

LØSNING: Eksamen 21. mai 2019

“MAT110 Statistikk 1”, vår 2019

Versjon 02

Oppgave 1: (telling av lakselus)

- a) Utfallsrommet Ω har et endelig antall utfall.
Derfor er utfallsrommet diskret.
- b) Hvert forsøk har 2 mulige utfall. Et eksperiment består av en sekvens med n forsøk.
Utfallsrommet Ω har dermed 2^n mulige utfall.¹

c) Eksperimentet:

$$l_1, l_2, l_3, \bar{l}_4 \quad (1)$$

betyr at:

- den tilfeldig valgte fisken fra merd 1 har **lakselus**
- den tilfeldig valgte fisken fra merd 2 har **lakselus**
- den tilfeldig valgte fisken fra merd 3 har **lakselus**
- den tilfeldig valgte fisken fra merd 4 har **ikke lakselus**

¹Med $n = 4$ får vi $2^4 = 16$, som stemmer med Ω oppgitt i oppgaven.

Med $p = 0.5$:

$$\underline{\underline{P(l_1, l_2, l_3, \bar{l}_4)}} = p^3(1-p)^1 \quad (2)$$

$$= 0.5^3 \cdot (1-0.5)^1 = 0.5^4 = \underline{\underline{0.0625}} \quad (3)$$

d) Ved å se på ligningen for begivenheten S som oppgitt i ligningen så innser vi at:

$$\begin{aligned} \underline{\underline{p(S)}} &= P(l_1, l_2, l_3, l_4) \\ &+ P(\bar{l}_1, l_2, l_3, l_4) + P(l_1, \bar{l}_2, l_3, l_4) \\ &+ P(l_1, l_2, \bar{l}_3, l_4) + P(l_1, l_2, l_3, \bar{l}_4) \\ &+ P(l_1, l_2, (\bar{l}_3, \bar{l}_4)) + P(l_1, \bar{l}_2, l_3, \bar{l}_4) + P(l_1, \bar{l}_2, \bar{l}_3, l_4) \\ &+ P(\bar{l}_1, \bar{l}_2, l_3, l_4) + P(\bar{l}_1, l_2, \bar{l}_3, l_4) + P(\bar{l}_1, l_2, l_3, \bar{l}_4) \end{aligned} \quad (4)$$

$$= \underline{\underline{p^4 + 4p^3(1-p)^1 + 6p^2(1-p)^2}} \quad (5)$$

- e) Oppgaven ber oss da finne den betingede sannsynligheten: ²

$$\boxed{P(A|S) = \frac{P(A \cap S)}{P(S)} \quad (6)}$$

hvor

$$P(A \cap S) = 6 \cdot 0.5^4 \quad (\text{ oppgitt i oppgaven }) \quad (7)$$

Fra oppgave **1d** kjenner vi $P(S)$, se lign.(5).
Med $p = 0.5$:

$$P(S) = 11 \cdot 0.5^4 \quad (8)$$

Den betingede sannsynligheten i lign.(6) blir da:

$$\underline{\underline{P(A|S)}} = \frac{P(A \cap S)}{P(S)} \quad (9)$$

$$= \frac{6 \cdot \cancel{0.5^4}}{11 \cdot \cancel{0.5^4}} = \frac{6}{\underline{\underline{11}}} \quad (= 0.5454) \quad (10)$$

■

²Det er dette som er litt vanskelig med dette oppgaven, å innse at det er den betingede sannsynligheten som oppgaven spør om.

Oppgave 2: (stokastiske variabler og forventningverdi - telling av lakselus)

a) Tolkning:

$$\begin{aligned} \underline{\underline{P(X = j)}} &= \text{sannsynligheten for at en tilfeldig valgt fisk} \\ &\text{fra et oppdrettsanlegg med lusetall mellom 0.4 og 0.6} \\ &\text{har } j \text{ antall lakselus} \end{aligned} \quad (11)$$

b) Sannsynlighetene $P(X = j)$ med $j = 0, 1, 2, 3$ utgjør en gyldig sannsynlighetsfordeling fordi den er normert til 1:

$$\underline{\underline{\sum_{j=0}^3 P(X = j)}} = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \quad (12)$$

