

Laboration2

2022-11-02

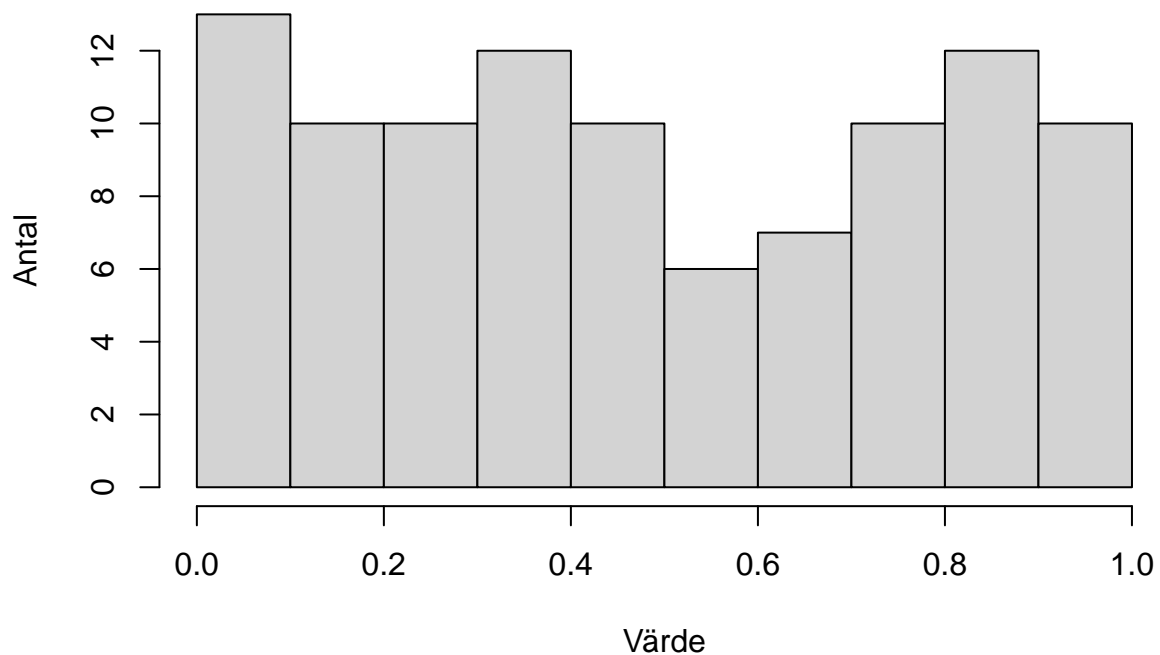
Uppgift 1

1. Eftersom intervallet delas upp i k lika delar så blir första respektive sista delintervallen; $I_1 = [a, \frac{b}{k}]$, $I_k = [\frac{k-1}{k}, 1]$.
2. Längden på varje delintervall blir $1 - (1 - \frac{1}{k}) = \frac{1}{k}$.
3. Låt $P(X_i)$ beteckna sannolikheten av att ett slumpstal hamnar i det i :e intervallet. Då är sannolikheten, eftersom att det är en likformig Bernoullifördelning, för $P(X_1) = \frac{1}{k}$. Alltså blir sannolikheten att det inte hamnar i det första intervallet; $P(X_1^c) = 1 - \frac{1}{k} = \frac{k-1}{k}$. Dessa är också parametrarna i fördelningen, i.e. $p = \frac{1}{k}$, $q = 1 - p = \frac{k-1}{k}$.
4. För $j \in [0, n]$
subset
 \mathbb{N} , $P(x_1 = j) = \binom{n}{j} \cdot p^j \cdot (1 - p)^{n-j}$ Vilket följer binomialfördelningen med parameter $p = \frac{1}{k}$.
5. Ja, antalet slumpstal i de andra delintervallen har samma sannolikhetsfördelning som X_1 , men de är inte oberoende eftersom att antalet dragna slumpstal är givet, därför ändras sannolikheten för hur många som hamnar i önskat intervall beroende på hur många som hamnat i andra intervall. Alltså; $P(X_l | X_m) \neq P(X_l)$ för l och m mellan 1 och k .
6. $Y_1 = \frac{X_1}{n} \implies E[Y_1] = E[\frac{X_1}{n}] = \frac{1}{n} E[X_1] = \frac{1}{n} (np) = p = \frac{1}{k}$
& $V(Y_1) = V(\frac{X_1}{n}) = \frac{1}{n^2} V(X_1) = \frac{1}{n^2} q(np) = \frac{pq}{n} = \frac{p-p^2}{n} = \frac{\frac{1}{k}(1-\frac{1}{k})}{n} = \frac{k-1}{k^2 n}$
 $\implies D(Y_1) = \sqrt{\frac{k-1}{k^2 n}}$

