Statistics in Matlab (Evaluated Task Set 2)

Soft deadline: 2020-05-05 00:00:01. After this date there will be 10% penalty for each week late.

Hard deadline: 2020-05-30 00:00:01. After this date no submission will be accepted for evaluation.

Submission requirements: Same as for the Evaluated Task Set 1.

Due to the quarantine you will have to submit your solutions through e-learning platform. Archive all files in "Surname-ETS2" to a single zip archive (named "Surname-ETS2.zip"). Note that you might get some follow up questions before your submission is evaluated. Be sure to answer the questions promptly.

Evaluation criteria: Is the submission on time? Does submission follow the requirements? Is the code clear and does it follow good coding practices? Are the input prompts proper? Is the output properly formatted? How well the task is performed? How well the author can explain his code/idea/approach? Is the approach reasonable? How well the author can answer any additional questions regarding the task? Note that in this task set you are already encouraged to be as "efficient" in "Matlab way" as you can.

Tasks

1. Let X_i be a result of a roll of a standard six sided dice. Let X_1, X_2, X_3 and X_4 be results of independent rolls.
   - What is the distribution of $Y = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$?
   - What is the distribution of $Z = \min(X_1, X_2, X_3, X_4)$?
   - What is the distribution of $W = \max(X_1, X_2, X_3, X_4)$? How does it compare to the distribution of $7 - Z$?
2. Let X be a standard Gaussian random variable (with $\mu = 0$, $\sigma = 1$).
   - Estimate the probability that $X \in [-1, 1]$. Compare simulation and theoretical results. To get theoretical results you use a built–in function for the cumulative distribution function of the Gaussian distribution. Print the probability out in a nicely formatted sentence.
   - How does $P(X \in [-1, 1])$ depend on μ? Provide a plot showing how $P(X \in [-1, 1])$ depends on μ. Show both simulation and theoretical curves. Pick your own reasonable bounds for μ.
   - How does $P(X \in [-1, 1])$ depend on σ? Provide a plot showing how $P(X \in [-1, 1])$ depends on σ. Show both simulation and theoretical curves. Pick your own reasonable bounds for σ.
   - What is the mean and standard deviation of X^2. How do they depend on μ and σ^2 of X? Provide a plot showing how the mean of X^2 depends on μ. Provide a plot showing how the standard deviation of X^2 depends on σ. Pick your own reasonable bounds for μ and σ.
3. Initially place 1 red and 1 black ball into the box. Then repeatedly draw a ball from the box and observe ball's color. After observing color place the ball back into the box and also place another ball of the same color into the box. For example, if the first draw would have been a red ball, then before the second draw the box would contain 2 red and 1 black balls. What is the distribution of the number of red balls, R, after 10 draws? After 100 draws? Provide two plots of $P(R)$.
4. **(this task counts as 2 tasks)** "engsoccerdata" is a package created for football fans who want to learn R language. It is available from CRAN and GitHub (find the repository here: https://github.com/jalapic/engsoccerdata). While most of it is useless to us, but GitHub repository also has raw data in CSV file format (see "data-raw" directory). Pick and download one of these files stored in "data-raw" (each corresponds to data from top

