Beskrivande statistik (kap. 11) läses på egen hand.

Punktskattningar

Stickprovsundersökningar — observationer på delar av en population för att få kunskap om helheten.

Exempel: brinntid för ljus beskrivs av X med $m = \mathsf{E}[X]$ och $\sigma^2 = \mathrm{V}(X)$. Vad är m och σ? Med observationer skattas m och σ^2 med

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

respektive

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\overline{x}^2\right).$$

Exempel: Opinionsundersökning. Intervjua n personer och låt X räkna antalet socialdemokrater. Modell: X är $\mathrm{Bin}(n, p)$ med okänt p, andelen sympatisörer i populationen.

Här är det naturligt att våra gissningar (punktskattningar) av värden på p ges av observationer av $p^* = X/n$, den relativa frekvensen. $n = 1023$ intervjuer gav utfallet $x = 387$ vilket ger skattningen $p^* = 387/1023 \approx 37.8\%$. Vi skall senare se hur vi kan göra uttalanden som $0.3486 \le p \le 0.4080$ (95%).

Definition: Ett (slumpmässigt) *stickprov* (av storlek n) är en serie av observationer $x_1, \ldots, x_n$ av oberoende stokastiska variabler $X_1, \ldots, X_n$.

Definition: Ett (slumpmässigt) stickprov (av storlek n) på en stokastisk variabel X är en serie av observationer $x_1, \ldots, x_n$ av oberoende stokastiska variabler $X_1, \ldots, X_n$ med samma fördelning som X.

En punktskattning av en parameter θ är en funktion av ett stickprov som skall ge oss information om θ:

$$\theta^* = \theta^*(x_1, \ldots, x_n)$$

Skattningar modelleras med skattningsvariabeln (stickprovsvariabeln)

$$\theta^* = \theta^*(X_1, \ldots, X_n).$$

Skattningsvariabeln (estimatorn) $\theta^*(X_1, \ldots, X_n)$ ger skattningar (estimat) $\theta^*(x_1, \ldots, x_n)$.

Egenskaper:

Väntevärdesriktighet: Skattningen är väntevärdesriktig om $\mathsf{E}[\theta^*(X_1, \ldots, X_n)] = \theta$.

Effektivitet: $\mathrm{V}(\theta^*(X_1, \ldots, X_n))$ liten

Konsistens (asymptotisk precision): $\mathsf{P}(|\theta^*(X) - \theta| > \epsilon) \to 0$ då $n \to \infty$.

Notera att begreppet effektivitet är meningsfullt för jämförelser av olika skattningsmetoder av samma parameter θ. Mindre varians är bättre!

Exempel: Opinionsundersökning. Med skattningsvariabeln $p^*(X) = X/n$ får vi

$$\mathsf{E}[p^*(X)] = \mathsf{E}[X/n] = \frac{1}{n}\mathsf{E}[X] = \frac{1}{n}np = p$$

och

$$\mathrm{V}\left(p^*(X)\right) = \mathrm{V}\left(X/n\right) = \frac{1}{n^2}\mathrm{V}\left(X\right) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n} \to 0$$

då $n \to \infty$. Alltså är skattningen $p^*(x) = x/n$ en väntevärdesriktig skattning av p och den blir effektivare för större värden på n. Med Tjebychevs olikhet får vi

$$\mathsf{P}\left(|p^*(X) - p| > \epsilon\right) \leq \frac{\mathrm{V}\left(p^*(X)\right)}{\epsilon^2} = \frac{p(1-p)}{\epsilon^2 n} \to 0$$

då n$\to \infty$ att skattningen är konsistent.

Exempel: Opinionsundersökningen (forts.) Med $n = 1023$ interjuer erhölls $x = 387$ sympatisörer.

$$p^*(x) = \frac{x}{n} = \frac{387}{1023} = 0.3783.$$

Skattningen $p^*(x)$ modelleras med $p^*(X) = X/n$ som har

$$\mathrm{V}\left(p^*(X)\right) = \mathrm{V}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{V}\left(X\right) = \frac{p(1-p)}{n}.$$

Dvs,

$$\mathrm{D}\left(p^*(X)\right) = \sqrt{\frac{p(1-p)}{n}}$$

som skattas med

$$\sqrt{\frac{p^*(x)(1-p^*(x))}{n}} = \sqrt{\frac{0.378(1-0.378)}{1023}} = 0.0152.$$

Definition: En skattning av $\mathrm{D}\left(\theta^*(X_1, \ldots, X_n)\right)$ kallas *medelfelet* till $\theta^*(x_1, \ldots, x_n)$.