För att visualisera detta kan man börja med att generera slumpstal med hjälp av att så ett så kallat "frö", och därefter plotta ett histogram över slumpstalen. Histogrammet delas in i tio lika stora delintervall och visar hur många av de genererade slumpstalen som hamnat i respektive intervall (stapel).

```
set.seed(19930610)
slumpstal <- runif(100)

hist(slumpstal,
     breaks = seq(from = 0, to = 1, length.out = 10 + 1),
     main = "",
     ylab = "Antal",
     xlab = "Värde")
```



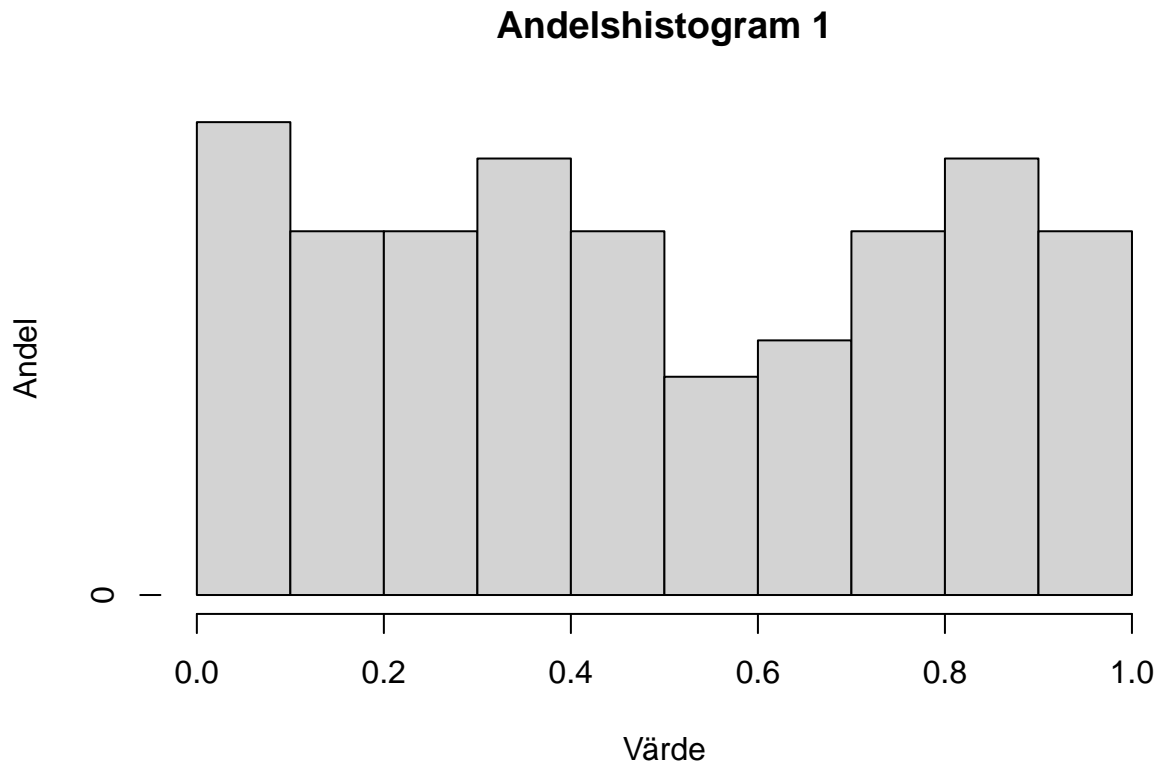
```
seq(from = 0, to = 1, length.out = 10 + 1)
```

```
## [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
```

