SOK-2012 Innlevering 1

2023-02-21

Oppgave 1.

Tabellen tar utgangspunkt i alle utfallene ved å kaste en mynt 3 ganger. Dette gir 8 forskjellige kombinasjoner av kron og mynt: $2^3 = 8$

De relevante myntkastene vi ser på er alle kast med minst ett kron før siste kast, av de 8 opprinnelige kombinasjonene er det 2 ikke passer inn: TTT og TTH. Av de 6 andre kombinasjonene er det 3 kombinasjoner som oppfyller kravet om en kron før siste kast, men ikke kravet om kron på påfølgende kast: THT, HTT og HTH. Det er en kombinasjon der halvparten av påfølgende kast etter en kron er ny kron: HHT. To kombinasjoner der alle påfølgende kast etter kron er ny kron: THH og HHH.

Om en regner ut andel blant disse 6 kombinasjonene får en:

$(0 + 0 + 0 + 1/2 + 1 + 1)/6 = 5/12$

Oppgave 2.

```
library(tidyverse)
rep <- 1e5
n <- 100
data <- array(sample(c(0,1), rep*n, replace=TRUE), c(rep,n))
prob <- rep(NA, rep)

for (i in 1:rep) {
  heads1 <- data[i,1:(n-1)]==1
  heads2 <- data[i,2:n]==1
  prob[i] <- sum(heads1 & heads2)/sum(heads1)
}

data <- mean(prob)
```

Sannsynligheten for at en observasjon $x_n = 1$ når den forrige observasjonen $x_{n-1} = 1$ blir 0.4949601

Oppgave 3.

```
rep <- 1e5
n <- 100
data <- array(sample(c(0,1), rep*n, replace=TRUE), c(rep,n))
prob <- rep(NA, rep)
intervals <- seq(0, 99, by=1)
mean_probs <- numeric(length(intervals))

for (j in 1:length(intervals)) {
  for (i in 1:rep){
    heads1 <- data[i,1:(n-1)][1:intervals[j]]==1
    heads2 <- data[i,2:n][1:intervals[j]]==1
    n_heads1 <- sum(heads1)
```

```
    if (n_heads1 > 0) {
      prob[i] <- sum(heads1 & heads2, na.rm=TRUE)/n_heads1
        }
      }
    mean_probs[j] <- mean(prob, na.rm=TRUE)
  }

mean50 <- mean_probs[50]
mean10 <- mean_probs[10]
mean5 <- mean_probs[5]
```

Utvalgstørrelsen med spesiell N' gir:

- a. $N' = 50$: 0.4898386
- b. $N' = 10$: 0.4454769
- c. $N' = 5$: 0.4092241

Fra de ulike gjennomsnittene kan en observere at de blir mindre når utvalgsstørrelsen blir mindre. Dette kommer av at sannsynligheten for å observere to 1ere på rad reduseres for færre observasjoner.

d. For å finne hvilke utvalgsstørrelse som gir minimal $P(x_n = 1 | x_n-1 = 1) brukes analyse av *mean_probs*.

