Løsningsforslag øving 1

1.1

Helene ble født i 2001. I 2020 er hun 19 år og har ikke fått barn. Fra tallene for 2018 vet vi at 5456 barn ble født med mødre mellom 20 og 24 år. Et estimat for sannsynligheten for at hun får barn før hun er 25 år er da $\hat{p}_m = \frac{5456}{55070} = 10\%$. Det er naturlig at det nye estimatet er lavere en 11%, da det har gått flere år hvor hun ikke har fått barn enda.

1.2

Ting endrer seg med tiden, og det er ikke slik at sannsynligheten for å få barn er identisk for alle år. Det er også mest sannsynlig slik at det å få barn påvirkes av flere faktorer en kun kjønn og alder, som f.eks yrkessituasjon, bosted, kultur, utdanning, osv. Fødsler av tvillinger og liknende ting vil også spille inn, da tabellen viser antall fødsler, ikke antall foreldre som får barn. Det er derfor viktig å huske på at estimatet vi gav i oppgave 1.1 kun er et *estimat*, og ikke en faktisk sannsynlighet. Så lenge man har dette i bakhodet, og forstår at det er store usikkerheter og lite data inne i bildet kan $\hat{p}_m$ virke som et helt greit sannsynlighetsestimat, gitt hva vi vet.

1.3

Det har ikke noe å si når Halvor er født, så lenge det ikke er alt for langt unna år 2018. Tallene i tabell 1.1 gir oss at antall fødsler hvor far er under 25 er $\mathrm{N_f} = 2653$. Da er et sannsynlighetsestimat $\hat{p}_f = 2653/55070 = 5\%$.

1.4

Vi bruker programmeringsspråket `R` til å regne ut et estimat for sannsynligheten for at Halvor blir far før han er 25 år.

```r
p_f <- pnorm(25, mean = 33.84, sd = 6.28)
print(p_f)
```

```
0.0796183577337692
```

1.5

Vi fortsetter antakelsen om normalfordeling og beregner sannsynligheten for at Halvor blir far etter at han er 50 år, eller før han er 25 år, men etter han er 15 år. Hendelsen at han blir far etter at han er 50 år er det motsatt som at han er far før han er 50 år. Vi kan derfor si at P(far etter 50) = 1 - P(far før 50). Sannsynligheten for at han blir far mellom 15 og 25 er det samme som sannsynligheten for å bli far før 25 minus sannsynligheten for å bli far før 15.

```
p1 <- 1 - pnorm(50, mean = 33.84, sd = 6.28)
print(p1)
```

```
0.00503744337295842
```

```
p2 <- pnorm(25, mean = 33.84, sd = 6.28) - pnorm(15, mean = 33.84, sd = 6.28)
print(p2)
```

```
0.0782684597021391
```

For Helene er det oppgitt estimatorene $\tilde{\mu}_m = 31.01$ og $\tilde{\sigma}_m = 5.05$. Da kan vi finne estimatorer for de samme sannsynlighetene for henne.

```
p1 <- 1 - pnorm(50, mean = 31.01, sd = 5.05)
print(p1)
```

```
8.48222830613921e-05
```

```
p2 <- pnorm(25, mean = 31.03, sd = 5.05) - pnorm(15, mean = 31.01, sd = 5.05)
print(p2)
```

```
0.115465949576631
```

1.6

Som oppgitt i teksten kan man estimere μ_f og σ_f ved hjelp av det empiriske middelet $\bar{x}$ og det empiriske standardavviket s. Vi følger oppskriften gitt i Tillegg A. Hvis det f.eks er oppgitt at 5456 personer var i intervallet 20-24, estimerer vi at alle 5456 er 22.5 år gamle, osv. Dette gir oss det empiriske middelet og det empiriske standardavviket sett under

```
alder_far <- rep(c(17 ,18.5 ,22.5 ,27.5 ,32.5 ,37.5 ,
                   42.5 ,47.5 ,52.5 ,57.5 ,62.5),
                 c(21, 96, 2536, 12189 ,  19025 ,
                   12701 ,  5163 ,1809 , 565, 185, 82))
```

```
x_bar <- mean(alder_far)
s <- sd(alder_far)

cat("x_bar =", x_bar, "\n")
cat("s =", s, "\n")

x_bar = 33.83673
s = 6.275608
```

1.7

Det er en sammenheng mellom N og M, ved at N kan sees på som en estimator for M. Det er en usikkerhet knyttet til M fordi ikke alle fødsler i Norge nødvendigvis har blitt registrert. I tillegg er det alltid en mulighet for at feil oppstår når noe registreres.