$$= 0.72 + 0.13 + 0.08 + 0.07 = \underline{\underline{1}} \quad (13)$$

c) Forventningsverdien $E[X]$:

$$\underline{\underline{E[X]}} = \sum_{j=0}^3 x_j P(X = j) \quad (14)$$

$$= \sum_{j=0}^3 j P(X = j) \quad (15)$$

$$= 0 \cdot 0.72 + 1 \cdot 0.13 + 2 \cdot 0.08 + 3 \cdot 0.07 \quad (16)$$

$$= \underline{\underline{0.5}} \quad (17)$$

d) Variansen $Var[X]$:

$$\underline{\underline{Var[X]}} = \sum_{j=0}^3 \left(x_j - E[X] \right)^2 P(X = j) \quad (18)$$

$$= \sum_{j=0}^3 \left(j - E[X] \right)^2 P(X = j) \quad (19)$$

$$\begin{aligned} = & \left(0 - 0.5 \right)^2 \cdot 0.72 + \left(1 - 0.5 \right)^2 \cdot 0.12 \\ & + \left(2 - 0.5 \right)^2 \cdot 0.08 + \left(3 - 0.5 \right)^2 \cdot 0.07 \end{aligned} \quad (20)$$

$$= \underline{\underline{0.83}} \quad (21)$$

med tilhørende standardavvik $\sigma[X_1]$:

$$\underline{\underline{\sigma[X]}} = \sqrt{Var[X]} \quad (22)$$

$$= \sqrt{0.83} = \underline{\underline{0.911}} \quad (23)$$

- e) Sannsynligheten for at gjennomsnittet $\overline{X}_t$ ligger over lusegrensen:

$$P(\overline{X}_t > 0.5) \quad (24)$$

Siden $X_{kt} \stackrel{\text{i.i.d.}}{\sim} X \sim P(X = j)$ med 50 variabler, dvs. $k = 1, 2, \dots, 50$, så er, ifølge sentralgrensesetningen, gjennomsnittet

$$\overline{X}_t = \frac{1}{n} \sum_{k=1}^n X_{kt} \quad (25)$$

normalfordelt:

$$\overline{X}_t \sim N \left[E[\overline{X}_t], \sigma[\overline{X}_t] \right] \quad (26)$$

hvor

$$E[\overline{X}_t] = \overbrace{E[X]}^{\text{oppg. 2c}} = 0.5 \quad (27)$$

$$\sigma[\overline{X}_t] = \frac{\overbrace{\sigma[X]}^{\text{oppg. 2d}}}{\sqrt{n}} = \frac{0.911}{\sqrt{50}} \quad (28)$$

Standardiserer:

$$\underline{P(\overline{X}_t > 0.5)} \stackrel{\text{standardiser}}{=} P \left(\frac{\overline{X}_t - E[\overline{X}_t]}{\sigma[\overline{X}_t]} > \frac{0.5 - E[\overline{X}_t]}{\sigma[\overline{X}_t]} \right) \quad (29)$$

$$= P \left(\underbrace{\frac{\overline{X}_t - E[X]}{\sigma[X]/\sqrt{n}}}_{\equiv Z} > \frac{0.5 - E[X]}{\sigma[X]/\sqrt{n}} \right) \quad (30)$$

$$= P \left(Z > \frac{0.5 - 0.5}{0.911/\sqrt{50}} \right) \quad (31)$$

$$= \underline{P(Z > 0)} \quad (32)$$

Siden ulikhetstegnet er ”feil” vi sammenlignet med tabellen i formelsamlingen så må vi:

$$\underline{P(\bar{X}_t > 0.5)} = P(Z > 0) \quad (33)$$

$$= \underline{1 - P(Z < 0)} \quad (34)$$

Via tabelloppslag i formelsamlingen finner vi nå:

$$\underline{\underline{P(\bar{X}_t > 0.5)}} = 1 - \overbrace{P(Z < 0)}^{= 0.5} \quad (35)$$

$$= 1 - 0.5 = \underline{\underline{0.5}} \quad (36)$$

- f) Siden sannsynligheten for at $\bar{X}_t > 0.5$ er 50 %, så er det helt tilfeldig om gjennomsnittet av de 50 tilfeldig valgte fiskene en gitt uke t er innenfor lusegrensen eller ikke.

