Page 363 / Exercise 9a

How large must n be before $S_n = X_1 + X_2 + \ldots + X_n$ is approximately normal? This number is often surprisingly small. Let us explore this question with a computer simulation. Choose n numbers from $[0, 1]$ with probability density $f(x)$, where $n = 3, 6, 12, 20$, and $f(x)$ is each of the densities in Exercise 7. Compute their sum S_n, repeat this experiment 1000 times, and make up a bar graph of 20 bars of the results. How large must n be before you get a good fit?

Probability density functions from Exercise 7:

    a. $f(x) = 1$

Solution

Per note on page 361, X can be simulated using $X = F^{-1}(rnd)$.

For $x \in [0, 1]$,

    a. $f(t) = 1$, $F(X) = \int_0^x 1 dt = x$, $F^{-1}(x) = x$
    b. $f(t) = 2t$, $F(X) = \int_0^x 2t dt = x^2$, $F^{-1}(x) = \sqrt{x}$
    c. $f(t) = 3t^2$, $F(X) = \int_0^x 3t^2 dt = x^3$, $F^{-1}(x) = \sqrt[3]{x}$

Plot experiments with various values for n for each density function.

Density Function $f(x) = 1$ (Uniform Distribution)

```
trials <- 1000

getsum <- function(n, trials) {
  sum <- rep(0, trials)
  for (i in 1:trials) {
    x <- runif(n,0,1)
    sum[i] <- sum(x)
  }
  return(sum)
}

par(mfrow=c(3,2))

n <- 1
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 1")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)

n <- 2
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 2")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)

n <- 3
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 3")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)

n <- 6
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 6")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)

n <- 12
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 12")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)

n <- 20
sum <- getsum(n, trials)
hist(sum, breaks=50, xlim=c(mean(sum)-3*sd(sum),mean(sum)+3*sd(sum)),
     prob=TRUE, xlab="", ylab="", main="n = 20")
curve(dnorm(x, mean=mean(sum), sd=sd(sum)),add=TRUE)
```
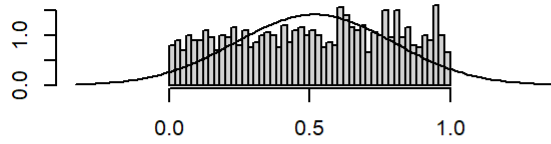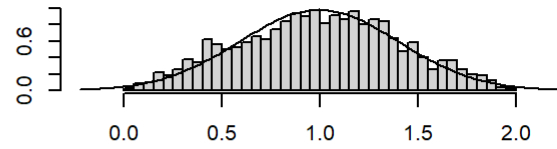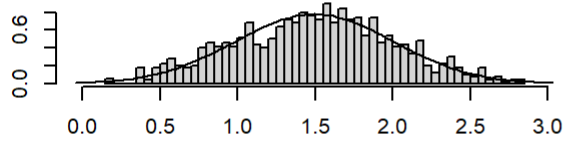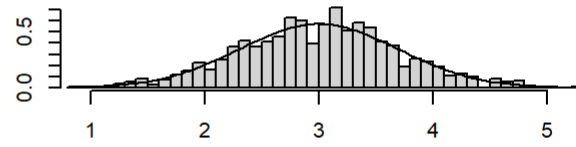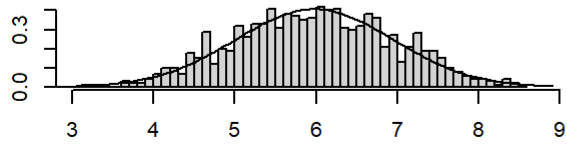
n = 1

n = 2

n = 3

n = 6

n = 12

n = 20