STK1110 Høsten 2026

Informasjon og MLEs egenskaper i store utvalg

Tilsvarer Avsnitt 7.4

Forrige uke og denne uka:

Forrige uke:

- Oversikt over kurset
- Repetisjon estimering - Avsnitt 7.1 og 7.2
  - Estimering generelt
  - Mean Square Error (MSE) og forventningsrette estimatorer
  - Standardfeil = estimert standardavvik for estimator
  - Bootstrapping: Parametrisk og ikke-parametrisk
  - Momentestimatorer
  - Maximum Likelihood Estimatorer

Denne uka: Avsnitt 7.4

- Egenskaper MLE i store utvalg
- Variansestimering for MLE i store utvalg
- Fisher informasjon
- Cramér-Rao-nedre grense for forventningsrette estimatorer
- Tilnærmet effisiens for MLE

Maksimum likelihood-estimering (MLE)

- La $f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$ være simultantettheten/-punktsannsynligheten til $X_1, \ldots, X_n$.
- Når vi ser på denne som en funksjon av $\theta_1, \ldots, \theta_m$ for gitte data $x_1, \ldots, x_n$, kalles den **likelihood-funksjonen**.

$$L(\theta_1, \ldots, \theta_m) = f(x_1, x_2, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

- **Maksimum likelihood-estimatene** (MLE) $\hat{\theta}_1, \ldots, \hat{\theta}_m$ gis ved at

$$L(\hat{\theta}_1, \ldots, \hat{\theta}_m) \geq L(\theta_1, \ldots, \theta_m)$$

- Tilsvarende fås maksimum likelihood-**estimatorer** (MLE) ved å erstatte observerte data med stokastiske variable $X_1, \ldots, X_n$

$$L(\hat{\theta}_1, \ldots, \hat{\theta}_m) = f(X_1, X_2, \ldots, X_n; \hat{\theta}_1, \ldots, \hat{\theta}_m) \geq L(\theta_1, \ldots, \theta_m)$$

Eksempel MLE: $Y \sim \text{Bin}(n, p)$

Likelihood for p:

$$L(p) = \binom{n}{Y} p^Y (1-p)^{n-Y}$$

Log-likelihood:

$$\ell(p) = \ln(L(p)) = \ln(\binom{n}{Y}) + Y \ln(p) + (n - Y) \ln(1-p)$$

Derivert av log-likelihood:

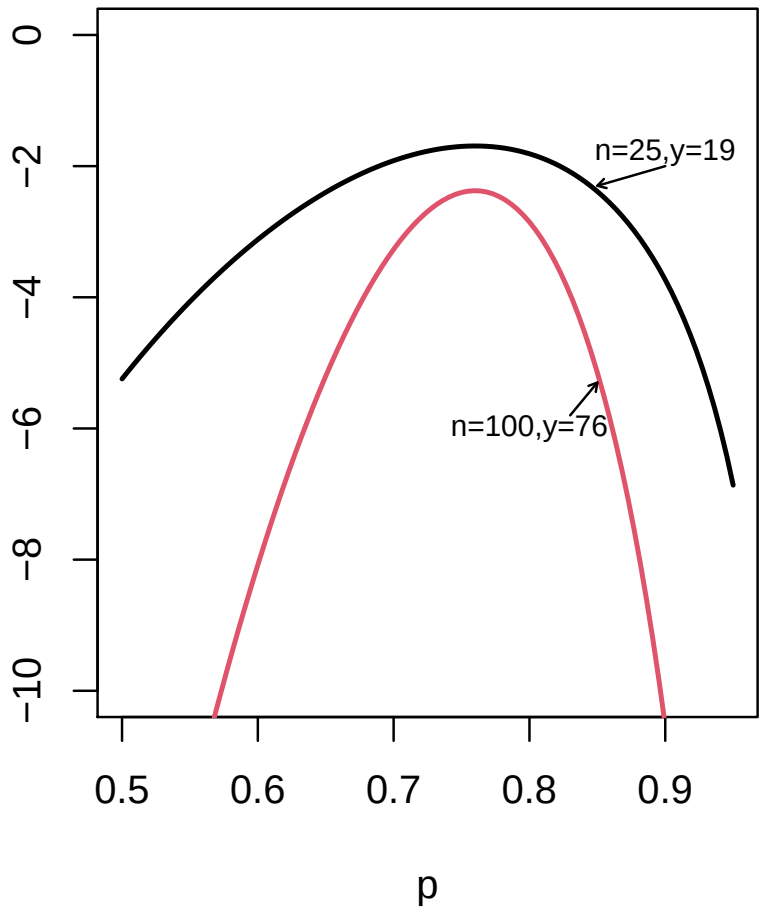
$$\ell'(p) = \frac{Y}{p} - \frac{n-Y}{1-p}$$

