Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

TMA4245 Statistikk
Vår 2015

Øving nummer 5, blokk I

Oppgave 1

En entreprenør har leid inn et transportfirma til å transportere masse bort fra en byggeplass. Det er to mulige veivalg fra byggeplassen til stedet der massen skal deponeres, gjennom eller utenom bykjernen. Av helse, miljø og sikkerhetshensyn velger entreprenøren at massen skal transporteres utenom bykjernen.

Når oppdraget er ferdig har transportfirmaet kjørt 1000 turer, og transportfirmaet informerer entreprenøren om at av de 1000 turene har 5 blitt kjørt gjennom bykjernen og 995 er blitt kjørt utenom bykjernen.

I løpet av transportperioden har entreprenøren ved 5 tilfeldig valgte transporter sjekket om transporten har skjedd gjennom bykjernen. La X være antall ganger, av de 5 transportene som ble sjekket, entreprenøren finner at massen er blitt transportert gjennom bykjernen.

Hvilken fordeling har X? Begrunn svaret.

Hvilken verdi av X har høyest punktsannsynlighet?

Nå viser det seg at entreprenøren fant at i 5 av de 5 tilfellene han sjekket så ble massen transportert gjennom bykjernen. Hva er sannsynligheten for dette, dvs. $P(X = 5)$?

Oppgave 2

Vi ser på dødsfall om natten ved sykehjemmet "Aftensol". Ved sykehjemmet er det tre sykepleiere i rene nattevaktstillinger, Anne, Bernt og Cecilie. Hver natt er en av dem på vakt gjennom hele natten, og det er da ingen andre ansatte tilstede ved hjemmet. Anne jobber i 100% nattevaktstilling, mens Bernt og Cecilie jobber i 50% nattevaktstillinger.

Vi ser på en tilfeldig valgt natt og definerer følgende hendelser:

A = Anne er på vakt,
B = Bernt er på vakt,
C = Cecilie er på vakt,
D = det skjer et dødsfall.

Anta at alle dødsfall er naturlige. Det er da rimelig å gå ut fra at sannsynligheten for dødsfall er den samme uansett hvilken sykepleier som er på vakt, dvs. at $P(D|A) = P(D|B) = P(D|C)$. Anta at den felles verdi for disse er 0.06.

a) Tegn de 4 hendelsene definert ovenfor i et venndiagram.

Hva er sannsynlighetene $P(A), P(B)$ og $P(C)$?

Finn $P(D)$. Er hendelsene D og C uavhengige? Begrunn svaret.

I den siste tiden har det vært 10 dødsfall om natten ved sykehjemmet, og hele 7 av disse har skjedd når Cecilie har vært på vakt. Det er derfor satt igang etterforskning for eventuelt å avdekke om Cecilie har noe med dødsfallene å gjøre.

Anta i det følgende at alle dødsfallene er naturlige, og at de har skjedd på forskjellige netter.

La X være en stokastisk variabel som beskriver antall av $n = 10$ naturlige dødsfall som skjer på Cecilies vakter.

b) Forklar hvorfor det kan antas at X er binomisk fordelt med $n = 10$ og $p = 0.25$. (Det er ikke tilstrekkelig å skrive opp de generelle forutsetningene for en binomisk fordeling, betingelsene må relateres direkte til situasjonen som er beskrevet.)

Hva er sannsynligheten for at 7 eller flere av 10 dødsfall om natten skjer på Cecilies vakter?

La oss tenke oss at det rundt om på sykehjem i Norge jobber 300 andre sykepleiere i tilsvarende stilling som Cecilie. Hva er sannsynligheten for at minst en av de 300 sykepleierne opplever at 7 eller flere av 10 naturlige dødsfall skjer på sine vakter?

Gir svarene i dette punktet grunn til å styrke mistanken mot Cecilie? Begrunn svaret.

Oppgave 3

På italiensk fjernsyn vises det et søndagsshow med Fabrizio Frizzi som programleder, der Frizzi sitter ved siden av en stor safe som inneholder 250 000 000 lire. Safen har en hemmelig firesifret kode. Seerne ringer inn og foreslår koder, og den første som gjetter riktig, får pengene som safen inneholder. Hvert av sifrene i den firesifrede koden er altså et av tallene $0, 1, 2, 3, 4, 5, 6, 7, 8$ eller 9. Anta foreløpig at ingen av seerne er i stand til å huske koder som er foreslått av tidligere innringere og forkastet, slik at hvert gjett kan betraktes som et tilfeldig forslag av en firesifret kode.

a) Hva er sannsynligheten for at en tilfeldig innringer gjetter riktig kode?

