STK-IN1050 - Statistikk for informatikere

12. Statistisk estimering av parameter

Matematisk institutt

2026-03-02

Hva er en parameter?

En **parameter** er et ukjent, men fast tall i en statistisk modell

Eksempel (binomisk modell):

- Vi har $X \sim \text{Bin}(n, p)$
- n: kjent (antall forsøk)
- p: ukjent parameter (sannsynlighet for suksess)

Målet i punktestimering:

- Estimere p fra observerte data

Estimat og estimator

Anta at vi har observert data $x_1, x_2, \ldots, x_n$. Et **estimat** av en ukjent parameter θ er en funksjon

$$\hat{\theta} = h(x_1, x_2, \ldots, x_n)$$

som bare avhenger av dataene.

Anta at $X_1, X_2, \ldots, X_n$ er stokastiske variabler. Da er den stokastiske variabelen

$$\hat{\theta} = h(X_1, X_2, \ldots, X_n)$$

en **estimator** for θ og $h(x_1, x_2, \ldots, x_n)$ er en realisering av denne stokastiske variabelen.

Estimat vs estimator

Det betyr at

- **Estimatoren** er selv en stokastisk variabel
- **Estimatet** er et tall

Eksempel (binomisk fordeling):

- $X \sim \text{Bin}(n, p)$
- Estimator for p:

$$\hat{p} = \frac{X}{n}.$$

- Observasjon $X = 7$ med $n = 10$ gir estimatet

$$\hat{p} = 0.7.$$

Estimator som funksjon

Hvis vi tenker ut fra en progammeringsvinkel, da er en estimator som en funksjon i kode:

$$\hat{\theta} = h(X_1, \ldots, X_n)$$

hvor vi har:

- Input: data (liste/array)
- Output: tall som estimerer parameter

Utvalgsfordeling

Estomatoren avhenger av tilfeldige data og er selv en stokastisk varibel.

Den har dermed også en egen sannsynlighetsfordeling.

Dvs. $\hat{\theta}$ er tilfeldig og har:

- Forventning $\mathbb{E}(\hat{\theta})$
- Varians $Var(\hat{\theta})$
- Ofte en kjent fordeling

Fordelingen til estimatoren kalles **utvalgsfordeling**.

Eksempel: Utvalgsfordelingen til $\hat{p}$ i $Bin(n, p)$

Vi har en binomisk modell, $X \sim \text{Bin}(n, p)$ hvor n er kjent og estimatoren for p er gitt ved

$$\hat{p} = \frac{X}{n}.$$

Vi kan nå studere:

- Forventningen $\mathbb{E}(\hat{p})$
- Variansen $Var(\hat{p})$
- Selve fordelingen til $\hat{p}$

Eksempel: Forventning og varians til $\hat{p}$

For $X \sim \text{Bin}(n, p)$ gjelder at $\mathbb{E}(X() = np$ og $Var(X) = np(1-p)$.

- Forventning:

$$\mathbb{E}(\hat{p}) = \mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n}\mathbb{E}(X) = \frac{1}{n} \cdot np = p$$

- Varians:

$$\text{Var}(\hat{p}) = \text{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\text{Var}(X) = \frac{1}{n^2} \cdot np(1-p) = \frac{p(1-p)}{n}$$

$Var(\hat{p})$ avtar når n øker: utvalgsfordelingen til $\hat{p}$ blir mer «konsentrert» rundt p.

Eksempel: Fordelingene til X og $\hat{p}$

X har verdimengde $\mathcal{X} = \{0, 1, \ldots, n\}$ og diskret sannsynlighetsfordeling

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

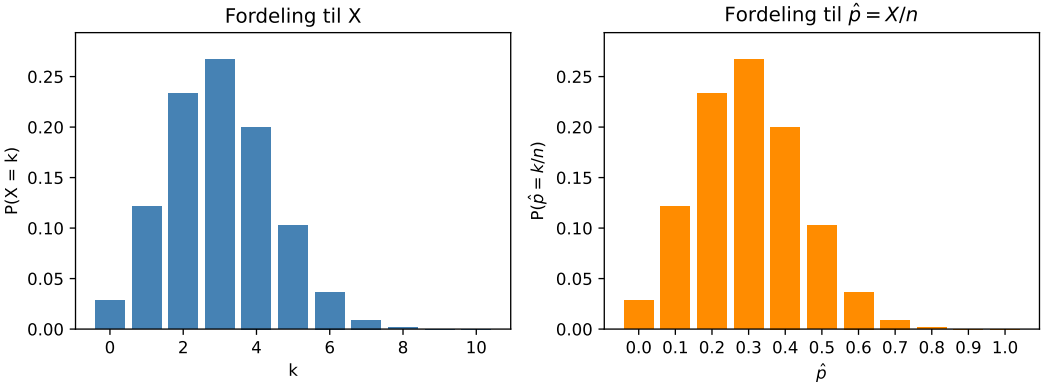
Estimatoren $\hat{p} = \frac{X}{n}$ har dermed verdimengde $\mathcal{P} = \left\{\frac{0}{n}, \frac{1}{n}, \ldots, \frac{n}{n}\right\} = \left\{0, \frac{1}{n}, \ldots, 1\right\}$ og sannsynlighetsfordeling

$$P\left(\hat{p} = \frac{k}{n}\right) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n.$$

Sannsynlighetene er de samme, men «flyttet» til skalaen $[0, 1]$.

