TMA4240 Statistikk
Høst 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Anbefalt øving 4
Løsningsskisse

Oppgave 1 Mureren

La X være mengden mørtel mureren bruker i løpet av en tilfeldig valgt arbeidsdag. Da er X en tilfeldig variabel med sannsynlighetstetthet

$$f(x) = \begin{cases} 1/3 & 4 < x \leq 7 \\ 0 & \text{ellers} \end{cases}$$

a) Sannsynligheten for at det en tilfeldig dag går med mer enn 6 hl mørtel er

$$P(X > 6) = \int_6^\infty f(x)dx = \int_6^7 \frac{1}{3}dx = \underline{\underline{\frac{1}{3}}}.$$

La m være mengden mørtel som kjøpes inn. Vi vil finne verdien $m_{0.05}$ som er slik at sannsynligheten for at forbruket X overstiger den innkjøpte mengden kontrolleres på fem prosent, altså har vi

$$P(X > m_{0.05}) = 0.05 \Leftrightarrow \int_{m_{0.05}}^\infty f(x)dx = 0.05 \Leftrightarrow \int_{m_{0.05}}^7 \frac{1}{3}dx = 0.05 \Leftrightarrow m_{0.05} = \underline{\underline{6.85}}.$$

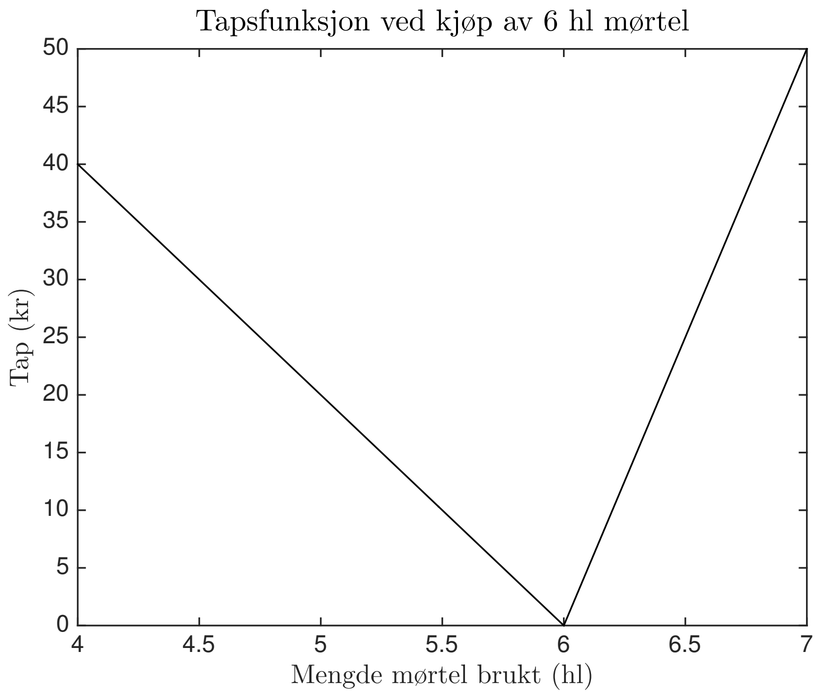
b) Fra **a)** vet vi at $P(X > 6) = \frac{1}{3}$. La nå Z være antall dager han får for lite mørtel om han kjøper inn 6 hl. Sannsynligheten for at han får for lite mørtel minst én av dagene er

$$P(Z \geq 1) = 1 - P(Z = 0) = 1 - \{P(X < 6)\}^4 = 1 - \left\{\frac{2}{3}\right\}^4 = \underline{\underline{0.8}}.$$

c) Mureren taper 20 kr per hl for mye mørtel og 50 kr per hl for lite mørtel. Vi definerer en funksjon $g(x)$ for å representere tap ved x hl mørtel. Hvis mureren har for mye mørtel ($4 < x \leq 6$), vil det være igjen $(6-x)$ hl mørtel som gir $20 \cdot (6-x)$ kr tap. Hvis mureren har for lite mørtel ($6 < x \leq 7$) vil tapet være på $50 \cdot (x-6)$ kr. For alle andre verdier av x er tapet 0 kr. Vi skriver dermed tapsfunksjonen

$$g(x) = \begin{cases} 20(6-x) & 4 < x \leq 6 \\ 50(x-6) & 6 < x \leq 7 \\ 0 & \text{ellers} \end{cases}$$

```
tap = @(x) 20.*(6-x).*(x<=6) + 50.*(x-6).*(x>6);
```