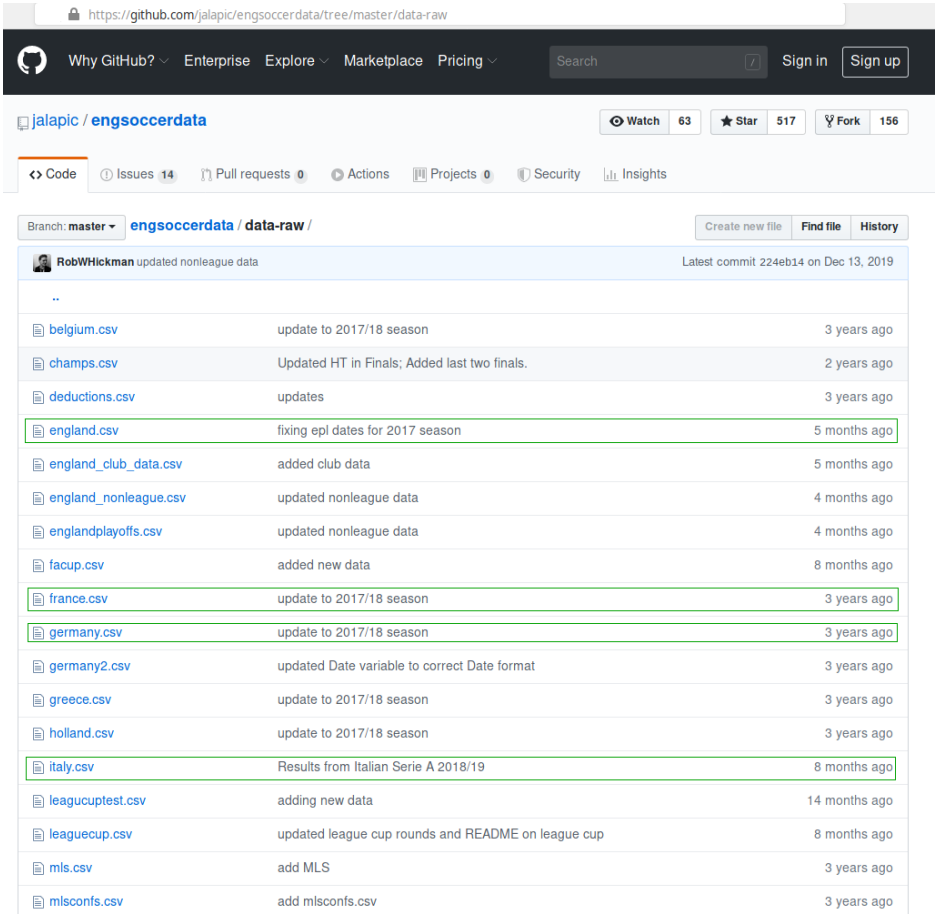

Figure 1: Screenshot of the "data-raw" folder with four of five files for your choice highlighted.

European football league): “england.csv”, “france.csv”, “germany.csv”, “italy.csv” or “span.csv” (see Fig. 1 for the screen you’ll likely see when choosing what to download). Write a program, which could be used to generate a nicely formatted report (recall lecture 3), which would answer the following questions:

- For the games in this millennium (since January 1st, 2000) in the top division:
  - Which game was the top scoring game?
  - How many unique teams has played games?
  - Which team has scored the most goals? Conceded the most goals? Has the best goal differential (scored minus conceded)? Has the best goal ratio (scored divided by conceded)?
  - Which team has scored the most goals on average? Conceded the most goals on average?
  - Do home teams have an advantage?
  - How many games ended as 0:0 draw? In how many games one team has scored more than 3 goals? In how many games loosing team has scored more than 3 goals?
- Which teams have kept their place in the top division through out all seasons of this millennium (take into account only those seasons which started after January 1st, 2000)? Provide plots comparing those teams in the following regards:
  - Cumulative points scored.
  - Cumulative goals scored.
  - Cumulative goals conceded.
  - Cumulative goal differential.
- How would the last (available in the data set) season’s final table change if we would use a different scoring method? Instead of the usual 3 points for victory, 1 point for draw and 0 for loss, use the following:
  - Winning team gets 10 points for the victory itself and a bonus 1 point for each goal scored in the game (by both teams). For example, team A has beaten another team 4:3, then team A gets 17 points.
  - Loosing team gets 1 point for the loss itself and a penalty of 1 point for each additional goal winning team has scored on top of the goals scoring by loosing team. For example team B has lost to another team 0:3, then team B gets -2 points.
  - Teams who drawn get 3 points for the draw itself plus a bonus 1 point for each goal scored. Unless it was a 0:0 draw, then both teams get 1 point. For example, teams C and D played 2:2 game, then both teams get 5 points each.