Eksempel: Fordelingene til X og $\hat{p}$

Fordelingene til X og $\hat{p}$ når $n = 10$ og $p = 0.3$:



Eksempel: Utvalgsgjennomsnittet

La $X_1, \ldots, X_n$ være uavhengige og identisk fordelte med

$$\mathbb{E}(X_i) = \mu, \qquad Var(X_i) = \sigma^2.$$

Vi definer **utvalgsgjennomsnittet** som

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

Utvalgsgjennomsnittet har forventning

$$\mathbb{E}(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n} \cdot n\mu = \mu.$$

Eksempel: Utvalgsgjennomsnittet

Variansen til utvalgsgjennomsnittet er gitt ved

$$Var(\bar{X}) = \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}$$

Det betyr at når n øker:

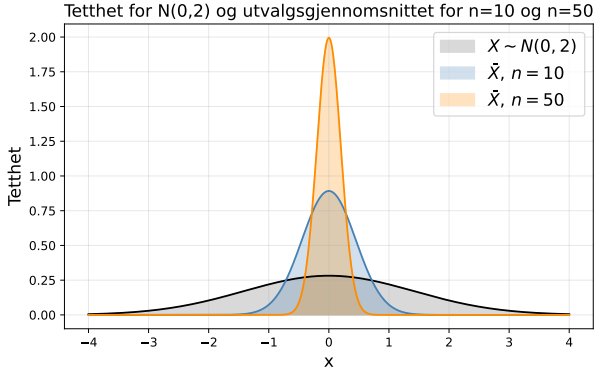
- $\text{Var}(\bar{X})$ blir mindre
- Estimatoren blir mer presis

Eksempel: Utvalgsgjennomsnittet

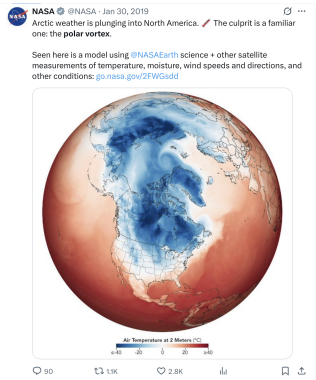
Fordelingen til utvalgsgjennomsnittet avhenger av fordelingen til de opprinnelige $X_1, \ldots, X_n$.

For eksempel, hvis $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, da gjelder

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right).$$



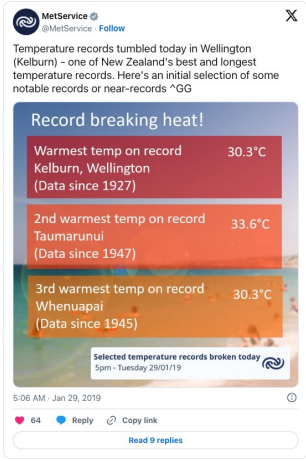
Eksempel: Lokale vs. globale klimaendringer



I starten av 2019 opplevde USA en ekstrem kuldeperiode forårsakt av en såkaldt polarvirvel.

Det førte til flere diskusjoner om «mangel på global oppvarming».

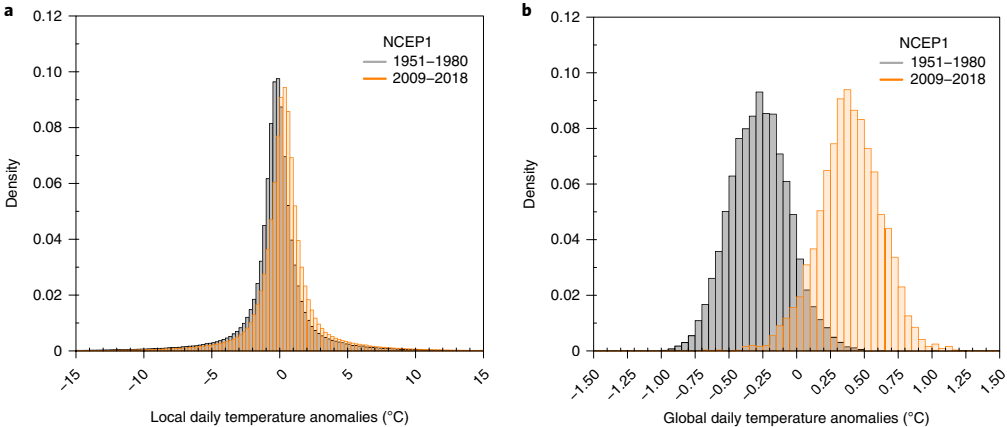
Eksempel: Lokale vs. globale klimaendringer



En annen hendelse fikk mindre oppmerksomhet i media: New Zealand opplevde på samme tidspunkt en ekstrem hetebølge med flere nye varmerekorder.

Dette førte noen klimaforskere i Sveits til å gjennomføre en studie, hvor de sammenlignet observerte lokale og globale klimaendringer.

Eksempel: Lokale vs. globale klimaendringer



Kilde: Sippel, S., Meinshausen, N., Fischer, E. M., Székely, E., & Knutti, R. (2020). Climate change now detectable from any single day of weather at global scale. *Nature Climate Change*, 10(1), 35-41.

Forventningsrett estimator

En estimator $\hat{\theta}$ for parameter θ er **forventningsrett** hvis

$$\mathbb{E}(\hat{\theta}) = \theta,$$

og hvis $\mathbb{E}(\hat{\theta}) \neq \theta$ sier vi at estimatoren er **forventningsskjev**.

Intuisjon:

- Hvis vi gjentar eksperimentet uendelig mange ganger, vil gjennomsnittet av alle estimatene for en forventningsrett estimator være den sanne θ.

Eksempel:

- $\bar{X}$ med $\mathbb{E}(\bar{X}) = \mu$ er en forventningsrett estimator for μ.
- $\hat{p} = X/n$ med $\mathbb{E}(\hat{p}) = p$ er en forventningsrett estimator for p i $Bin(n, p)$.

