TMA4245 Statistikk
Vår 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 Togforsinkelsen — Eksamen desember 2003, oppgave 1 av 3

I denne oppgaven kan du bruke uten å vise det at

$$\int_0^\infty x^r e^{-ax} dx = \frac{r!}{a^{r+1}} \text{ når } a > 0 \text{ og } r \text{ er et heltall} \geq 0$$

Vi betrakter ankomst- og oppholdstider for et bestemt lokaltog på en jernbanestasjon. Toget skal etter rutetabellen ankomme hver hverdag klokka 8:00, men kommer alltid etter dette tidspunktet.

La X (minutter) betegne togets forsinkelse på en tilfeldig valgt hverdag. Vi antar at X er en stokastisk variabel med sannsynlighetstetthet

$$g(x) = \begin{cases} kxe^{-2x} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases}$$

der $k > 0$ er en konstant.

a) Vis at $k = 4$.

Hva er den forventede forsinkelse for toget?

Vis at sannsynligheten for at toget er mer enn 2 minutter forsinket er tilnærmet lik 0.09.

b) La V være antall ganger i løpet av en måned (= 22 hverdager) at toget er mer enn 2 minutter forsinket. Foreslå en sannsynlighetsfordeling for V og sett opp de forutsetninger som ligger til grunn for denne.

Hva er sannsynligheten for at toget er mer enn 2 minutter forsinket minst 2 ganger i løpet av en måned (= 22 hverdager)?

Hva er (tilnærmet) sannsynligheten for at toget er mer enn 2 minutter forsinket mer enn 30 ganger i løpet av 220 hverdager?

La Y (minutter) være den tiden toget står på stasjonen. Oppholdstiden Y vil være influert av forsinkelsen, og vi antar at den betingede sannsynlighetstetthet $f(y|x)$ for Y, gitt at forsinkelsen X er lik x (> 0), er gitt ved

$$f(y|x) = \begin{cases} (x/2)\, e^{-xy/2} & \text{for } y > 0 \\ 0 & \text{for } y \leq 0 \end{cases}$$

c) Hvilken fordeling har oppholdstiden Y når det er gitt at forsinkelsen er 2 minutter?

Hva er forventet oppholdstid når forsinkelsen er 2 minutter?

Sett opp simultantettheten $f(x, y)$ for X og Y.

Finn sannsynlighetstettheten $h(y)$ for oppholdstiden Y.

Oppgave 3 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