TMA4245 Statistikk
Vår 2013

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Øving nummer 4, blokk I
Løsningsskisse

Oppgave 1

X og Y er uavhengige stokastiske variabler med med sannsynlighetstetthet

$$g(x) = \begin{cases} \frac{8}{x^3}, & x > 2 \\ 0, & \text{ellers,} \end{cases}$$

og

$$h(y) = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{ellers,} \end{cases} .$$

Skal beregne forventningsverdien til $Z = X\,Y$. Da X og Y er uavhengige stokastiske variabler kan vi bruke teorem 4.8, slik at

$$E(Z) = E(X\,Y) = E(X)\,E(Y).$$

Beregner

$$E(X) = \int_{-\infty}^{\infty} x\,g(x)\,dx = \int_{2}^{\infty} x\,\frac{8}{x^3}\,dx = 4$$

$$E(Y) = \int_{-\infty}^{\infty} y\,h(y)\,dy = \int_{0}^{1} y\,2y\,dy = \frac{2}{3}$$

som gir

$$E(Z) = E(X)\,E(Y) = 4 \times \frac{2}{3} = \frac{8}{3}.$$

Oppgave 2

a) Hvis vi tenker oss at alle soldater i en k gruppe blir kontrollert, vil vi ha k forsøk, alle uavhengige, og samme sannsynlighet p for smitte i hvert forsøk. Da er antall smittede binomisk fordelt med parametre k og p. Sannsynligheten for positiv reaksjon i blodprøveblandingen er da

$$1 - P(\text{ingen smittede blant de } k) = 1 - (1-p)^k.$$

b)

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Sett A = "19 Haugen positiv" og B = "k-gruppe positiv". Da er

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{p}{1-(1-p)^k}.$$

c) Den tilfeldige variable Y = "antall grupper som må analyseres på nytt" er binomisk fordelt med forventning $m[1-(1-p)^k]$. Antall analyser X er lik $m + kY$. Dermed får vi

$$\begin{aligned} E(X) &= E(m+kY) = m + kE(Y) \\ &= m + km[(1-(1-p)^k] = mk\left(1 + \frac{1}{k} - (1-p)^k =\right). \end{aligned}$$

Vi må ha at $E(X) < mk$ for at metoden skal lønne seg i det lange løp. Det gir

$$4m\left(1 + \frac{1}{4} - (1-p)^4\right) < 4m,$$

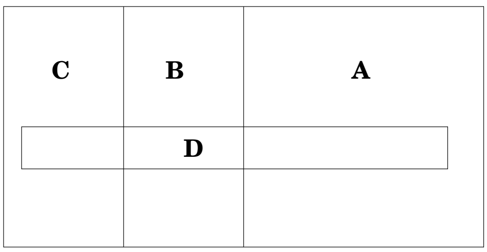
som er oppfylt hvis $p < 1 - \sqrt[4]{1/4} \approx 0{,}29$.

Oppgave 3

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.
Og antar at alle dødsfall skjer naturlig.

a) Venndiagram for de fire hendelsene:



Siden det bare er Anne, Bernt og Cecilie som jobber på sykehjemmet om natten vil hendelsene A, B og C utgjøre en partisjon av utfallsrommet, og vi må ha at $P(A) + P(B) + P(C) = 1$. Dette ser vi også av venndiagrammet. Siden Bernt og Cecilie jobber like ofte må $P(B) = P(C)$. Siden Anne jobber dobbelt så ofte som hver av Bernt og Cecilie må $P(A) = 2 \cdot P(B) = 2 \cdot P(C)$. Vi utrykker alt ved $P(B)$.

$$\begin{aligned} P(A) + P(B) + P(C) &= 1 \\ 2 \cdot P(B) + P(B) + P(B) &= 1 \\ P(B) &= 0.25 \end{aligned}$$

Dermed har vi at

$$\begin{aligned} P(A) &= 0.5 \\ P(B) &= 0.25 \\ P(C) &= 0.25 \end{aligned}$$

For å regne ut $P(D)$ kan vi bruke setningen om total sannsynlighet. Vi vet at A, B, C er en partisjon av utfallsrommet.

$$\begin{aligned} P(D) &= P(D \cap A) + P(D \cap B) + P(D \cap C) \\ &= P(D|A) \cdot P(A) + P(D|B) \cdot P(B) + P(D|C) \cdot P(C) \\ &= 0.06 \cdot (0.5 + 0.25 + 0.25) = \underline{\underline{0.06}} \end{aligned}$$

Definisjonen av uavhengighet sier at C og D er to uavhengige hendelser hvis og bare hvis $P(D|C) = P(D)$, dvs. at "tilleggsinformasjon ikke endrer bildet". Vi ser fra utregningene over at $P(D|C) = P(D) = 0.06$, og C og D er dermed uavhengige hendelser.

