Laboration 3

Sebastijan Babic

2024-10-26

Uppgift 1

Stora talens lag är en grundläggande idé som säger att då antalet observationer/samples ökar så konvergerar medelvärdet differansen mellan medelvärdet för en slumpvariabel X och dess medelvärde mot 0. Det vill säga konvergerar vi mot medelvärdet givet tillräckligt stort antal observationer n. Matematiskt säger vi givet slumpvariabler $X_1, X_2, \ldots$ som är obereoende och likformigt fördelade med väntevärde μ, då gäller att för varje $\epsilon > 0$ att

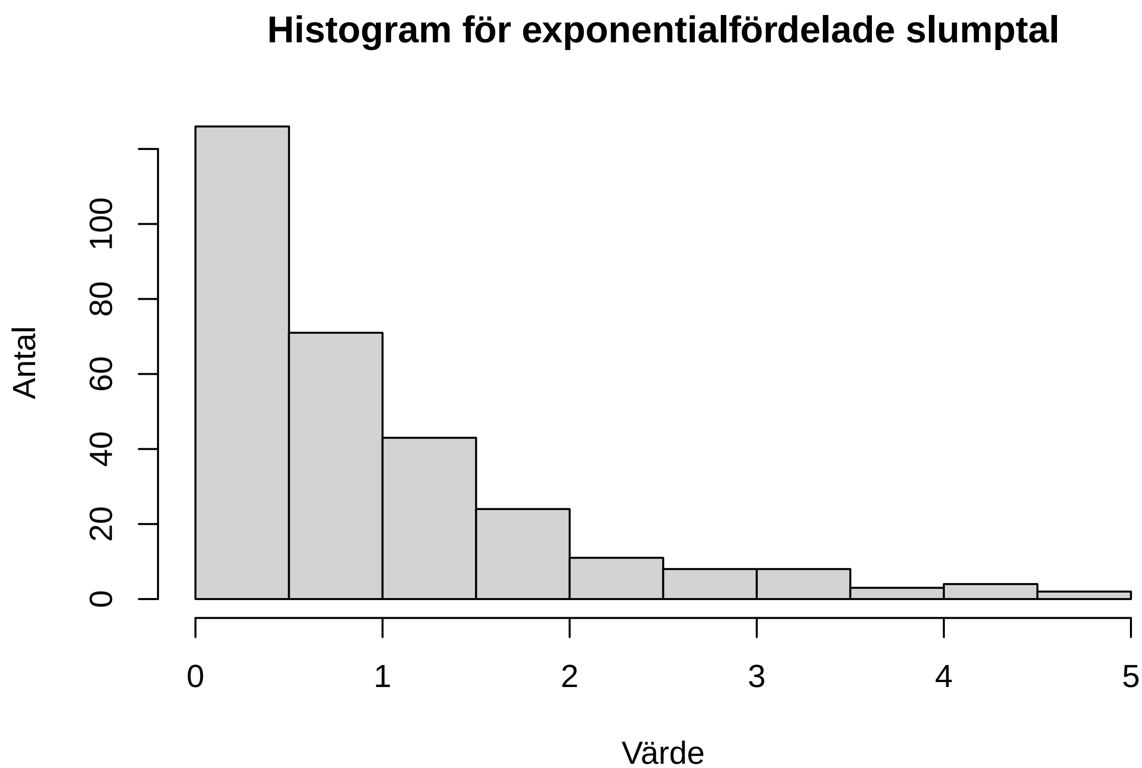
$$P(\mid \overline{X}_n - \mu \mid > \epsilon) \to 0 \qquad n \to \infty$$