MLE $\hat{p}$ finnes som løsning av $\ell'(\hat{p}) = 0$, dvs.

$$\ell'(\hat{p}) = \frac{Y}{\hat{p}} - \frac{n-Y}{1-\hat{p}} = 0 \quad \Leftrightarrow \quad \hat{p} = \frac{Y}{n}$$

Plott av log-likelihood for $Y \sim \text{Bin}(n, p)$

Anta at vi enten observerer $Y = 19$ med $n = 25$ eller $Y = 76$ for $n = 100$. Da vil log-likelihood $\ell(p)$ se ut som i figuren:



- Mer spiss funksjon for stor n
- Relatert til mindre variasjon!
- Pga. mer informasjon

Egenskaper $\hat{p} = \frac{Y}{n}$

- Forventningsrett $\mathrm{E}(\hat{p}) = p$
- Varians $\mathrm{V}(\hat{p}) = \frac{\mathrm{V}(Y)}{n^2} = \frac{p(1-p)}{n}$
- $\hat{p}$ er tilnærmet normalfordelt, $\mathrm{N}(p, \sqrt{\mathrm{V}(\hat{p})})$
  (sentralgrenseteoremet, $Y = \sum_{i=1}^{n} X_i$ der X_i-ene er Bernoulli)

Det er verdt å sammenligne variansen til $\hat{p}$ med forventningen til minus 2.derivert av log-likelihood. Først negativ 2. derivert:

$$-\ell''(p) = -\frac{\partial}{\partial p}\ell'(p) = -\frac{\partial}{\partial p}(\frac{Y}{p} - \frac{n-Y}{1-p}) = \frac{Y}{p^2} + \frac{n-Y}{(1-p)^2}$$

som har forventning

$$\mathrm{E}(-\ell''(p)) = \frac{\mathrm{E}(Y)}{p^2} + \frac{\mathrm{E}(n-Y)}{(1-p)^2} = \frac{n}{p} + \frac{n}{1-p} = \frac{n}{p(1-p)} = \frac{1}{\mathrm{V}(\hat{p})}$$

Generelt for regulære maximum likelihood estimatorer!!!

Ytterligere besnærende likhet

Vi har faktisk også følgende sammenheng for binomisk log-likelihood

$$\mathrm{V}(\ell'(p)) = \frac{n}{p(1-p)} = \frac{1}{\mathrm{V}(\hat{p})}$$

Utledning:

$$\mathrm{V}(\ell'(p)) = \mathrm{V}(\frac{Y}{p} - \frac{n-Y}{1-p}) = \mathrm{V}(\frac{Y-np}{p(1-p)}) = \frac{np(1-p)}{(p(1-p))^2} = \frac{1}{\mathrm{V}(\hat{p})}$$

Dette er også en generell egenskap for (regulære) likelihooder. Ganske generelt har vi identiteten

$$\mathrm{V}(\ell'(p)) = \mathrm{E}(-\ell''(p))$$

og denne er invers av variansen til MLE - i det minste i store utvalg.

Egenskaper for regulær MLE $\hat{\theta}$

Hvis $\hat{\theta}$ er MLE for skalar θ i en passelig "pen" likelihood $\ell(\theta)$ har vi

$$\hat{\theta} \sim \mathrm{N}(\theta, 1/\sqrt{I_n(\theta)}) \qquad \text{(tilnærmet)}$$

Her er $I_n(\theta)$ såkalt Fisher-informasjon - *for hele datasettet med n observasjoner* - definert ved

$$I_n(\theta) = -\mathrm{E}(\ell''(\theta)) = \mathrm{V}(\ell'(\theta))$$

Vi kan videre estimere tilnærmet varians for $\hat{\theta}$ ved $1/I_n(\hat{\theta})$, evt. ved $1/(-\ell''(\hat{\theta}))$ = invers "observert informasjon".

