STK1110 Høsten 2021

Informasjon og MLEs egenskaper i store utvalg

Tilsvarer Avsnitt 7.4

Ingrid Hobæk Haff
Matematisk institutt
Universitetet i Oslo

Fisher-informasjon

- Vi antar at vi har uavhengige, identisk fordelte stokastiske variabler $X_1, \ldots, X_n$ med punktsannsynlighet/sannsynlighetstetthet $f(x; \theta)$, som avhenger av én (ukjent) parameter θ.
- Hver X_i bidrar med informasjon om θ.
- Ett mål for denne informasjonen er **Fisher-informasjonen**, gitt ved

$$I(\theta) = \mathrm{V}(U) \quad \text{med} \quad U = \frac{\partial}{\partial \theta} \log f(X; \theta).$$

Eksempel

$X \sim \mathit{Bernoulli}(p)$ og vi vil finne Fisher-informasjonen $I(p)$ om p i én observasjon.

Fisher-informasjon (forts.)

- Under visse betingelser gjelder følgende:

$$\mathrm{E}(U) = 0 \quad \text{og} \quad I(\theta) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2}\log f(X;\theta)\right).$$

- Den viktigste betingelsen for at uttrykkene over skal gjelde er at vi kan partiellderivere

$$\int f(x;\theta)dx \quad \text{hvis } X \text{ er kontinuerlig}$$

$$\sum_x f(x;\theta) \quad \text{hvis } X \text{ er diskret}$$

  to ganger mhp. θ ved å bytte rekkefølgen på derivasjon og integrasjon/summasjon (se s. 676-679 i MAT1100-boka for detaljer).

- En viktig forutsetning er at mengden $\{x : f(x;\theta) > 0\}$ ikke avhenger av θ.

Eksempler

Eksempel

$X \sim \textit{Bernoulli}(p)$ og vi vil finne $\{x : f(x; p) > 0\}$.

Eksempel

$X \sim \textit{Eksponensiell}(\lambda)$ og vi vil finne $\{x : f(x; \lambda) > 0\}$.

Eksempel

$X \sim U(0, \theta)$ og vi vil finne $\{x : f(x; \theta) > 0\}$.

Eksempel

$X \sim \textit{Bernoulli}(p)$ og vi vil finne $I(p)$ ved hjelp av formelen på forrige foil.

Fisher-informasjon i et tilfeldig utvalg

- Vi antar at vi har et tilfeldig utvalg $X_1, \ldots, X_n$, dvs. at de er uavhengige, identisk fordelte stokastiske variabler.
- Videre antar vi at de har punktsannsynlighet/ sannsynlighetstetthet $f(x; \theta)$, som avhenger av én parameter θ.
- Likelihood-funksjonen og log-likelihood-funksjonen er da

$$f(X_1, \ldots, X_n; \theta) = \prod_{i=1}^{n} f(X_i; \theta)$$

og

$$\log f(X_1, \ldots, X_n; \theta) = \sum_{i=1}^{n} \log f(X_i; \theta).$$

Fisher-informasjon i et tilfeldig utvalg (forts.)

- **Score-funksjonen** er da definert som

$$\frac{\partial}{\partial \theta} \log f(X_1, \ldots, X_n; \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i; \theta).$$

- **Fisher-informajonen $I_n(\theta)$ til utvalget** er da lik variansen til score-funksjonen, dvs.

$$\begin{aligned} I_n(\theta) &= \mathrm{V}\left(\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i; \theta)\right) \\ &= \sum_{i=1}^{n} \mathrm{V}\left(\frac{\partial}{\partial \theta} \log f(X_i; \theta)\right) = nI(\theta). \end{aligned}$$

Cramér-Raos ulikhet og effisiens

- Vi antar at vi har uavhengige, identisk fordelte stokastiske variabler $X_1, \ldots, X_n$ med punktsannsynlighet/sannsynlighetstetthet $f(x;\theta)$, som avhenger av én parameter θ.
- Videre antar vi at $\{x : f(x;\theta) > 0\}$ ikke avhenger av θ.
- La $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ være en forventningsrett estimator for θ.
- I henhold til Cramér-Raos ulikhet er da

$$\mathrm{V}(\hat{\theta}) \geq \frac{1}{I_n(\theta)} = \frac{1}{nI(\theta)}.$$

- Videre kalles $\frac{1/nI(\theta)}{\mathrm{V}(\hat{\theta})}$ for **effisiensen** til $\hat{\theta}$, og $\hat{\theta}$ sies å være **effisient** dersom den har effisiens lik 1.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} \mathit{Bernoulli}(p)$ og vi vil finne effisiensen til $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n}$.