Intuitivt vil uavhengighet av C og D følge av antagelsen om naturlig død.

b) X er en stokastisk variabel som beskriver antall av $n = 10$ naturlige dødsfall som skjer på Cecilies vakter.

Betingelser for at X er binomisk fordelt:

- Vi ser på $n = 10$ dødsfall.
- For hver dødsfall sjekker vi om Cecilie var på vakt eller ikke.
- Sannsynligheten for at Cecilie er på vakt gitt at det har skjedd et dødsfall er $P(C|D) = P(C) = 0.25$, og denne sannsynligheten er det samme for alle de n dødsfallene.
- De n dødsfallene er uavhengige siden de er naturlige (og vi antar dermed at det ikke er snakk om smittsomme sykdommer eller epidemier).

Under disse 4 betingelsene er X=”antall naturlig dødsfall på Cecilies vakter” binomisk fordelt med parametere $n = 10$ og $p = 0.25$. Dermed er sannsynlighetsfordelingen til X gitt ved punktsannsynligheten $f(x)$,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, ..., n$$

Sannsynligheten for at 7 eller flere av 10 dødsfall om natten skjer på Cecilies vakter finner vi enklest ved tabelloppslag (s 13 i formelsamlingen),

$$P(X \geq 7) \quad = \quad = 1 - P(X \leq 6) = 1 - 0.996 = \underline{\underline{0.004}}$$

La Y være en stokastisk variabel som angir antall sykepleiere blant 300 sykepleiere som opplever flere enn 7 dødsfall på sine vakter av totalt 10 dødsfall. Y vil dermed være binomisk fordelt med $n = 300$ og $p = 0.004$.

Sannsynligheten for at minst en av de 300 sykepleierne opplever at 7 eller flere av 10 naturlige dødsfall skjer på sine vakter er gitt som $P(Y \geq 1)$.

$$\begin{aligned} P(Y \geq 1) &= 1 - P(Y = 0) = 1 - \binom{300}{0} 0.004^0 (1 - 0.004)^{300-0} \\ &= 1 - 0.996^{300} = 1 - 0.3 = \underline{\underline{0.7}} \end{aligned}$$

Selv om det er lite sannsynlig (bare 4 promille) at det skjer 7 av 10 naturlige dødsfall på Cecilies vakter, er det svært sannsynlig (70 prosent) at minst 7 av 10 dødsfall kan skje på vakten til en av sykepleierne i Norge som jobber i samme stillingstype som Cecilie. Disse observasjonene styrker ikke mistanken mot Cecilie.

Analogi: Hver uke er det (som regel) noen som får 7 rette i Lotto, selv om dette har en forsvinnende lav sannsynlighet for hver Lotto-spiller.

Oppgave 4 Antar $E(X_i) = 0, \quad i = 1, \ldots, n$

$$
\begin{aligned}
Var(\sum_{i=1}^{n} a_i X_i) &= E[(\sum_{i=1}^{n} a_i X_i - E(\sum_{i=1}^{n} a_i X_i))^2] \\
&= E[(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i E(X_i))^2] \\
&= E[(\sum_{i=1}^{n} a_i X_i)^2] \\
&= E[\sum_{i=1}^{n} a_i X_i * \sum_{i=1}^{n} a_i X_i] \\
&= E[\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j X_i X_j] \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j E(X_i X_j) \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j Cov(X_i, X_j) \\
&= \sum_{i=1}^{n} a_i^2 Var(X_i) + 2 \sum_{i=1}^{n} \sum_{j<i} a_i a_j Cov(X_i, X_j)
\end{aligned}
$$