Dessuten er $1/I_n(\theta)$ den minste varians en forventningsrett (regulær) estimator kan ha!

I tillegg er (regulære) MLE $\hat{\theta}$ tilnærmet forventningsrette, så de kan oppfattes som effisiente!!

Eksempel: Eksponensialfordelte data

Antar $X_i, i = 1, \ldots, n$ er uif med tetthet $f(x; \lambda) = \lambda \exp(-\lambda x)$

Beregnet sist

- Likelihood $L(\lambda) = \prod_{i=1}^{n} \lambda \exp(-\lambda X_i) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} X_i)$
- log-likelihood $\ell(\lambda) = n \ln(\lambda) - \lambda \sum_{i=1}^{n} X_i$
- derivert av log-likelihood $\ell'(\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} X_i$
- MLE $\hat{\lambda} = 1/\bar{X} = \frac{n}{\sum_{i=1}^{n} X_i}$

Vi har da tilnærmet - henhold til teori om MLE -

$$\hat{\lambda} \sim \mathrm{N}(\lambda, 1/\sqrt{\mathrm{V}(\ell'(\lambda))})$$

der $\mathrm{V}(\ell'(\lambda)) = \mathrm{V}(\frac{n}{\lambda} - \sum_{i=1}^{n} X_i) = \mathrm{V}(\sum_{i=1}^{n} X_i) = \frac{n}{\lambda^2}$.

Dermed får vi at standardfeilen til $\hat{\lambda}$ (dvs. estimert standardavvik) blir gitt ved

$$se(\hat{\lambda}) = \frac{\hat{\lambda}}{\sqrt{n}}$$

Skisse av utledning asymptotisk fordeling av MLE

- Anta (for enkelthet) kun én parameter.
- MLE: $\hat{\theta} = \arg\max_\theta \ell(\theta)$
- Løsning funnet ved $\ell'(\hat{\theta}) = 0$
- Ved 1. ordens Taylor: $0 = \ell'(\hat{\theta}) = \ell'(\theta) + (\hat{\theta} - \theta)\ell''(\theta) +$ rest
- Denne ligningen leder til tilnærmingen

$$\hat{\theta} - \theta \approx \frac{\ell'(\theta)}{-\ell''(\theta)} \Rightarrow \sqrt{n}(\hat{\theta} - \theta) \approx \frac{\sqrt{n}\ell'(\theta)}{-\ell''(\theta)} = \frac{\frac{1}{\sqrt{n}}\ell'(\theta)}{\frac{1}{n}[-\ell''(\theta)]}$$

- Vil vise at
  - Teller konvergerer mot en normalfordeling - med forventning 0
  - Nevner konvergerer mot en konstant

Asymptotisk fordeling av MLE - teller

Vi antar at $X_i, i = 1, \ldots, n$ er uavhengige identisk fordelte variable med punktsannsynligheter (tetthet) $f(x, \theta)$.

Da er også $U_i = \frac{\partial}{\partial \theta} \ln(f(X_i, \theta))$ uif og vi kan skrive

$$\frac{1}{\sqrt{n}} \ell'(\theta) = \frac{1}{\sqrt{n}} \left[\sum_{i=1}^{n} \frac{\partial}{\partial \theta} \ln f(X_i; \theta) \right] = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i$$

faktisk med

$$\mathrm{E}(U_i) = \mathrm{E}[\frac{\partial}{\partial \theta} \log f(X_i; \theta)] = 0$$

som vist på neste slide.

