



Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4240 Statistikk
Høst 2017

Anbefalt øving 3

Disse oppgavene passer til pensum i tredje uke med forelesninger, og dreier seg om tilfeldige variable og sannsynlighetsfordelinger.

Oppgave 1 X er kontinuerlig fordelt med sannsynlighetstetthet

$$f_X(x) = \begin{cases} nx^{n-1} & \text{dersom } 0 < x \leq 1 \\ 0 & \text{ellers.} \end{cases}$$

- a) Finn fordelingsfunksjonen $F(x)$ til X . Hva er sannsynligheten for at X ligger mellom $\frac{1}{4}$ og $\frac{3}{4}$ når $n = 1$ og når $n = 2$? Finn medianen til X , dvs. den verdi av a som er slik at $P(X \leq a) = \frac{1}{2}$, når $n = 1$ og når $n = 2$. Finn forventningsverdien til X når $n = 1$ og når $n = 2$. Sammenlign og diskuter tallverdiene for forventningsverdi og median.

I datafilen `data30.txt` finnes det 30 observasjoner som antas å være trukket fra fordelingen $f_X(x)$ med $n = 2$.

- b) Bruk Matlab til å lage et histogram av observasjonene. Sammenlign histogrammet med sannsynlighetstettheten $f_X(x)$ ved å plotte $f_X(x)$ i samme figur.

Plott videre den empiriske kumulative fordelingsfunksjonen $\hat{F}(x)$ i samme figur som den kumulative fordelingen $F(x)$. Her kan man for eksempel bruke funksjonen `ecdf`.

Estimer sannsynligheten for at X ligger mellom $\frac{1}{4}$ og $\frac{3}{4}$ ved å finne andelen av observasjonene i `data30.txt` som er innenfor dette intervallet. Finn også medianen og forventningsverdien til observasjonene, og sammenlign resultatene med svarene i deloppgave a).

Vi vil nå undersøke hva som skjer når størrelsen på datasettet øker. Vi antar derfor nå at vi har 300 observasjoner som er trukket fra fordelingen $f_X(x)$ med $n = 2$. De nye observasjonene er lagret i datafilen `data300.txt`.

- c) Gjennta spørsmålene i oppgave b), denne gangen ved bruk av det utvidede datasettet `data300.txt`. Sammenlign og diskuter resultatene.

Oppgave 2 Eva går hjemmefra på et tilfeldig tidspunkt mellom klokka 7:30 og 8:00. La X betegne antall minutter etter 7:30 Eva går hjemmefra. Vi antar at X er en kontinuerlig variabel med sannsynlighetstetthet

$$g(x) = \begin{cases} \frac{1}{30} & \text{for } 0 < x < 30 \\ 0 & \text{ellers} \end{cases}$$

La Y betegne tiden Eva bruker til jobb etter at hun har gått hjemmefra. Vi antar at Eva bruker mellom 39 og 52 minutter på turen til jobb, det vil si Y er en kontinuerlig variabel med sannsynlighetstetthet

$$h(y) = \begin{cases} \frac{1}{13} & \text{for } 39 < y < 52 \\ 0 & \text{ellers} \end{cases}$$

Vi antar at klokkeslettet da Eva forlater hjemmet er uavhengig av hvor lang tid hun bruker på turen til jobb, det vil si at X og Y er uavhengige variabler.

Hva er sannsynligheten for at Eva er på jobb før klokken 8:30?

Fasit