TMA4245 Statistikk
Vår 2020

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

**Anbefalt øving 5**
**Løsningsskisse**

**Oppgave 1**

Vi lar $X$ være antall tankskip som ankommer havnen i løpet av en dag. Vi har fått oppgitt at $X \sim \text{poisson}(\lambda)$ med

$$\lambda = E(X) = 2.$$

Videre vet vi at havnen maksimalt kan betjene 3 tankskip per dag.

**a)** Da $X$ er poissonfordelt har vi at

$$P(X=x) = \frac{\lambda^x}{x!}e^{-\lambda} = \frac{2^x}{x!}e^{-2}, \quad x = 0, 1, 2, \dots.$$

Med innsatte verdier for $x$ har vi følgende punktsannsynligheter:

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P(X=x)$ | 0.1353 | 0.2707 | 0.2707 | 0.1804 | 0.0902 | 0.0361 |

Vi ser dermed at det er størst sannsynlighet for at det ankommer ett eller to tankskip en bestemt dag. Tankskip må dirigeres til andre havner dersom det ankommer mer enn tre tankskip en dag. Sannsynligheten for at ett eller flere tankskip må omdirigeres er dermed

$$\begin{aligned} P(\text{omdirigering}) &= P(X > 3) = 1 - P(X \le 3) \overset{\text{tabell}}{=} 1 - 0.8571 \\ &= \underline{\underline{0.1429}}. \end{aligned}$$

**b)** Vi lar nå $Y$ være antall skip som betjenes ved havnen en dag. Havnens begrensede kapasitet gjør at $Y \le 3$, slik at

$$P(Y=y) = \begin{cases} P(X=y) & \text{for } y = 0,1,2 \\ P(X \ge 3) = 1 - P(X \le 2) & \text{for } y = 3 \\ 0 & \text{for alle andre verdier av } y. \end{cases}$$

Forventet antall skip som betjenes en gitt dag blir dermed

$$\begin{aligned} E(Y) &= \sum_{y=0}^{3} y \cdot P(Y=y) \\ &= 0 \cdot P(X=0) + 1 \cdot P(X=1) + 2 \cdot P(X=2) + 3 \cdot (1 - P(X \le 2)) \\ &= 0.2707 + 2 \cdot 0.2707 + 3 \cdot (1 - 0.6767) \\ &= \underline{\underline{1.782}}. \end{aligned}$$

**c)** Vi lar $k$ være havnens kapasitet, altså maksimalt antall skip som kan betjenes på en dag. Vi ønsker å finne $k$ slik at $P(X \leq k) \geq 0.90$. Fra tabellen for poissonfordelingen har vi disse kumulative sannsynlighetene:

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $P(X \leq k)$ | 0.1353 | 0.4060 | 0.6767 | 0.8571 | 0.9474 | 0.9834 |

Dermed ser vi at

$$P(X \leq 3) < 0.90 \quad \text{mens} \quad P(X \leq 4) > 0.90.$$

Havnen trenger altså en kapasitet på $\underline{\underline{k = 4}}$ skip per dag for med minst 90% sannsynlighet å kunne betjene samtlige skip som ankommer en gitt dag.

## Oppgave 2

**a)** Vi har at $X$ er antall A feil på en tilfeldig valgt PC i løpet av $t$ år. Det er kjent fra oppgavebeskrivelsen at $X$ er poissonfordelt med parameter $0.25t$. Fra Tabeller og formler i statistikk er det kjent at forventningsverdien til en poissonfordeling med parameter $\lambda t$ er lik $\lambda t$. Forventet antall A feil på en tilfeldig valgt PC i løpet av $t = 4$ år blir derfor

$$E[X] = 0.25 \cdot 4 = \underline{\underline{1}}.$$

Hendelsen at antall A feil på en tilfeldig PC er 3 eller mer over 4 år er komplementet til at antall A feil er mindre enn 3. Vi har da

$$\begin{aligned}
P(X \geqslant 3) &= 1 - P(X < 3) \\
&= 1 - \sum_{i=0}^{2} P(X = i) \\
&= 1 - \sum_{i=0}^{2} \frac{(0.25 \cdot 4)^i e^{-0.25 \cdot 4}}{i!} \\
&= 1 - 0.9197 \\
&= \underline{\underline{0.0803}}
\end{aligned}$$

**b)** $N_X$ er det samlede antallet A feil på alle 10 PC-ene. Vi kan uttrykke $N_X$ som summen av antall feil på hver PC, $N_X = \sum_{i=1}^{10} X_i$.