MLEs egenskaper for store utvalg

- Vi antar at vi $X_1, \ldots, X_n$ er uif med punktsannsynlighet/sannsynlighetstetthet $f(x;\theta)$, som avhenger av θ.
- Videre antar vi at $\{x : f(x;\theta) > 0\}$ ikke avhenger av θ.
- La $\hat{\theta}$ være maksimum likelihood-estimatoren for θ.
- En kan da vise at

$$\lim_{n\to\infty} \mathrm{P}(\sqrt{n}(\hat{\theta}-\theta) \leq x) = \mathrm{P}(W \leq x), \quad \text{med} \quad W \sim N\left(0, \frac{1}{I(\theta)}\right).$$

- Det skrives gjerne $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} W$, og betyr at når n er tilstrekkelig stor, har $\sqrt{n}(\hat{\theta}-\theta)$ og W tilnærmet samme fordeling.
- Det gjelder også $\hat{\theta}$ og $W/\sqrt{n}+\theta$, slik at for tilstrekkelig stor n er

$$\hat{\theta} \overset{tiln.}{\sim} N\left(\theta, \frac{1}{nI(\theta)}\right).$$

Eksempler

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} \mathit{Bernoulli}(p)$ og vi vil finne fordelingen til MLE for p for store n.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} \mathit{Eksponensiell}(\lambda)$ og vi vil finne fordelingen til MLE for λ for store n.

MLEs egenskaper for store utvalg ved flere parametere

- Vi antar at vi $X_1, \ldots, X_n$ er uif med punktsannsynlighet/sannsynlighetstetthet $f(x;\theta_1,\theta_2)$, som avhenger av θ_1 og θ_2.
- Videre antar vi at $\{x : f(x;\theta_1,\theta_2) > 0\}$ ikke avhenger av θ_1 og θ_2.
- La $\hat{\theta}_1$ og $\hat{\theta}_2$ være maksimum likelihood-estimatorene for θ_1 og θ_2.
- Når n er tilstrekkelig stor, er da

$$\hat{\theta}_j \overset{tiln.}{\sim} N\left(\theta_j, \frac{1}{n} I^{jj}(\theta_1,\theta_2)\right), \quad j = 1, 2,$$

der $I^{jj}(\theta_1,\theta_2)$ er j-te diagonalelement i den inverse Fisher-informasjonsmatrisa $I(\theta_1,\theta_2)$ for én observasjon (se neste foil).

MLEs egenskaper for store utvalg ved flere parametere (forts.)

- Fisher-informasjonsmatrisa $I(\theta_1, \theta_2)$ for én observasjon er

$$I(\theta_1, \theta_2) = \begin{pmatrix} I_{11}(\theta_1, \theta_2) & I_{12}(\theta_1, \theta_2) \\ I_{21}(\theta_1, \theta_2) & I_{22}(\theta_1, \theta_2) \end{pmatrix}$$

  med $I_{ij}(\theta_1, \theta_2) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(X; \theta_1, \theta_2) \right)$.

- Den inverse av denne matrisa, $I(\theta_1, \theta_2)^{-1}$, er da gitt ved

$$\begin{pmatrix} I_{11}(\theta_1, \theta_2) & I_{12}(\theta_1, \theta_2) \\ I_{21}(\theta_1, \theta_2) & I_{22}(\theta_1, \theta_2) \end{pmatrix}^{-1} = \begin{pmatrix} I^{11}(\theta_1, \theta_2) & I^{12}(\theta_1, \theta_2) \\ I^{21}(\theta_1, \theta_2) & I^{22}(\theta_1, \theta_2) \end{pmatrix}.$$

- Tilsvarende resultat gjelder for flere parametere.

Eksempel

$X_1, \ldots, X_n \overset{uif}{\sim} Gamma(\alpha, \beta)$ og vi vil finne fordelingen til MLE for α og β for store n.