Figur 1: Grafen til tapsfunksjonen når mureren kjøper inn 6 hl mørtel.

```
xx = linspace(4,7,100);

figure()
plot(xx,tap(xx),'k-');
xlabel('Mengde m{\o}rtel brukt (hl)', ...
    'Interpreter','Latex','FontSize',16)
ylabel('Tap (kr)','Interpreter','Latex','FontSize',16)
title('Tapsfunksjon ved kj{\o}p av 6 hl m{\o}rtel', ...
    'Interpreter','Latex','FontSize',18)
set(gca,'FontSize',14);
box on
```

d) Siden forbruket X er en tilfeldig variabel, vil også tapet $T = g(X)$ være en tilfeldig variabel. Med *forventet tap* menes forventningsverdien til T. Denne er definert som

$$E(T) = E(g(X)) = \int_{-\infty}^{\infty} g(x)f(x)dx.$$

Forventet tap ved kjøp av 6 hl blir

$$\begin{aligned}
E(T;\text{kjøper 6 hl}) &= 20\int_4^6 (6-x)f(x)dx + 50\int_6^7 (x-6)f(x)dx \\
&= \frac{20}{3}\int_4^6 (6-x)dx + \frac{50}{3}\int_6^7 (x-6)dx \\
&= \underline{\underline{21.7}}.
\end{aligned}$$

Vi skal plotte forventet tap $E(T)$ som en funksjon av mengde mørtel som kjøpes inn. La igjen m være mengden mørtel som kjøpes inn. Murerens faktiske tap T en tilfeldig dag er en funksjon av både mørtelforbruket X denne dagen, og av mengden innkjøpt mørtel m. Det forventede tapet $E(T)$ vil imidlertid kun være en funksjon av m, siden den tilfeldige variabelen X integreres ut når forventningsverdien beregnes.

Forventet tap når det kjøpes inn m hl mørtel er

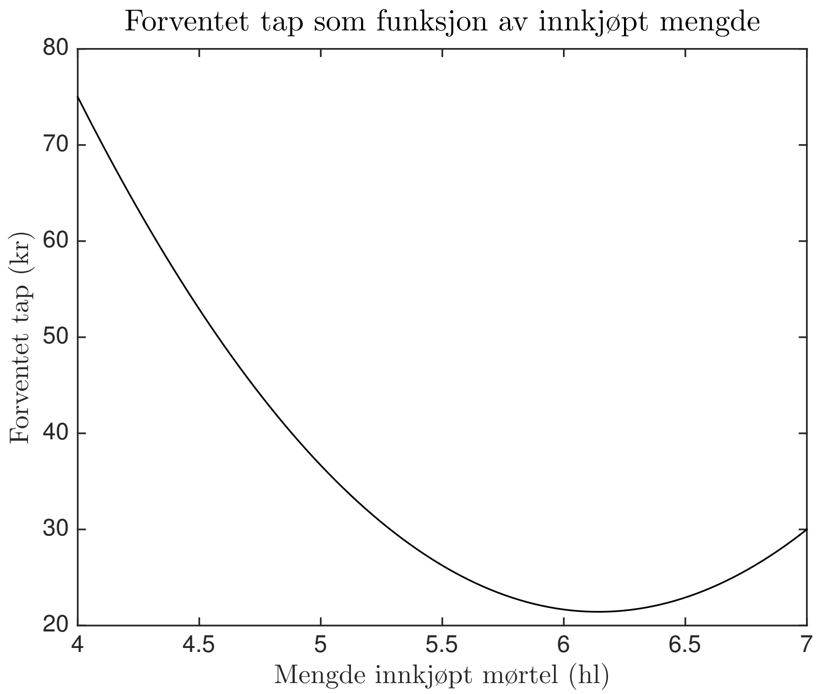
$$\begin{aligned}
E(T;\text{kjøper } m \text{ hl}) &= 20\int_4^m (m-x)f(x)dx + 50\int_m^7 (x-m)f(x)dx \\
&= \frac{20}{3}[mx - \frac{1}{2}x^2]_4^m + \frac{50}{3}[\frac{1}{2}x^2 - mx]_m^7 \\
&= \frac{20}{3}[m^2 - \frac{1}{2}m^2 - 4m + 8] + \frac{50}{3}[\frac{49}{2} - 7m - \frac{1}{2}m^2 + m^2] \\
&= \frac{70}{6}m^2 - \frac{430}{3}m + \frac{1385}{3}.
\end{aligned}$$

```
Etap = @(m) (70/6).*m.^2 - (430/3).*m + 1385/3;

xx = linspace(4,7,100);

figure()
plot(xx,Etap(xx),'k-');
xlabel('Mengde innkj{\o}pt m{\o}rtel (hl)', ...
    'Interpreter','Latex','FontSize',16)
ylabel('Forventet tap (kr)', ...
    'Interpreter','Latex','FontSize',16)
title('Forventet tap som funksjon av innkj{\o}pt mengde', ...
    'Interpreter','Latex','FontSize',18)
set(gca,'FontSize',14)
box on
```