Kriteriet $\bar{X}_t > 0.5$ er derfor ikke et godt kriterium for å avgjøre om lusegresen er brutt eller ikke.

Nei, Mattilsynet bør derfor ikke konkludere om lusegrensen er brutt eller ikke basert på gjennomsnittet $\bar{X}_t > 0.5$.

■

Oppgave 3: (statistisk modell og estimatorer)

a) Siden $X \sim \text{Poi}[\lambda]$ så er:

$$E[X_i] = \lambda \quad (37)$$

for alle $i = 1, 2, \dots, N$ fiskene. Forventningen av estimatoren $\hat{\lambda}$ blir da:

$$\underline{\underline{E[\hat{\lambda}]}} = E[\overline{X}] \quad (38)$$

$$= E\left[\frac{1}{N} \sum_{i=1}^N X_i\right] \quad (39)$$

$$= \frac{1}{N} \left(\underbrace{E[X_1]}_{\lambda} + \underbrace{E[X_2]}_{\lambda} + \underbrace{E[X_3]}_{\lambda} + \dots + \underbrace{E[X_N]}_{\lambda} \right) \quad (40)$$

$$= \frac{N\lambda}{N} = \underline{\underline{\lambda}} \quad (41)$$

dvs. estimatoren $\hat{\lambda}$ er forventingsrett.

- b)** Siden vi skal regne ut et *asymptotisk* konfidensintervall, dvs. for tilfellet $N \rightarrow \infty$, for estimatoren λ gitt ved:

$$\hat{\lambda} = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i \quad (42)$$

må vi bruke **sentralgrensesetningen** til å konstruere konfidensintervallet.

Oppgitt i oppgaven:

$$X_1, X_2, X_3, \dots, X_N \stackrel{\text{i.i.d.}}{\sim} X \sim \text{Poi}[\lambda] \quad (43)$$

Også oppgitt i oppgaven:

$$E[\hat{\lambda}] = E[\bar{X}] = \lambda \quad (\text{se oppgave } \mathbf{3b}) \quad (44)$$

$$\sigma[\hat{\lambda}] = \sigma[\bar{X}] = \sqrt{\frac{\lambda}{N}} \quad (45)$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er standard normalfordelt:

$$\underline{Z} = \frac{\bar{X} - E[\bar{X}]}{\sigma[\bar{X}]} = \frac{\bar{X} - \lambda}{\sqrt{\frac{\lambda}{N}}} \stackrel{n \rightarrow \infty}{\sim} N[0, 1] \quad (46)$$

Sentralgrensesetningen gir at

$$\lim_{N \rightarrow \infty} P(z_{\alpha/2} \leq Z \leq z_{1-\alpha/2}) = 1 - \alpha \quad (47)$$

hvor (se figur 1)

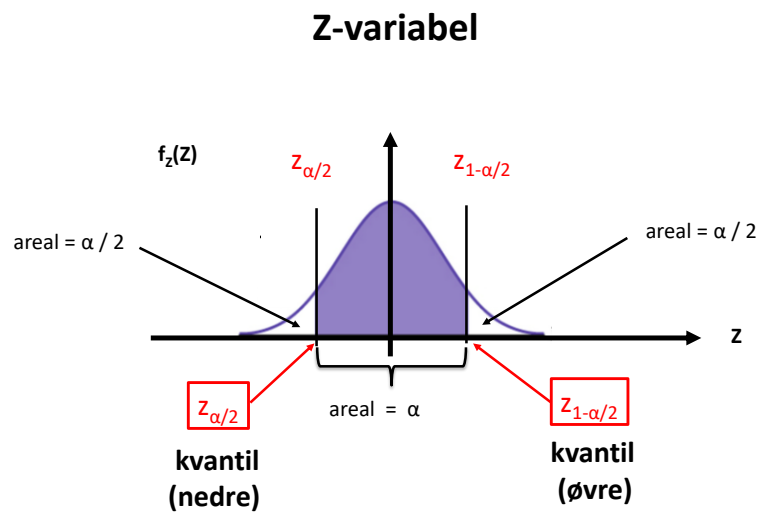
$$z_{\alpha/2} = \text{nedre kvantil til normalfordelingen} \quad (48)$$

$$z_{1-\alpha/2} = \text{øvre kvantil til normalfordelingen} \quad (49)$$

Vi må transformere lign.(47) over på formen

$$\lim_{N \rightarrow \infty} P_{\lambda}(LB_N^{\mu} \leq \lambda \leq UB_N^{\mu}) = 1 - \alpha \quad (50)$$

som er definisjonen av et $(1 - \alpha)$ 100 %-konfidensintervall for λ .