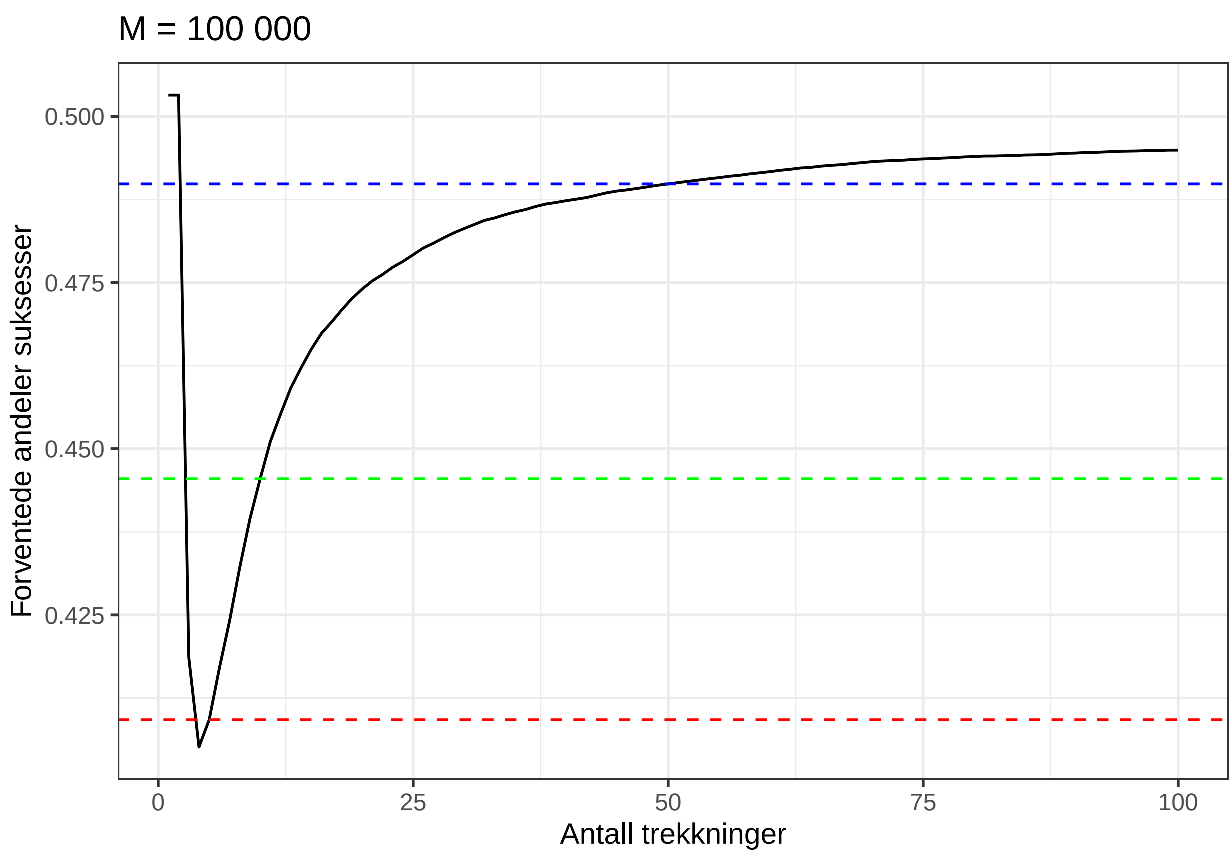
```
lowest <- which.min(mean_probs)
```

Denne funksjonen gir at laveste punkt på grafen er 4

Oppgave 4.

```
table <- data.frame(pulls = intervals+1, success = mean_probs)

ggplot(table, aes(pulls, success)) +
  geom_line() +
  geom_hline(yintercept = mean50, linetype = "dashed", color = "blue") +
  geom_hline(yintercept = mean10, linetype = "dashed", color = "green") +
  geom_hline(yintercept = mean5, linetype = "dashed", color = "red") +
  labs(x = "Antall trekkninger", y = "Forventede andeler suksesser",
  title = "M = 100 000") +
  theme_bw()
```



Oppgave 5.

a.

