Løsningsforslag øving 12

11.15

1. Vi trekker to kuler, som enten er sorte eller hvite. Utfallet kan beskrives ved de to tilfeldige variablene X og Y som beskriver verdien av trekkingen av den første og den andre kulen, henholdsvis. Hvis en kule er sort, er X eller Y lik 0. Ellers er de lik 1. Vi kan skrive

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1|X = 1).$$

   Her er $P(X = 1) = (37/100)$, mens $P(Y = 1|X = 1) = (36/99)$. Dette gir

$$P(X = 1, Y = 1) = (37/100) \cdot (36/99).$$

2. En urne inneholder 100 kuler, hvorav 37 er hvite. Hvis vi trekker én hvit kule ut av urnen kan dette gjøres på 37 forskjellige måter. Hvis vi deretter trekker enda en hvit kule ut av urnen kan dette gjøres på 36 forskjellige måter. Det totale antallet måter man kan trekke to kuler ut av urnen er lik $100 \cdot 99$. Derfor er sannsynligheten for å trekke to hvite kuler lik $(37 \cdot 36)/(100 \cdot 99)$.

3. En urne inneholder 100 kuler, hvorav 37 er hvite. Vi trekker 2 hvite kuler ut av urnen. Dette kan gjøres på $\binom{37}{2}$ forskjellige måter. Antall måter å trekke 2 kuler ut av urnen er $\binom{100}{2}$. Sannsynligheten for å trekke ut to hvite kuler er derfor lik $\binom{37}{2}/\binom{100}{2} = (\frac{37 \cdot 36}{2})/(\frac{100 \cdot 99}{2}) = (37 \cdot 36)/(100 \cdot 99)$.

11.18

At forventningsverdien $\mathrm{E}X$ er en estimand τ, betyr at vi ønsker å estimere verdien $\tau = \mathrm{E}X$. Hvis vi bruker estimatoren $\hat{\tau} = X$ så er dette en forventningsrett estimator, fordi $\mathrm{E}\hat{\tau} = \mathrm{E}X = \tau$.

11.21

Vi vet at for en Poisson-fordeling med parameter λ så er forventningsverdien lik λ. Da kan vi for eksempel bruke estimatoren $\hat{\lambda} = \bar{X} \approx 1.33$, som vil være en forventningsrett estimator. Det er mange andre estimatorer vi også kan bruke, om vi skulle ønske dette.

Én av dem er rimelighetsestimatoren. For å finne rimelighetsestimatoren må vi først finne rimelighetsfunksjonen:

$$L(\lambda) = \prod_{i=1}^{n} f(x_i; \lambda) = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)} e^{-n\lambda} = \frac{\lambda^{n\bar{x}}}{\prod_{i=1}^{n}(x_i!)}$$

For å maksimere rimelighetsfunksjonen finner vi først den deriverte av rimelighetsfunksjonen:

$$l(\lambda) = \ln L(\lambda) = n\bar{x}\ln\lambda - \sum_{i=1}^{n} \ln x_i! - n\lambda.$$

Vi deriverer med hensyn på λ og finner

$$l'(\lambda) = \frac{n\bar{x}}{\lambda} - n.$$

Vi ser at verdien av λ som gir null i dette uttrykket er $\hat{\lambda} = \bar{x}$. Den andrederiverte av l er $-n\bar{x}/\lambda^2$, som alltid er negativ. Derfor er $\hat{\lambda} = \bar{x}$ et maksimumspunkt av rimelighetsfunksjonen, og vi har funnet rimelighetsestimatoren. I dette tilfellet finner vi altså at rimelighetsestimatoren også er lik $\bar{X}$, som betyr at dette virker som et godt valg for en estimator.

11.27

Vi har et tilfeldig utvalg $X_1, \ldots, X_n$ fra en ukjent fordeling, og ønsker å estimere fordelingens forventningsverdi $\mathrm{E}X = \mu$ ved en lineær forventningsrett estimator. Alle lineære estimatorer kan skrives på formen $\hat{\mu} = a_1X_1 + \ldots a_nX_n + c$. Vi ønsker at estimatoren skal være forventningsrett, som betyr at vi må ha

$$\mu = \mathrm{E}\hat{\mu} = \mathrm{E}[a_1X_1 + \ldots a_nX_n + c] = \sum_{i=1}^{n} a_i\mathrm{E}[X_i] + c = \sum_{i=1} a_i\mu + c,$$

siden $\mathrm{E}X_i = \mu$ for $i = 1, \ldots, n$. For at denne likningen skal holde ser vi at vi må ha $c = 0$ og $\sum_{i=1}^{n} a_i = 1$. Vi kan skrive om dette kravet som at $a_n = 1 - \sum_{i=1}^{n-1} a_i$.