Figur 1: Kvantil.

Setter inn for Z fra lign.(46) inn i lign.(47):

$$\lim_{N \rightarrow \infty} P\left(z_{\alpha/2} \leq \frac{\overline{X} - \lambda}{\sqrt{\frac{\lambda}{N}}} \leq z_{1-\alpha/2}\right) = 1 - \alpha \quad (51)$$

Deretter multipliserer vi uttrykket i $P(\dots)$ med $\sqrt{\frac{\lambda}{N}}$:

$$\lim_{N \rightarrow \infty} P\left(z_{\alpha/2} \sqrt{\frac{\lambda}{N}} \leq \overline{X} - \lambda \leq z_{1-\alpha/2} \sqrt{\frac{\lambda}{N}}\right) = 1 - \alpha \quad (52)$$

Deretter multipliserer vi uttrykket i $P(\dots)$ med -1 hvor vi må huske at retningen på ulikhetegnene snus:

$$\lim_{N \rightarrow \infty} P\left(-z_{\alpha/2} \sqrt{\frac{\lambda}{N}} \geq -\overline{X} + \lambda \geq -z_{1-\alpha/2} \sqrt{\frac{\lambda}{N}}\right) = 1 - \alpha \quad (53)$$

Deretter summerer vi uttrykket i $P(\dots)$ med $\overline{X}$ for å sette λ alene i midten:

$$\lim_{N \rightarrow \infty} P\left(\overline{X} - z_{\alpha/2} \sqrt{\frac{\lambda}{N}} \geq \lambda \geq \overline{X} - z_{1-\alpha/2} \sqrt{\frac{\lambda}{N}}\right) = 1 - \alpha \quad (54)$$

Deretter snur vi høyre og venstre side i $P(\dots)$ hvor vi må huske at retningen på ulikhetegnene snus:

$$\lim_{N \rightarrow \infty} P\left(\overline{X} - z_{1-\alpha/2} \sqrt{\frac{\lambda}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2} \sqrt{\frac{\lambda}{N}}\right) = 1 - \alpha \quad (55)$$

Deretter erstatter vi λ med $\overline{X}$ på høyre og venstre side i $P(\cdots)$, dvs. setter $\lambda = \overline{X}$:

$$\lim_{N \rightarrow \infty} P\left(\overline{X} - z_{1-\alpha/2} \sqrt{\frac{\overline{X}}{N}} \leq \lambda \leq \overline{X} - z_{\alpha/2} \sqrt{\frac{\overline{X}}{N}}\right) = 1 - \alpha \quad (56)$$

Lign.(56) er et uttrykk på formen gitt i lign.(50) med

$$LB_N^\lambda = \overline{X} - z_{1-\alpha/2} \sqrt{\frac{\overline{X}}{N}} \quad (57)$$

$$UB_N^\lambda = \overline{X} - z_{\alpha/2} \sqrt{\frac{\overline{X}}{N}} \quad (58)$$

Vi har dermed vist at

$$\underline{\underline{[LB_N^\lambda, UB_N^\lambda] = \left[\overline{X} - z_{1-\alpha/2} \sqrt{\frac{\overline{X}}{N}}, \overline{X} - z_{\alpha/2} \sqrt{\frac{\overline{X}}{N}}\right] \text{ , q.e.d.}}} \quad (59)$$

er et $(1 - \alpha) 100\%$ asymptotisk konfidensintervall for λ .

c) Fra tabell 1 i oppgaven har vi at:

$$\bar{x} = 0.499 \quad , \quad s_x^2 = 0.894 \quad (60)$$

altså

$$\bar{x} < s_x^2 \quad (61)$$

dvs. det er en tendens til overspredning i datanene.

d) I oppgaven er det oppgitt at:

$$E[X] = \lambda \quad (62)$$

$$Var[X] = \lambda + \frac{\lambda^2}{\theta} \quad (63)$$

Da innser vi at:

$$E[X] < Var[X] \quad (\text{overspredning}) \quad (64)$$

siden $\lambda > 0$ og $\theta > 0$, dvs. positive. Derfor er negativ binomial fordelingen en passende modell dersom dataene har overspredning.