La $u(m)$ være forventet tap som funksjon av mengde innkjøpt mørtel. Figur 2 viser grafen til $u(m)$ for verdier av m mellom 4 og 7. Siden u er et kvadratisk polynom med positiv koeffisient i det første leddet, har den et globalt minimumspunkt som vi finner



Figur 2: Forventet tap når det kjøpes inn m hl mørtel, for $4 \leq m \leq 7$.

ved å derivere, og sette $u'(m) = 0$,

$$\begin{aligned}
u'(m) &= \frac{70}{3}m - \frac{430}{3} \\
u'(m^*) &= 0 \\
\frac{70}{3}m^* &= \frac{430}{3} \\
m^* &= \frac{430}{70} = \underline{\underline{6.1429}}.
\end{aligned}$$

Siden $4 \leq m^* \leq 7$ kan mureren minimere sitt forventede tap ved å kjøpe inn $m^* \approx 6.14$ hl mørtel.

e) Utledningen i **c)** og **d)** svarer til tilfellet $c = 20$. Vi gjentar utledningen med c som en ukjent variabel, og får da tapsfunksjonen

$$g_c(x) = \begin{cases} c(6-x) & 4 < x \leq 6 \\ 50(x-6) & 6 < x \leq 7 \end{cases}.$$

Videre blir forventet tap ved kjøp av m hl mørtel lik

$$
\begin{aligned}
u_c(m) &= E(g_c(X); \text{kjøper m hl}) \\
&= c\int_4^m (m-x)f(x)dx + 50\int_m^7 (x-m)f(x)dx \\
&= \frac{c}{3}[mx - \frac{1}{2}x^2]_4^m + \frac{50}{3}[\frac{1}{2}x^2 - mx]_m^7 \\
&= \frac{c}{3}[m^2 - \frac{1}{2}m^2 - 4m + 8] + \frac{50}{3}[\frac{49}{2} - 7m - \frac{1}{2}m^2 + m^2] \\
&= \frac{1}{6}cm^2 - \frac{4}{3}cm + \frac{8}{3}c + \frac{1225}{3}m + \frac{25}{3}m^2.
\end{aligned}
$$

```
Etapc = @(m,c) (1/6).*c.*m.^2 - (4/3).*c.*m + (8/3).*c + ...
    1225/3 - (350/3).*m + (25/3).*m.^2;

xx = linspace(4,7,100);

figure()
hold on
plot(xx,Etapc(xx,20),'k-');
plot(xx,Etapc(xx,25),'k--');
plot(xx,Etapc(xx,30),'k-.');
hleg = legend('$c=20$','$c=25$','$c=30$');
set(hleg,'Interpreter','Latex','FontSize',16);
xlabel('Mengde innkj{\o}pt m{\o}rtel (hl)', ...
    'Interpreter','Latex','FontSize',16)
ylabel('Forventet tap (kr)', ...
    'Interpreter','Latex','FontSize',16)
title('Forventet tap som funksjon av innkj{\o}pt mengde', ...
    'Interpreter','Latex','FontSize',18)
set(gca,'FontSize',14)
box on
```

Fra figur 3 kan vi lese av at når $c = 25$ får vi minimalt forventet tap ved å velge $\underline{\underline{m \approx 6}}$, og når $c = 30$ minimeres det forventede tapet når vi velger $\underline{\underline{m \approx 5.88}}$.

