University of Manitoba
Department of Statistics

STAT 4600: Computational Statistics

Take-home Part of the Final Exam

Instructor: Alexandre Leblanc
Date: Thursday, April 12, 2018.
Due date: Wednesday, April 18, 12:00 pm (noon).

Instructions:

- This is an exam: there should be absolutely no collaboration between students.
- Show as much detail as you can: please pay attention to the quality and clarity of the work you turn in.
- Pay particular attention to the clarity of your R code.
- Justifications/explanations should be short and clear. They should not be very lenghty.
- All work is to be turned in electronically and done within RMarkdown.
- The take-home part of the exam contains four questions on nine pages (including the cover page).

Good Luck!

Question 5: Refer to Question 2 from the first part of the Exam.

Consider once again the integral

$$I = \int_{-1}^{1} (1+x)\log(1+x)\,dx,$$

and recall that our goal was to approximate this integral using different Monte Carlo schemes.

In each of the following cases, approximate the sampling distribution of the resulting Monte Carlo estimator by repeating the Monte Carlo experiment 500 times for a simulation size of n=1000. Then, make a histogram of the 500 values thus obtained and obtain a 95% confidence interval for I based on the quantiles of the simulated sampling distributions. Make the required comparisons based on this output.

(A) Approximate the above integral based on the approach outlined in Question 2 (A).

(B) Approximate I based on the antithetic variable strategy outlined in Question 2 (B).

(C) Approximate I by using the randomized trapezoidal rule based on (3).

(D) How do the three approaches compare to each other in terms of performance?

Randomized Riemann sums: The basic Monte Carlo approximation to an integral can be viewed as a random version of a Riemann sum. Indeed, the standard Riemann approximation of

$$J = \int_{a}^{b} f(y)\,dy,$$

is given by

$$J_R = \frac{(b-a)}{n}\sum_{i=0}^{n} f(y_k), \tag{2}$$

where $a = y_0 < y_1 < \cdots < y_n = b$ form a fixed grid of evenly spaced points. Then, the standard Monte Carlo approximation of J simply consists in replacing the fixed grid $y_0, y_1, \ldots, y_n$ by points that are randomly uniformly generated on (a, b).

An improved and more general deterministic approach, when comparing to (2), is to use the trapezoidal rule which approximates J by

$$J_T = \frac{1}{2}\sum_{k=0}^{n-1}(y_{k+1} - y_k)\big[f(y_k) + f(y_{k+1})\big], \tag{3}$$

based on a grid $a = y_0 < y_1 < y_2 < \ldots < y_{n-1} < y_n = b$, even if the points are not equally spaced. A Monte Carlo approximation of J can be constructed based on this and using points that are uniformly generated on (a, b).

Question 6: Refer to Question 3 from the first part of the Exam.

Consider once again the change-point model for the *British coal mining disaster data* for the years 1851 to 1962. The observed data can be defined in R with:

```
counts <- c(4, 5, 4, 1, 0, 4, 3, 4, 0, 6, 3, 3, 4, 0, 2, 6, 3, 3, 5, 4, 5, 3,
1, 4, 4, 1, 5, 5, 3, 4, 2, 5, 2, 2, 3, 4, 2, 1, 3, 2, 2, 1, 1, 1, 1, 3, 0, 0,
1, 0, 1, 1, 0, 0, 3, 1, 0, 3, 2, 2, 0, 1, 1, 1, 0, 1, 0, 1, 0, 0, 0, 2, 1, 0,
0, 0, 1, 1, 0, 2, 3, 3, 1, 1, 2, 1, 1, 1, 1, 2, 4, 2, 0, 0, 0, 1, 4, 0, 0, 0,
1, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1)

years <- 1851:1962
```

We here wish to carry out the analysis started earlier.

(A) Graph the profile log-likelihood for k and display the MLE of k on that graph.

What year corresponds to the change-point that you have identified?

(B) To complete the analysis of the coal mining data, make a scatterplot of `counts` versus `years`.

To this graph, add lines displaying the estimated Poisson means before and after the change-point ($\hat{\lambda}_1$ for $i = 1, \ldots, \hat{k}$ and $\hat{\lambda}_2$ for $i = \hat{k} + 1, \ldots, 112$), and the Poisson mean that would be obtained from fitting a Poisson model with no change-point to the data.

Question 7: Refer to Question 4 from the first part of the Exam.

