Introduction to Bayesian Data Analysis

Assignment 1

Linus Hof

Info

The assignment comprises of 10 tasks. Tasks 1-4 each give 1 point. Tasks 5-10 each give 2 points (correct approaches and form give points). In total, 16 points can be obtained.

By the end of the deadline (17 March, 8pm), upload one `R` script that produces the answers to Moodle:

- Indicate which task the code belongs to by adding a line `#tasknumber` over the code for the respecitve task.
- `# comment` your code by describe what the code is supposed to do.
- At the beginning of the script, specify which packages you used (`library(package)`)

2 Dice

Assume you throw 2 dice.

1. Create 2 vectors representing the sample spaces of both dice throws. (1 point)
2. Use the 2 vectors and the `expand_grid()` function from the *tidyverse* package to generate a data frame representing all 36 possible outcomes of throwing *both* dice. Each row should represent one of the possible combinations, i.e.: (1 point)

```
  d1  d2
1  1   1
2  1   2
3  1   3
4  1   4
5  1   5
6  1   6
7  2   1
```

```
8   2   2
9 ... ...
```

3. Add a column to the data frame that specifies the probability of each possible outcome (combination). (1 point)
4. Add a column to the data frame that specifies the sum across both dice for each possible outcome (combination). (1 point)

Tip for Tasks 5-7: Use the function `subset()` (or `filter()` from the *tidyverse* package) to create subsets of the data.

5. Compute the probability that the sum is ≥ 7 , *given* the first dice shows `3`. (2 points)
6. Compute the probability that the sum lies between 4 and 9. (2 points)
7. What is the probability of the most probable sum? (2 points)

Probability of Delay

Assume you take 10 trains, each of which is either delayed ("L") or on time ("O"). Assume the possible probabilities of delay are p=0, p=0.1, p=0.2, p=0.3, p=0.4, p=0.5, p=0.6, p=0.7, p=0.8, p=0.9, p=1.

8. For each possible probability of delay, compute the probability distribution over all possible numbers of delays using the `dbinom()` function. (2 points)

Bonus: Store all probability distributions in one data frame, where columns are the possible numbers of delay and rows are the possible probabilities of delay (or vice versa).

9. Use the code below to generate train ride data. Use the `dbinom()` function to compute the likelihood of the data (observed number of delays), given p=0, p=0.1, p=0.2, p=0.3, p=0.4, p=0.5, p=0.6, p=0.7, p=0.8, p=0.9, p=1. (2 points)

```r
sim_rides <- function(N, p){
  sample(c("L", "O"), size=N, replace=TRUE, prob=c(p, 1-p))
}

set.seed(1237)
obs <- sim_rides(10, .3)
obs
```

```
[1] "O" "L" "O" "O" "O" "L" "L" "O" "L" "L"
```

10. Assume the below prior probabilities. Use the Bayes' rule to calculate the posterior probability using the prior probabilities and the likelihoods that you computed. (2 points)

```
prior <- c(0.000, 0.004, 0.041, 0.123, 0.209, 0.246, 0.209, 0.123, 0.041, 0.004, 0.000)
```