SOK-2009 Mappeinnlevering 1

28

```r
rm(list=ls())
library(tidyverse)
```

```
Warning: package 'tidyverse' was built under R version 4.2.2

Warning: package 'ggplot2' was built under R version 4.2.2

Warning: package 'tidyr' was built under R version 4.2.2

Warning: package 'readr' was built under R version 4.2.2

Warning: package 'purrr' was built under R version 4.2.2

Warning: package 'dplyr' was built under R version 4.2.2

Warning: package 'stringr' was built under R version 4.2.2

Warning: package 'forcats' was built under R version 4.2.2

Warning: package 'lubridate' was built under R version 4.2.2

-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.0     v readr     2.1.4
v forcats   1.0.0     v stringr   1.5.0
v ggplot2   3.4.1     v tibble    3.1.8
v lubridate 1.9.2     v tidyr     1.3.0
v purrr     1.0.1
-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

Oppgave 3

1.

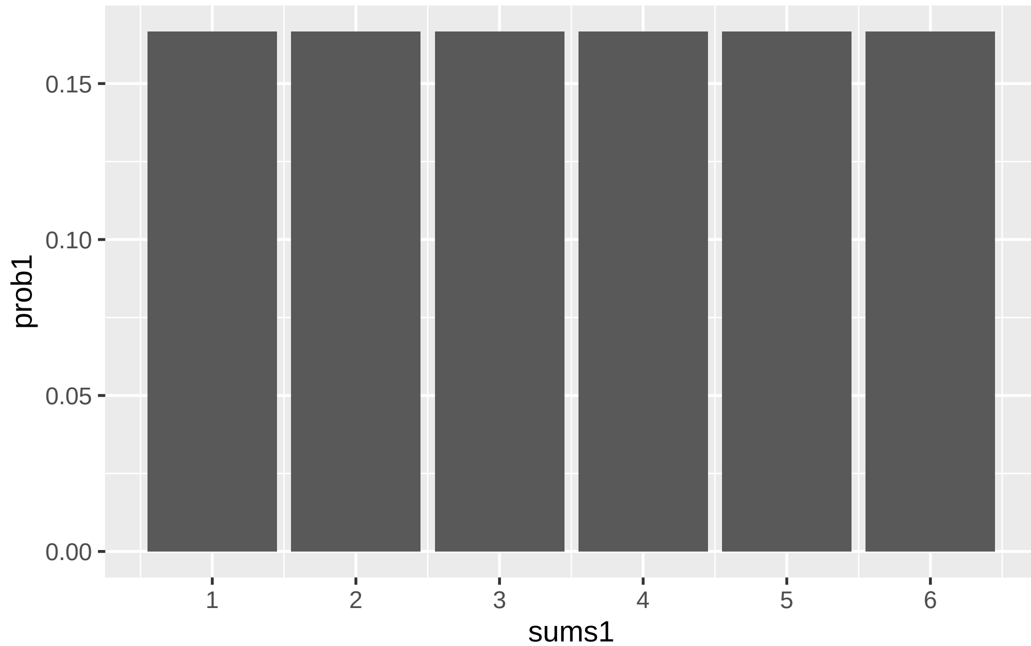
Definer T som antall prikker p ˚ a en seksidet terning og 2T som antall prikker p ˚ a to sekssidede terninger.

1. Et av utfallsrommene T og 2T har uniform sannsynlighet det andre har ikke uniform sannsynlighetsfordeling.

Forklar hva det betyr og hvorfor det er slik. Tegn opp grafen for f (T), f (2T) og den kumulative sannsynlighets-fordelingen F (2T), F (T)?

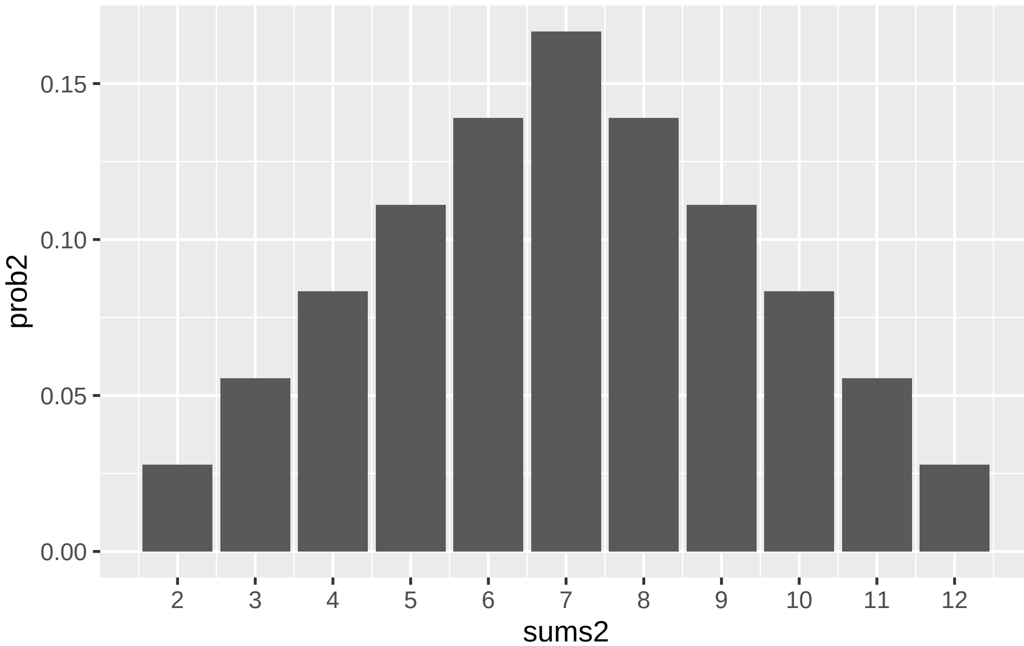
```
T <- 6