För att undersöka om andelarna förhåller sig inom ramen av felmarginalen på 2 till 3 standardavvikelser från väntevärdet, så kan detta histogram vara behjälpligt:

```
# x: vektor med slumpstal (eller annan data)
prop_hist <- function(x, xlab = "Värde") {
  p <- hist(x, plot = FALSE)
  p$counts <- p$counts / sum(p$counts) # Konvertera antal till andelar
  plot(p,
    main = "Andelshistogram 1",
    ylab = "Andel",
    xlab = xlab)
}
```

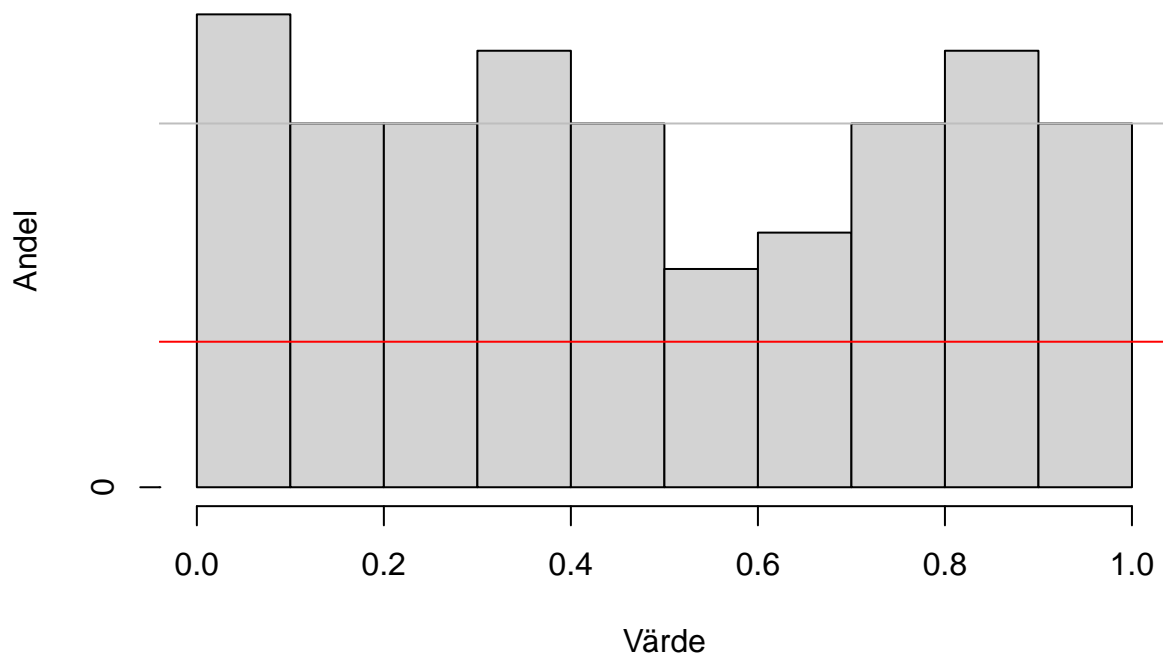
```
prop_hist(slumptal)
```



Standardavvikelsen för $n = 100$ när intervallet är uppdelat i 10 (väntevärde $1/10$) blir; $\sqrt{\frac{10-1}{10^2 \cdot 100}} = 0.03$. Med den informationen kan ännu ett histogram plottas, där även ± 2 gånger standardavvikelsen ifrån väntevärdet tas med för att lätt se om andelarna faller utanför intervallet.

```
E <- 0.1
D <- 0.03
prop_hist(slumptal)
abline(a = E, b = 0, col = "grey") # Väntevärdet
abline(a = E + 2 * D, b = 0, col = "red") # Väntevärdet + 2 standardavvikelser
abline(a = E - 2 * D, b = 0, col = "red") # Väntevärdet - 2 standardavvikelser
```

Andelshistogram 1



För att jämföra mellan olika storlekar på n , så kan exempelvis 10 och 10000 vara lämpliga jämförelser för att se avvikelser/skillnader de emellan:

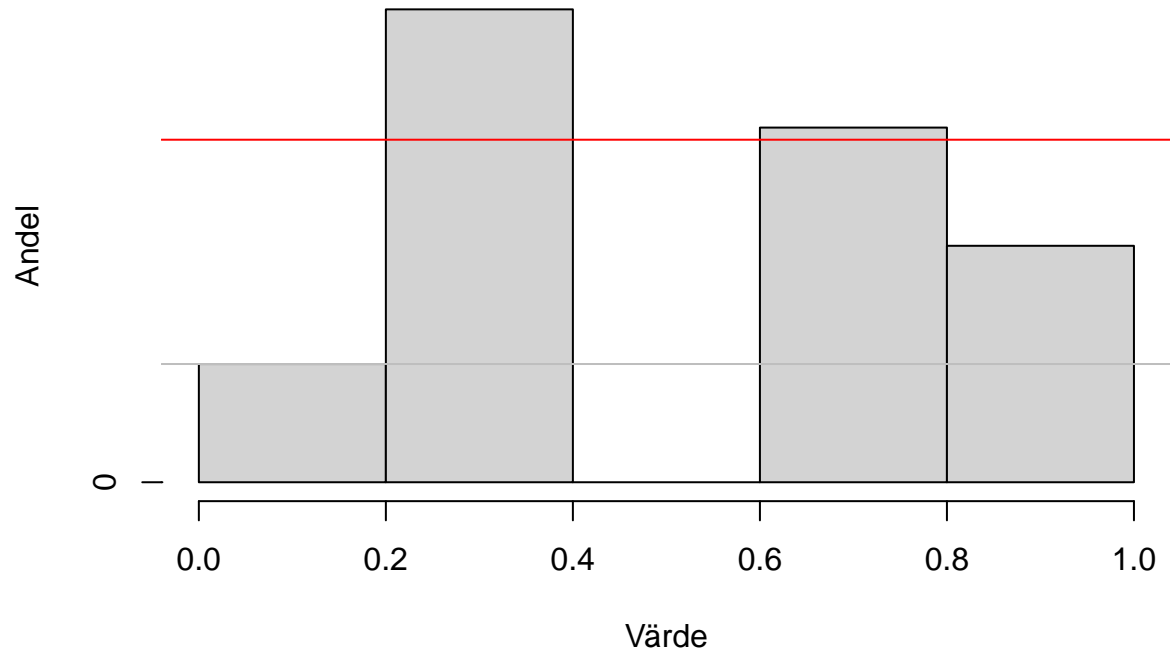
För $n = 10$; $\sigma = \sqrt{\frac{10-1}{10^2 \cdot 10}} = 0.0948683$