First, recall that the number of goals scored by Chelsea in home games is assumed to satisfy

$$X_{1i}|\lambda_C, \theta_C \ \sim \text{Poisson}(\lambda_C + \theta_C),$$

for $i = 1, 2, \ldots, 19$, and in away games,

$$X_{2i}|\lambda_C \ \sim \text{Poisson}(\lambda_C),$$

for $i = 1, 2, \ldots, 19$. Again, random scores are considered to be conditionally independent, given their respective rates and the prior adopted for Λ_C and Θ_C is given by (1). The number of goals scored by Chelsea, playing at home and away, was:

```
home.chelsea <- c(2, 3, 1, 3, 4, 5, 2, 1, 3, 4, 2, 3, 3, 4, 1, 2, 4, 3, 5)
away.chelsea <- c(2, 2, 0, 2, 2, 1, 3, 1, 1, 0, 3, 1, 1, 2, 2, 3, 0, 3, 1)
```

(A) Find the (joint) posterior mode $(\hat{\lambda}_{C,\text{MAP}}, \hat{\theta}_{C,\text{MAP}})'$.

(B) Obtain an approximation to the Bayes estimators of λ_C and θ_C by running the Monte Carlo experiment you described in part (A) of Question 4.

(C) Obtain an approximation to the Bayes estimators of λ_C and θ_C by running the Monte Carlo experiment you described in part (B) of Question 4.

Now, assume also that for Tottenham,

$$Y_{1i}|\lambda_T, \theta_T \ \sim \text{Poisson}(\lambda_T + \theta_T),$$

for $i = 1, 2, \ldots, 19$ when playing at home, and

$$Y_{2i}|\lambda_T \ \sim \text{Poisson}(\lambda_T),$$

for $i = 1, 2, \ldots, 19$ while playing on the road, with the same independence assumption and prior distributions. The number of goals scored by Tottenham in all their games was:

```
home.tottenham <- c(1, 1, 1, 2, 1, 3, 5, 3, 2, 2, 4, 1, 4, 3, 2, 4, 4, 2, 2)
away.tottenham <- c(1, 4, 2, 1, 0, 1, 1, 0, 4, 4, 2, 0, 0, 1, 2, 3, 6, 0, 7)
```

(D) Create an R function that, based only one simulation, returns estimates of the posterior probability

$$p_d = \mathbb{P}(|\Theta_C - \Theta_T| < d \,|\, x, y)$$

for values of d to be passed as argument (as a vector). Your function should also take an optional argument `GRAPH` (with default set to `TRUE`) that produces a graph of p_d for values of $d \in (0, 1)$.

Finally, use your function to approximate p_d for $d = 0.05, 0.1, \ldots, 0.5$ and to produce the graph of p_d.

Question 8:

The results to the 1271 regular season games played (each team played 82 games during the regular season) in 2017-18 are available in the file `NHL_1718.csv`.

(A) Write an R function `season.summary` that produces a table summary of the league results as shown below in Figures 1, 2 and 3. That function should take the following two arguments:

- `scores`: the data frame containing the results to all the games,
- `display.by`: defaulted to `league` to produce a summary of the results for the league as a whole, but that could also take the values `division` or `conference`.

Also, note that teams are always ordered (overall or within conferences or divisions)

- first according to total points (two for each win, one for each loss in overtime or shootout), displayed as `PTS` in summary tables,
- then, in case of a tie, in terms of the number of wins excluding shootout wins, displayed as `ROW` in the summary tables,
- finally, if two teams are still tied, in terms of goal difference, displayed as `GD` in the summary tables.

This is important as the ordering between some of the teams would be wrong otherwise. Finally, the composition of the conferences and divisions in the NHL is given in Table 1 and all variables to be displayed in the different summaries are defined in Table 2.

(B) In order to determine the league champion, at the end of the regular season, the best teams enter the *playoffs* to determine the league champion and winner of the Stanley Cup. The teams are selected to participate in the playoffs according to the following scheme:

- the top three teams in each division automatically qualify,
- within each conference, two 'wild card' teams are identified (these are the best two teams of the conference that did not qualify as top three from one of either divisions so that it is possible for a division to hold both wild card spots),
- in total, each conference sees 8 teams entering the playoffs.

Write a second R function that, given a data frame `scores` as used in part (A), returns a list of teams entering the playoffs:

- `$atlantic` : the top three teams that qualify from the Atlantic division,
- `$metro` : the top three teams that qualify from the Metropolitan division,
- `$east.wild.cards` : the two wild card teams from the Eastern Conference,
- `$central` : the top three teams that qualify from the Central division,
- `$pacific` : the top three teams that qualify from the Pacific division,
- `$west.wild.cards` : the two wild card teams from the Western Conference.