Vi regner nå ut variansen til estimatoren:

$$\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(a_1X_1 + \ldots a_nX_nc) = \sum_{i=1}^{n} a_i^2\mathrm{Var}(X_i) = \sum_{i=1}^{n} a_i^2\sigma^2.$$

Vi ønsker at denne skal være så liten som mulig, slik at $\sum_{i=1}^{n} a_i = 1$. Vi erstatter derfor a_n med $1 - \sum_{i=1}^{n-1}$ og partiellderiverer mhp. a_j-ene.

$$\frac{\mathrm{Var}(\hat{\mu})}{\sigma^2} = \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n-1} a_i^2 + \left(1 - \sum_{i=1}^{n-1} a_i\right)^2.$$

Dette gir

$$\frac{\partial}{\partial a_i}\frac{\mathrm{Var}(\hat{\mu})}{\sigma^2} = 2a_i + 2\left(1-\sum_{j=1}^{n-1} a_j\right)^2 \cdot (-1), \quad i = 1, \ldots, n-1.$$

Vi setter de partiellderiverte lik null og får

$$\frac{\partial}{\partial a_i}\frac{\mathrm{Var}(\hat{\mu})}{\sigma^2} = 0 \implies a_i = \left(1-\sum_{j=1}^{n-1} a_j\right)^2 = a_n, \quad i = 1, \ldots, n-1.$$

Dette betyr at variansen er minst når alle a_j-er er like. Siden summen av alle a_j-er må være lik 1 får vi $a_j = 1/n$ for alle $j = 1, \ldots, n$.

Merk: Dette minimeringsproblemet kunne vært enklere løst ved å bruke Lagrange-multiplikatoren, for de som kjenner til denne. For å vite at vi har funnet et minimumspunkt og ikke et maksimumspunkt må vi egentlig også regne ut determinanten til Jacobian-matrisen, men vi dropper det her, da dette er en såpass enkel funksjon at vi kan skjønne at $a_j = 1/n$ må være et minimumspunkt.

11.29

Vi har $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. For å vise at S^2 er en forventningrett estimator for σ^2 må vi regne ut forventningsverdien. Vi får

$$\begin{aligned}
\mathrm{E}S^2 &= \mathrm{E}\left[\frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2\right] \\
&= \frac{1}{n-1}\sum_{i=1}^n \mathrm{E}X_i^2 - 2\mathrm{E}X_i\bar{X} + \mathrm{E}\bar{X}^2 \\
&= \frac{n}{n-1}\left(\mathrm{E}X_1^2 - 2\mathrm{E}X_1\bar{X} + \mathrm{E}\bar{X}^2\right),
\end{aligned}$$

hvor det siste steget kan gjøres hvis vi antar at alle X-ene har lik fordeling. Vi finner hver av de tre forventningsverdiene fra likningen ovenfor. Fra definisjonen av varians har vi

$$\mathrm{Var}(X) = \mathrm{E}X^2 - (\mathrm{E}X)^2 \implies \mathrm{E}X^2 = \mathrm{Var}(X) + (\mathrm{E}X)^2 = \sigma^2 + \mu^2.$$

For det andre leddet kan vi skrive

$$\mathrm{E}X_1\bar{X} = \mathrm{E}\left[X_1 \frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n}\mathrm{E}\left[X_1^2 + \sum_{i=2}^n X_1 X_i\right].$$

Hvis vi antar at alle X-ene er uavhengige av hverandre, eller i alle fall har kovarians lik 0, kan vi forenkle dette uttrykket til

$$\mathrm{E}X_1\bar{X} = \frac{1}{n}\left(\mathrm{E}X_1^2 + (n-1)\mathrm{E}[X_1]^2\right) = \frac{1}{n}\left(\mu^2 + \sigma^2 + (n-1)\mu^2\right).$$

Det siste leddet kan skrives som

$$\mathrm{E}\bar{X}^2 = \frac{1}{n^2}\mathrm{E}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} X_i X_j\right] = \frac{1}{n^2}\left(n\mathrm{E}X_1^2 + n(n-1)\mathrm{E}[X_1]^2\right) = \frac{1}{n^2}\left(n(\mu^2+\sigma^2) + n(n-1)\mu^2\right).$$