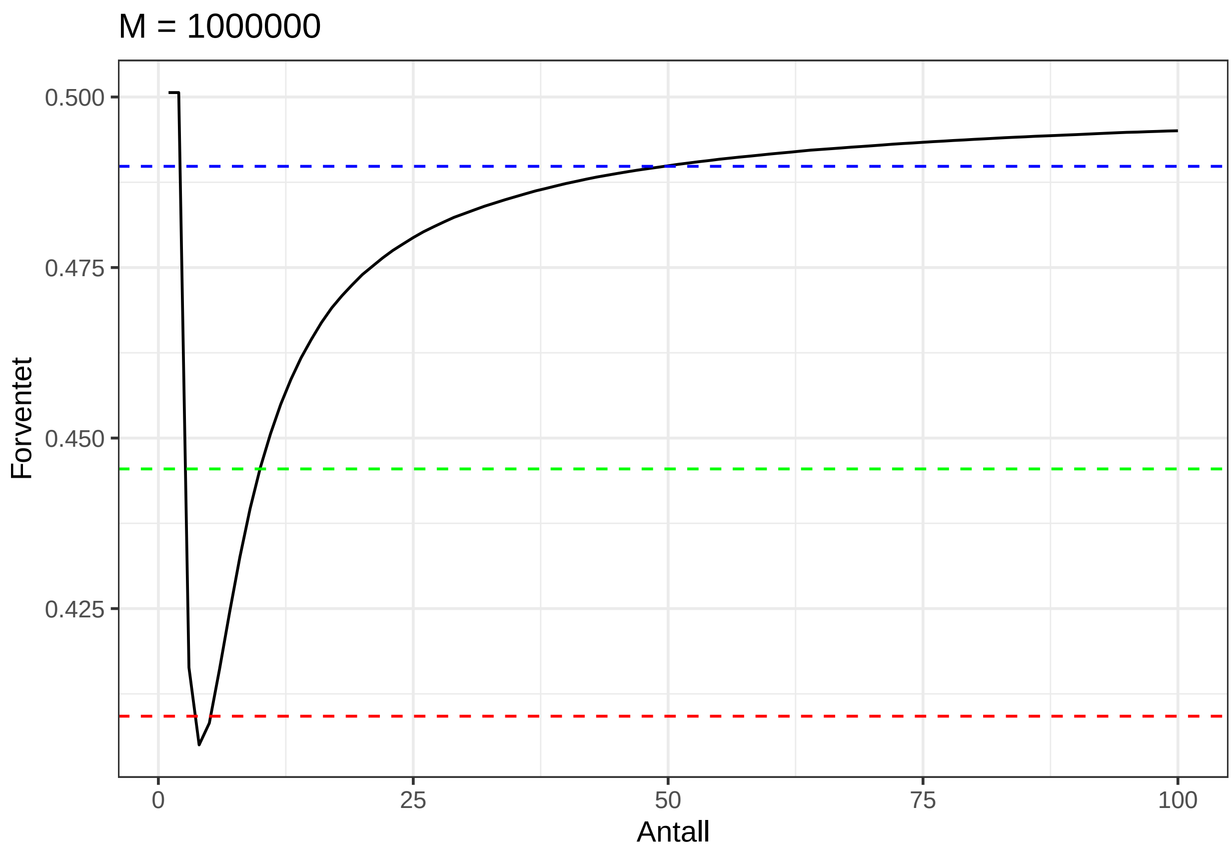
```
rep <- 1e6
n <- 100
data <- array(sample(c(0,1), rep*n, replace=TRUE), c(rep,n))
prob <- rep(NA, rep)
intervals <- seq(0, 99, by=1)
mean_probs <- numeric(length(intervals))

for (j in 1:length(intervals)) { for (i in 1:rep){
  heads1 <- data[i,1:(n-1)][1:intervals[j]]==1
  heads2 <- data[i,2:n][1:intervals[j]]==1
  n_heads1 <- sum(heads1)
  if (n_heads1 > 0) {
    prob[i] <- sum(heads1 & heads2, na.rm=TRUE)/n_heads1 }
    }
  mean_probs[j] <- mean(prob, na.rm=TRUE)
}

table1e6 <- data.frame(pulls = intervals+1, success = mean_probs)

ggplot(table1e6, aes(pulls, success)) +
  geom_line() +
  geom_hline(yintercept = mean50, linetype = "dashed", color = "blue") +
```

```
  geom_hline(yintercept = mean10, linetype = "dashed", color = "green") +
  geom_hline(yintercept = mean5, linetype = "dashed", color = "red") +
  labs(x = "Antall", y = "Forventet",
  title = "M = 1000000") +
  theme_bw()
```



