

STK1110 Høsten 2021

Estimering

Repetisjon av Avsnitt 7.1

Ingrid Hobæk Haff
Matematisk institutt
Universitetet i Oslo

Bakgrunn

- Vi antar at vi har observasjoner x_1, \dots, x_n av de stokastiske variablene X_1, \dots, X_n .
- X_1, \dots, X_n kommer fra en sannsynlighetsfordeling med ukjent parameter θ .
- For å estimere θ bruker vi en **estimator** $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$.
- Et **estimat** får en ved å sette inn dataene: $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$.

MSE

- En ønsker at $\hat{\theta}$ skal være så nær θ som mulig.
- Ett mål for avstanden mellom dem er minste kvadraters feil, eller “mean square error” (MSE) på engelsk.
- MSE er gitt ved

$$MSE(\hat{\theta}) = E((\hat{\theta} - \theta)^2) = V(\hat{\theta}) + (Bias(\hat{\theta}))^2,$$

der $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Forventningsrett estimator

- En estimator $\hat{\theta}$ er **forventningsrett** (“unbiased” på engelsk) for parameteren θ dersom

$$E(\hat{\theta}) = \theta.$$

- **Skjevheten** (“bias” på engelsk) til $\hat{\theta}$ er gitt ved

$$Bias(\hat{\theta}) = E(\hat{\theta}) - \theta,$$

og er 0 dersom $\hat{\theta}$ er forventningsrett.

- Da er også $MSE(\hat{\theta}) = V(\hat{\theta})$.

Eksempel

$$Y \sim Bin(n, p)$$

La $\hat{p} = \frac{Y}{n}$ være en estimator for p .

Uavhengige, identisk fordelte variabler

- Anta at X_1, \dots, X_n er uavhengige og identisk fordelt (uif) med forventning μ og varians σ^2 .
- Da er

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

forventningsrett for μ og

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

forventningsrett for σ^2 .

- Merk at S ikke er forventningsrett for σ .

Standardfeil

- Ethvert estimat er beheftet med usikkerhet.
- Når en rapporterer et estimat, bør en derfor også si noe om usikkerheten.
- Ett mål for usikkerheten er **standardfeilen** til den tilsvarende estimatoren, dvs.

$$\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}.$$

- Av og til har en et uttrykk for $\sigma_{\hat{\theta}}$ som funksjon av parameterne i modellen.
- En kan da få et estimat $s_{\hat{\theta}}$ for $\sigma_{\hat{\theta}}$ ved å sette tilsvarende parameterestimer inn i uttrykket.

Eksempel

$Y \sim \text{Bin}(n, p)$ med $\hat{p} = \frac{Y}{n}$.

Bootstrapping

- Dersom en ikke har noe uttrykk for standardfeilen, kan bootstrapping være et alternativ.
- La $f(x; \theta)$ være sannsynlighetstettheten/punktsannsynligheten til X . Da er prosedyren:
 - ① For $b = 1, \dots, B$
 - ② Trekk $x_{b1}^*, \dots, x_{bn}^*$ fra $f(x; \hat{\theta})$
 - ③ La $\hat{\theta}_b^* = \hat{\theta}(x_{b1}^*, \dots, x_{bn}^*)$.
- La så $s_{\hat{\theta}} = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\theta}^*)^2}$, der $\bar{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*$.
- Dette er **parametrisk bootstrap**, der en trekker data fra den estimerte modellen $f(x; \hat{\theta})$.
- Alternativt kan en bruke **ikke-parametrisk bootstrap**.
- Da bytter en ut steg 2 over med:
 - ② Trekk $x_{b1}^*, \dots, x_{bn}^*$ fra x_1, \dots, x_n med tilbakelegging