Übung 2

Explorative Datenanalyse und Visualisierung

Wintersemester 2019

S. Döhler (FBMN, h_da)

Name: Valentina Cisternas Seeger & Roman Kessler

Aufgabe 3. *Erzeugen Sie in R die Folge: 1 2 3 1 2 3 1 2 3 1 2 3*

(Folgende R -Befehle können hilfreich sein: `rep, seq`*)*

Lösung

```
rep(seq(3),4)
#>  [1] 1 2 3 1 2 3 1 2 3 1 2 3
```

Anmerkungen/Korrektur

Aufgabe 4. *Compute using R the 0.99 quantile of the standard normal distribution.*

Lösung

```
qnorm(0.99)
#> [1] 2.326348
```

Anmerkungen/Korrektur

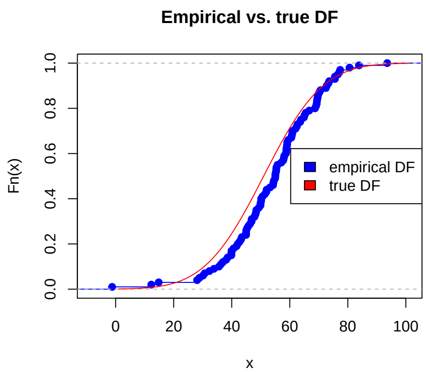
Aufgabe 5. *Simulate 100 observations from $N(50, 16)$. Plot the empirical distribution function for this sample and overlay the true distribution function F.*

(Folgender R -Befehle kann hilfreich sein: `ecdf`*)*

Lösung

```
#100 samples N(50,16)
samples <- rnorm(n=100, mean=50, sd=16)
#empirical distribution function
Fn <- ecdf(samples)
#true distribution function F
```

```
Fnt <- pnorm(seq(from=0, to=100, by=1), mean=50, sd=16)

plot(Fn, col='blue', main="Empirical vs. true DF")
lines(Fnt, col='red')
legend("right",c("empirical DF","true DF"), fill=c("blue","red"))
```



Anmerkungen/Korrektur

Aufgabe 6. *Schreiben Sie eine R -Funktion* `diff.med`*, die zu einem gegebenen Vektor x die Differenz des median und des arithmetischen Mittels ausgibt.*

(a) *Simulieren Sie 100 Beobachtungen von $N(50, 16)$ und wenden Sie* `diff.med` *auf diesen Vektor an.*

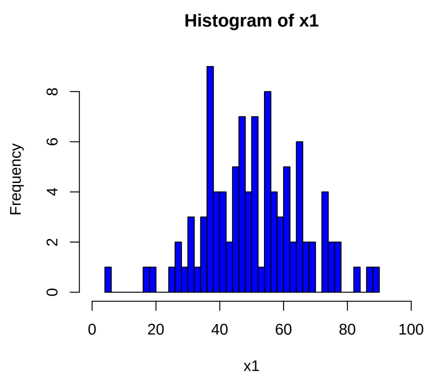
(b) *Simulieren Sie 100 Beobachtungen von $LN(1, 2)$ (log-normal) und wenden Sie* `diff.med` *auf diesen Vektor an.*

Vergleichen Sie die beiden Ergebnisse. Was würden Sie erwarten? Führen Sie ggf. mehr Simulationen durch und untersuchen Sie die Histogramme, die zu den Daten aus (a) und (b) gehören.
(Folgende R -Befehle können hilfreich sein: `mean, median, hist`*)*

