är en fördelning som liknar $\mathrm{N}(0,1)$ men med lite ”bredare” form.

Ex 2 Vi löser (a) och (b) men med σ okänt. Med samma observerade stickprov som i föregående exempel men med σ punktskattar vi σ^2 med (7) eller med direkt med (8). Man får den v.v.r. skattningen av σ^2, som man drar roten ur för att få σ[1].
Beräkning ger observerad punktskattning s av σ som

$$s = 1.22474\ldots$$

Med ett stickprov av storleken $n = 4$ talar man om antal "frihetsgrader" $n-1 = 3$.

(a) Ett tvåsidigt 95% intervall Vi ser i tabell för $t-$fördelningen att motsvarade kvantil (fraktil i boken) är $t_{\alpha/2}(3) = 3.19$. Vi har samma medelvärde 49.5. Det ger det tvåsidiga konfidensintervallet

$$[47.5465, 51.4535].$$

(b) Vi får kvantilen $t_\alpha = t_{0.05} = 2.35$ och motsvarande konfidensintervall

$$[48.0609,\ \infty)\,.$$

■

1.2.2 Intervallskattning av σ^2 och σ

Här utgår vi från samma typ av stickprov med $\xi_k \in \mathrm{N}(\mu, \sigma)$. Punktskattningen av σ^2 är (7). Oftast är det ett uppåt begränsat intervall som söks, d.v.s. av typ $[0, a)$. Man utnyttjar då χ^2-fördelningen. Vi har två fall

- μ känd. Frekvensfunktionen för χ^2-fördelningen med n "frihetsgrader" är

$$f_n(x) = \begin{cases} 0, & \text{om } x < 0 \\ k_n \cdot x^{(n-1)/2} e^{-x/2}, & \text{om } x \geq 0, \end{cases}$$

$$\psi_n := \frac{n\,\sigma^{2*}}{\sigma^2} \in \chi^2_n, \text{ en "chitvå"-fördelning.}$$

- μ okänd. och måste punktskattas med $\overline{\xi}$ och utnyttjar

$$\psi_{n-1} := \frac{(n-1)\sigma^{2*}}{\sigma^2} \in \chi^2_{n-1}, \text{ d.v.s. en "chitvå"-fördelning.}$$

  Frekvensfunktionen för χ^2-fördelningen med $n-1$ "frihetsgrader" är

$$f_{n-1}(x) = \begin{cases} 0, & \text{om } x < 0 \\ k_{n-1} \cdot x^{(n-3)/2} e^{-x/2}, & \text{om } x \geq 0, \end{cases}$$

  där k_{n-1} gör att $\displaystyle\int_0^\infty f_{n-1}(x)dx = 1$. Denna fördelning är inte symmetrisk som normal- och $t-$fördelningarna.

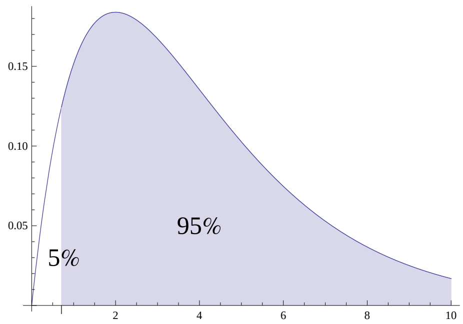
[1]Ej v.v.r. men konvergerar mot v.v.r. då $n \longrightarrow \infty$.

Vi studerar intervall av typ $[0, a)$, ett uppåt begränsat intervall, så att intervallet innehåller σ^2 med sannolikhet $1 - \alpha$. Vi ser att σ^2, som skall uppskattas, finns i nämnaren. Därför söker vi

$$\psi_{n-1} \geq \chi^2_{1-\alpha}(n-1)$$

där $\chi^2_{1-\alpha}(n-1)$ är den kvantil, sådan att $1 - \alpha$ av arean under kurvan $y = f_{n-1}(x)$ ligger t.h. om $x = \chi^2_{1-\alpha}(n-1)$. Olikheten kan skrivas

$$(0 \leq)\sigma^2 \leq \frac{(n-1)\sigma^{2^*}}{\chi^2_{1-\alpha}(n-1)}.$$



Grafen av av $\chi^2(4)$ fördelningens frekvensfunktion $y = f_4(x)$.

Ex 3 Vi antar att vi har samma observerade värden, som i exempel 2, d.v.s. samma observerade stickprov från en normalfördelning.

Bestäm ett 99% uppåt begränsat konfidensintervall för σ^2 och σ

(a) om vi känner $\mu(= 50.0)$ och

(b) μ okänd och måste punktskattas med $\overline{\xi}$.

Lösning

(a) Vi känner $\mu(= 50.0)$.
Här är

$$n = 4,\ \sigma^{2^*} = 1.5\ldots \text{ och } \chi^2_{0.99}(4) = 0.3.$$

Övre gräns för σ^2 är $\dfrac{4 \cdot 1.5}{0.3} = 20.0$ och för σ får vi $\sqrt{20.0} = 4.47$, alltså

$$[0, 20.0] \text{ respektive } [0, 4.47].$$

(b) μ okänd och måste punktskattas med $\overline{\xi}$.
Sätt $\alpha = 0.01$. Här är

$$n - 1 = 3,\ \sigma^{2^*} = 1.5\ldots \text{ och } \chi^2_{0.99}(3) = 0.11.$$

Intervallets övre gräns är

$$\frac{(n-1)\sigma^{2^*}}{\chi^2_{1-\alpha}(n-1)} = \frac{3 \cdot 1.5}{0.11} = 40.9091\,.$$

Konfidensintervallet för σ erhålls genom att dra roten ur undre och övre gräns:

$$[0, 6.4\ldots].$$

Ett uppåt begränsat 95% intervall för σ kräver bara kvantilen

$$x = \chi^2_{0.95}(3) = 0.352.$$

Intervallets övre gräns är ≈ 3.6, alltså $[0, 3.58]$.

■

Ett exempel med CGS

Ex 4

CGS En långtradare kan last 12 ton. En last består av tunnor med var och en med v.v. $\mu = 150$ kg och $\sigma = 10$, enhet kg.

(a) Vad är sannolikheten att total vikt av 81 tunnor överskrider 12 ton?

(b) Hur många tunnor kan lastas så att sannolikheten att 12 ton överkrids är 5%?

Lösning

(a) Sätt ξ_j = tunna nr j:s vikt och $\zeta := \sum_{j=1}^{81} \xi_j$. Vi söker sannolikheten

$$= P\left(\zeta \geq 12000\right) = 1 - P(\zeta \leq 12000) =: 1 - p.$$

$$\zeta \in \mathrm{N}(150 \cdot 81, 9 \cdot 10) \text{ så att } p = \Phi\left(\frac{12000 - 150 \cdot 81}{9 \cdot 10}\right).$$

$$1 - p = \Phi\left(\frac{150}{9 \cdot 10}\right) = \Phi(5/3) = 0.952\ldots$$

Sannolikheten att total vikt av 81 tunnor överskrider 12 ton är 95.2%.

(b) $x :=$ antal tunnor. Vi får sannolikheten

$$\Phi\left(\frac{150x - 12000}{\sqrt{x} \cdot 10}\right) = 0.05$$

eller ekvivalent

$$\frac{150x - 12000}{10\sqrt{x}} = -1.6455 \iff x = 79.0$$

Långtradaren maximalt lasta 79 tunnor.

■