För $n = 10000$; $\sigma = \sqrt{\frac{10-1}{10^2 \cdot 10^4}} = 0.003$

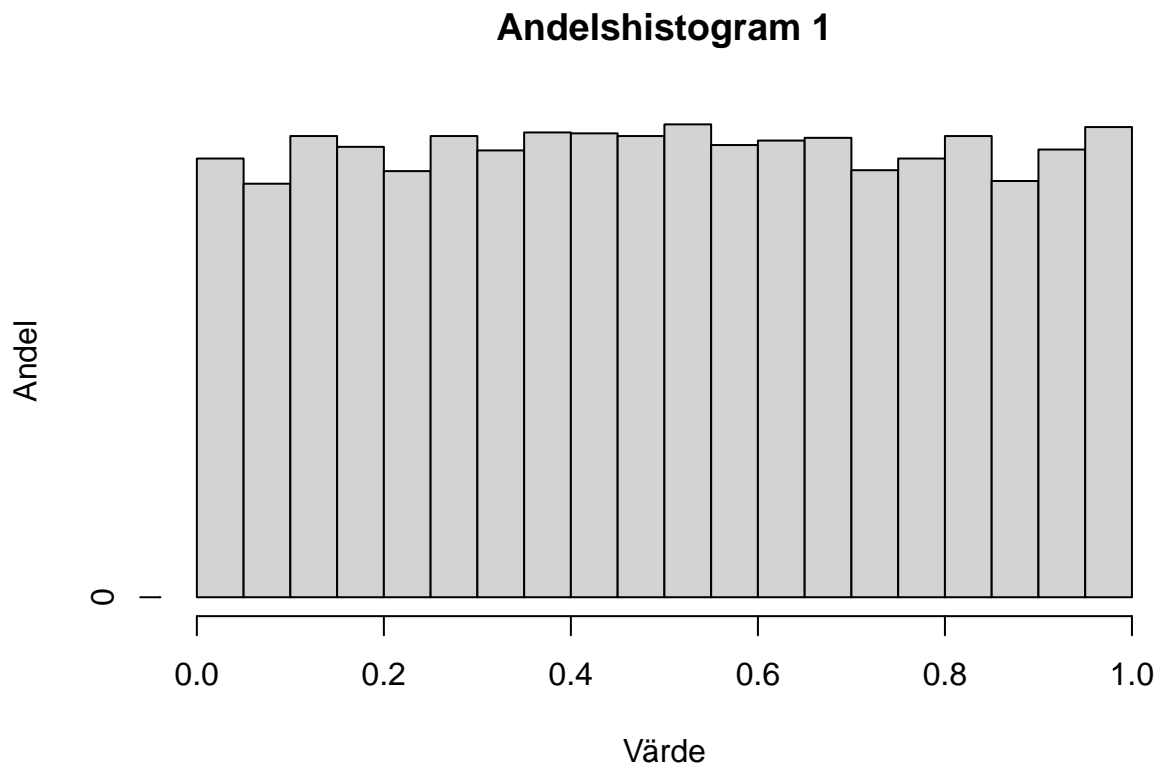
Så histogrammen för $n = 10$ och $n = 10000$ blir då, respektive:

```
set.seed(19930610)
E <- 0.1
D <- 0.0948683
prop_hist(runif(10))
abline(a = E, b = 0, col = "grey") # Väntevärdet
abline(a = E + 2 * D, b = 0, col = "red") # Väntevärdet + 2 standardavvikelser
abline(a = E - 2 * D, b = 0, col = "red") # Väntevärdet - 2 standardavvikelser
```

Andelshistogram 1



```
set.seed(19930610)
E <- 0.1
D <- 0.003
prop_hist(runif(10000))
abline(a = E, b = 0, col = "grey") # Väntevärdet
abline(a = E + 2 * D, b = 0, col = "red") # Väntevärdet + 2 standardavvikelser
abline(a = E - 2 * D, b = 0, col = "red") # Väntevärdet - 2 standardavvikelser
```

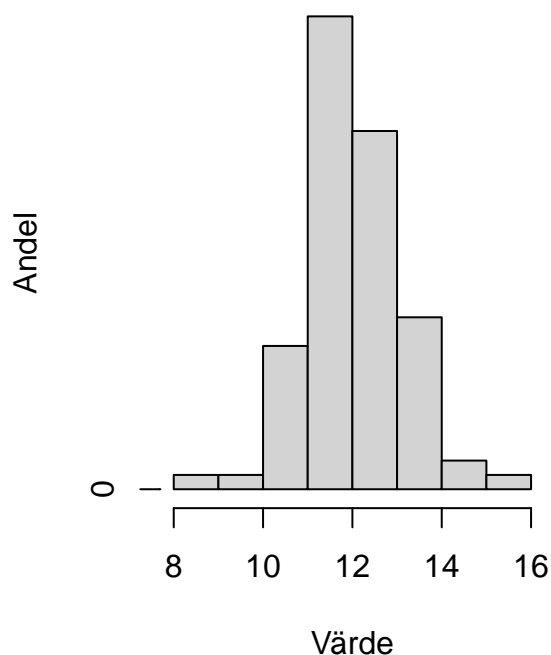


Som man kan se så verkar fördelningen bli jämnare och jämnare ju högre antal genererade slumpstal en använder sig av, och går mer mot en likformig fördelning. Så för ett givet väldigt stort n , ser det ut som en likformig fördelning, som väntat. Med andra ord, en kan även se att avvikelserna blir mindre och mindre desto större n , vilket även den beräknade standardavvikelsen förutspådde. Summa summarum får slumpstalsgeneratorn i R godkänt. # Uppgift 2