T1 <- expand.grid(1:T)
T2 <- expand.grid(1:T, 1:T)

names(T2)[1] = "T"
names(T2)[2] = "T2"
```

```
T2$T2 <- rowSums(T2)


sums1 <- rowSums(T1)
sums2 <- rowSums(T2)


prob1 <- table(sums1) / length(sums1)
prob2 <- table(T2$T2) / 36

prob.dist1 <- data.frame(sums1 = as.integer(names(prob1)),
                        prob1 = as.numeric(prob1))


prob.dist2 <- data.frame(sums2 = as.integer(names(prob2)),
                        prob2 = as.numeric(prob2))

ggplot(prob.dist1, aes(x = sums1, y = prob1)) +
  geom_col() +
  scale_x_continuous(breaks = seq(1, 6, by = 1),
                     labels = seq(1, 6, by = 1))
```



Uniform sansynlighet kommer av at det er like stor sannsynlighet for å få alle tallene da utfallet ikke er avhengi av en annen terning.

```
ggplot(prob.dist2, aes(x = sums2, y = prob2)) +
  geom_col() +
  scale_x_continuous(breaks = seq(2, 12, by = 1),
                     labels = seq(2, 12, by = 1))
```



Ikke unform sannsynlighet visualiseres godt her da vi ser at det er mer sannsynlig at to terninger ender opp med enkelte tall en andre, dette fordi noen tall har flere kombinasjoner på to terninger en andre.

2.

Kan du regne ut korrelasjon, kovarians, varians og standardavik for T og 2T ?

```
#Varians
T2$var <- apply(T2[, c("T","T2")],1,var)


#Standardavvik
T2$Sd <- apply(T2[, c("T","T2")],1,sd)


#Korvans
T2$cov <- cov(T2$T, T2$T2)

#Korelasjon
T2$kor <- T2$cov / (T2$T * T2$T2)
```

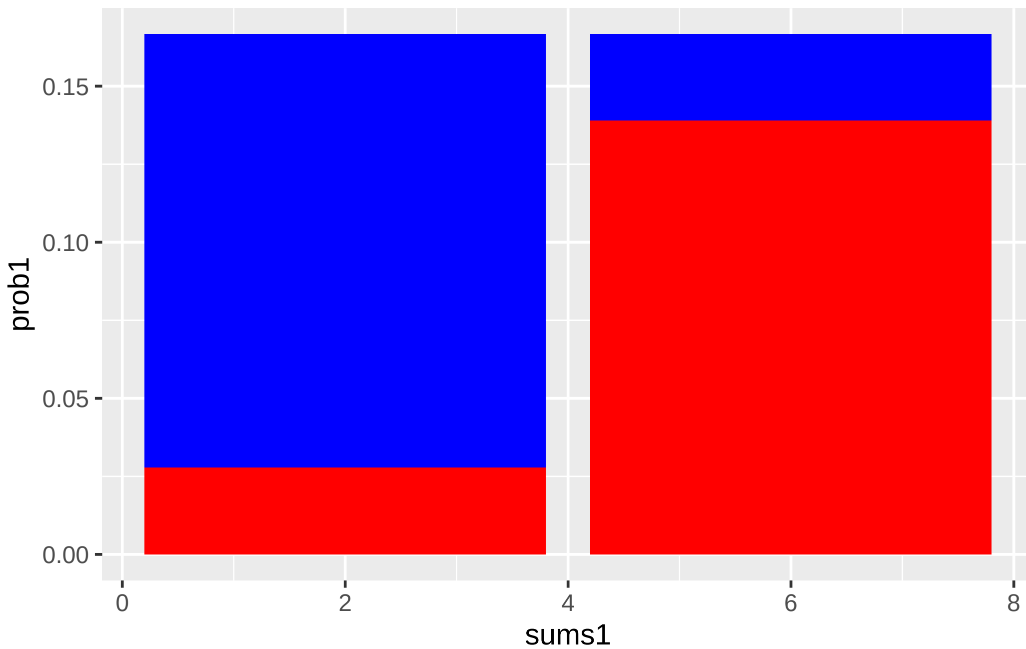
```
T2
```