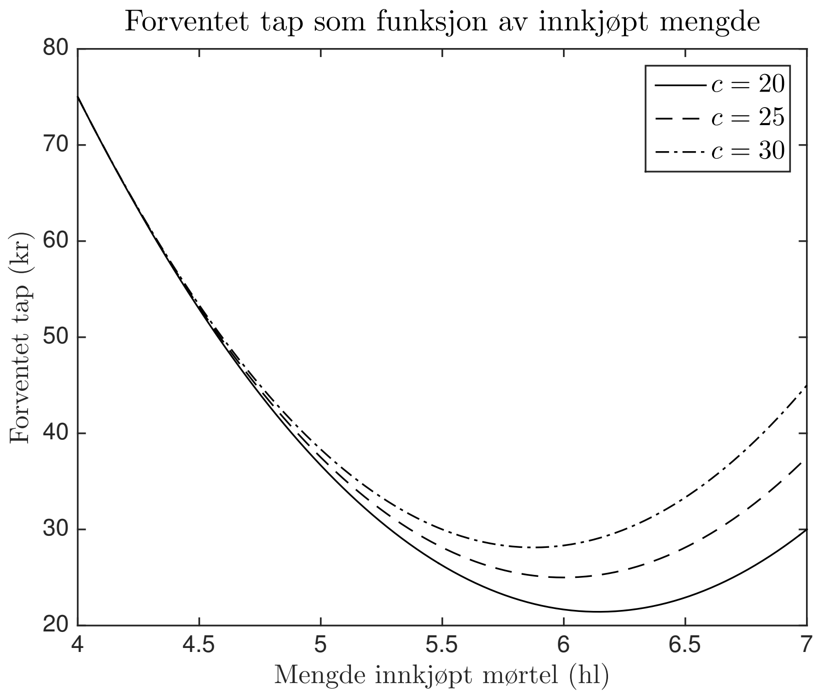
Oppgave 2

a) Utfallsrommet til X_1 er $\{1, 2, 3, 4, 5, 6\}$. Sannsynlighetsfordelingen til X_1 er den den diskrete uniforme fordelingen på dette utfallsrommet, dvs. X har punktsannsynlighet

$$P(X_1 = x) = \frac{1}{6}, \quad \text{for } x \in \{1, 2, 3, 4, 5, 6\}.$$

Forventingsverdien til X_1 er

$$E(X_1) = \sum_{x=1}^{6} x \cdot P(X = x) = \sum_{x=1}^{6} \frac{x}{6} = \frac{21}{6} = \underline{\underline{\frac{7}{2}}}.$$



Figur 3: Forventet tap for $c = 20$, $c = 25$ og $c = 30$.

b) Sannsynligheten for at $Y_2 = 1$ er lik sannsynligheten for å få seksere på første og andre kast,

$$P(Y_2 = 1) = P(X_1 = 6 \cap X_2 = 6) = P(X_1 = 6)P(X_2 = 6) = \frac{1}{6^2}.$$

Tilsvarende har vi, for Y_3,

$$P(Y_3 = 1) = P(X_2 = 6 \cap X_3 = 6) = P(X_2 = 6)P(X_3 = 6) = \frac{1}{6^2}.$$

Den simultane sannsynligheten for at både $Y_2 = 1$ og $Y_3 = 1$ er

$$\begin{aligned} P(Y_2 = 1 \cap Y_3 = 1) &= P(X_1 = 6 \cap X_2 = 6 \cap X_3 = 6) \\ &= P(X_2 = 6)P(X_2 = 6)P(X_3 = 6) = \frac{1}{6^3}. \end{aligned}$$

Dette er ikke det samme som produktet av sannsynlighetene $P(Y_2 = 1)$ og $P(Y_3 = 1)$. Vi har altså

$$P(Y_2 = 1 \cap Y_3 = 1) = \frac{1}{6^3} \neq \frac{1}{6^4} = P(Y_2 = 1)P(Y_3 = 1),$$

som betyr at de tilfeldige variablene Y_2 og Y_3 ikke er uavhengige av hverandre.