La X være antall innringinger til (og med) den personen som vinner pengene. Gjør kort rede for hvorfor X er geometrisk fordelt.

Bestem sannsynligheten for at akkurat innringer nummer 300 er den første som gjetter riktig.

Anta så at det i begynnelsen av programmet opplyses at sifferet 7 forekommer nøyaktig to ganger i den hemmelige koden (det er i denne formen showet faktisk blir vist) og la m betegne antall koder innringerne nå kan velge blant.

b) Vis at $m = 486$ (dvs. forklar hvordan denne verdien fremkommer).

Dersom vi antar at Frizzi snakker så fort at han rekker å ta i mot to forslag per minutt, hva blir nå forventet tid til noen vinner de 250 000 000 lire?

Anta så at seerne skriver ned alle koder som blir foreslått slik at ingen innringere foreslår en kode som tidligere er blitt foreslått. La Y betegne antall innringinger til førstemann gjetter riktig kode i dette tilfellet. Anta fremdeles at det opplyses at sifferet 7 opptrer nøyaktig to ganger i koden.

c) Bestem svarene på følgende spørsmål uttrykt som funksjon av m:

Hva er utfallsrommet til Y?

Utled sannsynlighetsfordelingen til Y.

Hva blir nå forventet tid til noen vinner de 250 000 000 lire (hvis Frizzi fremdeles rekker å ta i mot to forslag per minutt)?

Oppgave 4

En produsent av et kjent frokostblandingsprodukt ønsker å profilere det høye rosininnholdet i frokostblandingen. Produsenten garanterer at hver porsjon (av en bestemt størrelse) en person spiser av frokostblandingen skal inneholde minst *en* rosin med en (høy) sannsynlighet β. En pakke med frokostblanding har innhold til m porsjoner og inneholder n rosiner som vi antar er uniformt fordelt i pakken. La X_1, X_2, ..., X_m være det antall rosiner en person får i porsjon nummer $1, 2, \ldots, m$.

a) Hva er fordelingen til X_1? (Begrunn svaret.) Hva er fordelingen til X_2 gitt at $X_1 = x_1$? Hvor mange rosiner n må det minst tilsettes for at det er minst en sannsynlighet på β for at den første porsjonen skal inneholde minst en rosin?

b) Er X_1 og X_2 uavhengige? (Begrunn svaret.) Hva er fortegnet på korrelasjonskoeffisienten mellom X_1 og X_2? (Begrunn svaret.) Finn simultanfordelingen til X_1 og X_2.

Produsenten ønsker å undersøke hvor mange av de m porsjonene av frokostblandingen som er uten rosiner. La W betegne dette antall. For å forenkle beregningen av $\mathrm{E}(W)$ og $\mathrm{Var}(W)$, uttrykker vi W ved hjelp av tellevariable $Y_1, \ldots, Y_m$,

$$W = \sum_{i=1}^{m} Y_i, \qquad \text{der} \quad Y_i = \begin{cases} 1, & X_i = 0, \\ 0, & X_i > 0. \end{cases}$$

c) Finn $\mathrm{E}(Y_1)$ og $\mathrm{E}(W)$. Finn $\mathrm{Var}(Y_1)$, $\mathrm{Cov}(Y_1, Y_2)$ og $\mathrm{Var}(W)$.

Fasit

1. $P(X = 5) = 1.21 \cdot 10^{-13}$

2. **a**) P(D)=0.06 **b**) 0.004, 0.7

3. **a**) $1/10000, 9.7 \cdot 10^{-5}$ **b**) 4 timer 3 minutter **c**) 121 minutter 45 sekunder

4. **a**) $n \geq \frac{\ln(1-\beta)}{\ln(1-1/m)}$ **c**) $\mathrm{E}(Y_1) = (1-1/m)^n$, $\mathrm{E}(W) = m \cdot (1-1/m)^n$, $\mathrm{Var}(Y_1) = (1-1/m)^n - (1-1/m)^{2n}$, $\mathrm{Cov}(Y_1, Y_2) = \left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}$, $\mathrm{Var}(W) = m \cdot \left[\left(\frac{m-1}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right] + m(m-1) \cdot \left[\left(\frac{m-2}{m}\right)^n - \left(\frac{m-1}{m}\right)^{2n}\right]$