Det er kjent at antall A feil i løpet av $t$ år på én PC er poissonfordelt med forventningsverdi $0.25 \cdot t$. Når vi undersøker summen av antall A feil på to forskjellige PC-er får vi to uavhengige stokastiske variabler som begge er poissonfordelte med forventningsverdi $0.25 \cdot t$. Vi vet at antall hendelser i en poissonprosess er uavhengige mellom alle ikke-overlappende tidsintervaller. Det er derfor mulig å tenke seg at summen av antall A feil på de to PC-ene kan beskrives ved hjelp av én poissonprosess over $2t$ år, istedenfor to poissonprosesser over $t$ år. Dette gjør at summen av antall A feil på to forskjellige PC-er er en stokastisk poissonfordelt variabel med forventningsverdi $0.25 \cdot 2t$. Fordelingen for det totale antallet A feil på alle 10 PC-ene blir derfor en stokastisk poissonfordelt variabel med forventningsverdi $0.25 \cdot 10t = 2.5t$. Samme argumentasjon kan brukes for å se at $N_Y$ er poissonfordelt med forventningsverdi $0.15 \cdot 10t = 1.5t$.

Punktsannsynlighetene til $N_X$ og $N_Y$ blir

$$\underline{\underline{f_{N_X}(n) = \frac{(2.5t)^n e^{-2.5t}}{n!} \text{ og } f_{N_Y}(n) = \frac{(1.5t)^n e^{-1.5t}}{n!}}}.$$

Kommentar: Dette resultatet kan også vises ved hjelp av momentgenererende funksjoner, som skal undervises senere i faget.

**c)** Det er kjent at feilene som oppstår på to forskjellige PC-er er uavhengige av hverandre. Sannsynligheten for at en gitt PC er feilfri i 4 år er lik

$$\begin{aligned} P(\text{feilfri}) &= P(X = 0 \cap Y = 0) \\ &= P(X = 0) \cdot P(Y = 0) \\ &= e^{-0.25 \cdot 4} \cdot e^{-0.15 \cdot 4} = e^{-1.60}. \end{aligned}$$

Denne sannsynligheten er identisk for alle 10 PC-ene. Dette betyr at vi kan betrakte problemet vårt som 10 uavhengige forsøk med lik sannsynlighet for suksess. Da er antall PC-er som er feilfrie i 4 år binomisk fordelt.

$$P(Z = z) = \binom{10}{z} p^z (1-p)^{10-z},$$

med $p = e^{-1.60}$.

Vi får da

$$\begin{aligned} P(Z > 2) &= 1 - P(Z \leqslant 2) \\ &= 1 - P(Z = 0) - P(Z = 1) - P(Z = 2) \\ &= 1 - 0.672 \\ &= \underline{\underline{0.328}}. \end{aligned}$$

$A'$ er hendelsen at det ikke skjer A feil på en PC i et visst tidsrom. Vi har da $P(A') = 1 - P(A)$. Det samme kan sies for $B'$. For at to stokastiske variabler $A'$ og $B'$ skal være uavhengige må vi ha at $P(A' \cap B') = P(A')P(B')$. Vi har at

$$P(A')P(B') = (1 - P(A))(1 - P(B)) = 1 - P(A) - P(B) + P(A)P(B).$$

Hendelsen at det verken blir A feil eller B feil er det motsatte av at det enten blir A feil eller B feil. Vi får derfor

$$P(A' \cap B') = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B)).$$

Det er kjent at $A$ og $B$ er uavhengige. Dette kan derfor forenkles til

$$P(A' \cap B') = 1 - P(A) - P(B) + P(A)P(B).$$

Vi har dermed vist at $P(A' \cap B') = P(A')P(B')$, og dermed at $A'$ og $B'$ er uavhengige.