Dette fører til uttrykket

$$\mathrm{E}S^2 = \frac{n}{n-1}\left(\sigma^2 + \mu^2 - \frac{2}{n}\left(\mu^2+\sigma^2+(n-1)\mu^2\right) + \frac{1}{n}\left(\mu^2+\sigma^2+(n-1)\mu^2\right)\right) = \frac{n}{n-1}(\sigma^2(1-\frac{1}{n})) = \sigma^2.$$

Da har vi altså vist at S^2 er en forventningsrett estimator for σ^2 dersom alle X-ene er identisk fordelt og ukorrelerte.

11.31

- Observatorer er definert på side 52 i notatet. At W er en observator for Ω betyr at W er en funksjon fra Ω til et annet utfallsrom Ω_W slik at $(W \in A)$ er en hendelse i Ω for enhver hendelse $A \in \Omega_W$.
- Dersom W er en estimator betyr det at W er en "gjetning" for en parameter vi ønsker å estimere. Mer spesifikt vil det si at W er en observator for Ω med verdier i en mengde Ω_W, som er slik at parametermengden Ω_Γ er inneholdt i Ω_W.
- Dersom τ er en estimand betyr det at τ er en parameter som vi ønsker å estimere. Mer spesifikt så er en estimand en funksjon av modellparametrene, $\tau : \Omega_\theta \to \Omega_\tau$.
- Dersom W er en estimator for τ betyr det at vi bruker verdien av W for å estimere parameteren τ, og at $\Omega_\tau \subseteq \Omega_W$.
- Dersom W er en estimator for τ basert på observatoren $(X_1, \ldots, X_n)$ betyr det at W er en funksjon av observatoren $(X_1, \ldots, X_n)$, altså en statistikk, som brukes for å estimere parameteren τ.

11.43

Her er det to relevante statistiske modeller som kan brukes. Den ene er en binomisk fordeling, og den andre er en Poisson-prosess. For $n = 100000$ og data av størrelsesorden ~ 1 vil den binomiske modellen være tilnærmet lik en Poisson-prosess. Vi beskriver derfor forekomsten av fingeravtrykk ved en Poisson-prosess med intensitet-parameter λ. Vi sier at antall spesielle fingeravtrykk N funnet i et arkiv med 100 000 fingeravtrykk har sannsynlighetsfordeling

$$P(N = n) = \frac{\lambda^n}{n!}e^{-\lambda}.$$

Det er kun én parameter å estimere, så ved å bruke momentmetoden trenger vi kun å regne ut det første momentet. Da setter vi $\mathrm{E}N = \bar{n}$ som gir oss $\lambda \approx 2.167$. Denne estimatoren er forventningsrett, siden $\mathrm{E}\bar{N} = \lambda$.

11.47

Vi kan anta at sannsynligheten for at en tilfeldig kvinne sier ja er identisk for alle kvinner som blir spurt, og at studenten slutter å spørre nye kvinner etter at han har fått sitt første ja. Det betyr at hvis har inviterer f.eks 3 kvinner så har han først fått to nei, og deretter ett ja. Dette kan modelleres ved en geometrisk fordeling, som beskriver antall forsøk før den første suksessen. Antagelsen at alle kvinner som blir invitert sier ja eller nei med samme sannsynlighet kan virke ganske tvilsom, men er nødvendig for å kunne bruke en enkel geometrisk modell. Antall invitasjoner før en suksess, X vil da ha sannsynlighetsfordeling $P(X = x) = (1 - p)^{x-1}p$, hvor p er sannsynligheten for at en kvinne takker ja til invitasjonen.

- Utvalgets varians er $S^2 = 7.7$. Variansen til en geometrisk fordeling med parameter p er $\sigma^2 = (1 - p)/p^2$. Vi kan derfor estimere p ved å sette $S^2 = (1 - \hat{p})/\hat{p}^2$, som gir $\hat{p} \approx 0.3$ og $\hat{\mu} = 1/\hat{p} \approx 3.32$.

- Vi bruker momentmetoden og estimerer $\hat{p} = \bar{n}^{-1} = 5/24 \approx 0.2$. Dette gir $\hat{\mu} = 1/\hat{p} = 24/5 = 4.8.$ og $\hat{\sigma}^2 = (1 - \hat{p})/\hat{p}^2 = 18.24$.