Estimatorer for varians

Vi har tidligere i kurset diskutert at standard estimator for variansen til en stokastisk variabel er gitt ved

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

Vi skal nå vise at denne er forventningsrett, dvs. at $\mathbb{E}(S^2) = \sigma^2$.

Husk at vi har $\mathbb{E}(X_i) = \mathbb{E}(\bar{X}) = \mu, \quad Var(X_i) = \sigma^2, \quad Var(\bar{X}) = \sigma^2/n$

1. Vi skriver S^2 om:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right].$$

2. Vi tar forventning på begge sider:

$$\begin{aligned}\mathbb{E}(S^2) &= \frac{1}{n-1}\left(\mathbb{E}\Big[\sum_{i=1}^{n}(X_i - \mu)^2\Big] - n\mathbb{E}\big[(\bar{X} - \mu)^2\big]\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^{n}\mathbb{E}\big[(X_i - \mu)^2\big] - n\mathbb{E}\big[(\bar{X} - \mu)^2\big]\right).\end{aligned}$$

3. Vi beregner de to forventningene:

$$\mathbb{E}[(X_i - \mu)^2] = Var(X_i) = \sigma^2 \quad \text{for hver } i$$
$$\mathbb{E}[(\bar{X} - \mu)^2] = Var(\bar{X}) = \sigma^2/n$$

4. Vi setter inn:

$$\mathbb{E}(S^2]) = \frac{1}{n-1}\left(n\sigma^2 - n \cdot \frac{\sigma^2}{n}\right)$$
$$= \frac{1}{n-1}(n\sigma^2 - \sigma^2) = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2.$$

$\Rightarrow S^2$ er en forventningsrett estimator for σ^2.

Intuisjon om $n - 1$ i nevneren

- Vi bruker $\bar{X}$ til å estimere μ
- Dermed «bruker vi opp» én **frihetsgrad**
- Det gir en systematisk undervurdering av variansen hvis vi deler på n
- Korrigering med $n - 1$ gir forventningsrett estimator

Forventningsrett «arves» ikke nødvendigvis

Vi har vist at S^2 er en forventningsrett estimator for σ^2:

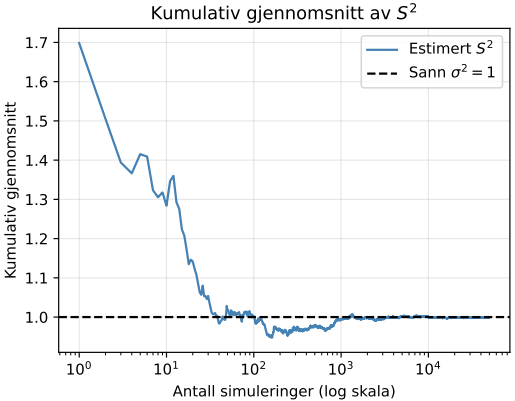
$$\mathbb{E}(S^2) = \sigma^2.$$

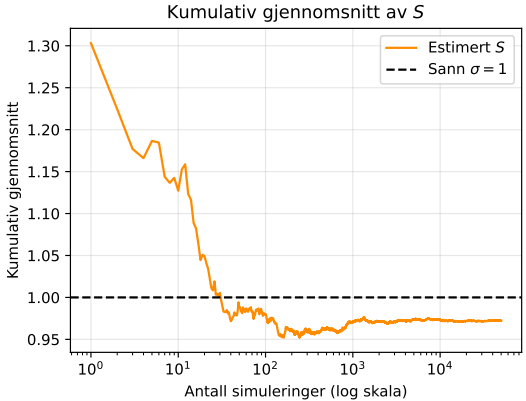
Men det tilsvarande standardavviket $S = \sqrt{S^2}$ er ikke en forventningsrett estimator for σ:

$$\mathbb{E}(S) \neq \sigma.$$

Generelt: Hvis $\hat{\theta}$ er forventningsrett for θ, da er ikke nødvendigvis $g(\hat{\theta})$ forventningsrett for $g(\theta)$ (men det gjelder hvis g er en lineær funksjon).

Eksempel: Estimering av $\mathbb{E}(S^2)$ og $\mathbb{E}(S)$ for $N(0,1)$ og $n = 10$





Varians til estimatorer

Mest effisient estimator

Av flere forventningsrette estimatorer for en parameter sier vi at den med minst varians er **mest effisient**.

Intuisjon:

$\hat{\theta}_1$ er mer presis enn $\hat{\theta}_2$ hvis

$$Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$$

Eksempel: To estimatorer for μ

La $X_1, \ldots, X_n$ være uavhengig og identisk fordelt med $\mathbb{E}(X_i) = \mu$, $Var(X_i) = \sigma^2$.

Vi har to estimatorer for μ.

- Estimator A:

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

- Estimator B:

$$\hat{\mu}_B = X_1$$

Eksempel: Sammenlikning av $\hat{\mu}_A$ og $\hat{\mu}_B$

- Begge er forventningsrette:

$$E[\hat{\mu}_A] = \mu, \quad E[\hat{\mu}_B] = \mu$$

- De har forskjellig varians:

$$\text{Var}(\hat{\mu}_A) = \frac{\sigma^2}{n}, \quad \text{Var}(\hat{\mu}_B) = \sigma^2$$

- For $n > 1$ gjelder at:

$$\frac{\sigma^2}{n} < \sigma^2$$

$\Rightarrow \bar{X}$ er mer presis enn X_1

Forventningsskjevhet og kvadratisk feil

Selv om det er bra med en forventningsrett estimator, så bør ytelsen til en estimator generelt bedømmes ut fra måten estimatene sprer seg rundt parameteren θ som skal estimeres.

