Exercise Sheet 4

Familiarity with R is a prerequisite for solving these exercises – please do not start on them before you have completed the tutorial!

1. Old Faithful, a geysir in Yellowstone National Park, is one of the best-known hot springs. As such, the time between eruptions and the duration of these eruptions are of great interest to both spectators and the National Park Service alike.

The file http://stat.ethz.ch/Teaching/Datasets/geysir.dat contains measurements taken from August 1st to 8th, 1978, with 3 columns indicating the day ("Tag"), waiting time ("Zeitspanne") and duration ("Eruptionsdauer") for eruptions.

a) Plot histograms of the time between successive eruptions:

Was is evident from these plots? How do the three histograms differ? **Remark:**

When the number of classes for a histogram is given by **breaks=20**, this is treated merely as a "suggestion" which may still be changed internally.

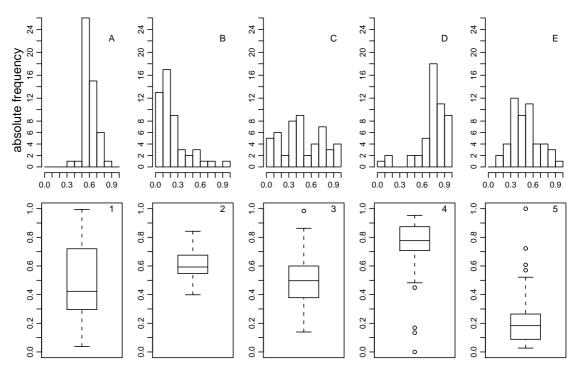
- b) Draw histograms (varying the number of classes) of the duration of eruptions.
 > hist(geysir[,"Eruptionsdauer"], ...)
 What do you notice here? Compare this to the first part of the question.
- 2. In a random experiment, 3 dice are rolled simultaneously. Instead of analyzing this experiment as we did in Problem 1 of Exercise Sheet 1, we would now like to acquaint ourselves with some of its properties by means of simulation.
 - a) Use R to simulate 100 samples of 3 dice throws each, compute their sum each time, and save the results in vectors die1,die2,die3 and diceSum.

b) Plot a histogram of the dice sums thus generated, and compute their average and standard error (whose corresponding theoretical quantities are the expectation and standard deviation).

```
> hist(diceSum,breaks=2.5:18.5,freq=FALSE,ylab="rel. frequency")
```

- > mean(diceSum); sd(diceSum)
- c) Raise the number of samples to 10000 and repeat steps a) and b). What do you notice here?

- d) A casino offers the following game: Three dice are rolled simultaneously; if the sum of dice is greater than 12, the gambler wins \$ 12, otherwise, he loses \$ 1. Simulate 100 repeats of this game and compute the average gain to the gambler.
 > gain <- ifelse(diceSum>12,2,-1)
 > mean(gain)
- e) Plot a sample of how the gain to the gambler develops over the course of 100 repeats of this game. Any comment?
 - > cuGain <- cumsum(gain)</pre>
 - > plot(1:100,cuGain,xlab="Game no.",ylab="Total gain")
- **3.** A histogram and a box plot are drawn for each of five samples of size n = 100. Assign each box plot to the histogram of the same sample. Justify each assignment!



Cumulative gain