där $\overline{X}_n$ är stickprovsmedelvärdet, dvs. $\overline{X}_n = \frac{1}{n}(X_1, \ldots, X_n)$.

```
# Först ger vi ett "frö" (seed) till slumptalsgeneratorn genom funktionen set.seed
# Detta gör att vi får exakt samma slumptal varje gång vi kör koden.
set.seed(200409110) # seed 1


m <- 1 # väntevärde
slumptal1 <- rexp(300, rate = 1/m) # Notera att argumentet rate = 1 / väntevärde.
# Hade vi velat generera slumptal från en exponentialfördelning med
# väntevärde 2 så hade vi alltså skrivit rate = 1 / 2 istället.
medel1 <- cumsum(slumptal1) / 1:length(slumptal1) # ger en vektor med 300 element (se slumptal rexp(300
# cumsum : adderar alla element i en vektor :
# cumsum(1:10)
# [1]  1  3  6 10 15 21 28 36 45 55


hist(slumptal1,
    main = "Histogram för exponentialfördelade slumptal",
    ylab = "Antal",
    xlab = "Värde")
```

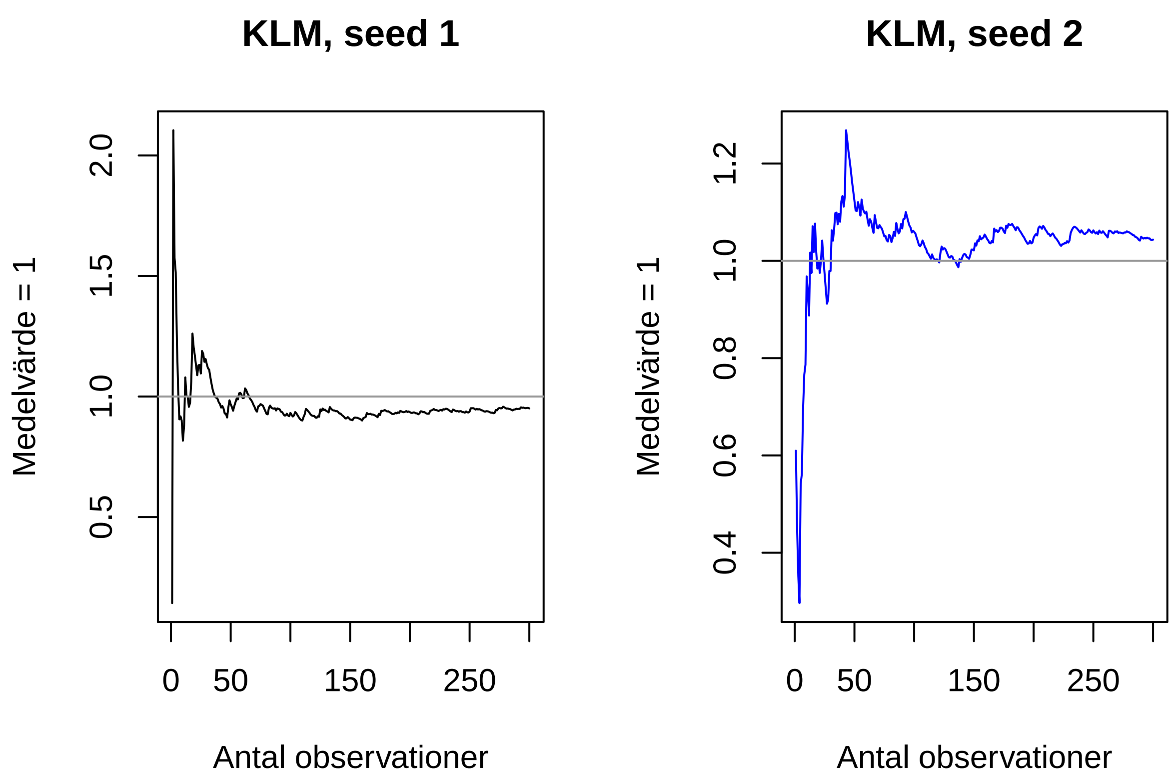
Exponentialfördelning, medelvärde = 1

```
par(mfrow = c(1, 2))
plot(medel1,
    type = "l", # plotta en linje
    main = "KLM, seed 1",
```