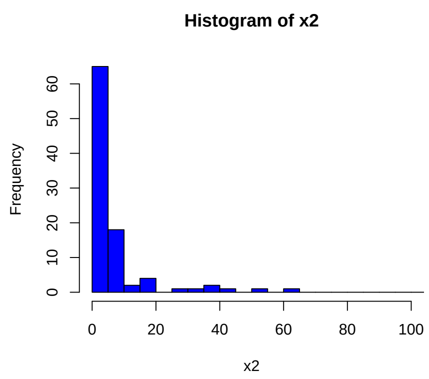
Lösung

Die Differenz zwischen Median und Mittelwert ist vom Betrag her bei der logarithmischen Normalverteilung (LN) größer als bei der Normalverteilung (N). Die Normalverteilung ist annähernd symmetrisch, und Mean und Median sollten ungefähr gleich sein, während die LN rechtsschief ist. **Beobachtung:** Bei einer Logarithmischen Verteilung liegen Median und Mittelwert viel weiter auseinander als bei einer Normalverteilung.

```r
diff.med <- function(x){
  d <- median(x) - mean(x)
  return(d)
}

# part a
x1 <- rnorm(n = 100, mean = 50, sd = 16)
diff.med(x1)
#> [1] -0.5182871

# part b
x2 <- rlnorm(100, meanlog = 1, sdlog = 2)
diff.med(x2)
#> [1] -10.8165

hist(x1, breaks = 50, col = "blue", xlim = c(0,100))
```



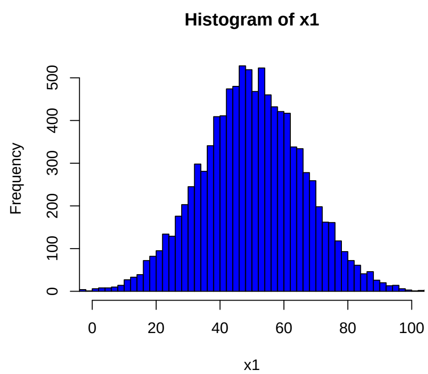

```r
hist(x2, breaks = 50, col = "blue", xlim = c(0,100))
```



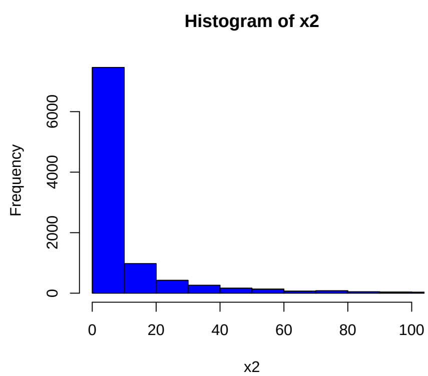
Hier wiederholen wir das ganze nochmal

mit mehr Simulationen:

```r
# part a
x1 <- rnorm(n = 10000, mean = 50, sd = 16)
diff.med(x1)
#> [1] -0.1622742

# part b
x2 <- rlnorm(10000, meanlog = 1, sdlog = 2)
diff.med(x2)
#> [1] -15.80907

hist(x1, breaks = 50, col = "blue", xlim = c(0,100))
```




```r
hist(x2, breaks = 500, col = "blue", xlim = c(0,100))
```



Anmerkungen/Korrektur

Aufgabe 7. *Ein Risiko V hängt von einem anderen Risiko X folgendermaßen ab:*

$$V = e^X + 2X^2 \qquad \text{mit} \qquad X \sim N(0,1).$$

Schreiben Sie eine R -Funktion `Sim.V`*, die zu einem gegebenen Niveau $\alpha \in (0,1)$ und gegebener Anzahl N_{Sim} von Simulationen von V das α-Quantil der Verteilung von V schätzt.*

(Folgende R -Befehle können hilfreich sein: `quantile`*)*

Lösung

```
Sim.V <- function(alpha, Nsim){
  X <- rnorm(Nsim, mean = 0, sd = 1)
  V <- exp(X) + 2 * X**2
  q <- quantile(V, probs = alpha)
  return(q)
}
```

Hier ein Beispiel:

```
Sim.V(c(0.01,0.1,0.25,0.5,0.75,0.9,0.99), 100000)
#>         1%        10%        25%        50%        75%        90%
#>  0.8991021  0.9388150  1.1576547  2.0634207  4.3571196  8.3094362
#>        99%
#> 21.1760934
```

Anmerkungen/Korrektur