Ved Sentralgrenseteoremet får vi dermed

$$\frac{1}{\sqrt{n}} \ell'(\theta) \to N(0, \mathrm{sd}(\frac{\partial}{\partial \theta} \ln f(X_i; \theta)))$$

Utledning av $\mathrm{E}(U_i) = \mathrm{E}[\frac{\partial}{\partial\theta} \ln f(X_i;\theta)] = 0$

Utledningen gjøres i det diskrete tilfellet med $f(x,\theta)$ punktsannsynligheter.

$$
\begin{aligned}
E[\tfrac{\partial}{\partial\theta} \ln f(X_i;\theta)] &= \textstyle\sum_x f(x;\theta) \tfrac{\partial}{\partial\theta} \ln f(x;\theta) \\
&= \textstyle\sum_x f(x;\theta) \frac{\frac{\partial}{\partial\theta} f(x;\theta)}{f(x;\theta)} \\
&= \textstyle\sum_x \frac{\partial}{\partial\theta} f(x;\theta) \\
&= \tfrac{\partial}{\partial\theta} \textstyle\sum_x f(x;\theta) = \tfrac{\partial}{\partial\theta}(1) = 0
\end{aligned}
$$

siden $\sum_x f(x;\theta) = 1$ og ikke avhenger av θ.

(Tilsvarende for kontinuerlig fordelte variable med tetthet $f(x;\theta)$ byttes $\sum_x$ med et integral over mulige utfall.)

Asymptotisk fordeling av MLE - nevner

Som foran er $X_i; i = 1, \ldots, n$ er uavhengige identisk fordelte variable. Da er også

$$W_i = \frac{\partial^2}{\partial \theta^2} \ln f(X_i; \theta)$$

uavhengige identisk fordelte variable. Dermed ved store talls lov

$$\frac{1}{n}[-\ell''(\theta)] \quad = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta^2} \ln f(X_i; \theta) = -\frac{1}{n} \sum_{i=1}^{n} W_i$$

$$\rightarrow -\mathrm{E} W_i = -\mathrm{E} \frac{\partial^2}{\partial \theta^2} \ln f(X_i; \theta)$$

Har i tillegg

$$\mathrm{E}\left(-\frac{\partial^2}{\partial \theta^2} \ln f(X_i; \theta)\right) = \mathrm{V}\left(\frac{\partial}{\partial \theta} \ln f(X_i; \theta)\right)$$

som er viktig for variansuttrykk for MLE (identiteten vises senere)

Asymptotisk fordeling av MLE - forts

Har da

$$\sqrt{n}(\hat{\theta}-\theta) \approx \frac{\frac{1}{\sqrt{n}}\ell'(\theta)}{\frac{1}{n}[-\ell''(\theta)]}$$

$$\frac{1}{\sqrt{n}}\ell'(\theta) \to N\left(0, \mathrm{sd}(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))\right)$$

$$\frac{1}{n}[-\ell''(\theta)] = \frac{1}{n}\left[-\sum_{i=1}^{n}\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right] \to \mathrm{V}(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))$$

Kombinert får vi da at

$$\sqrt{n}(\hat{\theta}-\theta) \to N\left(0, 1/\sqrt{\mathrm{V}[\frac{\partial}{\partial\theta}\ln f(X_i;\theta)]}\right)$$

Da har vi brukt regelen $\mathrm{V}(aY) = a^2\mathrm{V}(Y)$ med
$Y \sim N\left(0, \mathrm{sd}(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))\right)$ og $a = 1/\mathrm{V}(\frac{\partial}{\partial\theta}\ln f(X_i;\theta))$.

Fisher-informasjon

Definerer så Fisher informasjon *for en enkelt observasjon* X_i som

$$I(\theta) = V\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right) = \mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right)$$

- Merk at den siste likheten er en generell identitet.
- Det har som konsekvens at vi kan bruke både $V\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right)$ og $\mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right)$ når vi skal finne variansen til MLE $\hat{\theta}$.
- Det varierer fra situasjon til situasjon hvilken variant som er enklest å benytte
- Siden denne ikke referer til noen n er $I(\theta)$ egnet til å uttrykke den asymptotiske variansen til $\sqrt{n}(\hat{\theta}-\theta)$.
- Tilnærmet varians til $\hat{\theta}$ blir $1/I_n(\theta)$ der $I_n(\theta) = nI(\theta)$ som uttrykker samlet informasjon for alle uif $X_i; i = 1,\ldots,n$.