Figur 1: Histogram som visar en exponentialfördelning av slumptal.

```
    ylab = "Medelvärde = 1",
    xlab = "Antal observationer")
abline(a = m, b = 0, col = "gray60") # Horisontell linje vid y = 1 = väntevärdet



set.seed(200409111) # seed 2

slumptal2 <- rexp(300, rate = 1/m)
medel2 <- cumsum(slumptal2) / 1:length(slumptal2)

plot(medel2,
    type = "l", # plotta en linje
    main = "KLM, seed 2",
    ylab = "Medelvärde = 1",
    xlab = "Antal observationer",
    col = "blue"
    )
abline(a = m, b = 0, col = "gray60") # Horisontell linje vid y = 1 = väntevärdet
```

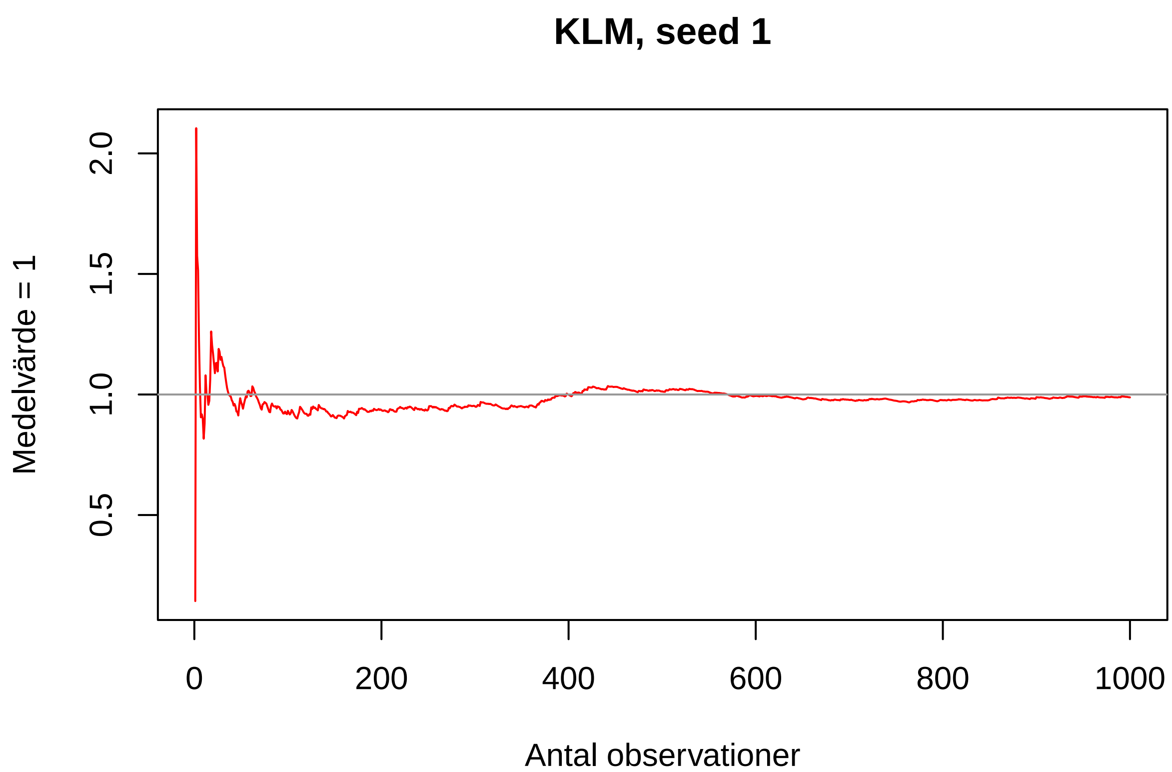


Figur 2: KLM: Kumulativt löpande medelvärde. Figur som visar två medelvärdesplottar hur medelvärdet beter sig givet 300 element av en annan slumpgenererad vektor än innan (höger) respektive samma slumpgenererade vektorn från figur 1 (vänster)

Vi ser att medelvärdet verkar stabilisera sig kring medelvärdet $m = 1$, dvs. den horisonetlla linjen som ritats. Vi har från början en väldigt kaotisk beteende då n litet men då n blir större verkar vi konvergera mot medelvärdet.