Gjennomsnittlig kvadratisk feil

Anta at $\hat{\theta}$ er en estimator for parameteren θ. Den **gjennomsnittlige kvadratiske feilen** (mean squared error eller MSE) til $\hat{\theta}$ er gitt ved

$$MSE(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \theta)^2\right].$$

Forventningsskjevhet og kvadratisk feil

Det gjelder at

$$\begin{aligned}
\text{MSE}(\hat{\theta}) &= \mathbb{E}[(\hat{\theta} - \theta)^2] \\
&= \mathbb{E}\big[(\hat{\theta} - \mathbb{E}(\hat{\theta}) + \mathbb{E}(\hat{\theta}) - \theta)^2\big] \\
&= \mathbb{E}\big([\hat{\theta} - \mathbb{E}(\hat{\theta})]^2\big) + 2[\mathbb{E}(\hat{\theta}) - \theta]\mathbb{E}[\hat{\theta} - \mathbb{E}(\hat{\theta})] + [\mathbb{E}(\hat{\theta}) - \theta]^2 \\
&= Var(\hat{\theta}) + [\mathbb{E}(\hat{\theta}) - \theta]^2
\end{aligned}$$

Dvs. den gjennomsnittlige kvadratiske feilen kombinerer variansen og den kvadratiske forventningsskjevheten til $\hat{\theta}$.

Vi kan da utvide konseptet om effisiens til å inkludere forventningskjeve estimatorer og si at $\hat{\theta}_1$ er mer **effisient** enn $\hat{\theta}_2$ hvis $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$.

Eksempel: En krympet estimator for forventningen μ

En standard estimator for forventningen μ til uavhengige og identisk fordelte stokastiske varibler $X_1, \ldots, X_n$ er gitt ved gjennomsnittet:

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

Estimatoren $\hat{\mu}$ er forventningsrett

$$\mathbb{E}(\hat{\mu}) = \mu$$

og har varians

$$Var(\hat{\mu}) = \frac{\sigma^2}{n}.$$

MSE til $\hat{\mu}$ er dermed også gitt ved

$$MSE(\hat{\mu}) = \frac{\sigma^2}{n}.$$

Eksempel: En krympet estimator for forventningen μ

Definer for en konstant c med $0 < c < 1$ en ny estimator:

$$\tilde{\mu} = c\bar{X}.$$

Den nye estimatoren $\tilde{\mu}$ er forventningsskjev med

$$\mathbb{E}(c\tilde{\mu}) = c\mu$$

og varians

$$Var(\tilde{\mu}) = c^2 Var(\bar{X}) = c^2 \frac{\sigma^2}{n}.$$

MSE er her gitt ved

$$MSE(\tilde{\mu}) = Var(\tilde{\mu}) + [\mathbb{E}(\tilde{\mu}) - \mu]^2 = c^2 \frac{\sigma^2}{n} + (c-1)^2 \mu^2.$$

Eksempel: En krympet estimator for forventningen μ

Vi kan nå vurdere når $\tilde{\mu}$ er mer effisient enn $\hat{\mu}$, dvs. når

$$MSE(\tilde{\mu}) < MSE(\hat{\mu}).$$

Det gjelder når

$$c^2\frac{\sigma^2}{n} + (c-1)^2\mu^2 < \frac{\sigma^2}{n}.$$

Hvis μ er liten (for eksempel hvis vi vet på forhånd at μ er nær 0), kan dette lett skje.

Eksempel: En krympet estimator for forventningen μ

For eksempel, anta at den sanne forventningen er gitt ved $\mu = 0$ og la $c = 0.8$.

I dette tilfelle er $\tilde{\mu}$ forventningsrett dersom $\mathbb{E}(c\tilde{\mu}) = c\mu = 0$ og dermed:

$$MSE(\tilde{\mu}) = 0.8^2 \frac{\sigma^2}{n} = 0.64 \frac{\sigma^2}{n} < \frac{\sigma^2}{n} = MSE(\hat{\mu}).$$

En (svakt) forventningsskjev estimator kan ha lavere MSE enn en forventningsrett estimator hvis den reduserer variansen nok til at det veier opp mot en liten forventningsskjevhet.

Denne type krymping er ideen bak mange regulariseringsmetoder i statistikk og maskinlæring.

Eksempel: To estimatorer for p i binomisk fordeling

Anta at $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ og $Y = \sum_{i=1}^{n} X_i \sim \text{Bin}(n, p)$. Da gjelder

- $\mathbb{E}(X_i) = p$ and $Var(X_i) = 1(1-p)$ for alle $i = 1, \ldots, n$
- $\mathbb{E}(Y) = np$ og $Var(Y) = np(1-p)$.

Vi definerer to estimatorer for p:

$$\hat{p}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{Y}{n}$$

$$\hat{p}_2 = \frac{1}{2}(X_1 + X_2)$$

Oppgave: Sammenlign effisiensen til p_1 og p_2 ved bruk av MSE.

(Husk at $\text{MSE}(\hat{\theta}) = Var(\hat{\theta}) + [\mathbb{E}(\hat{\theta}) - \theta]^2$)

Eksempel: To estimatorer for p i binomisk fordeling

Både $\hat{p}_1$ og $\hat{p}_2$ er forventningsrette:

$$\mathbb{E}(\hat{p}_1) = p, \quad \mathbb{E}(\hat{p}_2) = p$$

For variansene gjelder at

$$\text{Var}(\hat{p}_1) = \frac{1}{n^2} Var(Y) = \frac{p(1-p)}{n}$$

$$\text{Var}(\hat{p}_2) = \frac{1}{4}\big(\text{Var}(X_1) + \text{Var}(X_2)\big) = \frac{1}{4}(2p(1-p)) = \frac{p(1-p)}{2}$$

$\hat{p}_1$ er mer effektiv enn $\hat{p}_2$ for $n > 2$:

$$MSE(\hat{p}_1) = \frac{p(1-p)}{n} < \frac{p(1-p)}{2} = MSE(\hat{p}_2).$$

Eksempel: Hvor mange elementer?

