TMA4245 Statistikk
Vår 2018

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalt øving 9
Løsningsskisse

Oppgave 1

a) Vi lar her Y være antall fugler som kolliderer med vindmølla i løpet av den gitte perioden på $t = 5$ uker. Siden Y er poissonfordelt med intensitet $\lambda = 1$ fugl/uke, har Y punktsannsynlighet

$$P(Y = y) = \frac{(\lambda t)^y}{y!} e^{-\lambda t} = \frac{5^y}{y!} e^{-5}, \quad y = 0, 1, 2, \ldots .$$

Figur 1 illustrerer denne punktsannsynligheten for verdier av y mellom 0 og 14. Sannsynligheten for at mer enn 10 fugler kolliderer i løpet av vedkommende periode er da

$$P(Y > 10) = 1 - P(Y \leq 10) = 1 - \sum_{y=0}^{10} \frac{5^y}{y!} e^{-5} = 1 - 0.986 = \underline{\underline{0.014}}.$$

Sannsynligheten for at færre enn fem fugler kolliderer er

$$P(Y < 5) = P(Y \leq 4) = 0.44,$$

og den betingede sannsynligheten for at ingen fugler kolliderer, gitt at færre enn fem gjør det, er

$$P(Y = 0 | Y < 5) = \frac{P(Y = 0 \cap Y < 5)}{P(Y < 5)} = \frac{P(Y = 0)}{P(Y < 5)} = \frac{e^{-5}}{0.44} = \underline{\underline{0.015}}.$$

b) Rimelighetsfunksjonen er sannsynligheten for å få $Y = 261$ med parameter $4 \cdot 52\lambda = 208\lambda$. Rimelighetsfunksjonen er en funksjon av parameteren λ.

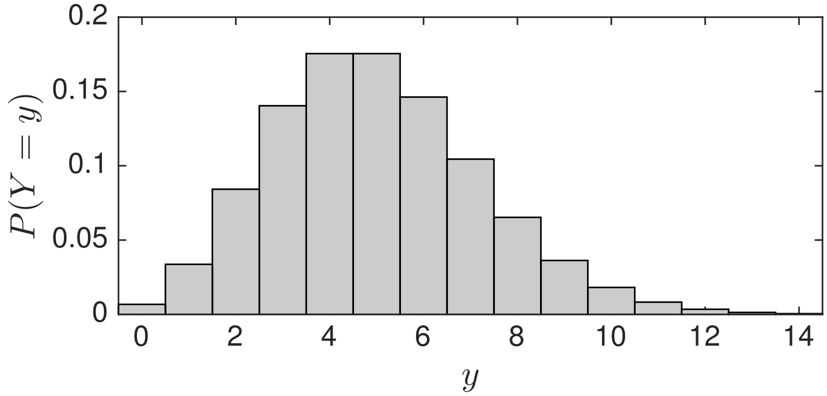
Vi tar logaritmen til rimelighetsfunksjonen for å forenkle utregningene,

$$l(\lambda) = \log P(Y = 261; \lambda) = 261 \log(208\lambda) - \log(261!) - 208\lambda$$

Vi finner maksimum ved å derivere med hensyn på λ,

$$l'(\lambda) = \frac{261}{\lambda} - 208 = 0 \Rightarrow \lambda = \frac{261}{208}.$$

Dette gir $\hat{\lambda} = 261/208 = 1.25$.



Figur 1: Punktsannsynligheten til den stokastiske variabelen Y, som er poissonfordelt med forventningsverdi $\mu = \lambda \cdot t = 5$.

c) Momentgenererende funksjon til en sum av to uavhengige variabler er produktet av de momentgenererende funksjonene for de to variablene. Momengenererende funksjon for en Poisson-fordelt variabel er

$$M_X(t) = \sum_{y=0}^{\infty} e^{ty} \frac{\lambda^y}{y!} e^{-\lambda} = \sum_{y=0}^{\infty} \frac{(e^t \lambda)^y}{y!} e^{-\lambda} = e^{\lambda(e^t - 1)} \sum_{y=0}^{\infty} \frac{(e^t \lambda)^y}{y!} e^{-e^t \lambda} = e^{\lambda(e^t-1)}$$

hvor den siste summen er 1 fordi det er en sum over en Poisson-fordelt variabel med parameter $e^t \lambda$. Da er

$$M_Z(t) = M_X(t) M_Y(t) = e^{\lambda(e^t-1)} e^{\nu(e^t-1)} = e^{(\lambda+\nu)(e^t-1)}$$

Vi kjenner igjen formen på funksjonen som en momentgenererende funksjon. Dette er momentgenererende funksjon til en Poisson-fordelt tilfeldig variabel med parameter $\lambda + \nu$.

Oppgave 2

a) $S = \sum_{i=1}^{4} X_i$. i) X_i er enten suksess, 1, om ras skjer, eller ikke-suksess, 0, om ras ikke skjer. ii) Uavhengige X_i-er. iii) Konstant suksess sannsynlighet $p = 0.15$. S er binomisk fordelt. $P(S = 0) = (1 - p)^4 = 0.52$

$P(S = 1|S \geq 1) = P(S = 1)/P(S \geq 1) = p(1 - p)^3 4/(1 - 0.52) = 0.77$.

