TMA4245 Statistikk
Vår 2018

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

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

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

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

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

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

Forventet antall skip som betjenes en gitt dag blir dermed

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

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) X er en stokastisk variabel som beskriver antall korrekte svar på $n = 20$ spørsmål på midtveiseksamen (flervalgsoppgave).

Betingelser for at X er binomisk fordelt:

- Vi ser på $n = 20$ svar.
- For hvert svar sjekker vi om svaret er korrekt eller ikke.
- Sannsynligheten for at svaret er korrekt er $\frac{1}{m}$ fordi Ole tipper, og det er bare ett svar som er korrekt blant m mulig svar. Denne sannsynligheten er den samme for alle n svarene.
- De n svarene er uavhengige av hverandre, siden Ole tipper svaret på hvert spørsmål.

Under disse 4 betingelsene er X binomisk fordelt med parametere $n = 20$ og $p = \frac{1}{m}$. 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 å svare korrekt på minst 8 spørsmål finner vi enklest ved tabelloppslag (s 17 i formelsamlingen). Vi gjør dette for de tre verdiene av m som er oppgitt, $m = 2, 4, 5$. (Grunnen til at disse tre verdiene er valgt er kun pga at de finnes i tabellen...)

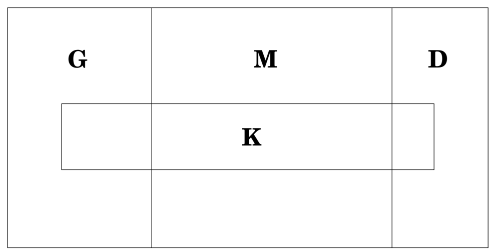
$$\begin{aligned}
m = 2 : P(X \geq 8) &= 1 - P(X \leq 7 | p = 0.5, n = 20) = 1 - 0.132 = \underline{\underline{0.87}} \\
m = 4 : P(X \geq 8) &= 1 - P(X \leq 7 | p = 0.25, n = 20) = 1 - 0.898 = \underline{\underline{0.10}} \\
m = 5 : P(X \geq 8) &= 1 - P(X \leq 7 | p = 0.2, n = 20) = 1 - 0.968 = \underline{\underline{0.03}}
\end{aligned}$$

Forventningen til X er $\mathrm{E}(X) = np$. Forventet antall korrekte svar for $m = 2, 4, 5$ blir da:

$$\begin{aligned}
m = 2 : \mathrm{E}(X) &= 20 \cdot \frac{1}{2} = 10 \\
m = 4 : \mathrm{E}(X) &= 20 \cdot \frac{1}{4} = 5 \\
m = 5 : \mathrm{E}(X) &= 20 \cdot \frac{1}{5} = 4
\end{aligned}$$

b) G = spørsmålet er fra den delen av pensum som Ole kan godt,
M = spørsmålet er fra den delen av pensum som Ole kan middels godt,
D = spørsmålet er fra den delen av pensum som Ole kan dårlig,
K = Ole svarer korrekt på spørsmålet.

Venndiagram for de fire hendelsene:



Sannsynligheten for at Ole svarer korrekt på et tilfeldig valgt spørsmål, $P(K)$, finner vi vet å bruke setningen om total sannsynlighet. Vi vet at G, M, D er en partisjon av utfallsrommet (det ser vi lett av venndiagrammet og at summen av sannsynlighetene er 1).

$$\begin{aligned} P(K) &= P(K \cap G) + P(K \cap M) + P(K \cap D) \\ &= P(K|G) \cdot P(G) + P(K|M) \cdot P(M) + P(K|D) \cdot P(D) \\ &= 0.8 \cdot 0.3 + 0.4 \cdot 0.5 + 0.2 \cdot 0.2 = \underline{\underline{0.48}} \end{aligned}$$

Bayes regel kan benyttes til å finne sannsynligheten for at spørsmålet var fra den delen av pensum som Ole kan dårlig, gitt at Ole svarte korrekt på spørsmålet.

$$\begin{aligned} P(D|K) &= \frac{P(K \cap D)}{P(K)} \\ &= \frac{P(K|D) \cdot P(D)}{P(K)} = \frac{0.2 \cdot 0.2}{0.48} = \underline{\underline{0.08}} \end{aligned}$$

Oppgave 3

a) For hver deltaker har vi følgende situasjon:

- Deltakeren får en serie oppgaver.
- Hver runde har to mulige utfall: Deltakeren klarer ikke oppgaven og går ut av konkurransen (hendelse A), eller han/hun klarer oppgaven og går videre til neste runde (hendelse A').
- Sannsynligheten for ikke å klare oppgaven, $p = P(A)$, er lik i hver runde.
- Resultatene fra hver runde er uavhengige.

Denne situasjonen svarer til en Bernoulli-forsøksrekke, der vi ikke bestemmer antall forsøk på forhånd, men repeterer forsøket (gir nye oppgaver) inntil første gang hendelsen A (klarer ikke oppgaven) inntreffer. Siden X er antall forsøk inntil A inntreffer *første* gang (deltakeren første gang ikke klarer oppgaven), er det rimelig å anta at X er geometrisk fordelt.

Sannsynligheten for at deltakeren går ut i første runde:

$$P(X=1)=f(1)=p(1-p)^{1-1}=p=\underline{\underline{0.10}}$$

Sannsynligheten for at deltakeren fortsatt er med etter fem runder:

$$P(X>5)=1-P(X\leq 5)=1-F(5)=1-(1-(1-p)^5)=(1-p)^5=0.90^5=\underline{\underline{0.59}}.$$

Sannsynligheten for at deltakeren ikke klarer oppgaven i niende runde ($X=9$), dersom deltakeren klarer oppgavene til og med femte runde ($X>5$): Her bruker vi betinget sannsynlighet, og resultatet fra forrige spørsmål.

$$\begin{aligned} P(X=9 \mid X>5) &= \frac{P(X=9\cap X>5)}{P(X>5)} = \frac{P(X=9)}{P(X>5)} = \frac{f(9)}{1-F(5)} \\ &= \frac{p(1-p)^{9-1}}{(1-p)^5} = p(1-p)^3 = 0.10\cdot 0.90^3 = \underline{\underline{0.073}} \end{aligned}$$

b) Vi har følgende situasjon for hver oppgavelager:

- Resultater for et visst antall (n_1 eller n_2) deltakere blir registrert
- To mulig utfall: Deltakeren klarer færre enn fem oppgaver (hendelse C), eller ikke (dvs. klarer fem eller flere, hendelse C').
- Sannsynligheten for C er lik i for hver deltaker.
- Resultatene for hver deltaker er uavhengige.

Dette svarer til et binomisk forsøk, og Z_1 og Z_2 er dermed binomisk fordelte, med parametre som gitt i oppgaven.

Oppgave 4

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 ønsker 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 $Y_1 \sim \text{Poisson}(\lambda s(1-p))$.

Oppgave 5

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

$$\begin{aligned} 4m\left(1 + \frac{1}{4} - (1-p)^4\right) &< 4m \\ 1 + \frac{1}{4} - (1-p)^4 &< 1 \\ (1-p)^4 &> \frac{1}{4} \\ 1 - p &> \left(\frac{1}{4}\right)^{1/4} \\ p &< 1 - \left(\frac{1}{4}\right)^{1/4}. \end{aligned}$$

For at metoden skal lønne seg trenger vi altså at $\underline{\underline{p < 1 - (1/4)^{1/4} \approx 0.293}}$.