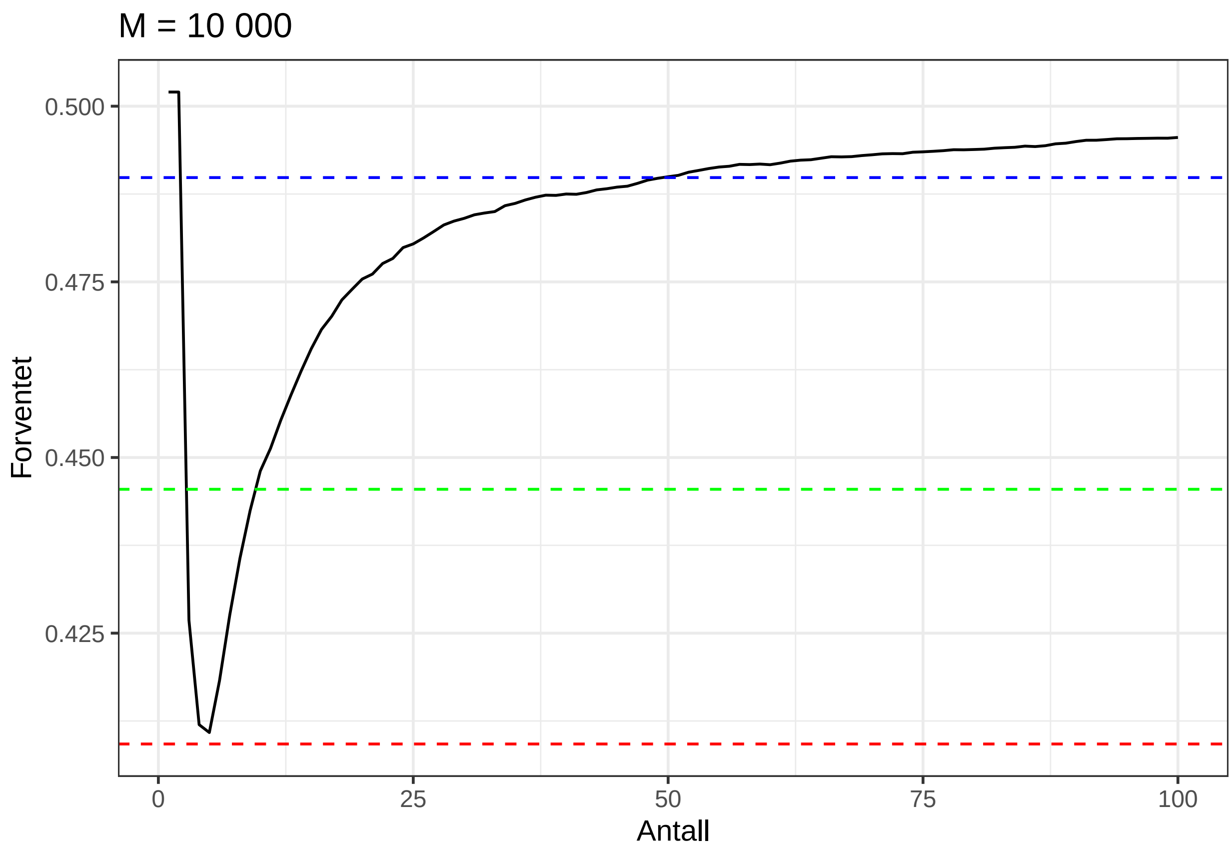
b.

```
rep <- 1e4
n <- 100
data <- array(sample(c(0,1), rep*n, replace=TRUE), c(rep,n))
prob <- rep(NA, rep)
intervals <- seq(0, 99, by=1)
mean_probs <- numeric(length(intervals))

for (j in 1:length(intervals)) { for (i in 1:rep){
  heads1 <- data[i,1:(n-1)][1:intervals[j]]==1
  heads2 <- data[i,2:n][1:intervals[j]]==1
  n_heads1 <- sum(heads1)
  if (n_heads1 > 0) {
    prob[i] <- sum(heads1 & heads2, na.rm=TRUE)/n_heads1 }
    }
  mean_probs[j] <- mean(prob, na.rm=TRUE)
  }
```

```r
table1e4 <- data.frame(pulls = intervals+1, success = mean_probs)

ggplot(table1e4, aes(pulls, success)) +
  geom_line() +
  geom_hline(yintercept = mean50, linetype = "dashed", color = "blue") +
  geom_hline(yintercept = mean10, linetype = "dashed", color = "green") +
  geom_hline(yintercept = mean5, linetype = "dashed", color = "red") +
  labs(x = "Antall", y = "Forventet",
  title = "M = 10 000") +
theme_bw()
```



c.