Anta at det finnes et ukjent antall elementer som er sekvensielt nummerert fra 1 til N.

Anta at $X_1, X_2, \ldots, X_n$ er et tilfeldig utvalg trukket uten tilbakelegging fra disse tallene $1, 2, \ldots, N$.

Vi ønsker å estimere N fra de observerte tallene $x_1, x_2, \ldots, x_n$.

Eksempel: German tank problem

Under andre verdenskrig brukte de allierte denne statistiske metoden for å anslå hvor mange tyske stridsvogner som var produsert hver måned.

Det gjorde de ved å analysere serienummer på forskjellige deler av fangede stridsvogner.



Tyske Panzer V stridsvogner i 1943
Bilde: Bundesarchiv, Bild 183-H26258 (wikipedia)

Eksempel: German tank problem

Vi ser på to naturlige estimatorer for N:

1. Maksimumsbasert estimator $\widehat{N}_{\text{max}}$ basert på største serienummer:

$$M = \max(X_1, \ldots, X_n).$$

2. Gjennomsnittsbasert estimator $\widehat{N}_{\text{mean}}$ basert på utvalgsgjennomsnittet:

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

Vi ønsker forventningsrette estimatorer med minst mulig varians.

Eksempel: German tank problem

Vi finner først forventningen til utvalgsgjennomsnittet $\bar{X}$.

Selv om vi trekker uten tilbakelegging, så kan det vises at hver X_i har marginal fordeling

$$P(X_i = k) = \frac{1}{N} \quad \text{for } k = 1, 2, \ldots, N.$$

Forventningen til hver X_i er dermed gitt ved

$$\mathbb{E}(X) = 1 \cdot \frac{1}{N} + \cdots + N \cdot \frac{1}{N} = \frac{1 + \cdots + N}{N} = \frac{\frac{1}{2}N(N+1)}{N} = \frac{N+1}{2}.$$

Forventningen er lineær, også for avhengige variabler, som gir $\mathbb{E}(\bar{X}) = \mathbb{E}(X) = (N+1)/2$. Hvis vi løser for N får vi en forventningsrett estimator for N basert på $\bar{X}$:

$$\widehat{N}_{\text{mean}} = 2\bar{X} - 1.$$

Eksempel: German tank problem

For M kan det vises at sannsynlighetsfordelingen er gitt ved

$$P(M = m) = \frac{\binom{m-1}{n-1}}{\binom{N}{n}}, \quad m = n, \ldots, N.$$

Det følger at forventningen er lik

$$\mathbb{E}(M) = \sum_{m=n}^{N} m\, P(M = m) = \frac{n}{n+1}(N+1).$$

Hvis vi løser denne ligningen for N får vi igjen en forventningsrett estimator

$$\widehat{N}_{\max} = \frac{n+1}{n} M - 1.$$

Eksempel: German tank problem

$Var(\widehat{N}_{\text{mean}})$:

For å finne $Var(\widehat{N}_{\text{mean}}) = Var(2\overline{X} - 1)$ trenger vi $Var(\overline{X})$.

For identisk fordelte X_1, X_2 gjelder

$$Var(X_1 + X_2) = 2Var(X_1) + 2Cov(X_1, X_2).$$

Den mer generelle regelen er

$$Var(X_1 + \cdots X_n) = nVar(X_1) + n(n-1)Cov(X_1, X_2).$$

For å finne $Var(\overline{X})$ trenger vi altså $Var(X_1)$ og $Cov(X_1, X_2)$.

Eksempel: German tank problem

$Var(X_1)$:

Vi har funnet

$$\mathbb{E}(X) = \frac{N+1}{2}$$

og vi har

$$\mathbb{E}(X^2) = 1^2 \cdot \frac{1}{N} + \cdots + N^2 \cdot \frac{1}{N} = \frac{1^2 + \cdots + N^2}{N} = \frac{\frac{1}{6}N(N+1)(2N+1)}{N} = \frac{(N+1)(2N+1)}{6}.$$

Dermed blir

$$\begin{aligned} Var(X_1) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \frac{(N+1)(2N+1)}{6} - \frac{(N+1)^2}{4} \\ &= \frac{2(2N^2+3N+2) - 3(N^2+2N+1)}{12} = \frac{(N+1)(N-1)}{12} \end{aligned}$$

Eksempel: German tank problem

$Cov(X_1, X_2)$:

Vi har at

$$Var(X_1 + X_2 + \cdots + X_N) = 0$$

(dersom summen av alle er et fast tall og ikke tilfeldig). Regel for varians gir

$$0 = NVar(X_1) + N(N-1)Cov(X_1, X_2).$$

Vi løser for $Cov(X_1, X_2)$:

$$Cov(X_1, X_2) = -\frac{1}{N-1}Var(X_1) = -\frac{1}{12}(N+1).$$

Eksempel: German tank problem

$Var(\widehat{N}_{\text{mean}})$:

Vi har $\widehat{N}_{\text{mean}} = 2\overline{X} - 1$ og dermed:

$$
\begin{aligned}
Var(\widehat{N}_{\text{mean}}) &= 4Var(\overline{X}) \\
&= \frac{4}{n^2} Var(X_1 + \cdots X_n) \\
&= \frac{4}{n^2} \left[nVar(X_1) + n(n-1)Cov(X_1, X_2) \right] \\
&= \frac{4}{n^2} \left[\frac{n(N+1)(N-1)}{12} - \frac{n(n-1)(N+1)}{12} \right] \\
&= \frac{(N+1)(N-n)}{3n}.
\end{aligned}
$$

Eksempel: German tank problem

$Var(\widehat{N}_{\max})$:

Det kan vises at variansen til M er gitt ved

$$Var(M) = \frac{n(N+1)(N-n)}{(n+1)^2(n+2)}.$$

Vi har

$$\widehat{N}_{\max} = \frac{n+1}{n}M - 1,$$

og bruker regneregler for varians til å finne variansen til $\widehat{N}_{\max}$:

$$Var(\widehat{N}_{\max}) = \left(\frac{n+1}{n}\right)^2 Var(M) = \left(\frac{n+1}{n}\right)^2 \frac{n(N+1)(N-n)}{(n+1)^2(n+2)} = \frac{(N+1)(N-n)}{n(n+2)}.$$

Eksempel: German tank problem

Begge estimatorene er forventningsrette, så vi sammenlikner variansene for å finne hvilken estimator er mest effektiv:

$$Var(\widehat{N}_{\text{mean}}) = \frac{(N+1)(N-n)}{3n},$$
$$Var(\widehat{N}_{\text{max}}) = \frac{(N+1)(N-n)}{n(n+2)}.$$

Den **relative effisiensen** er definert som forholdet mellom variansene:

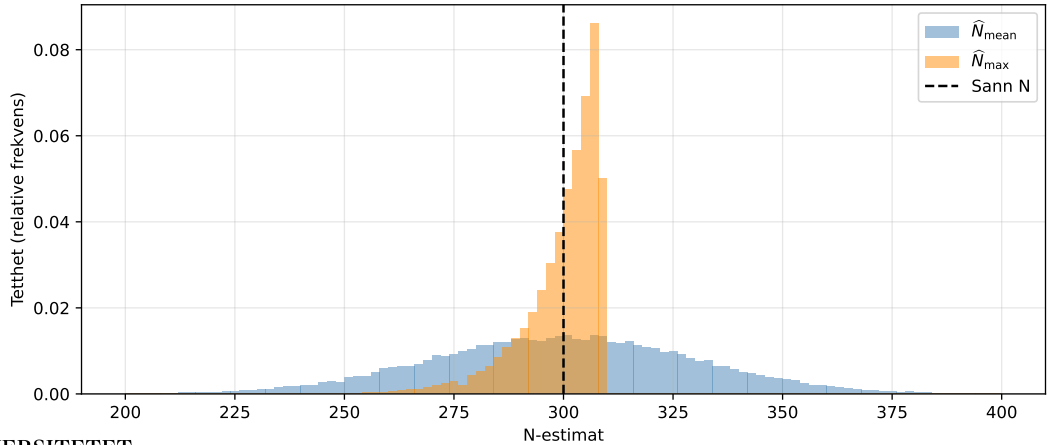
$$\frac{Var(\widehat{N}_{\text{mean}})}{Var(\widehat{N}_{\text{max}})} = \frac{n+2}{3}.$$

Vi ser at $Var(\widehat{N}_{\text{max}}) < Var(\widehat{N}_{\text{mean}})$ for alle N og hvis $n \geq 2$.

$\widehat{N}_{\text{max}}$ er dermed altid mer effisient enn $\widehat{N}_{\text{mean}}$ medmindre $n = 1$.

Eksempel: German tank problem

Vi kan simulere fordelingene til $\widehat{N}_{\text{mean}}$ og $\widehat{N}_{\text{max}}$, f.eks. for $N = 300$, $n = 30$:



Eksempel: German tank problem

Husk at vi har definert:

$$\widehat{N}_{\text{mean}} = 2\bar{X} - 1,$$
$$\widehat{N}_{\text{max}} = \frac{n+1}{n} M - 1.$$

Anta at $n = 5$ og at vi har observert

$$7 \quad 3 \quad 10 \quad 45 \quad 15.$$

Finn verdien av estimatorene $\widehat{N}_{\text{max}}$ og $\widehat{N}_{\text{mean}}$ for disse dataene.

Virker noe rart her?

Tilfeldig utvalg

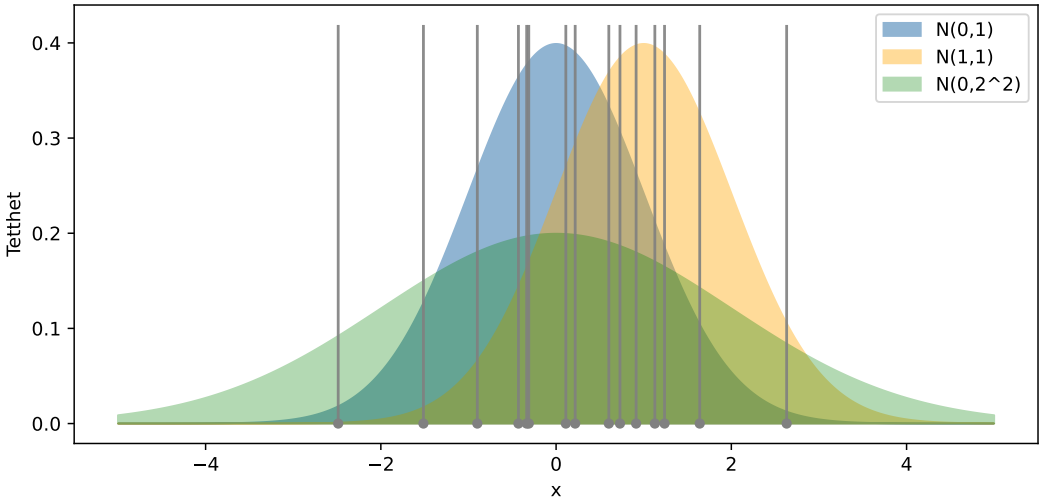
Tilfeldig utvalg

Anta at $X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte stokastiske variabler med marginalfordeling gitt ved tettheten $f(x)$.