Ser en mycket annorlunda beteende ty elementen i vektorn är annorlunda men ändå har vi något som ser ut som konvergerar mot medelvärdet $m = 1$ vilket är det förväntade resultatet.

```
set.seed(200409110) # seed 1

slumptal1 <- rexp(1000, rate = 1/m)
medel3 <- cumsum(slumptal1) / 1:length(slumptal1)

plot(medel3,
    type = "l", # plotta en linje
    main = "KLM, seed 1",
    ylab = "Medelvärde = 1",
    xlab = "Antal observationer",
    col = "red"
    )
abline(a = m, b = 0, col = "gray60") # Horisontell linje vid y = 1 = väntevärdet
```

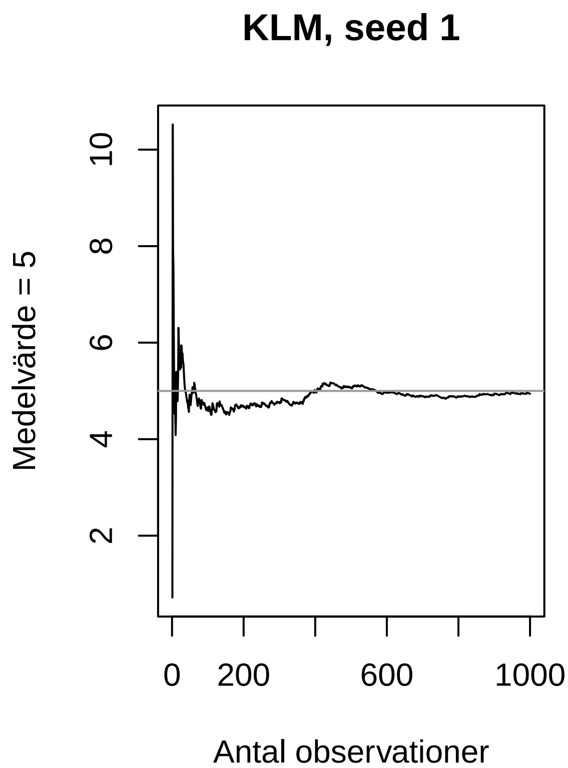


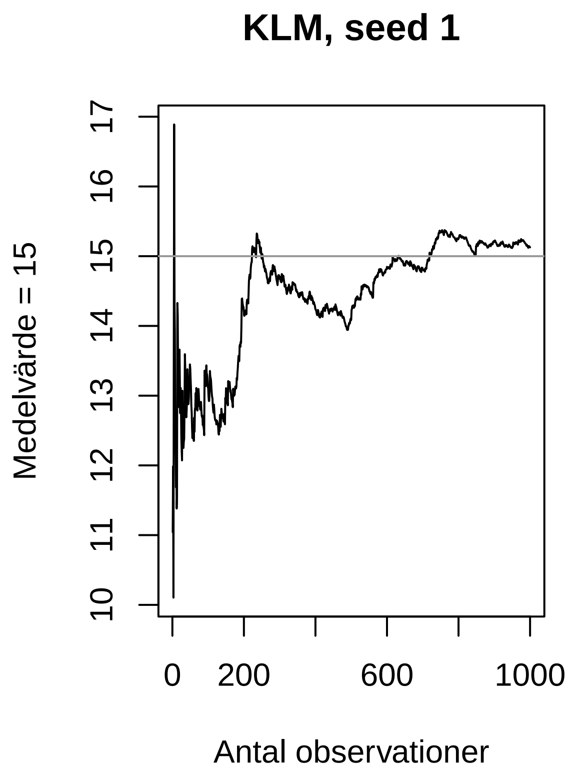
Figur 3: Exponentialfördelad slumptal av storlek n = 1000.

Ser att medelvärdet verkligen verkar konvergera mot medelvärdet $m = 1$ då n blirt stort som är en förväntad resultat ty Stora talens lag.