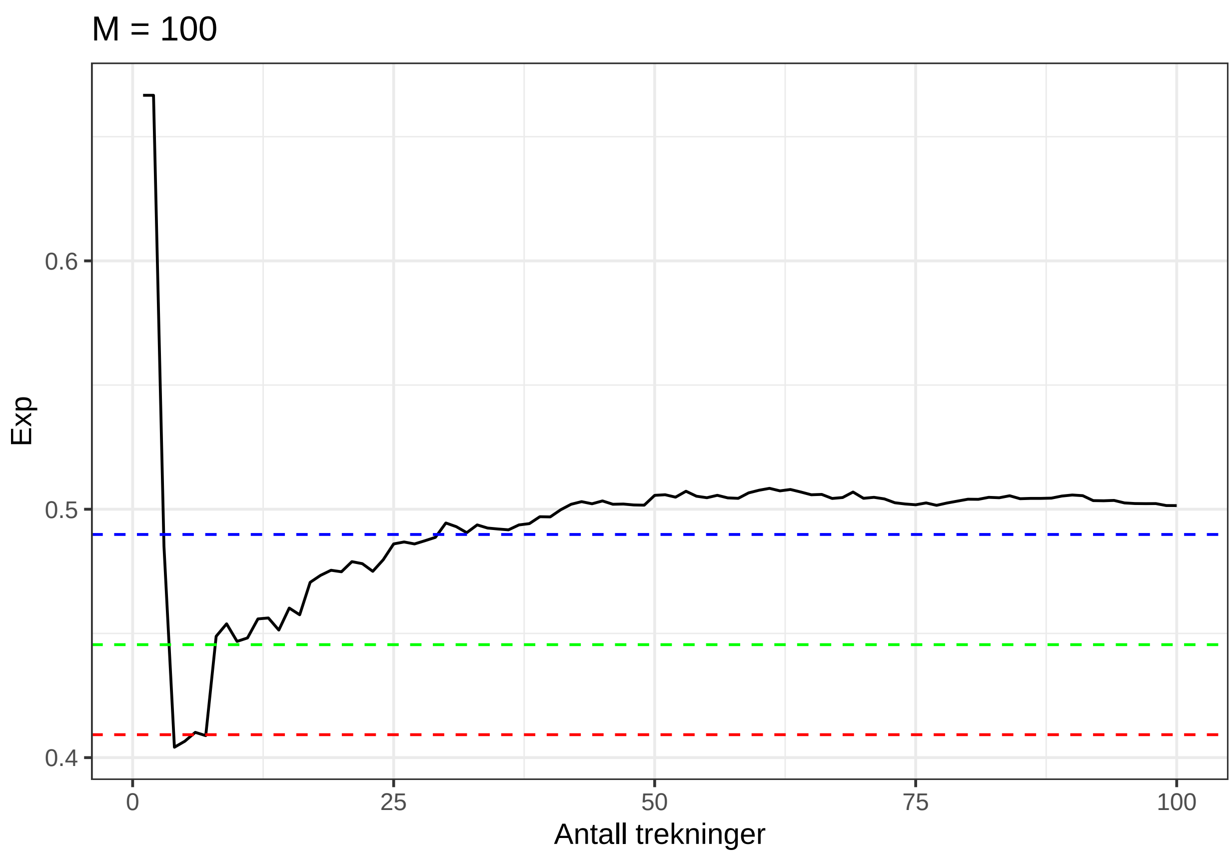
```r
rep <- 100
n <- 100
data <- array(sample(c(0,1), rep*n, replace=TRUE), c(rep,n))
prob <- rep(NA, rep)
intervals <- seq(0, 99, by=1)
mean_probs <- numeric(length(intervals))

for (j in 1:length(intervals)) { for (i in 1:rep){
  heads1 <- data[i,1:(n-1)][1:intervals[j]]==1
  heads2 <- data[i,2:n][1:intervals[j]]==1
  n_heads1 <- sum(heads1)
  if (n_heads1 > 0) {
```

```
    prob[i] <- sum(heads1 & heads2, na.rm=TRUE)/n_heads1 }
    }
  mean_probs[j] <- mean(prob, na.rm=TRUE)
  }

table100 <- data.frame(pulls = intervals+1, success = mean_probs)

ggplot(table100, aes(pulls, success)) +
  geom_line() +
  geom_hline(yintercept = mean50, linetype = "dashed", color = "blue") +
  geom_hline(yintercept = mean10, linetype = "dashed", color = "green") +
  geom_hline(yintercept = mean5, linetype = "dashed", color = "red") +
  labs(x = "Antall trekninger", y = "Exp",
  title = "M = 100") +
  theme_bw()
```



Fra grafene kan en se at flere trekninger fører til mer statistisk nøyaktighet og at grafene blir mer gjevn.

6. Grafene fra Miller og Sanjurjo er data på sansynligheten for å treffe på et basketballskudd avhengig av hvor mange forsøk som gjennomføres. Sannsynligheten for å treffe reduseres avhengig av hvor mange spilleren har truffet på rad.

Grafen i oppgave 4 viser sannsynligheten for å suksess på avhengig av antall trekninger. Grafen i oppgave 4 har kun en sannsynlighet som baseres på hvor mange trekninger som gjøres. Grafen til Miller og Sanjurjo har ulike verdier avhengig av sannsynligheten en spiller har til å treffe samt også hvor mange treff den har på rad.