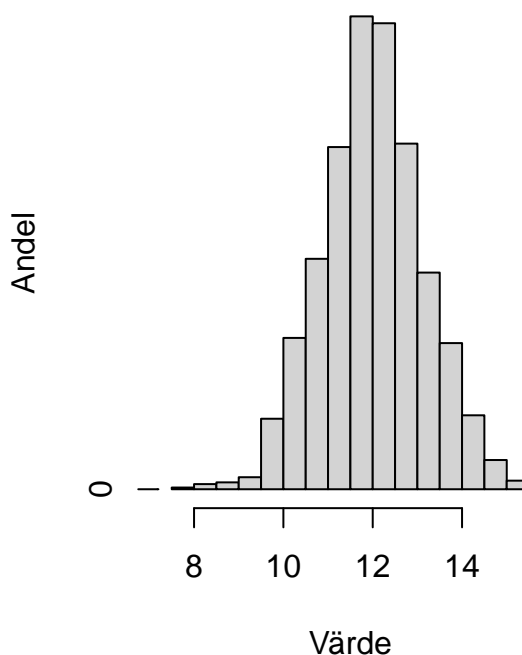
Här plottas två för normalfördelningen, med medelvärde/väntevärde 12 och en varians på 1.2^2 alltså standardavvikelse på 1.2, kurvor där den ena är tänkt att ha 85 observationer av något objekt medan den andra har 1500 observationer. Skillnaden de emellan, samt vad som utmärker de för att vara just normalfördelade diskuteras.

```
set.seed(19930610)
par(mfrow = c(1, 2))
prop_hist(rnorm(85, mean = 12, sd = 1.1)) # n = 85
prop_hist(rnorm(1500, mean = 12, sd = 1.1)) # n = 1500
```

Andelshistogram 1



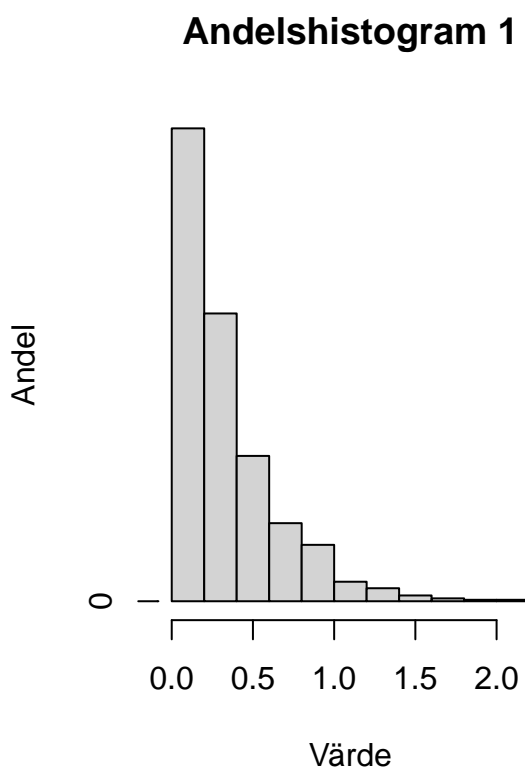
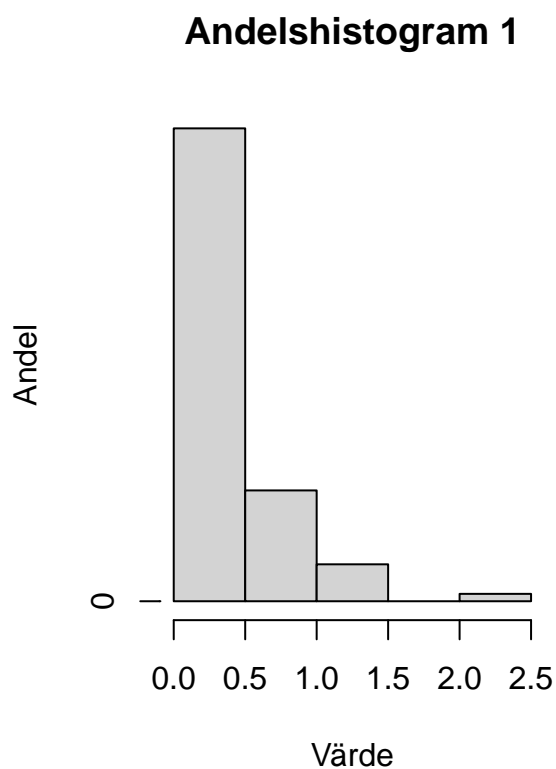
Andelshistogram 1



Dessa två histogram, som ju ska vara normalfördelade, ser också normalfördelade ut, eftersom de båda håller sig symmetriskt kring medelvärde/väntevärde och har samma omega-formade struktur på funktionen som normalfördelningskurvan, täthetsfunktionen. Ju fler observationer desto mer komprimeras kurvan in runt väntevärdet, i x-led, vilket är typiskt för normalfördelningen, medan man kan se att få observationer ger ett mer osäkert underlag för hur den normala fördelningen ser ut.