Oppsummering large sample MLE, så langt

- Anta 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.
- Fisher-informasjonen blir da et mål på denne informasjonen:

$$I(\theta) = \mathrm{V}(U_i) \quad \text{med} \quad U_i = \frac{\partial}{\partial\theta} \ln f(X_i;\theta).$$

- Argumenterte for at asymptotisk fordeling til ML estimator $\hat{\theta}$ gis ved:

$$\sqrt{n}(\hat{\theta} - \theta) \to N\left(0, 1/\sqrt{I(\theta)}\right)$$

- og altså

$$\hat{\theta} \approx \mathrm{N}\left(\theta, 1/\sqrt{nI(\theta)}\right)$$

Angående betingelser for resultatet

- Resultatet om asymptotisk fordeling for MLE $\hat{\theta}$ benytter

$$\mathrm{E}(U_i) = 0 \quad \text{og} \quad V\left(\frac{\partial}{\partial\theta} \ln f(X_i;\theta)\right) = -\mathrm{E}\left(\frac{\partial^2}{\partial\theta^2} \ln f(X_i;\theta)\right).$$

- Viktigste betingelse: Vi kan partiellderivere uttrykk som

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

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

  to ganger mht θ (for aktuelle $g(x;\theta)$)
- Dessuten må det være lov å bytte om rekkefølgen på derivasjon og integrasjon/summasjon i de aktuelle tilfellene.
- Ytterligere viktig forutsetning: Mengden $\{x : f(x;\theta) > 0\}$ avhenger ikke av θ.

Eksempler

Eksempel

$X \sim \textit{Bernoulli}(p)$ og $\{x : f(x; p) > 0\} = \{0, 1\}$ og avhenger ikke av p. Så regulær MLE.

Eksempel

$X \sim \textit{Eksponensiell}(\lambda)$ og $\{x : f(x; \lambda) > 0\} = \{x : x > 0\}$ og avhenger ikke av λ. Også regulær MLE.

Eksempel

$X \sim U(0, \theta)$ og $\{x : f(x; \theta) > 0\} = \{x : 0 < x < \theta\}$ som avhenger av θ. Altså ikke regulær MLE.

Utledning av $V\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right) = \mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right)$

Mer først at $\mathrm{E}\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right) = 0$, så

$$V\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right) = \mathrm{E}\left([\frac{\partial}{\partial\theta}\ln f(X_i;\theta)]^2\right)$$

Vil vise at $\mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right)$ er lik høyre side i uttrykket over.

Men

$$\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta) = \frac{\partial}{\partial\theta}\left[\frac{\partial \ln f(X_i;\theta)}{\partial\theta}\right] = \frac{\partial}{\partial\theta}\left[\frac{\frac{\partial f(X_i;\theta)}{\partial\theta}}{f(X_i;\theta)}\right]$$

$$= \frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2} f(X_i;\theta) - \left(\frac{\partial f(X_i;\theta)}{\partial\theta}\right)^2}{f(X_i;\theta)^2} = \frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2}}{f(X_i;\theta)} - \frac{\left(\frac{\partial f(X_i;\theta)}{\partial\theta}\right)^2}{f(X_i;\theta)^2}$$

$$= \frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2}}{f(X_I;\theta)} - \left(\frac{\partial \ln(f(X_i;\theta))}{\partial\theta}\right)^2$$

Utledning fortsetter!

Dette gir oss at høyre side i formelen

$$\mathrm{E}\left(-\frac{\partial^2}{\partial\theta^2}\ln f(X_i;\theta)\right) = -\mathrm{E}\left[\frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2}}{f(X_i;\theta)} - \left(\frac{\partial \ln(f(X_i;\theta))}{\partial\theta}\right)^2\right]$$

$$= -\mathrm{E}\left[\frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2}}{f(X_i;\theta)}\right] + \mathrm{E}\left[\left(\frac{\partial \ln(f(X_i;\theta))}{\partial\theta}\right)^2\right]$$

$$= 0 + V\left(\frac{\partial}{\partial\theta}\ln f(X_i;\theta)\right), \text{ altså venstre side!!!}$$