(C) The data that you were provided in the file `NHL_1718.csv` also contains the attendance at each game throughout the season.

Write an R function that, given a data frame `scores` as used above, returns the average attendance for the home games of each NHL team. As a side effect, the function should produce a graph displaying these numbers in a clear way. The graph should also display the League average attendance per game.

Table 1: Composition of conferences and divisions in the NHL, 2017-18 season

Eastern Conference

Atlantic Division	Metropolitan Division
Boston Bruins	Carolina Hurricanes
Buffalo Sabres	Columbus Blue Jackets
Detroit Red Wings	New Jersey Devils
Florida Panthers	New York Islanders
Montreal Canadiens	New York Rangers
Ottawa Senators	Philadelphia Flyers
Tampa Bay Lightning	Pittsburgh Penguins
Toronto Maple Leafs	Washington Capitals

Western Conference

Central Division	Pacific Division
Chicago Blackhawks	Anaheim Ducks
Colorado Avalanche	Arizona Coyotes
Dallas Stars	Calgary Flames
Minnesota Wild	Edmonton Oilers
Nashville Predators	Los Angeles Kings
St. Louis Blues	San Jose Sharks
Winnipeg Jets	Vancouver Canucks
	Vegas Knights

Table 2: Variables to be displayed in summary

`W` : Total number of wins
`L` : Total number of losses in regular time
`OSL` : Total number of losses in overtime and shootout
`PTS` : Total number of points
`ROW` : Total number of wins in regular time and overtime (excluding shootout)
`GF` : Total number of goals scored
`GA` : Total number of goals allowed
`GD` : Goal difference `GF - GA`

Figure 1: Overall season summary that should be produced for Part (A)

```
> season.summary(hockey)
                       W  L OSL PTS ROW  GF  GA  GD
Nashville Predators   53 18  11 117  47 267 211 +56
Winnipeg Jets         52 20  10 114  48 277 218 +59
Tampa Bay Lightning   54 23   5 113  48 296 236 +60
Boston Bruins         50 20  12 112  47 270 214 +56
Vegas Knights         51 24   7 109  47 272 228 +44
Washington Capitals   49 26   7 105  46 259 239 +20
Toronto Maple Leafs   49 26   7 105  42 277 232 +45
Minnesota Wild        45 26  11 101  42 253 232 +21
Anaheim Ducks         44 25  13 101  40 235 216 +19
Pittsburgh Penguins   47 29   6 100  45 272 250 +22
San Jose Sharks       45 27  10 100  40 252 229 +23
Los Angeles Kings     45 29   8  98  43 239 203 +36
Philadelphia Flyers   42 26  14  98  40 251 243  +8
Columbus Blue Jackets 45 30   7  97  39 242 230 +12
New Jersey Devils     44 29   9  97  39 248 244  +4
Florida Panthers      44 30   8  96  41 248 246  +2
Colorado Avalanche    43 30   9  95  41 257 237 +20
St. Louis Blues       44 32   6  94  41 226 222  +4
Dallas Stars          42 32   8  92  38 235 225 +10
Calgary Flames        37 35  10  84  35 218 248 -30
Carolina Hurricanes   36 35  11  83  33 228 256 -28
NY Islanders          35 37  10  80  32 264 296 -32
Edmonton Oilers       36 40   6  78  31 234 263 -29
NY Rangers            34 39   9  77  31 231 268 -37
Chicago Blackhawks    33 39  10  76  32 229 256 -27
Vancouver Canucks     31 40  11  73  31 218 264 -46
Detroit Red Wings     30 39  13  73  25 217 255 -38
Montreal Canadiens    29 40  13  71  27 209 264 -55
Arizona Coyotes       29 41  12  70  27 208 256 -48
Ottawa Senators       28 43  11  67  26 221 291 -70
Buffalo Sabres        25 45  12  62  24 199 280 -81
```