Exponentialfördelning med medelvärde 5 och 15

```r
set.seed(200409110) # seed 1

slumptal3 <- rexp(1000, rate = 1/5) # m = 5 = väntevärde

medel4 <- cumsum(slumptal3) / 1:length(slumptal3)

par(mfrow = c(1, 2))
plot(medel4,
    type = "l", # plotta en linje
    main = "KLM, seed 1",
    ylab = "Medelvärde = 5",
    xlab = "Antal observationer",
    )
abline(a = 5, b = 0, col = "gray60") # Horisontell linje vid y = 5 = väntevärdet



slumptal4 <- rexp(1000, rate = 1/15) # m = 15 = väntevärde

medel5 <- cumsum(slumptal4) / 1:length(slumptal4)

plot(medel5,
    type = "l", # plotta en linje
    main = "KLM, seed 1",
    ylab = "Medelvärde = 15",
    xlab = "Antal observationer",
    )
abline(a = 15, b = 0, col = "gray60") # Horisontell linje vid y = 15 = väntevärdet
```





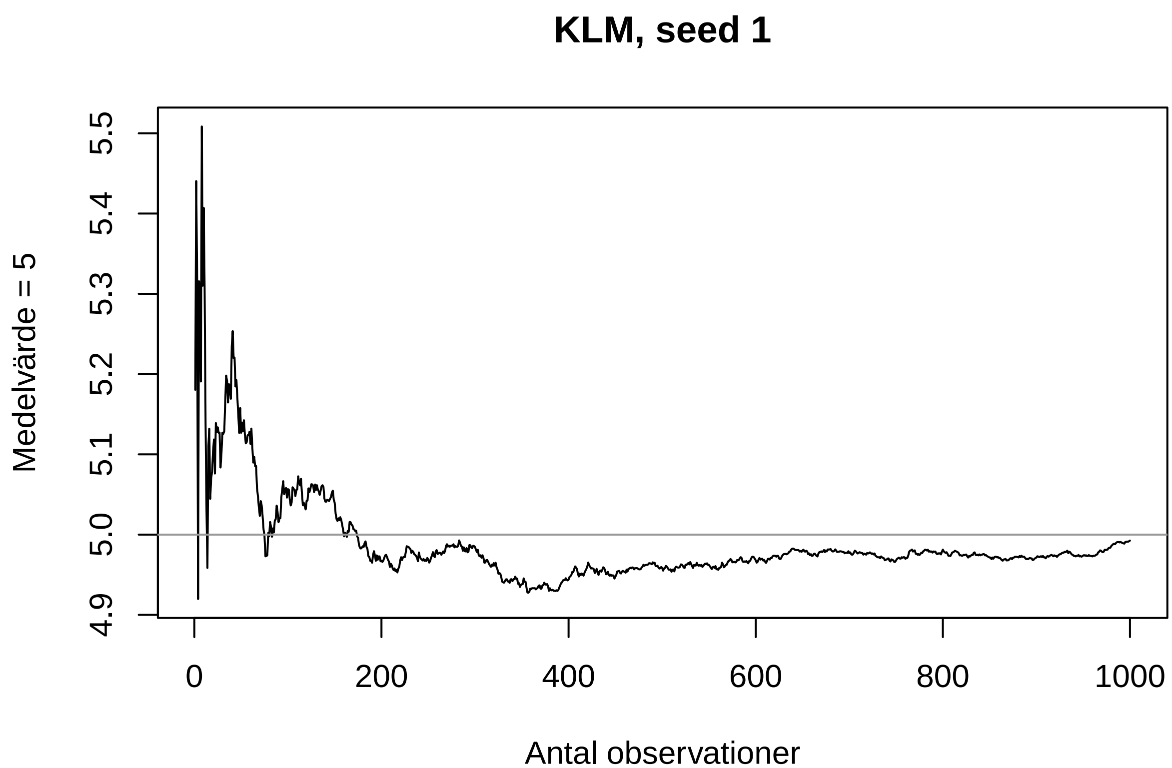
Även här, med ett annat m-värde ser vi konvergens mot just denna m ty Stora talens lag. Vi ser att det eventuellt tar längre att konvergera till $m = 15$ än $m = 5$.