Sannsynligheten for at $Y_4 = 1$ er lik sannsynligheten for å få seksere i tredje og fjerde kast,

$$P(Y_4 = 1) = P(X_3 = 6 \cap X_4 = 6) = \frac{1}{6^2},$$

og den simultane sannsynligheten for at både Y_2 og Y_4 er lik 1, er

$$\begin{aligned}P(Y_2 = 1 \cap Y_4 = 1) &= P(X_1 = 6 \cap X_2 = 6 \cap X_3 = 6 \cap X_4 = 6) \\ &= P(X_1 = 6)P(X_2 = 6)P(X_3 = 6)P(X_4 = 6) = \frac{1}{6^4}.\end{aligned}$$

I dette tilfellet er det likhet mellom den simultane sannsynligheten og produktet av de to marginale sannsynlighetene, $P(Y_2 = 1)$ og $P(Y_4 = 1)$. Det vil si

$$P(Y_2 = 1 \cap Y_4 = 1) = \frac{1}{6^4} = \frac{1}{6^2} \cdot \frac{1}{6^2} = P(Y_2 = 1)P(Y_4 = 1),$$

hvilket betyr at Y_2 og Y_4 *er* uavhengige av hverandre.

Korrelasjonskoeffisienten mellom Y_2 og Y_3 er positiv, siden

$$P(Y_3 = 1|Y_2 = 1) = \frac{P(Y_2 = 1 \cap Y_3 = 1)}{P(Y_2 = 1)} = \frac{1/6^3}{1/6^2} = \frac{1}{6} > \frac{1}{6^2} = P(Y_3 = 1).$$

Med andre ord er sannsynligheten for å få poeng på tredje kast gitt at man allerede har fått poeng på andre kast, større enn den marginale sannsynligheten for å få poeng på tredje kast. Korrelasjonskoeffisienten mellom Y_2 og Y_4 blir null, siden de to er uavhengige tilfeldige variable.

c) Kovariansen mellom Y_2 og Y_3 er

$$\begin{aligned}\mathrm{Cov}(Y_2, Y_3) &= E(Y_2Y_3) - E(Y_2)E(Y_3) \\ &= \sum_{y_2=0}^{1}\sum_{y_3=0}^{1} y_2y_3P(Y_2 = y_2 \cap Y_3 = y_3) - P(Y_2 = 1)P(Y_3 = 1) \\ &= P(Y_2 = 1 \cap Y_3 = 1) - P(Y_2 = 1)P(Y_3 = 1) = \tfrac{1}{6^3} - \tfrac{1}{6^2} \cdot \tfrac{1}{6^2} \\ &= \tfrac{1}{6^3} - \tfrac{1}{6^4} = \underline{\underline{0.0039}}.\end{aligned}$$

Forventningsverdien til $Z = Y_2 + Y_3 + \ldots + Y_{10}$ er

$$E(Z) = E\left(\sum_{j=2}^{10} Y_j\right) = \sum_{j=2}^{10} E(Y_j) = \sum_{j=2}^{10} P(Y_j = 1) = 9 \cdot \frac{1}{6^2} = \underline{\underline{\frac{1}{4}}},$$

og variansen er (siden $\mathrm{Cov}(Y_k, Y_{k+h}) = 0$ for $|h| > 1$)

$$\begin{aligned}\mathrm{Var}(Z) = \mathrm{Var}\left(\sum_{j=2}^{10} Y_j\right) &= \sum_{j=2}^{10} \mathrm{Var}(Y_j) + 2\sum_{j=2}^{9} \mathrm{Cov}(Y_j, Y_{j+1}) \\ &= 9\mathrm{Var}(Y_2) + 2 \cdot 8\mathrm{Cov}(Y_2, Y_3) \\ &= 9 \cdot \tfrac{1}{6^2}(1 - \tfrac{1}{6^2}) + 16 \cdot (\tfrac{1}{6^3} - \tfrac{1}{6^4}) \\ &= \underline{\underline{0.3048}},\end{aligned}$$

hvor vi bruker at variansen til Y_2 er

$$\mathrm{Var}(Y_2) = E(Y_2^2) - E(Y_2)^2 = P(Y_2 = 1) - P(Y_2 = 1)^2 = \tfrac{1}{6^2}(1 - \tfrac{1}{6^2}).$$