Exempel: Skattning av väntevärden och varianser Låt $x_1, \ldots, x_n$ vara ett stickprov på X med väntevärde $m = \mathsf{E}\left[X_i\right]$ och varians $\mathrm{V}\left(X\right) = \sigma^2$. Skattningen

$$m^*(x_1, \ldots, x_n) = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

som beskrivs av

$$m^*(X_1, \ldots, X_n) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

är väntevärdesriktig ty

$$\mathsf{E}\left[m^*(X_1, \ldots, X_n)\right] = \mathsf{E}\left[\overline{X}\right] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathsf{E}\left[X_i\right]}_{=m} = m.$$

Vidare så är

$$\mathrm{V}\left(m^*(X_1, \ldots, X_n)\right) = \mathrm{V}\left(\overline{X}\right) = \mathrm{V}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \{\text{oberoende}\} = \frac{1}{n^2}\sum_{i=1}^{n}\underbrace{\mathrm{V}\left(X_i\right)}_{=\sigma^2} = \frac{\sigma^2}{n}$$

vilket med Tjebychevs olikhet ger att skattningen är konsistent.

Skattningen av σ^2, s^2, beskrivs av S^2 som ger väntevärdesriktiga skattningar: $\mathsf{E}\left[S^2\right] = \sigma^2$. Det är en nyttig övning att visa detta med räknelagarna!

Generella skattningsmetoder

Hur väljer man sina skattningar?

- Maximum likelihood-metoden (maximi-metoden, ML-metoden)
- Minsta kvadrat-metoden (MK-metoden)

I fortsättningen, låt $x_1, \dots, x_n$ vara ett stickprov. Antag att i fördelningen för $(X_1, \dots, X_n)$ ingår parametrarna $\theta_1, \dots, \theta_r$.

Maximum likelihood-metoden Skattningarna $\theta_1^*, \dots, \theta_r^*$ är de värden på $\theta_1, \dots, \theta_r$ som maximerar *trolighetsfunktionen*

$$L(\theta_1, \dots, \theta_r) = f_{X_1,\dots,X_n}(x_1, \dots, x_n)$$

vid kontinuerliga fördelningar, eller

$$L(\theta_1, \dots, \theta_r) = \mathsf{P}\left(X_1 = x_1, \dots, X_n = x_n\right)$$

vid diskreta.

Filosofin bakom Maximum likelihood-metoden är att vårt utfall (stickprovet) bestäms av parametrarna i de ingående fördelningarna. Skattningen av dessa parametrar är de värden som gör det observerade så troligt som möjligt.

Tips! Ofta är det enklare att maximera $\log L(\theta_1, \dots, \theta_r)$.

Exempel: Binomialfördelningen. Låt X är $\mathrm{Bin}(n, p)$ med observerat utfall x.

$$L(p) = \mathsf{P}\left(X = x\right) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Logaritmering ger

$$\ln L(p) = \ln\left(\binom{n}{x}\right) + x \ln p + (n-x)\ln(1-p).$$

Derivering bestämmer maximum:

$$\frac{d}{dp} \ln L(p) = 0 + x\frac{1}{p} - (n-x)\frac{1}{1-p} = 0$$

ger skattningen $p = x/n$. Detta gäller för alla utfall x och $p^*(x) = x/n$ och vi skriver skattningsvariablen som $p^*(X) = X/n$.

(Skall detta göras ordentligt skall man bestämma 2:a-derivatan för att se att det är ett maximum man fått fram, samt även undersöka randpunkterna.)

Minsta kvadrat-metoden Kräver inga fördelningsantaganden på $X_1, \dots, X_n$, bara att vi kan modellera deras väntevärden. Skattningarna $\theta_1^*, \dots, \theta_r^*$ de värden på $\theta_1, \dots, \theta_r$ som minimerar *minsta kvadrat-avstånden* mellan observationer och väntevärden:

$$Q(\theta_1, \dots, \theta_r) = \sum_{i=1}^{n} (x_i - \mathsf{E}\left[X_i\right])^2.$$

Även här minimeras Q oftast med hjälp av derivering.

Exempel: Binomialfördelning, X är $\mathrm{Bin}(n, p)$, med observerat värde x. Här blir

$$Q(p) = (x - np)^2$$

och man ser direkt $p = x/n$ ger minimum. Skattningen är således $p^*(x) = x/n$. Detta gäller för alla utfall x så vi får skattningsvariabeln $p^*(X) = X/n$.