

Løsningsforslag PrøveEksamen 2 i ISTx100y

Oppgave 1

Merk: Denne oppgaven er basert på Oppgave 3 på eksamen i TALM1005, kont 2015.

a) Generelt vet vi at i en diskret sannsynlighetsfordeling må summen av alle punkt-sannsynlighetene være lik 1. Her må derfor

$$\begin{aligned}\sum_{x=0}^3 a \left(\frac{1}{2}\right)^x &= 1 \\ \Downarrow \\ a \sum_{x=0}^3 \left(\frac{1}{2}\right)^x &= 1 \\ \Downarrow \\ a \left(\left(\frac{1}{2}\right)^0 + \left(\frac{1}{2}\right)^1 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^3 \right) &= 1 \\ \Downarrow \\ a &= \frac{1}{1.875} \approx 0.533\end{aligned}$$

b) Forventningsverdien til X er gitt ved

$$E(X) = \sum_{x=0}^3 x \cdot P(X=x) = 0 \cdot a \left(\frac{1}{2}\right)^0 + 1 \cdot a \left(\frac{1}{2}\right)^1 + 2 \cdot a \left(\frac{1}{2}\right)^2 + 3 \cdot a \left(\frac{1}{2}\right)^3 \approx 0.73$$

Dersom vi tar som utgangspunkt at forventningsverdien er 0.73, så er variansen til X er gitt ved

$$\begin{aligned}\text{Var}(X) &= \sum_{x=0}^3 (x - 0.73)^2 \cdot P(X=x) \\ &= (0 - 0.73)^2 \cdot a \left(\frac{1}{2}\right)^0 + (1 - 0.73)^2 \cdot a \left(\frac{1}{2}\right)^1 + (2 - 0.73)^2 \cdot a \left(\frac{1}{2}\right)^2 + (3 - 0.73)^2 \cdot a \left(\frac{1}{2}\right)^3 \\ &\approx 0.86\end{aligned}$$

c) Vi lar X betegne antall ganger studenten dupper av, og vi vet at $X \geq 1$. Vi regner dermed ut

$$P(X=3|X \geq 1) = \frac{P(X=3 \cap X \geq 1)}{P(X \geq 1)} = \frac{P(X=3)}{P(X \geq 1)} = \frac{P(X=3)}{1 - P(X=0)} = \frac{a \cdot 0.5^3}{1 - a \cdot 0.5^0} \approx 0.143$$

d) i. Fra sentralgrenseteoremet vet vi at en sum av 300 slike uavhengige stokastiske variabler med samme fordeling er tilnærmet normalfordelt. Videre er

$$E(X_{\text{tot}}) = \sum_{i=1}^{300} E(X_i) \approx 300 \cdot 0.73 = 219$$

og

$$\text{Var}(X_{\text{tot}}) = \sum_{i=1}^{300} \text{Var}(X_i) \approx 300 \cdot 0.86 = 258$$

Dermed er X_{tot} tilnærmet normalfordelt med forventning 219 og standardavvik $\sqrt{258}$.

ii. Ved å bruke normaltilnærming regner vi ut

$$\begin{aligned} P(200 < X_{\text{tot}} \leq 250) &\approx P\left(\frac{200 - 219}{\sqrt{258}} < Z \leq \frac{250 - 219}{\sqrt{258}}\right) \\ &\approx P(-1.18 < Z \leq 1.93) \\ &= \Phi(1.93) - \Phi(-1.18) \\ &= 0.9732 - 0.1190 = 0.8542 \end{aligned}$$

Oppgave 2

Merk: Denne oppgaven er basert på Oppgave 2 på eksamen i TALM1005, høst 2019.

a) Med en Poissonprosess antar vi at hendelser inntreffer uavhengig av hverandre, med konstant rate, og at det er svært usannsynlig at to hendelser inntreffer på samme tidspunkt. For denne oppgaven innebærer det at