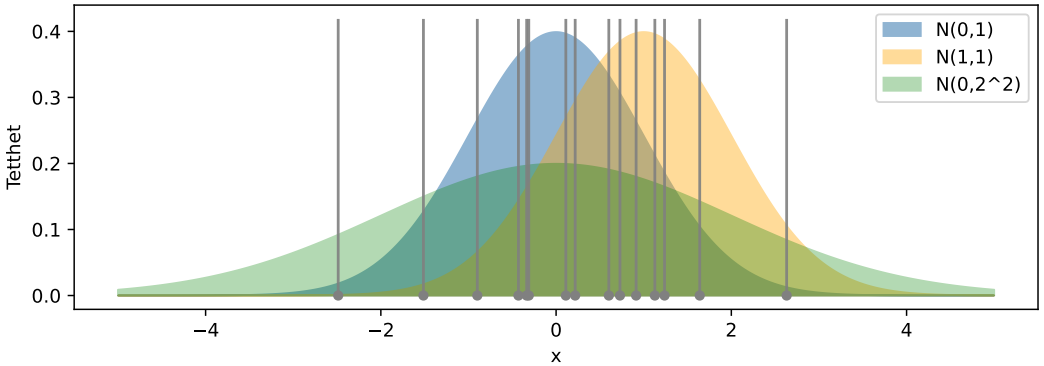
Da sier vi at $X_1, X_2, \ldots, X_n$ er et **tilfeldig utvalg** fra $f(x)$-populasjonen.

Hvis vi har observert verdier $x_1, x_2, \ldots, x_n$ til hver av de stokastiske variablene $X_1, X_2, \ldots, X_n$, da sier vi at $x_1, x_2, \ldots, x_n$ er et **observert tilfeldig utvalg**.

Fra hvilken fordeling kommer dataene?



Fra hvilken fordeling kommer dataene?




```
Gjennomsnittet til utvalget: 0.2154
Standardavviket til utvalget: 1.2869
```

Likelihoodfunksjonen

Likelihoodfunksjonen

Anta at $X_1, X_2, \ldots X_n$ er stokastiske varibler med simultan tetthetsfunksjon $f(x_1, x_2, \ldots, x_n; \theta)$, der θ er en eller flere parametre.

Anta videre at en formel for simultanfordelingen $f(x_1, x_2, \ldots, x_n; \theta)$ er kjent, men at verdien til θ er ukjent.

Hvis man observerer verdier $x_1, x_2, \ldots, x_n$ til hver av de stokastiske variablene $X_1, X_2, \ldots, X_n$ er **likelihoodfunksjonen** gitt som

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n; \theta).$$

Likelihoodfunksjonen

Merk at $f(x_1, x_2, \ldots, x_n; \theta)$ er et utrykk for hvor sannsynlig det er å observere det man har observert.

Her er

- $f(x_1, x_2, \ldots, x_n; \theta)$ en funksjon av $x_1, x_2, \ldots, x_n$ mens θ har en fasthold verdi.
- $L(\theta; x_1, x_2, \ldots, x_n)$ en funksjon av θ mens $x_1, x_2, \ldots, x_n$ har en fasthold verdi.

Likelihoodfunksjonen

Hvis vi antar at $X_1, X_2, \ldots, X_n$ er et tilfeldig utvalg fra $f(x;\theta)$-populasjonen, da er simultanfordelingen til $X_1, X_2, \ldots, X_n$ lik produktet av de n marginalfordelingene slik at vi får

$$L(\theta; x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

I praksis kan det være upraktisk å jobbe med at produkt av mange små tall. Derfor bruker vi ofte **log-likelihoodfunksjonen**:

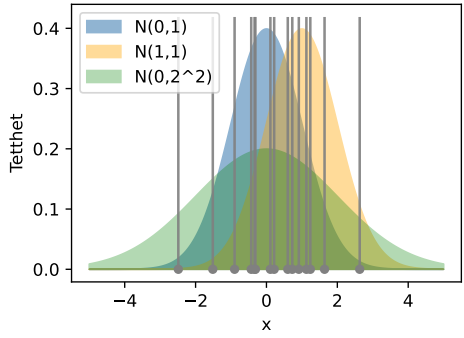
$$\begin{aligned}\ell(\theta; x_1, x_2, \ldots, x_n) &= \log\left(L(\theta; x_1, x_2, \ldots, x_n)\right) \\ &= \sum_{i=1}^{n} \log\left(f(x_i; \theta)\right).\end{aligned}$$

Fra hvilken fordeling kommer dataene?