Här jämförs två exponentialfördelningar med intensitet på 3 vardera.

```
set.seed(19930610)
par(mfrow = c(1, 2))
prop_hist(rexp(85, rate = 3)) # väntevärde 1/3
prop_hist(rexp(1500, rate = 3)) # väntevärde 1/3
```



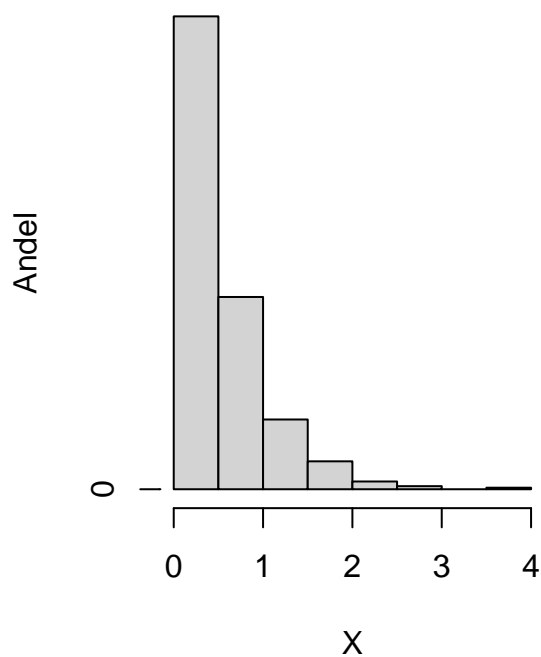
Här kan observeras att histogrammen tenderar mot exponentialfördelningens täthetsfunktion mer och mer jämnt desto större y (mer tid eller vad man nu använder sig av) som anges.

Uppgift 3

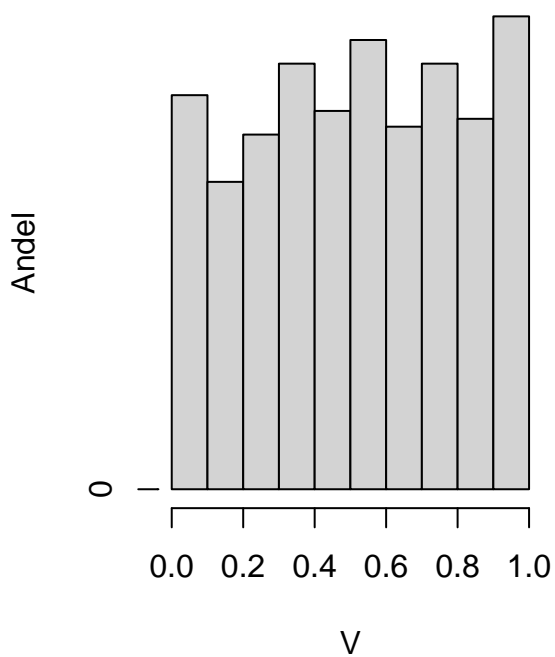
X exponentialfördelad med väntevärde $1/2$, och $V = 1 - \exp(-2X)$:

```
set.seed(19930610)
x <- rexp(500, rate = 2)
v <- 1 - exp(-2 * x)
# Plotta x och v bredvid varandra
par(mfrow = c(1, 2))
prop_hist(x, xlab = "X")
prop_hist(v, xlab = "V")
```


Andelshistogram 1



Andelshistogram 1



X-plotten tenderar mot en kontinuerlig exponentialfunktion för större n , likt exemplen ovan. Bokens påstående verkar också stämma, eftersom vid större n ser v-plotten ut att bli mer och mer lik en likformig fördelning i intervallet $[0,1]$.