Normafördelningen med medelvärde = 5

```
set.seed(200409110) # seed 1

slumptal5 <- rnorm(1000, mean = 5) # m = 5 = väntevärde, N(0,1)

medel4 <- cumsum(slumptal5) / 1:length(slumptal5)

plot(medel4,
    type = "l", # plotta en linje
    main = "KLM, seed 1",
    ylab = "Medelvärde = 5",
    xlab = "Antal observationer",
    )
abline(a = 5, b = 0, col = "gray60") # Horisontell linje vid y = 5 = väntevärdet
```

Ser återigen konvergens mot medelvärdet m=5. Eventuellt lite långsammare konvergens men ser ändå ut som om medelvärdet konvergerar mot 5.

1. Hur utvecklas medelvärdet för de exponentialfördelade slumptalen?



Figur 4: Medelvärdesplot för en standardnormalfördelning på slumpvariabel med medelvärde = 5

- Svar: Stabilisering av den kumulativa löpande medelvärdet, mer specifikt stabilisering mot det medelvärdet.

2. Vad tycks det konvergera mot?

- Svar: Mot medelvärdet.

3. Gäller samma sak för normal- och likformigt fördelade slumptal?

- Svar: Ja.

Slutsats

Alla diagram som har visats, dvs. figur 2-5 har en gemensam egenskap vilket är att, givet tillräckligt stort n så kommer den kumulativa löpande medelvärdet verkligen att konvergera mot medelvärdet självt. Det är en slutsats som vi kan dra både från det lite experimentella vi har precis gjort eller genom att tänka på Stora talens lag som säger precis det som våra grafer visar resultat för.

Uppgift 2

Centrala gränsvärdessatsen säger att om tillräckligt stort antal observationer n så har vi gränsvärdet

$$P\left(\frac{\sum_{k=1}^{n}\overline{X}_k n\mu}{\sigma\sqrt{n}} \leq x\right) \longrightarrow \Phi(x) \qquad n \to \infty, x \in \mathbb{R}$$

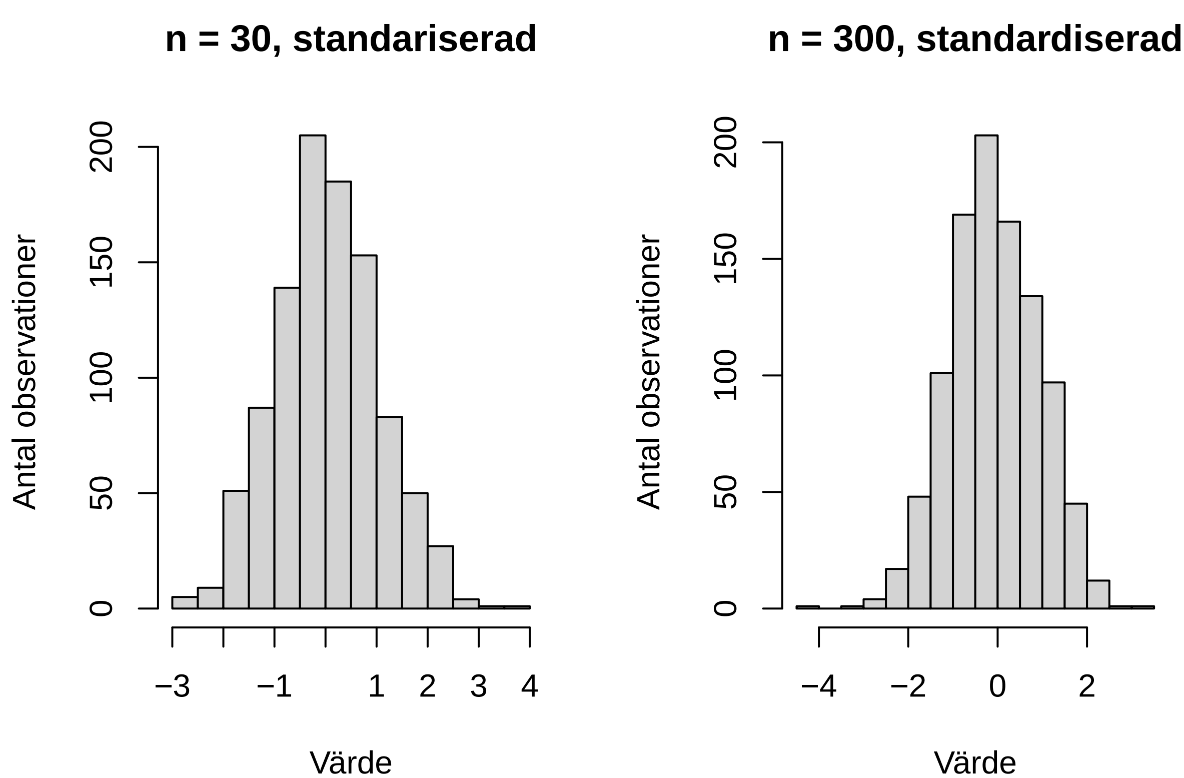
Där $\overline{X}_k$ representerar medelvärdet, dvs $\overline{X}_k = \frac{1}{k}(X_1+, \ldots, +X_k)$, för $k = 1, 2, 3, \ldots$. Med andra ord så kommer vi givet en standariserad slumpvariabel X gå mot en normalfördelning oberoende på den ursprungliga fördelningen givet tillräckligt stort n.