Dette fordi

$$\mathrm{E}\left[\frac{\frac{\partial^2 f(X_i;\theta)}{\partial\theta^2}}{f(X_i;\theta)}\right] = \sum_x \left[\frac{\frac{\partial^2 f(x;\theta)}{\partial\theta^2}}{f(x;\theta)}\right] f(x;\theta) = \frac{\partial^2 \sum_x f(x;\theta)}{\partial\theta^2} = 0$$

hvor 2. derivasjonsoperatoren er flyttet utenfor summasjonen over mulige x.

Fisher-informasjon i et tilfeldig utvalg (forts.)

- Anta $X_1, ..., X_n \overset{uif}{\sim} f(X;\theta)$
- Likelihood-funksjonen og log-likelihood-funksjonen er da

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

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

- Score-funksjonen defineres videre som

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

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

$$I_n(\theta) = \mathrm{V}\left(s(\theta)\right) \overset{uavh.}{=} \sum_{i=1}^{n} \underbrace{\mathrm{V}(U_i)}_{I(\theta)} = nI(\theta).$$

Cramér-Rao's ulikhet og effisiens

- Anta $X_1, \ldots, X_n$ uif med punktsannsynlighet/ sannsynlighetstetthet $f(x;\theta)$, avhengig av én parameter θ.
- Anta $\{x : f(x;\theta) > 0\}$ ikke avhenger av θ.
- La $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$ være en forventningsrett estimator for θ.
- Da får vi Cramér-Raos ulikhet:

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

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

Eksempel Cramér-Rao's ulikhet og effisiens

Anta at $X_1, \ldots, X_n \overset{uif}{\sim}$ Bernoulli(p).

Da er $\hat{p} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{Y}{n}$ MLE (se slide 3).

Vi fant at $E(\hat{p}) = p$, så forventningrett.

Videre er $V(\hat{p}) = \frac{p(1-p)}{n} = \frac{1}{I_n(p)} =$ Cramér-Rao's nedre grense.

Altså er $\hat{p}$ effisient!

Utledning av Cramér-Rao's ulikhet

Minner om følgende ulikhet: For to stokastiske variable V og W med endelige varianser σ_V^2 og σ_W^2 gjelder

$$\mathrm{Cov}(V, W)^2 \leq \sigma_V^2 \sigma_W^2$$

som følger fra side 299 i DBC.

Anta $X_1, \ldots, X_n$ er uif med tetthet/punktsannsynligheter $f(x; \theta)$. La videre $\hat{\theta} = T(X_1, \ldots, X_n) = T$ være forventningsrett for θ.

Vi har da (bevises på neste slide) når $s(\theta)$ er score-funksjonen for θ,

$$\mathrm{Cov}(T, s(\theta)) = 1$$

Dermed får vi

$$1 = \mathrm{Cov}(T, s(\theta))^2 \leq \mathrm{V}(T)\mathrm{V}(s(\theta)) = \mathrm{V}(T) n I(\theta)$$

og det følger at $\mathrm{V}(\hat{\theta}) = \mathrm{V}(T) \geq \frac{1}{nI(\theta)}$ = Cramér-Rao's nedre grense.

Utledning $\mathrm{Cov}(T, s(\theta)) = 1$

La $f(x_1, \ldots, x_n; \theta)$ være tetthet/pkt.sh for $X_1, \ldots, X_n$ og $\mathrm{E}(T) = \mathrm{E}(T(X_1, \ldots, X_n)) = \theta$.

Da er $\mathrm{Cov}(T, s(\theta)) = \mathrm{E}(Ts(\theta)) - \mathrm{E}(T)\mathrm{E}(s(\theta)) = \mathrm{E}(Ts(\theta)) - 0$ siden $\mathrm{E}(s(\theta)) = 0$.