■

Oppgave 4: (krav som indikerer overskridelse av lusegrensa)

a) Kvantilen $b_{0.96}$ til $\bar{X}_t$ er definert ved:

$$P(\bar{X}_t > b_{0.96}) = 0.04 \quad (65)$$

Siden (lign.(66) er oppgitt i oppgaven)

$$X_{1t}, X_{2t}, \dots, X_{nt} \stackrel{i.i.d.}{\sim} X \sim \text{Poi}[\lambda] \quad (66)$$

hvor $n = 50$, dvs. variablene er Poisson-fordelt, så er:

$$E[\bar{X}_t] = \lambda \quad (67)$$

$$\sigma[\bar{X}_t] = \sqrt{\frac{\lambda}{n}} \quad (68)$$

Sentralgrensesetningen sier da at den skalerte variabelen Z er tilnærmet standard normalfordelt fordi $n = 50$:³

$$\underline{Z} = \frac{\bar{X} - \textcolor{red}{E}[\bar{X}_t]}{\textcolor{blue}{\sigma}[\bar{X}_t]} = \frac{\bar{X}_t - \textcolor{red}{\lambda}}{\textcolor{blue}{\sqrt{\frac{\lambda}{n}}}} \stackrel{n \rightarrow \infty}{\sim} N[0, 1] \quad (69)$$

³En *tommelfingerregel* er at det må være $\gtrsim 30$ forsøk for at sentralgrensesetningen skal gjelde.

Standardierer lign.(65):

$$P\left(\underbrace{\frac{\bar{X}_t - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{= Z} > \underbrace{\frac{b_{0.96} - \lambda}{\sqrt{\frac{\lambda}{n}}}}_{= z_{0.96}}\right) = 0.04 \quad (70)$$

hvor

$$z_{0.96} = \frac{b_{0.96} - \lambda}{\sqrt{\frac{\lambda}{n}}} \quad (71)$$

er 96 %-kvantilen til den standardiserte normalfordelingen.
Løser lign.(71) mhp. $b_{0.96}$:

$$\underline{b_{0.96} = \lambda + z_{0.96} \sqrt{\frac{\lambda}{n}}} \quad (72)$$

Tabeloppslag:

$$z_{0.96} = 1.75 \quad (73)$$

Numerisk verdi:

$$\underline{b_{0.96}} = \lambda + z_{0.96} \sqrt{\frac{\lambda}{n}} = 0.49 + 1.75 \cdot \sqrt{\frac{0.49}{50}} = \underline{0.663} \quad (74)$$

Siden

$$b_{0.96} = 0.663 < \bar{x}_t = 0.68 \quad (75)$$

konkluderer vi med at kravet er overskridet,
dvs. laksegrensa er overskredet ihht. krav 1.

b) Krav 2:

$$P(B) = P(\overline{X}_t > a \text{ og } \overline{X}_{t+1} > a \text{ og } \overline{X}_{t+2} > a \text{ og } \overline{X}_{t+3} > a) \quad (76)$$

Siden lusetellingene er uavhengige av hverandre fra uke til uke, har vi at:

$$\underline{\underline{P(B)}} = P\left(\overline{X}_t > a \text{ og } \overline{X}_{t+1} > a \text{ og } \overline{X}_{t+2} > a \text{ og } \overline{X}_{t+3} > a\right) \quad (77)$$

$$= P(\overline{X}_t > a) P(\overline{X}_{t+1} > a) P(\overline{X}_{t+2} > a) P(\overline{X}_{t+3} > a) \quad (78)$$

$$= 0.50^4 \quad (79)$$

$$= \underline{\underline{0.0625}} \quad (80)$$

■