Det finnes mange mulige eksperimenter for å undersøke sannsynligheten for at en faktisk fødsel blir registrert. Hvis registrene er søkbare kan man f.eks undersøke om alle i familien har blitt registrert, og lage en estimator utifra dette. En annen mulighet er å holde oppsyn med et spesifikt sykehus og undersøke hvor mange av fødslene der som blir registrerte. Det er også mulig å lage estimatorer ved å undersøke liknende estimatorer fra andre land. En kan tenke seg å sammenlikne fødselsregistre med andre lister, som barnehageregistre eller skoleregistre, og på en eller annen måte utvikle en estimator utifra dette. Denne oppgaven er generelt sett veldig åpen. Finn på ditt eget eksperiment. For å utvikle en estimator for hvordan denne sannsynligheten påvirkes av alder kan man gjøre alle de nevnte forsøkene, men også undersøke hva alderen på foreldre som ikke blir registrert er.

1.8

Det er mulig å lage flere forskjellige parametre knyttet til modeller for Helene. En mulig parameter er sannsynligheten p_m for at Helene skal bli mor før hun er 25. En annen mulig parameter er f.eks sannsynligheten p_{2m} for at Helene blir mor mellom en alder av 20 og 25. Vi kan også knytte parametre til en modell som antar at sannsynligheten for når Helene blir mor er normalfordelt. Da har vi både parametrene μ og σ. En siste parameter kan f.eks være sannsynligheten p_{3m} for om Helene blir mor eller ikke.

1.9

Vi kan definere mange statistikker knyttet til Tabell 1.1. En statistikk kan f.eks være antallet fødsler hvor moren er under 25 år gammel. Dette vil da være en estimator for sannsynligheten p_m fra oppgaven over. En annen statistikk er antallet fødsler hvor moren er mellom 20 og 25 år gammel. To andre statistikker er den empiriske middelverdien $\bar{x}$ og det empiriske standardavviket s. En femte statistikk fra Tabell 1.1 er f.eks differansen på antallet kvinner og menn under 25 som har blitt registrert i tabellen.

1.10 - 1.12

Se eget ark.

1.13

$A_n = [0, 1/n]$ for alle n. Da er

$$B = \cup_{n=1}^{10} A_n = [0, 1],$$

$$C = \cap_{n=1}^{10} A_n = [0, 1/10],$$

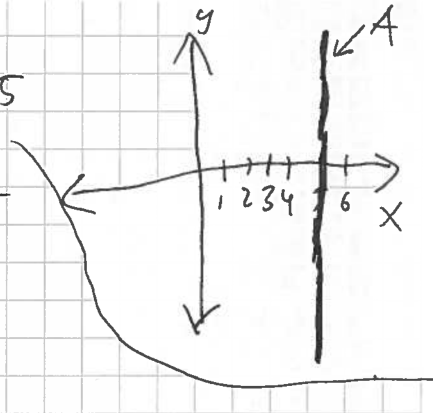
$$D = \cap_{n=1}^{\infty} A_n = \{0\},$$

$$E = \cup_{n=1}^{\infty} A_n = [0, 1].$$

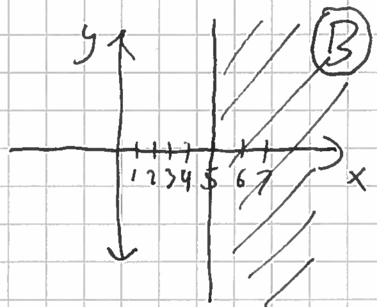
1.14

Se eget ark.

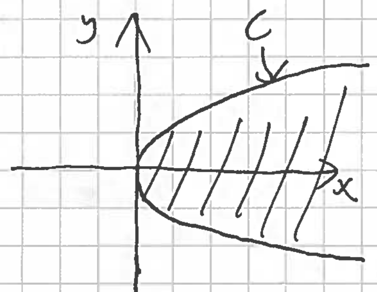
1.10 $A = \{(x,y) ; x = 5\}$ er den rette linjen ved $x = 5$



$B = \{(x,y) ; x \geq 5\}$ er hele planet til høyre for $x=5$

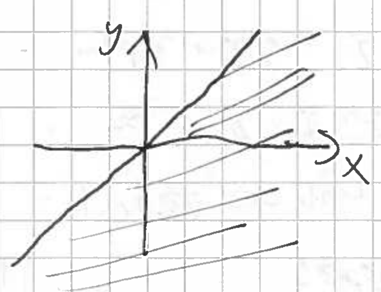


$C = \{(x,y) ; x \geq y^2\}$ er området til høyre for parabelen $x = y^2$



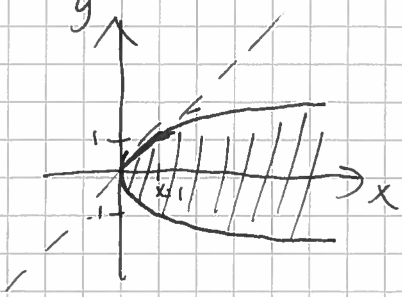
$D = C \setminus B$ er delene av C som ikke er i B

$E = \{(x,y) ; x \geq y\}$



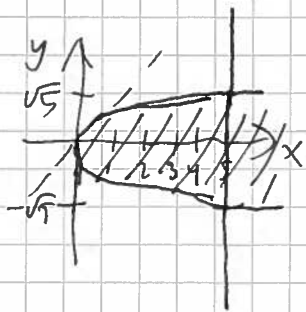
1.10 $F = C \cap E$ er ~~mengde~~ punktene i både C og E.