Likformiga slumptal standariserad

Vi skapar först en dataset med 300 element från en likformig fördelning.

```
set.seed(20040911) # seed 0


mu <- 1/2 # väntevärdet
sigma2 <- 1/12 # variansen
M <- 1000 # antalet kolumner

n1 = 30 # antalet rader
likf1 <- runif(30) # 30 observationer
likf_matris1 <- matrix(runif(n1 * M), nrow = n1, ncol = M) # n rader=30, M kolonner=1000
summa1 <- colSums(likf_matris1) # summerade kolonner

par(mfrow = c(1, 2))
stand1 <- (summa1- n1 * mu) / sqrt(n1 * sigma2) # standardisera observationerna
hist(stand1,
    main = "n = 30, standariserad",
    ylab = "Antal observationer",
```

```
    xlab = "Värde")



n2 = 300 # antalet rader
likf2 <- runif(300) # 300 observationer
likf_matris2 <- matrix(runif(n2 * M), nrow = n2, ncol = M) # n rader=300, M kolonner=1000
summa2 <- colSums(likf_matris2) # summerade kolonner

stand2 <- (summa2- n2 * mu) / sqrt(n2 * sigma2) # standardisera observationerna
hist(stand2,
    main = "n = 300, standardiserad",
    ylab = "Antal observationer",
    xlab = "Värde")
```



Ser en normalfördelning fast vi utgår ifrån en likformig fördelning ty vi har standariserad respektive slumpvariabler.

Exponential fördelning standariserad

```
set.seed(20040911)

M <- 1000
```

```r
mu <- 1/300 # Väntevärdet
sigma2 <- 1/300^2 # Variansen


n1 <- 30
exp1 <- rexp(n1, 1/mu)
exp_matris1 <- matrix(runif(n1 * M), nrow = n1, ncol = M)
summa1 <- colSums(exp_matris1)

par(mfrow = c(1, 2))
stand1 <- ((summa1 - n1 * mu) / sqrt(n1 * sigma2)) # n = 30
hist(stand2,
    main = "n = 30, standardiserad",
    ylab = "Antal observationer",
    xlab = "Värde")




n2 <- 300
exp2 <- rexp(n2, 1/mu)
exp_matris2 <- matrix(runif(n2 * M), nrow = n2, ncol = M)
summa2 <- colSums(exp_matris2)

stand2 <- ((summa2 - n2 * mu) / sqrt(n2 * sigma2)) # n = 300
hist(stand2,
    main = "n = 300, standardiserad",
    ylab = "Antal observationer",
    xlab = "Värde")
```

Vi har fortfarande histogram som alla ser normal-fördelade ut fastän vi utgick ifrån en exponentiell fördelning. Ser dessutom att givet att n blir större, så får vi en histogram som liknar en normalfördelning mer och mer.

