

TMA4240 - Høst 2023

Tema for uka

- ▶ Parameterestimering
- ▶ Evaluering av estimatorer
- ▶ Rimelighetsfunksjonen og sannsynlighetsmaksimeringsestimator

Del 2 i kurset

Sannsynlighetsregning

- Kjenner $f(x; \theta)$ og verdinen til parameter θ
- Vi regner sannsynlighet for forskjellige hendelser

Statistisk inferens

- Kjenner $f(x; \theta)$ (oftest) men **ikke** verdinen til parameter θ
- Ønsker å si noe om hele populasjon basert på et tilfeldig utvalg.
- Bakgrunn: vår kunnskap i sannsynlighetsregning.



Statistisk Inferens

Statistisk inferens er prosessen å ta konklusjoner om **populasjonsparametrene** ut i fra et **utvalg** av denne populasjonen

Definisjoner

- ▶ **Populasjon** En populasjon består av alle mulige enheter vi er interessert i. Kan være endelig eller uendelig
- ▶ **Utvalg** Et utvalg er en del av populasjon

Utvalget må være representativt av hele populasjonen!

Evaluering av estimatorer

Vi ser på

- ▶ Forventing av vår estimator
- ▶ Varians av vår estimator

Oppgave 1 - Mai 2016

Oppgave 3 Besøk på webside

For firmaet SeMeg er antall besøk på deres webside viktig. La X_i være antall besøk i løpet av t_i timer, og la $X_1, X_2, \dots, X_n$ være antall besøk i n ikke-overlappende tidsintervall. Vi antar at besøk på websiden er en poissonprosess med besøksintensitet λ . Dermed er $X_1, X_2, \dots, X_n$ uavhengige poissonfordelte stokastiske variabler med sannsynlighetsfordeling

$$f(x_i) = \frac{(\lambda t_i)^{x_i}}{x_i!} e^{-\lambda t_i} \quad \text{for } x_i = 0, 1, 2, \dots$$

Vi antar nå at besøksintensiteten λ er ukjent. SeMeg ønsker å estimere intensiteten λ fra data på antall besøk fra n ikke-overlappende tidsintervall. Det er foreslått tre estimatorer,

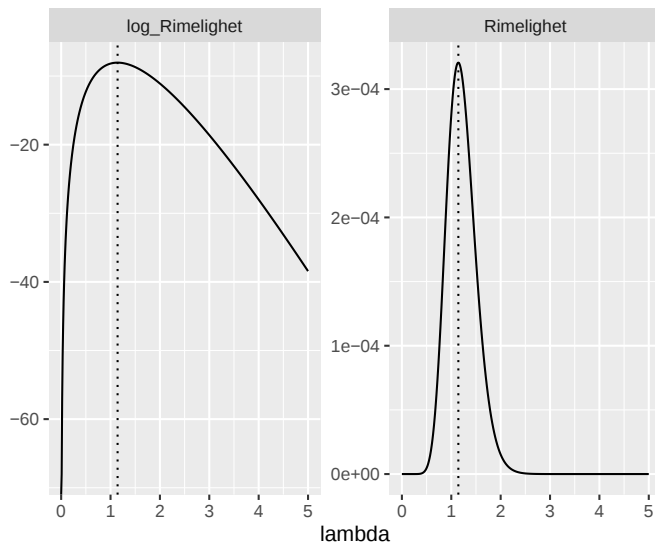
$$\tilde{\lambda} = \frac{1}{n} \sum_{i=1}^n X_i, \quad \hat{\lambda} = \frac{\sum_{i=1}^n X_i}{\sum_{i=1}^n t_i} \quad \text{og} \quad \hat{\hat{\lambda}} = \frac{1}{n} \sum_{i=1}^n \frac{X_i}{t_i},$$

og vi oppgir at $E[\hat{\lambda}] = \lambda$ og $\text{Var}[\hat{\lambda}] = \lambda / \sum_{i=1}^n t_i$.

- b) Hvilken av de tre estimatorene vil du foretrekke når $n = 5$ og $t_1 = 1, t_2 = 2, t_3 = 5, t_4 = 1, t_5 = 5$? Begrunn svaret.
- c) Utled sannsynlighetsmaksimeringsestimatoren (SME) for λ basert på $X_1, X_2, \dots, X_n$.

Oppgave 1 - Rimelighetsfunksjon

Vi har $\mathbf{t} = (1, 2, 5, 1, 5)$ og $\mathbf{x} = (1, 3, 8, 0, 4)$



Oppgave 2

Vi vil estimere Fruktbarhetstall (hvor mange barn som fødes per kvinne).

Anta at X = “antall barn per kvinne” er Poisson fordelt med parameter λ

Anta at $X_1, \dots, X_n$ er en tilfeldig utvalg fra populasjonen

- ▶ Finn SME for λ
- ▶ Finn forventnig of varians av estimatoren

Oppgave 3 - Nov 19

Eksamen i TMA4240 Statistikk, 30. november 2019

4(a) Binomisk fordelte variabler

Innledning: Anta at vi har to uavhengige stokastiske variabler X og Y , der $X \sim b(x; n, p)$ og $Y \sim b(y; n, 2p)$. X er altså antall suksesser i n uavhengige forsøk der hvert forsøk har sannsynlighet p for å gi suksess, mens Y er antall suksesser i n andre uavhengige forsøk der hvert forsøk gir suksess med sannsynlighet $2p$.

4(b) Innledning: I resten av oppgaven skal vi anta at verdien av parameteren $p \in [0, 0.5]$ er ukjent og skal estimeres basert på X og Y . Følgende tre estimatorer er foreslått,

$$\hat{p} = \frac{X+Y}{2n}, \quad \tilde{p} = \frac{X+Y}{3n} \quad \text{og} \quad p^* = \frac{X}{2n} + \frac{Y}{4n} \quad (1)$$

Oppgave:

- Hvilken av de tre estimatorene vil du foretrekke? Begrunn svaret.

Oppgave 4 - Nov 19

Eksamen i TMA4240 Statistikk, 30. november 2019

4(a) Binomisk fordelte variabler

Innledning: Anta at vi har to uavhengige stokastiske variabler X og Y , der $X \sim b(x; n, p)$ og $Y \sim b(y; n, 2p)$. X er altså antall suksesser i n uavhengige forsøk der hvert forsøk har sannsynlighet p for å gi suksess, mens Y er antall suksesser i n andre uavhengige forsøk der hvert forsøk gir suksess med sannsynlighet $2p$.

4(c) Innledning: Anta at vi også ønsker å estimere p ved å benytte sannsynlighetsmaksimeringsprinsippet, fremdeles basert på X og Y .

Oppgave:

- Finn et uttrykk for log-rimelighetsfunksjonen for p , $l(p)$.
- Bestem hva estimatet for p blir ved å benytte sannsynlighetsmaksimeringsprinsippet når $n = 25$, og man observerer $x = 3$ og $y = 8$.

Oppgave 5

La $X_1, \dots, X_n$ være uif men fordeling

$$f(x) = 1/\theta, \quad x \in (0, \theta)$$

- Finn SME for θ