Men vi har

$$\begin{aligned}
\mathrm{E}(Ts(\theta)) &= \sum_{(x_1,\ldots,x_n)} T(x_1, \ldots x_n) \frac{\partial \ln(f(x_1,\ldots,x_n;\theta))}{\partial\theta} f(x_1, \ldots, x_n; \theta) \\
&= \sum_{(x_1,\ldots,x_n)} T(x_1, \ldots x_n) \frac{\frac{\partial f(x_1,\ldots,x_n;\theta)}{\partial\theta}}{f(x_1,\ldots,x_n;\theta)} f(x_1, \ldots, x_n; \theta) \\
&= \frac{\partial}{\partial\theta} \sum_{(x_1,\ldots,x_n)} T(x_1, \ldots x_n) f(x_1, \ldots, x_n; \theta) = \frac{\partial(\mathrm{E}(T))}{\partial\theta} \\
&= \frac{\partial(\theta)}{\partial\theta} = 1
\end{aligned}$$

hvor vi enda en gang har benyttet å bytte om sum og derivasjon.

MLEs egenskaper for store utvalg litt mer presist

- Anta $X_1, \ldots, X_n$ er uif med punktsannsynlighet/ sannsynlighetstetthet $f(x;\theta)$, avhengig av θ.
- Videre antar vi at $\{x : f(x;\theta) > 0\}$ ikke avhenger av θ.
- La $\hat{\theta}$ være maksimum likelihood-estimatoren for θ.
- Grensefordelingen MLE $\hat{\theta}$ angis da mer presist som

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

- Dette skrives (noen ganger) ved $\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} W$, hvor d betyr konvergens i fordeling. Vi vil i STK1110 bare si 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

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

Eksempler

Eksempel

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

Eksempel

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

MLEs egenskaper for store utvalg ved flere parametere

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

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

der $I^{jj}(\theta_1, \theta_2)$ er j-te diagonalelement i den inverse av 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

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

Eksempel Gammafordelte data

Anta at $X_1, X_2, \ldots, X_n$ er uif med gammatetthet

$$f(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp(-\frac{x}{\beta}) \quad \text{for } x > 0$$

Vi har da

$$\ln(f(X_i;\alpha,\beta)) = -\alpha \ln(\beta)) - \ln(\Gamma(\alpha)) + (\alpha-1)\ln(X_i) - \frac{X_i}{\beta}$$

som har førstederiverte

$$\frac{\partial \ln(f(X_i;\alpha,\beta))}{\partial \alpha} = -\ln(\beta) - \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + \ln(X_i) = -\ln(\beta) - \psi(\alpha) + \ln(X_i)$$

$$\frac{\partial \ln(f(X_i;\alpha,\beta))}{\partial \beta} = -\frac{\alpha}{\beta} + \frac{X_i}{\beta^2}$$

der $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{\partial \ln(\Gamma(\alpha))}{\partial \alpha}$ er den såkalte "digamma-funksjonen".

Eksempel Gammafordelte data, forts.

Dermed blir de 2. deriverte

$$\frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\alpha^2} = -\psi'(\alpha) \quad , \quad \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\beta^2} = \frac{\alpha}{\beta^2} - 2\frac{X_i}{\beta^3}$$

og $$\frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\alpha\partial\beta} = \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\beta\partial\alpha} = -\frac{1}{\beta}$$

Disse 2. deriverte sammen med $\mathrm{E}(X_i) = \alpha\beta$ gir informasjonsmatrise for X_i lik

$$I(\alpha,\beta) = -\mathrm{E}\begin{bmatrix} \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\alpha^2} & \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\alpha\partial\beta} \\ \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\alpha\partial\beta} & \frac{\partial^2 \ln(f(X_i;\alpha,\beta))}{\partial\beta^2} \end{bmatrix} = \begin{bmatrix} \psi'(\alpha) & \frac{1}{\beta} \\ \frac{1}{\beta} & \frac{\alpha}{\beta^2} \end{bmatrix}$$

som har invers

$$I^{-1}(\alpha,\beta) = \begin{bmatrix} \frac{\alpha}{\beta^2} & -\frac{1}{\beta} \\ -\frac{1}{\beta} & \psi'(\alpha) \end{bmatrix} \frac{1}{\psi'(\alpha)\frac{\alpha}{\beta^2} - \frac{1}{\beta^2}}$$