1. Vilken fördelning verkar de standardiserade summorna av likformiga slumptal få?

Svar: Normalfördelning.

2. Gäller samma sak för exponentialfördelade slumptal?

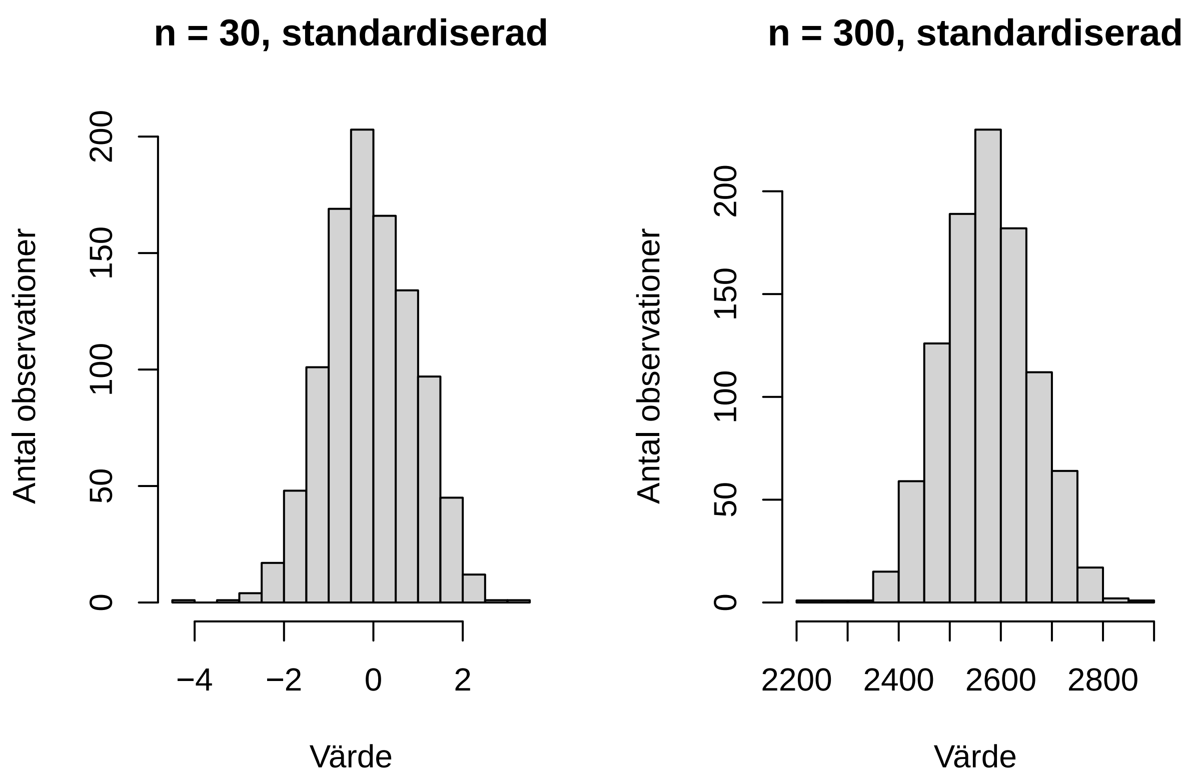
Svar: Ja.

3. Vad händer när ni ändrar antalet slumptal n som summeras?

Svar: Får en histogram som liknar en normalfördelning mer och mer då n blir större.

Slutsats

För större värden av n (t.ex. 300) tenderar fördelningen av summan av likformigt respektive exponentiell fördelade slumptal att närma sig en fördelning mycket likt normalfördelningen, i enlighet med Centrala gränsvärdessatsen. Med andra ord, vid standarisering får vi en normalfördelad slumptal.



Figur 5: Två histogram som visar utseendet på en exponential fördelad slumpvariabel med 30 och 300 element n från vänster till höger.