Figure 2: Season summary, presented by division, that should be produced for Part (A)

```
> season.summary(hockey, display.by='division')
$Atlantic
                      W  L OSL PTS ROW  GF  GA  GD
Tampa Bay Lightning 54 23   5 113  48 296 236 +60
Boston Bruins       50 20  12 112  47 270 214 +56
Toronto Maple Leafs 49 26   7 105  42 277 232 +45
Florida Panthers    44 30   8  96  41 248 246  +2
Detroit Red Wings   30 39  13  73  25 217 255 -38
Montreal Canadiens  29 40  13  71  27 209 264 -55
Ottawa Senators     28 43  11  67  26 221 291 -70
Buffalo Sabres      25 45  12  62  24 199 280 -81

$Metropolitan
                       W  L OSL PTS ROW  GF  GA  GD
Washington Capitals   49 26   7 105  46 259 239 +20
Pittsburgh Penguins   47 29   6 100  45 272 250 +22
Philadelphia Flyers   42 26  14  98  40 251 243  +8
Columbus Blue Jackets 45 30   7  97  39 242 230 +12
New Jersey Devils     44 29   9  97  39 248 244  +4
Carolina Hurricanes   36 35  11  83  33 228 256 -28
NY Islanders          35 37  10  80  32 264 296 -32
NY Rangers            34 39   9  77  31 231 268 -37

$Central
                      W  L OSL PTS ROW  GF  GA  GD
Nashville Predators 53 18  11 117  47 267 211 +56
Winnipeg Jets       52 20  10 114  48 277 218 +59
Minnesota Wild      45 26  11 101  42 253 232 +21
Colorado Avalanche  43 30   9  95  41 257 237 +20
St. Louis Blues     44 32   6  94  41 226 222  +4
Dallas Stars        42 32   8  92  38 235 225 +10
Chicago Blackhawks  33 39  10  76  32 229 256 -27

$Pacific
                     W  L OSL PTS ROW  GF  GA  GD
Vegas Knights     51 24   7 109  47 272 228 +44
Anaheim Ducks     44 25  13 101  40 235 216 +19
San Jose Sharks   45 27  10 100  40 252 229 +23
Los Angeles Kings 45 29   8  98  43 239 203 +36
Calgary Flames    37 35  10  84  35 218 248 -30
Edmonton Oilers   36 40   6  78  31 234 263 -29
Vancouver Canucks 31 40  11  73  31 218 264 -46
Arizona Coyotes   29 41  12  70  27 208 256 -48
```

Figure 3: Season summary, presented by conference, that should be produced for Part (A)

```
> season.summary(hockey, display.by='conference')
$Eastern
                       W  L OSL PTS ROW  GF  GA  GD
Tampa Bay Lightning   54 23   5 113  48 296 236 +60
Boston Bruins         50 20  12 112  47 270 214 +56
Washington Capitals   49 26   7 105  46 259 239 +20
Toronto Maple Leafs   49 26   7 105  42 277 232 +45
Pittsburgh Penguins   47 29   6 100  45 272 250 +22
Philadelphia Flyers   42 26  14  98  40 251 243  +8
Columbus Blue Jackets 45 30   7  97  39 242 230 +12
New Jersey Devils     44 29   9  97  39 248 244  +4
Florida Panthers      44 30   8  96  41 248 246  +2
Carolina Hurricanes   36 35  11  83  33 228 256 -28
NY Islanders          35 37  10  80  32 264 296 -32
NY Rangers            34 39   9  77  31 231 268 -37
Detroit Red Wings     30 39  13  73  25 217 255 -38
Montreal Canadiens    29 40  13  71  27 209 264 -55
Ottawa Senators       28 43  11  67  26 221 291 -70
Buffalo Sabres        25 45  12  62  24 199 280 -81

$Western
                     W  L OSL PTS ROW  GF  GA  GD
Nashville Predators 53 18  11 117  47 267 211 +56
Winnipeg Jets       52 20  10 114  48 277 218 +59
Vegas Knights       51 24   7 109  47 272 228 +44
Minnesota Wild      45 26  11 101  42 253 232 +21
Anaheim Ducks       44 25  13 101  40 235 216 +19
San Jose Sharks     45 27  10 100  40 252 229 +23
Los Angeles Kings   45 29   8  98  43 239 203 +36
Colorado Avalanche  43 30   9  95  41 257 237 +20
St. Louis Blues     44 32   6  94  41 226 222  +4
Dallas Stars        42 32   8  92  38 235 225 +10
Calgary Flames      37 35  10  84  35 218 248 -30
Edmonton Oilers     36 40   6  78  31 234 263 -29
Chicago Blackhawks  33 39  10  76  32 229 256 -27
Vancouver Canucks   31 40  11  73  31 218 264 -46
Arizona Coyotes     29 41  12  70  27 208 256 -48
```