Hvis $A'$ og $B'$ er disjunkte er også $P(A' \cap B') = 0$. Vi regner derfor ut

$$P(A' \cap B') = 1 - P(A) - P(B) + P(A)P(B).$$

Vi vet at sannsynligheten $P(A)$ for at det skjer én eller flere A feil på en tilfeldig PC i løpet av 4 år er det samme som 1 minus sannsynligheten for at det ikke skjer en eneste A feil. Vi får derfor

$$P(A) = P(X \geqslant 1) = 1 - P(X = 0) = 1 - e^{-0.25 \cdot 4}.$$

Vi finner også at

$$P(B) = P(Y \geqslant 1) = 1 - P(Y = 0) = 1 - e^{-0.15 \cdot 4}.$$

Da kan vi finne $P(A' \cap B')$. Vi får

$$\begin{aligned} P(A' \cap B') &= 1 - P(A) - P(B) + P(A)P(B) \\ &= 1 - (1 - e^{-1}) - (1 - e^{-0.6}) + (1 - e^{-1})(1 - e^{-0.6}) \\ &= 0.202 \neq 0. \end{aligned}$$

Derfor er ikke $A'$ og $B'$ disjunkte.

**d)** Hendelsen at det oppstår feil på en tilfeldig PC i løpet av 4 år er den motsatte av at en PC er feilfri i 4 år. Vi får derfor

$$P(\text{feil}) = 1 - P(\text{feilfri}) = 1 - e^{-1.60},$$

hvor sannsynligheten for en feilfri PC er funnet i oppgave c). Vi kan ved hjelp av samme argumentasjon som i oppgave c) skjønne at fordelingen til $V$ også vil være binomisk, denne gangen med suksessanssynlighet $1 - e^{-1.60}$.

Det er kjent at korrelasjonskoeffisienten mellom to stokastiske variable er

$$\rho(V, Z) = \frac{\text{Cov}(V, Z)}{\sqrt{\text{Var}(V) \cdot \text{Var}(Z)}}.$$

For en binomisk fordeling med $n$ individer og suksessanssynlighet $p$ vet vi at variansen til fordelingen er lik $np(1 - p)$. Vi får derfor

$$\text{Var}(Z) = 10 \cdot e^{-1.60} \cdot (1 - e^{-1.60}),$$

og

$$\text{Var}(V) = 10 \cdot (1 - e^{-1.60}) \cdot (1 - (1 - e^{-1.60})) = \text{Var}(Z).$$

Det er klart at en PC etter 4 år enten må ha opplevd en feil, eller være feilfri. Derfor kan vi se den enkle sammenhengen $V = 10 - Z$. For å finne kovariansen mellom $Z$ og $V$ kan vi derfor sette

$$\text{Cov}(V, Z) = \text{Cov}(10 - Z, Z) = \text{Cov}(10, Z) - \text{Cov}(Z, Z) = 0 - \text{Var}(Z).$$

Dette gir oss en korrelasjonskoeffisient på

$$\rho(V, Z) = \frac{\text{Cov}(V, Z)}{\sqrt{\text{Var}(V) \cdot \text{Var}(Z)}} = \frac{-\text{Var}(Z)}{\sqrt{\text{Var}(Z) \cdot \text{Var}(Z)}} = \underline{\underline{-1}}.$$

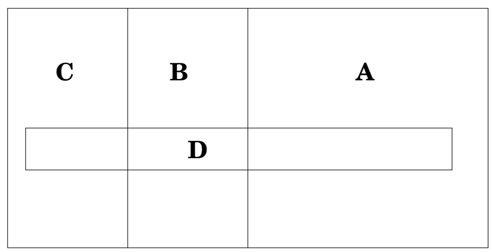
Dette kommer klart frem fra sammenhengen $V = 10 - Z$. De to variablene er perfekt negativt korrelert, fordi vi vet at summen av dem alltid må være lik 10. Dette betyr at når $V$ er stor må $Z$ være liten, og motsatt.

## 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

Den hypergeometriske fordelingen kan tilnærmes med den binomiske fordelingen i de tilfeller hvor trekning uten tilbakelegging er tilnærmet lik trekning med tilbakelegging. Dette skjer når vi trekker et lite antall $n$ fra et stort antall $N$, trekningene blir tilnærmet uavhengige selv om man trekker uten tilbakelegging. Vi trenger altså $n \ll N$.