Her mister vi en liten bit av C for $0 < x < 1$

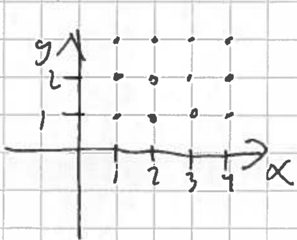


$G = F \cup A$ ~~[illegible]~~

Union betyr at vi bare tegner opp begge mengdene "oppå hverandre"



$H = \{x, y;\ x, y \in \mathbb{N}\}$ Dette er en mengde med punkter i planet. Mengden H er tellbar. Dette ser vi fordi det er mulig å liste opp alle elementene slik som dette: $(1,1), (1,2), (2,2), (2,1), (1,3), (2,3), (3,3), (3,2), (3,1), \ldots$



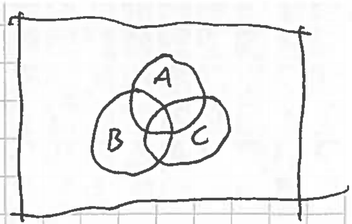
1.11 Det er en en-til-en-korrespondanse mellom H og Q_+ via funksjonen $f(x,y) = \frac{x}{y}$. Vi kan derfor telle gjennom Q_+ på samme måte som vi teller gjennom H:

$Q_+ = \{\frac{1}{1}, \frac{1}{2}, \frac{2}{1}, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, \frac{3}{1}, \ldots\}$. Q_+ er tellbar.

Da er også de rasjonale tall tellbare:

$Q = \{0, \frac{1}{1}, -\frac{1}{1}, \frac{1}{2}, -\frac{1}{2}, \frac{2}{1}, -\frac{2}{1}, \ldots\}$

1.12 Vi lager tre mengder i et Venn-diagram, og markerer områdene



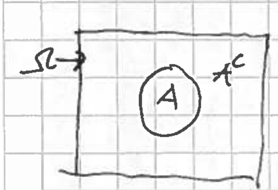
$A \cap (B \cup C)$ $\quad$ $(A \cap B) \cup (A \cap C)$

$=$

Fortsett slik for alle de forskjellige uttrykkene (trenger en D for ett av uttrykkene også)

Vi finner at alle reglene gjelder, og kan generaliseres til union/snitt av vilkårlige familier av mengder.

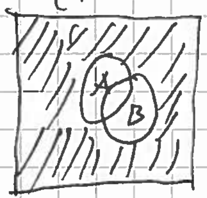
1.14 Vi tegner et Venn-diagram. I dette tilfellet er Ω hele "boksen".
Siden A^c er definert som alt i Ω som ikke er i A, kan vi se



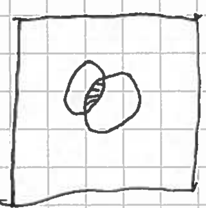
at $\Omega = A \uplus A^c$

Nå tegner vi opp $(A \cup B)^c$, $(A \cap B)$, $(A \setminus B)$ og $(B \setminus A)$

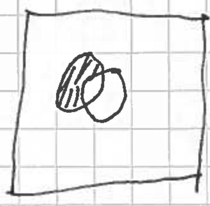
$(A \cup B)^c$



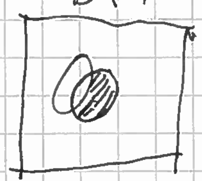
$A \cap B$



$A \setminus B$



$B \setminus A$



Vi ser at unionen av disse blir til hele Ω.
Vi kan generalisere til tre delmengder, A, B, C.
Da har vi

$$\Omega = A \uplus (B \setminus A) \uplus (C \setminus (A \cup B)) \uplus (A \cup B \cup C)^c$$

For n delmengder $A_i, i=1,\dots,n$ kan vi lage den disjunkte partisjonsmengden

$$\Omega = (A_1 \cup A_2 \cup \dots \cup A_n)^c \uplus A_1 \uplus A_2 \setminus A_1 \uplus \dots \uplus A_n \setminus (A_1 \cup A_2 \cup \dots \cup A_{n-1})$$

Det samme kan gjøres for et tellbart antall delmengder