Punktskattningar

Exempel: Låt $x_1, \ldots, x_n$ vara den uppmätta brinntiden för ljus. Vi modellerar dessa som utfall av oberoende och likafördelade stokastiska variabler $X_1, \ldots, X_n$, där $\mu = \mathsf{E}[X]$ och $\sigma^2 = \mathrm{V}(X)$. Väntevärdet μ skattas med

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

och variansen σ^2 med

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

Exempel: Opinionsundersökning. Intervjua n personer och låt x vara antalet socialdemokrater bland de intervjuade. Med $n = 2661$ och $x = 961$ fås skattningen $p^*_{\mathrm{obs}} = x/n = 961/2661 = 0.3611$ av p, andelen sympatisörer i populationen.

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

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

Oftast har $X_1, \ldots, X_n$ samma fördelning.

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

$$\theta^*_{\mathrm{obs}} = \theta^*(x_1, \ldots, x_n)$$

Skattningar modelleras med skattningsvariabeln (stickprovsvariabeln)

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

Skattningsvariabeln (estimatorn) θ^* ger skattningar (estimat) θ^*_{obs}.

Egenskaper: Skattningen θ^*_{obs} är

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

effektiv $\mathrm{V}(\theta^*) = \mathrm{V}(\theta^*(X_1, \ldots, X_n))$ är liten.

konsistent (asymptotisk precis) $\mathsf{P}(|\theta^* - \theta| > \epsilon) \to 0$ 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. Skattningen p^*_{obs} modelleras med $p^* = X/n$ där X är $\mathrm{Bin}(n, p)$ (oändlig population). Då är

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

dvs. skattningen är väntevärdesriktig, och

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

då $n \to \infty$. Med Tjebychevs olikhet får vi

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

då n$\to \infty$ att skattningen är konsistent. (Detta är stora talens lag i dess svaga form!)

Med siffror, $n = 2661$ interjuer och $x = 961$ sympatisörer, är skattningen

$$p^*_{\text{obs}} = \frac{x}{n} = \frac{961}{2661} = 0.3611.$$

Standardavvikelsen för p^* är

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

vilken skattas med

$$\sqrt{\frac{p^*_{\text{obs}}(1 - p^*_{\text{obs}})}{n}} = \sqrt{\frac{0.3611(1 - 0.3611)}{2661}} = 0.0093115.$$

Definition: En skattning av $\mathrm{D}\left(\theta^*\right)$, betecknad $d(\theta^*)$, kallas *medelfelet* för skattningen θ^*_{obs}.

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

$$\mu^*_{\text{obs}} = \overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

är väntevärdesriktig (v.v.r.) ty med

$$\mu^* = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

är

$$\mathsf{E}\left[\mu^*\right] = \mathsf{E}\left[\overline{X}\right] = \mathsf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n}\underbrace{\mathsf{E}\left[X_i\right]}_{=\mu} = \mu.$$

Vidare så är

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

vilket med Tjebychevs olikhet ger att skattningen μ^*_{obs} är en konsistent skattning av μ.

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

Felfortplantning

Hur kan man skatta $g(\theta)$ med hjälp av θ^*_{obs} och vad är medelfelet för skattningen?

Låt θ^*_{obs} vara en väntevärdesriktig skattning av θ, $\mathsf{E}\left[\theta^*\right] = \theta$, och $g(x)$ vara en funktion. Taylorutveckla g runt en punkt μ, $g(x) \approx g(\mu) + g'(\mu)(x - \mu)$. Då är

$$\mathsf{E}\left[g(X)\right] \approx \mathsf{E}\left[g(\mu) + g'(\mu)(X - \mu)\right] = g(\mu) + g'(\mu)(\mathsf{E}\left[X\right] - \mu)$$

Med $\mu = \mathsf{E}[X]$ är
$$\mathsf{E}[g(X)] \approx g(\mathsf{E}[X])$$
Med $X = \theta^*$ är
$$\mathsf{E}[g(\theta^*)] \approx g(\mathsf{E}[\theta^*]) = g(\theta)$$
så $g(\theta^*_{\text{obs}})$ är (approximativt) en väntevärdesriktig skattning av $g(\theta)$. Vidare så

$$\mathrm{V}(g(X)) \approx \mathrm{V}(g(\mu) + g'(\mu)(X - \mu)) = [g'(\mu)]^2\, \mathrm{V}(X)$$

så
$$\mathrm{D}(g(X)) \approx |g'(\mu)|\mathrm{D}(X)\,.$$
Detta ger att
$$\mathrm{D}(g(\theta^*)) \approx |g'(\theta)|\mathrm{D}(\theta^*)$$
så medelfelet för skattningen $g(\theta^*_{\text{obs}})$ fås approximativt som $d(g(\theta^*)) = |g'(\theta^*_{\text{obs}})|d(\theta^*)$.

Generaliseringar för funktioner av fler parametrar, se litteraturen.

Generella skattningsmetoder

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

Maximum likelihood-metoden Skattningarna $(\theta_1)^*_{\text{obs}}, \ldots, (\theta_r)^*_{\text{obs}}$ är de värden på $\theta_1, \ldots, \theta_r$ som maximerar *trolighetsfunktionen*

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

vid kontinuerliga fördelningar, eller

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

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.