- Vi bruker rimelighetsmetoden. Rimelighetsfunksjonen er lik

  $$L(p) = \prod_{i=1}^{n} p(1 - p)^{x_i - 1} = p^n(1 - p)^{n(\bar{x}-1)}.$$

  Vi maksimerer rimelighetsfunksjonen ved å derivere logaritmen og sette lik null.

  $$l = \ln L(p) = n \ln p + n(\bar{x} - 1)\ln(1 - p)$$

  Den deriverte er lik

  $$l' = \frac{n}{p} - \frac{n(\bar{x} - 1)}{1 - p}.$$

  Vi finner punktet hvor den deriverte er lik null:

  $$l'(\hat{p}) = 0 \implies \hat{p} = \frac{1}{\bar{x}}.$$

  Rimelighetsestimatoren er altså lik estimatoren fra momentmetoden, og vi får de samme estimatene for forventningsmetoden og variansen.

11.49

Vi beregner rimelighetsestimatoren for $[\theta_1, \theta_2]$. Rimelighetsfunksjonen er lik

$$L(\theta) = \begin{cases} \left(\frac{1}{\theta_2 - \theta_1}\right)^n, & \theta_1 \leq \min\{x_1, \ldots, x_n\} \leq \max\{x_1, \ldots, x_n\} \leq \theta_2 \\ 0, & ellers. \end{cases}$$

Vi kan se at rimelighetsfunksjonen blir større jo mindre intervallet $[\theta_1, \theta_2]$ blir. Rimelighetsestimatoren for intervallet er derfor den som fører til det minste mulige intervallet, slik at ikke rimelighetsfunksjonen er lik null, noe som skjer hvis en av de observerte tidene er utenfor det estimerte intervallet. Da får vi $[\hat{\theta}_1, \hat{\theta}_2] = [\min\{x_1, \ldots, x_n\}, \max\{x_1, \ldots, x_n\}] = [5{:}14, 5{:}29]$. Rimelighetsestimatet for den forventede avgangstiden blir da $\hat{\mu} = (\hat{\theta}_1 + \hat{\theta}_2)/2 =$ 5:21:30.

11.50

Vi beregner forventningsverdien EY og variansen Var(Y):

$$\mathrm{E}Y = \int_0^1 (\theta+1) y^{\theta+1} dy = \frac{\theta+1}{\theta+2}.$$

$$\mathrm{E}Y^2 = \int_0^1 (\theta+1) y^{\theta+2} dy = \frac{\theta+1}{\theta+3}.$$

$$\mathrm{Var}(Y) = \mathrm{E}Y^2 - (\mathrm{E}Y)^2 = \frac{\theta+1}{\theta+3} - \frac{(\theta+1)^2}{(\theta+2)^2}.$$

- Momentestimatoren for θ løser likningen $\bar{y} = (\hat{\theta}_m + 1)/(\hat{\theta}_m + 2)$. Vi får $\hat{\theta}_m = (2\bar{y} - 1)/(1 - \bar{y}) = 0.5$. Momentestimatoren gir estimatene $\hat{\mu}_m = 0.6$ og $\hat{\sigma}^2_m \approx 0.0686$.

- Rimelighetsestimatet finnes ved å maksimere rimelighetsfunksjonen,

$$L(\theta) = (\theta+1)^n \left(\prod_{i=1}^n y_i \right)^\theta .$$

  Vi finner logaritmen til rimelighetsfunksjonen og deriverer for å finne maksimum.

$$l = \ln L = n \ln(\theta+1) + \theta \sum_{i=1}^n \ln y_i$$

$$l' = \frac{n}{\theta+1} + \sum_{i=1}^n \ln y_i$$

$$l'(\hat{\theta}_r) = 0 \implies \hat{\theta}_r) = -\frac{n}{\sum_{i=1}^n \ln y_i} - 1 \approx 0.7495.$$

  Nå som vi har rimelighetsestimatoren for θ kan vi også estimere $\hat{\mu}_r = (\hat{\theta}_r + 1)/(\hat{\theta}_r + 2) \approx 0.6363.$ og $\hat{\sigma}^2_r \approx 0.0617$. Vi ser at selv om forskjellen i estimatene for θ virker ganske stor, så fører dette ikke til en veldig stor forskjell i estimatene for forventningsverdien og variansen til Y.