```
   T T2  var        Sd cov        kor
1  1  2  0.5 0.7071068   3 1.50000000
2  2  3  0.5 0.7071068   3 0.50000000
3  3  4  0.5 0.7071068   3 0.25000000
4  4  5  0.5 0.7071068   3 0.15000000
5  5  6  0.5 0.7071068   3 0.10000000
6  6  7  0.5 0.7071068   3 0.07142857
7  1  3  2.0 1.4142136   3 1.00000000
8  2  4  2.0 1.4142136   3 0.37500000
9  3  5  2.0 1.4142136   3 0.20000000
10 4  6  2.0 1.4142136   3 0.12500000
11 5  7  2.0 1.4142136   3 0.08571429
12 6  8  2.0 1.4142136   3 0.06250000
13 1  4  4.5 2.1213203   3 0.75000000
14 2  5  4.5 2.1213203   3 0.30000000
15 3  6  4.5 2.1213203   3 0.16666667
16 4  7  4.5 2.1213203   3 0.10714286
17 5  8  4.5 2.1213203   3 0.07500000
18 6  9  4.5 2.1213203   3 0.05555556
19 1  5  8.0 2.8284271   3 0.60000000
20 2  6  8.0 2.8284271   3 0.25000000
21 3  7  8.0 2.8284271   3 0.14285714
22 4  8  8.0 2.8284271   3 0.09375000
23 5  9  8.0 2.8284271   3 0.06666667
24 6 10  8.0 2.8284271   3 0.05000000
25 1  6 12.5 3.5355339   3 0.50000000
26 2  7 12.5 3.5355339   3 0.21428571
27 3  8 12.5 3.5355339   3 0.12500000
28 4  9 12.5 3.5355339   3 0.08333333
29 5 10 12.5 3.5355339   3 0.06000000
30 6 11 12.5 3.5355339   3 0.04545455
31 1  7 18.0 4.2426407   3 0.42857143
32 2  8 18.0 4.2426407   3 0.18750000
33 3  9 18.0 4.2426407   3 0.11111111
34 4 10 18.0 4.2426407   3 0.07500000
35 5 11 18.0 4.2426407   3 0.05454545
36 6 12 18.0 4.2426407   3 0.04166667
```

3.

Kan du lage en grafisk fremmstilling fra to utfallsrom til T til ett utfallsrom med 2T .

```
probT <- prob.dist1[c(2,6),]
probT2 <- prob.dist2[c(1,5),]

ggplot() +
  geom_col(data=probT, aes(x=sums1, y=prob1), fill='blue') +
  geom_col(data=probT2, aes(x=sums2, y=prob2), fill='red')
```



4.

For to terningkast, T2 hva er sannsynligheten for ˚a observere at to terninger er like, $t1 = t2$ definer dette somP (A)?

Hva er sannsynligheten for ˚a observere at summen er mellom 7 og 10 definer dette som P (B)?

Hva ersannsynligheten for ˚a observere at summen er 2, 7 eller 8 definer dette som P (C).

```r
#Av 36 totale utfall er det kun 6 utfall som gir like tall.

#1+1, 2+2, 3+3, 4+4, 5+5, 6+6.