Oppgave 3

a) Siden $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0$, så er hendelsene disjunkte.

Siden $P(A \cap B) = 0 \neq 0{,}3 \cdot 0{,}3 = P(A)P(B)$, så er hendelsene ikke uavhengige.

b) Hendelsene med sannsynligheter er tegnet inn i Venndiagrammet i Figur 4.

Sannsynligheten for olje på felt 1, dvs. hendelsen A, kan beregnes fra lov om total sannsynlighet,

$$P(A) = P(A \cap B) + P(A \cap B^{\mathcal{C}}) = 0{,}05 + 0{,}15 = 0{,}20.$$

Hendelsen olje på felt 1 gitt olje er funnet på felt 2 kan skrives $A|B$ og sannsynligheten beregnes fra Bayes lov,

$$\begin{aligned} P(A|B) &= \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(B \cap A) + P(B \cap A^{\mathcal{C}})} \\ &= \frac{0{,}05}{0{,}05 + 0{,}1} = \frac{1}{3}. \end{aligned}$$

Her er lov om total sannsynlighet brukt for å beregne $P(B)$.

Sannsynligheten for hendelsen ikke olje på felt 2 kan sees fra ligningen over

$$P(B^{\mathcal{C}}) = 1 - P(B) = 1 - 0{,}15 = 0{,}85.$$

Dette kan settes inn i Bayes lov for å få sannsynlighet for olje på felt 1 gitt ingen olje funnet på felt 2,

$$\begin{aligned} P(A|B^{\mathcal{C}}) &= \frac{P(A \cap B^{\mathcal{C}})}{P(B^{\mathcal{C}})} \\ &= \frac{0{,}15}{0{,}85} \approx 0{,}176. \end{aligned}$$

Fra de tidligere beregningene ser vi at $P(A|B) \neq P(A|B^{\mathcal{C}})$, dermed er A og B ikke uavhengige.

c) La F_A være den stokastiske variabelen "Fortjeneste fra felt 1". Da er

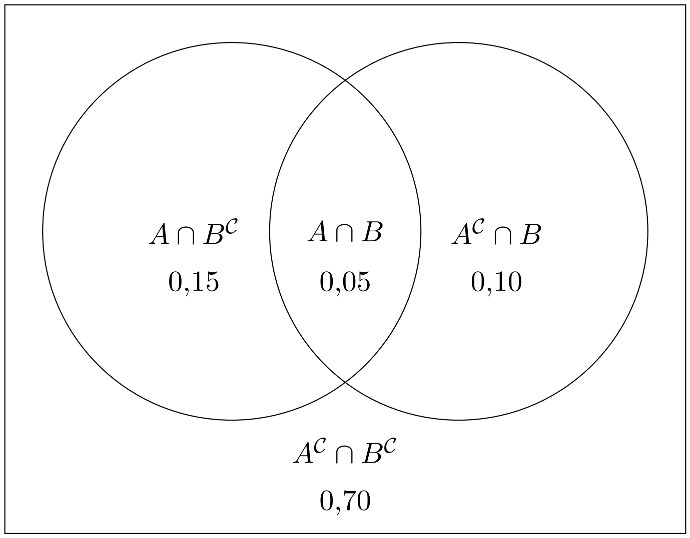
$$\mathrm{E}[F_A] = 400P(A) - 100P(A^{\mathcal{C}}) = 400 \cdot 0{,}20 - 100 \cdot 0{,}80 = 0.$$