$P(S > 1|S \geq 1) = 1 - P(S = 1|S \geq 1) = 1 - 0.71 = 0.23$

b) Z er kostnad, $Z = 40$ mill ved $X = 1$, dvs 'RAS'. $Z = 0$ ved $X = 0$, dvs 'IKKE RAS'. $P(Z = 40) = p = 0.15$. $P(Z = 0) = 1 - p = 0.85$

$$E(Z) = 40 \cdot 0.15 + 0 \cdot 0.85 = 40 \cdot 0.15 = 6$$

$$Var(Z) = E(Z^2) - E(Z)^2 = 40^2 * 0.15 - 6^2 = 14.3^2 = 204$$

$Std(Z) = \sqrt{Var(Z)} = 14.3$ millioner

Strategi A har forventet kostnad (6 millioner), mindre enn 7 millioner som man får ved strategi B. Ved kun å se på forventet verdi, vil vi velge A. Usikkerheten i kostnad er derimot stor, og hvis man ikke liker risiko om uforutsett utgift på 40 millioner, vil man velge B.

Enten er kostnad 7 millioner, dersom grundig undersøkelse svarer 'RAS'. Dette skjer med sannsynlighet 0.15. Eller er kostnad 0, dersom grundig undersøkelse svarer 'IKKE RAS'. Dette skjer med sannsynlighet 0.85. I tillegg kommer en kostnad til ekspertene på 5 millioner.

$$E(X) = 5 + (0.15 \cdot 7 + 0.85 \cdot 0) = 6.05$$

Forventet kostnad er 6.050.000 $>$ 6 mill. Dersom man bruker forventet kostnad som beslutningsgrunnlag, bør den grundige undersøkelsen ikke gjennomføres. Undersøkelsen har derimot mindre forventet kostnad enn stategi B, så hvis du har valgt strategi B over, er det kanskje igjen smart å gjennomføre undersøkelsen. Underøkelsen gir utfallsrom på kostnad: $\{5, 5+7=12\}$.

c) Indikatoren I_i er enten 0 (ved feil uttalelse, dvs $Y_i \neq X_i$) eller 1 (ved riktig uttalelse, dvs $Y_i = X_i$). Sannsynligheten for rett uttalelse, gitt sannheten X_i er:
$P(I_i = 1) = P(Y_i = X_i | X_i) = \gamma$. Ikke-suksess er $I_i = 0$, som skjer 5 ganger, mens suksess er $I_i = 1$ som skjer 10 ganger. Rimelighetsfunksjonen (likeilhood) er

$$L(\gamma) = \prod_{i=1}^{15} P(Y_i = y_i | X_i = x_i) = [\gamma^5 (1-\gamma)^{7-5}][\gamma^5 (1-\gamma)^{8-5}] = \gamma^{\sum_{i=1}^{15} I_i} (1-\gamma)^{15 - \sum_{i=1}^{15} I_i} \tag{2.1}$$

Log-likelihood er

$$l(\gamma) = \ln L(\gamma) = \sum_{i=1}^{15} I_i \ln \gamma + (15 - \sum_{i=1}^{15} I_i) \ln(1-\gamma)$$

Vi deriverer log likelihood og får: $l'(\hat{\gamma}) = \sum_{i=1}^{15} I_i / \hat{\gamma} - (15 - \sum_{i=1}^{15} I_i)/(1-\hat{\gamma}) = 0$. Løsningen er $\hat{\gamma} = \sum_{i=1}^{15} I_i / 15$. Så forslaget er SME.

Det er også mulig å løse oppgaven ved å tenke at $W = \sum_{i=1}^{15} I_i$ er binomisk fordelt, med parameter 15 og γ. Likelihood blir da

$$L(\gamma) = \binom{15}{w} \gamma^w (1-\gamma)^{15-w} \tag{2.2}$$

med samme løsning som over.

Innsetting: $\hat{\gamma} = \sum_{i=1}^{15} I_i / 15 = 10/15 = 2/3 = 0.67$.

d)

$$E(\hat{\gamma}) = \frac{\sum_i E(I_i)}{15} = 15\gamma/15 = \gamma$$

$$Var(\hat{\gamma}) = \frac{\sum_i Var(I_i)}{15^2} = 15(1-\gamma)\gamma/15^2 = (1-\gamma)\gamma/15$$

her er $E(I_i) = \gamma \cdot 1 + (1-\gamma) \cdot 0 = \gamma$, og $Var(I_i) = E(I_i^2) - E(I_i)^2 = \gamma - \gamma^2 = (1-\gamma)\gamma$.

e) Loven om total sannsynlighet:

$$P(Y=1) = P(Y=1|X=1)P(X=1) + P(Y=1|X=0)P(X=0)$$

Vi får: $P(Y=1) = 0.66 \cdot 0.15 + 0.33 \cdot 0.85 = 0.38$ $P(Y=0) = 1 - P(Y=1) = 0.62$
Dette gjør det lettere å bruke Bayes formel

$$P(X=1|Y=0) = \frac{P(Y=0|X=1)P(X=1)}{0.62} = 0.33 \cdot 0.15/0.62 = 0.08$$
$$P(X=0|Y=0) = 1 - 0.08 = 0.92$$

$$P(X=1|Y=1) = \frac{P(Y=1|X=1)P(X=1)}{0.38} = 0.66 \cdot 0.15/0.38 = 0.26$$
$$P(X=0|Y=1) = 1 - 0.26 = 0.74$$

Beslutningen blir billigste løsning. Dersom $E(Z|Y=y) < 7$ mill, velges strategi A. Dersom $E(Z|Y=y) > 7$ mill, velges strategi B.

$$E(Z|Y=0) = 40 \cdot 0.08 + 0 \cdot 0.92 = 3.2 < 7$$

$$E(Z|Y=1) = 40 \cdot 0.26 + 0 \cdot 0.74 = 10.4 > 7$$

$$C = 1 + E(Z|Y=0)P(Y=0) + 7P(Y=1) = 1 + 3.2 \cdot 0.62 + 7 \cdot 0.38 = 5.65$$

Siden forventet kostnad nå er mindre enn 6 mill, bør denne undersøkelsen gjennomføres.