'P(A)' <- sum(6/36)*100 # eller 1/6,

#Av 36 instanser ser vi at der finnes 6(7), 5(8), 4(9) og 3(10)

#dette kan vi bevise ved å printe ut instansene av disse
#tallene i de sammenlagte summene.

print(filter(T2, T2 == '7')) # 6
```

```
  T T2  var        Sd cov        kor
1 6  7  0.5 0.7071068   3 0.07142857
2 5  7  2.0 1.4142136   3 0.08571429
3 4  7  4.5 2.1213203   3 0.10714286
4 3  7  8.0 2.8284271   3 0.14285714
5 2  7 12.5 3.5355339   3 0.21428571
6 1  7 18.0 4.2426407   3 0.42857143
```

```r
print(filter(T2, T2 == '8')) # 5
```

```
  T T2  var       Sd cov     kor
1 6  8  2.0 1.414214   3 0.06250
2 5  8  4.5 2.121320   3 0.07500
3 4  8  8.0 2.828427   3 0.09375
4 3  8 12.5 3.535534   3 0.12500
5 2  8 18.0 4.242641   3 0.18750
```

```r
print(filter(T2, T2 == '9')) # 4
```

```
  T T2  var       Sd cov        kor
1 6  9  4.5 2.121320   3 0.05555556
2 5  9  8.0 2.828427   3 0.06666667
3 4  9 12.5 3.535534   3 0.08333333
4 3  9 18.0 4.242641   3 0.11111111
```

```
print(filter(T2, T2 == '10')) # 3
```

```
  T T2  var       Sd cov   kor
1 6 10  8.0 2.828427   3 0.050
2 5 10 12.5 3.535534   3 0.060
3 4 10 18.0 4.242641   3 0.075
```

```
'P(B)' <- sum((6+5+4+3)/36)*100 # 50 % eller 1/2

#2(1), 7(6), 8(5)

print(filter(T2, T2 == '2')) # 1
```

```
  T T2 var        Sd cov kor
1 1  2 0.5 0.7071068   3 1.5
```

```
print(filter(T2, T2 == '7')) # 6
```

```
  T T2  var        Sd cov        kor
1 6  7  0.5 0.7071068   3 0.07142857
2 5  7  2.0 1.4142136   3 0.08571429
3 4  7  4.5 2.1213203   3 0.10714286
4 3  7  8.0 2.8284271   3 0.14285714
5 2  7 12.5 3.5355339   3 0.21428571
6 1  7 18.0 4.2426407   3 0.42857143
```

```
print(filter(T2, T2 == '8')) # 5
```

```
  T T2  var       Sd cov     kor
1 6  8  2.0 1.414214   3 0.06250
2 5  8  4.5 2.121320   3 0.07500
3 4  8  8.0 2.828427   3 0.09375
4 3  8 12.5 3.535534   3 0.12500
5 2  8 18.0 4.242641   3 0.18750
```

```
'P(C)' <- sum((2+7+8)/36)*100 # 47 % eller 17/36
```

Hva er sannsynligheten for at alle tre A B og C skal skje i.e P (A   B   C)?

Hva er sannsynligheten for P (A   B)? Kan du forklare hvorfor du f ˚ar svarene?

#For at betingelsene til (A   B   C) skal kunne oppfylles må vi se hvilke tall innenfor disse rammene kan være med på å oppfylle krvene. Vi vet fra P(A) at det må være 2 like tall, vi vet fra P(B) at summen må være mellom 7 og 10, og i P(C) utelokkes 2 alternativer som gjør at vi kun sitter igjen med en mulighet, nemlig 8.

#Da har vi muligheten 4+4=8.Som regnet ut i et tidligere datasett er 13.88% sjanse for å få.

#For at betingelsene til (A   B   C) skal kunne oppfylles må vi se hvilke tall innenfor disse rammene kan være med på å oppfylle krvene. Vi vet fra P(A) at det må være 2 like tall, vi vet fra P(B) at summen må være mellom 7 og 10, og i P(C) utelokkes 2 alternativer som gjør at vi kun sitter igjen med en mulighet, nemlig 8.

#Da har vi muligheten 4+4=8.Som regnet ut i et tidligere datasett er 13.88% sjanse for å få.

```
'(A   B   C)' <- sum((0.13888889*100)) # = 13.888889
```

#Ved betingelsene til (A   B) trenger vi ikke å utelokke like mange tall, her sitter vi igjen med to muligheter. Vi kan få enten 4+4=8 eller 5+5=10. om vi legger sammen prosenten til begge mulighetene ser vi sansynligheten for A   B.

```
'(A   B)' <-  sum((0.13888889*100)+(0.08333333*100)) # = 22.22222
```