- Fartsoverskredelser skjer uavhengig av hverandre. Det at én person har blitt tatt i å kjøre for fort, påvirker ikke oppførselen til andre sjåførere.
- Forventet antall hendelser innen for et tidsintervall av en viss lengde er konstant. Det vil for eksempel si at hvis fartskontrollen varer fra klokka 8 til klokka 14, så forventer vi like mange fartsoverskredelser mellom klokka 8 og 9 som mellom klokka 13 og 14. Det er ingen "rushtid" med høyere rate av hendelser.
- Det er veldig lite sannsynlig at to eller flere fartsoverskredelser skjer samtidig.

b) Vi lar X være en poissonfordelt stokastisk variabel som teller antall hendelser i løpet av et tidsintervall av lengde $t = 1$. Vi vet at det i gjennomsnitt skjer 4 hendelser i løpet av én time, slik at raten $\lambda = 4$. Da kan vi regne ut

$$P(X > 5) = 1 - P(X \leq 5) = 1 - 0.7851 = 0.2149$$

der vi har slått opp den kumulative sannsynligheten i tabell ($\lambda t = 4.0$).

c) i. T er eksponensialfordelt med parameter $\lambda = 4$. ii. Fra formelark leser vi av at

$$E(T) = \frac{1}{\lambda} = \frac{1}{4} = 0.2500,$$

og

$$SD(T) = \sqrt{\text{Var}(T)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = \frac{1}{4} = 0.2500$$

d) i. “Mer enn 10 minutter” før første hendelse inntreffer oversetter vi til “mer enn 1/6 time”. Ved å bruke formel for kumulativ fordeling i eksponensialfordelingen finner vi at

$$P(T > 1/6) = 1 - P(T \leq 1/6) = e^{-4 \cdot 1/6} = 0.5134171 \approx 0.513$$

ii. Vi kan umiddelbart observere at en Poissonprosess er hukommelsesløs, som vil si at sannsynligheten for at det går ti minutter før en hendelse inntreffer er 0.513, uansett hva som har skjedd frem til nå.

Alternativt kan vi regne ut dette ved å se på sannsynligheten for at det totalt sett går 20 minutter (2/6 time) før første fartsoverskridelse, betinget på at det ikke skjer noen fartsoverskridelser i løpet av de første ti minuttene (1/6 time):

$$P(T > 2/6 | T > 1/6) = \frac{P(T > 2/6)}{P(T > 1/6)} = \frac{e^{-4 \cdot 2/6}}{e^{-4 \cdot 1/6}} = e^{-4 \cdot (2/6 - 1/6)} = e^{-4 \cdot 1/6} \approx 0.513$$

Oppgave 3

a) Her skal vi gjøre en venstresidig test. Med signifikansnivå 0.05 skal vi da forkaste nullhypotesen ($\mu = 400$) til fordel for alternativhypotesen ($\mu < 400$) dersom vi observerer $z < -1.645$. Her regner vi ut $z = -2.2136$ så vi forkaster nullhypotesen. Vi konkluderer med at dataene gir grunnlag for å hevde at forventet vekt (gjennomsnittsvikt av veldig mange poser med poteter) er mindre enn 400 gram.

b) Teststyrken gir oss sannsynligheten for at vi faktisk forkaster nullhypotesen dersom alternativhypotesen er sann, under de forutsetninger som er spesifisert for forsøket. Det vil i dette tilfellet si at vi veier 10 tilfeldig valgte poser med poteter og gjennomfører en test med signifikansnivå 0.05. Når $\mu = 391$ gram slik som her (alternativhypotesen er sann) er det 88% sjanse for at våre 10 observasjoner (vekt av potetposer) gir tilstrekkelig bevis til å forkaste nullhypotesen.

Oppgave 4

a) I et slikt kryssplott kan vi sjekke den viktige antagelsen om at det er en lineær sammenhengen mellom x og y , og vi kan forsøke å se om variansen til $Y|x$ er den samme uansett verdi av x . Her virker det rimelig å anta en lineær sammenheng mellom BMI og fettprosent. Det virker også også rimelig å anta at variasjonen i fettprosent er den samme for alle verdier av BMI.

b) For BMI = 20 er predikert fettprosent 11.3, og for BMI = 30 er predikert fettprosent 30.3.

c) Ingen av personene i studien har BMI på 30 eller høyere, mens flere av personene i studien har BMI nær 20. Derfor er vi mer usikre på prediksjonen for personer med BMI = 30.