En binomisk fordeling kan tilnærmes med en poissonfordeling når $n$ er stor og $p$ er liten.

### Oppgave 5

**a)** Vi har at en gjennomlesing av teksten tilsvarer $n$ repeterte forsøk, ett forsøk for hver skrivefeil i teksten. Hvert forsøk resulteterer i suksess (feilen oppdages) eller ikke-suksess (feilen oppdages ikke). Sannsynligheten for suksess er $p$, og denne er konstant for alle forsøkene. Vi må i tillegg anta at hvert ord leses uavhengig av alle andre ord i teksten, slik at forsøkene er uavhengige.

Vi har $\lambda = 2$ og $s = 8$ og ønsker å finne sannsynligheten for at antall trykkfeil, $N$, er større enn 10. Vi har $\mu = \lambda s = 2 \cdot 8 = 16$

$$
\begin{aligned}
P(N > 10) &= 1 - P(N \leq 10) \\
&= 1 - \sum_{i=1}^{10} P(N = i) \\
&= 1 - 0.0774 \\
&= 0.923.
\end{aligned}
$$

Vi har nå gitt $N = 12$ og $p = 0.6$ og œnsker å finne sannsynligheten for at korrekturleseren oppdager alle trykkfeilene.

$$
\begin{aligned}
P(X = 12 | N = 12) &= \binom{12}{12} \cdot 0.6^{12} \cdot 0.4^{0} \\
&= 0.6^{12} \\
&= 0.0022.
\end{aligned}
$$

**b)** $Y_k$ = antall trykkfeil som gjenstår etter $k$ uavhengige gjennomlesninger. Vi finner først simultanfordelingen til $Y_1$ og $N$.
Vi har
$$P(X = x | N = n) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, ..., n$$

Simultanfordelingen til $Y_1$ og $N$ er da gitt ved

$$
\begin{aligned}
P(Y_1 = u, N = n) &= P(Y_1 = u | N = n) \cdot P(N = n) \\
&= P(N - X = u | N = n) \cdot P(N = n) \\
&= P(X = n - u | N = n) \cdot P(N = n) \\
&= \binom{n}{n-u} p^{n-u} (1-p)^u \cdot P(N = n)
\end{aligned}
$$

for $u = 0, 1, \ldots$ og $n = u, u+1, \ldots$.

Vi finner deretter marginalfordelingen til $Y_1$.

$$
\begin{aligned}
P(Y_1 = u) &= \sum_{n=u}^{\infty} P(Y_1 = u, N = n) \\
&= \sum_{n=u}^{\infty} \binom{n}{n-u} p^{n-u}(1-p)^u e^{-\lambda s} \cdot \frac{(\lambda s)^n}{n!} \\
&= \sum_{n=0}^{\infty} \binom{n+u}{n} p^{n}(1-p)^u e^{-\lambda s} \cdot \frac{(\lambda s)^{n+u}}{(n+u)!} \\
&= \sum_{n=0}^{\infty} \frac{1}{n!u!} p^{n}(1-p)^u e^{-\lambda s} (\lambda s)^{n+u} \\
&= \frac{(\lambda s)^u e^{-\lambda s}}{u!}(1-p)^u \sum_{n=0}^{\infty} \frac{(\lambda p s)^n}{n!} \\
&= \frac{(\lambda s)^u e^{-\lambda s}}{u!}(1-p)^u e^{\lambda p s} \\
&= \frac{(\lambda s(1-p))^u}{u!} e^{-\lambda s(1-p)}.
\end{aligned}
$$

Vi ser at marginalfordelingen til $Y_1$ er $Y_1 \sim \text{Poisson}(\lambda s(1-p))$.

Siden fordelingen for $Y_1$ er den samme som for $N$ bortsett fra at parameteren er endret, vil mønsteret gjenta seg slik at også $Y_2$ blir poissonfordelt, og parameteren i fordelinga til $Y_2$ blir $\lambda s(1-p)(1-p) = \lambda s(1-p)^2$. Tilsvarende blir $Y_3 \sim \text{Poisson}(\lambda s(1-p)^3)$ og generelt $Y_k \sim \text{Poisson}(\lambda s(1-p)^k)$.