La $F_A|B$ være den stokastiske variabelen "Fortjeneste fra felt 1 gitt funn på felt 2". Da er

$$\mathrm{E}[F_A|B] = 400P(A|B) - 100P(A^{\mathcal{C}}|B) = 400 \cdot \frac{1}{3} - 100\frac{2}{3} = \frac{200}{3} \approx 66{,}67.$$

Vi begynner med de følgende to observasjonene. Forventningsverdien for hver strategi er bestemt av fortjenesten ved hver av hendelsene $A \cap B$, $A \cap B^{\mathcal{C}}$, $A^{\mathcal{C}} \cap B$ og $A^{\mathcal{C}} \cap B^{\mathcal{C}}$, og den beste strategien er å ikke lete noe sted, å begynne å lete på felt 1 eller å begynne å lete på felt 2. Det er dermed tre muligheter vi må vurdere.

Første mulighet. Vi leter ingen steder, dette gir umiddelbart forventningsverdi 0.



Figur 4: Venndiagram med hendelsene $A \cap B$, $A \cap B^{\mathcal{C}}$, $A^{\mathcal{C}} \cap B$ og $A^{\mathcal{C}} \cap B^{\mathcal{C}}$ og tilhørende sannsynligheter påtegnet.

Andre mulighet. Vi begynner med å lete på felt 1. Dette gir oss informasjon om felt 1 og vi kan velge mellom følgende strategier for felt 2. Ikke lete uansett om det er olje på felt 1 eller ikke, lete uansett om det er olje på felt 1 eller ikke, lete hvis og bare hvis vi finner olje på felt 1 eller lete hvis og bare hvis vi ikke finner olje på felt 1. Kall de stokastiske variablene som beskriver fortjenesten ved hver av disse strategiene for henholdsvis F_1, F_2, F_3 og F_4. Da har vi

$$\begin{aligned}
&\mathrm{E}[F_1] = \mathrm{E}[F_A] = 0\\
&\mathrm{E}[F_2] = 300P(A \cap B^{\mathcal{C}}) + 1400P(A \cap B) + 900P(A^{\mathcal{C}} \cap B) - 200P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 65,\\
&\mathrm{E}[F_3] = 300P(A \cap B^{\mathcal{C}}) + 1400P(A \cap B) - 100P(A^{\mathcal{C}} \cap B) - 100P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 35,\\
&\mathrm{E}[F_4] = 400P(A \cap B^{\mathcal{C}}) + 400P(A \cap B) + 900P(A^{\mathcal{C}} \cap B) - 200P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 30.
\end{aligned}$$

Tredje mulighet. Vi begynner med å lete på felt 2. Dette gir oss informasjon om felt 2 og vi kan velge mellom følgende strategier for felt 1. Ikke lete uansett om det er olje på felt 2 eller ikke, lete uansett om det er olje på felt 2 eller ikke, lete hvis og bare hvis vi finner olje på felt 2 eller lete hvis og bare hvis vi ikke finner olje på felt 2. Kall de stokastiske variablene som beskriver fortjenesten ved hver av disse strategiene for henholdsvis F_5, F_6, F_7 og F_8. Da har vi

$$\begin{aligned}
&\mathrm{E}[F_5] = 1000P(B) - 100P(B^{\mathcal{C}}) = 65\\
&\mathrm{E}[F_6] = 300P(A \cap B^{\mathcal{C}}) + 1400P(A \cap B) + 900P(A^{\mathcal{C}} \cap B) - 200P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 65,\\
&\mathrm{E}[F_7] = -100P(A \cap B^{\mathcal{C}}) + 1400P(A \cap B) + 900P(A^{\mathcal{C}} \cap B) - 100P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 75,\\
&\mathrm{E}[F_8] = 300P(A \cap B^{\mathcal{C}}) + 1000P(A \cap B) + 1000P(A^{\mathcal{C}} \cap B) - 200P(A^{\mathcal{C}} \cap B^{\mathcal{C}}) = 55.
\end{aligned}$$

Utregningene for hver av mulighetene viser at beste strategi er å lete først i felt 2 og så lete videre i felt 1 hvis og bare hvis man finner olje på felt 2. Dette gir en forventet fortjeneste på 75 millioner.