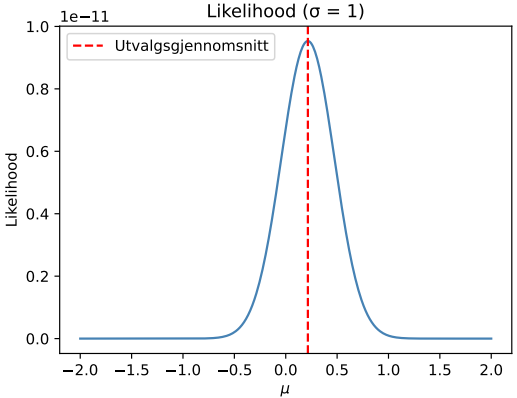
```
Gjennomsnittet til utvalget: 0.2154
Standardavviket til utvalget: 1.2869

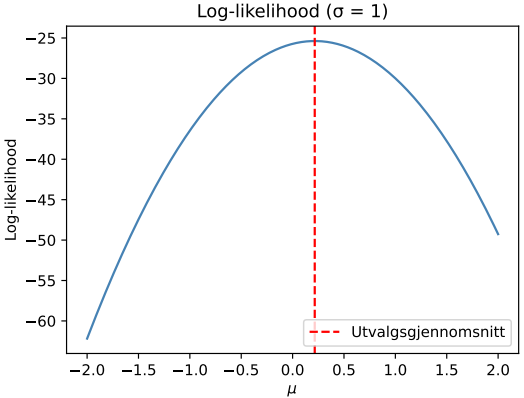
Likelihood:
  L(data | N(0,1)) = 6.731140e-12
  L(data | N(1,1)) = 9.419576e-14
  L(data | N(0,2)) = 1.591514e-12

Log-likelihood:
  log L(data | N(0,1)) = -25.724277
  log L(data | N(1,1)) = -29.993401
  log L(data | N(0,2)) = -27.166335
```



Fra hvilken fordeling kommer dataene?





Maksimum likelihood-prinsippet

Maksimum likelihood-prinsippet (sannsynlighetsmaksimeringsprinsippet)

Dersom man skal estimere verdien til en parameter θ ut fra observerte verdier $x_1, x_2, \ldots, x_n$, skal man velge som estimat den verdi av θ som gjør det mest sannsynlig å observere de verdiene man faktisk har observert.

Maksimum likelihood-prinsippet

Maksimum likelihood-estimator og estimat

La $h(x_1, x_2, \ldots, x_n)$ betegne estimatet for θ man får ved å benytte sannsynlighetsmaksimeringsprinsippet. **Maksimum likelihood-estimatoren** for θ for er da

$$\hat{\theta} = h(X_1, X_2, \ldots, X_n),$$

mens tilhørende observert verdi $\hat{\theta} = h(x_1, x_2, \ldots, x_n)$ der $x_1, x_2, \ldots, x_n$ er observerte verdier, kalles **maksimum likelihood-estimatet** (MLE).

MLE for binomisk fordeling

Anta at $X \sim \text{Bin}(n, p)$ og at vi observer $X = x$.

Likelihoodfunksjonen er gitt ved

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

og den tilsvarende log-likelihood er gitt ved

$$\ell(p; x) = x \log p + (n-x) \log(1-p) + \text{ konstant.}$$

MLE for binomisk fordeling

Vi finner maksimumsverdi ved å derivere med hensyn til p og løse for den deriverte lik null:

- Deriver $\ell(p; x)$:

$$\frac{d}{dp}\ell(p; x) = \frac{x}{p} - \frac{n-x}{1-p}$$

- Sett lik null:

$$\frac{x}{p} - \frac{n-x}{1-p} = 0 \Rightarrow \frac{x}{p} = \frac{n-x}{1-p} \Rightarrow x(1-p) = p(n-x) \Rightarrow x = np \Rightarrow \hat{p}_{\text{MLE}} = \frac{x}{n}$$

Dette er den samme estimatoren som tidligere: $\hat{p} = X/n$.

MLE for Poissonfordelingen

Anta at $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$ og at vi observer $x_1, \ldots, x_n$.

Likelihoodfunksjonen er gitt ved

$$L(\lambda; x_1, \ldots, x_n) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \lambda^{\sum x_i} e^{-n\lambda} \prod_{i=1}^{n} \frac{1}{x_i!}$$

og den tilsvarende log-likelihood er gitt ved

$$\ell(\lambda; x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} x_i \right) \log \lambda - n\lambda + \text{konstant}$$

MLE for Poissonfordelingen

Vi deriver:

$$\frac{d}{d\lambda}\ell(\lambda; x_1, \ldots, x_n) = \frac{\sum x_i}{\lambda} - n$$

Løser for den deriverte lik null:

$$\frac{\sum x_i}{\lambda} - n = 0 \Rightarrow \lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{X}$$

Finner MLE:

$$\hat{\lambda}_{\text{MLE}} = \bar{X}$$

MLE for normalfordeling

Anta at $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Vi vil vise i forelesningen at hvis både μ og σ^2 er ukjent, da gjelder:

$$\hat{\mu}_{\text{MLE}} = \bar{X}$$

$$\hat{\sigma}^2_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Merk at $\hat{\sigma}^2_{\text{MLE}}$ er forventningsskjev; forventningsrett estimator bruker normalisering $1/(n-1)$.

Egenskaper til MLE

Under svake antakelser er MLE ofte:

- **Konsistent**, dvs. at $\hat{\theta}_{\text{MLE}} \to \theta$ når $n \to \infty$
- **Asymptotisk normalfordelt**: For stor n er fordelingen til $\hat{\theta}_{\text{MLE}}$ nær normalfordeling
- **Asymptotisk effisient**: MLE har minst mulig varians (asymptotisk) blant mange estimatorer

Men:

- MLE kan være forventningsskjev for endelig n
- Det å finne MLE kan kreve numerisk optimering i mer kompliserte modeller

Oppsummering

Vi har diskutert

- estimator vs estimat
- utvalgsfordelinger
- forventningsrette og forveningsskjeve estimatorer
- effisiens
- maksimum likelihood