Biostatistics 615 Learning Exercise #6 (10 pts)

Due by October 1st 2024 (Tuesday) 11:59pm. Use Gradescope (via Canvas) to submit an R file.

- Your submission should only contain one `R` file named `stationaryDistribution.R` that contains a function named `stationaryDistribution(P)`.
- Your code will be evaluated in Gradescope using 10 different test cases using an automated script. Full credit will be given if your code passes all test cases.
- You are allowed to submit multiple times before the deadline, but only the last submission will be graded. Automated feedback will be provided for each submission.
- You need to implement the function to work with arbitrary (valid) input values beyond the 10 cases tested. If you tweak your implementation so that your functions works specifically for the test cases, you will not receive any credit.
- Implement your function as efficient as you can. If your program does not finish within the time limit for each test case, you will lose the points for those test cases. Note that the official solution finishes much faster than the test cases, so this should be a reasonable time limit.

Problem 1 - Find the stationary distribution of a discrete Markov chain (10 pts)

Write an R function `stationaryDistribution(P)` in the file `stationaryDistribution.R` so that it returns the stationary distribution of a discrete Markov chain $\{X_t\}_{t\geq 0}$ with n states: $\{1,\ldots,n\}$.

Specifically, let $\boldsymbol{P} = (p_{i,j})_{n\times n}$ be the transition probability matrix with

$$p_{i,j} = \Pr(X_{t+1} = i \mid X_t = j), \qquad \text{for } t \geq 0,$$

where $\sum_{i=1}^{n} p_{i,j} = 1$ for $j = 1,\ldots,n$. The stationary distribution $\boldsymbol{\pi} = (\pi_1,\ldots,\pi_n)^T$ is defined as $\pi_i = \lim_{t\to\infty} \Pr(X_t = i)$ for $i = 1,\ldots,n$.

Suppose the Markov chain is irreducible and positive recurrent, which implies that $\boldsymbol{\pi}$ is the solution of the following linear system

$$\boldsymbol{P}\boldsymbol{\pi} = \boldsymbol{\pi},$$

subject to $\sum_{i=1}^{n} \pi_i = 1$.

Your task is to implement a function `stationaryDistribution(P)` that takes $\boldsymbol{P}$ as input and returns $\boldsymbol{\pi}$ as a numeric vector.

The input argument `P` is transition probability matrix. Example input matrices are provided as RDS format on the Canvas homework page. An example result is provided as follows.

```
> P = readRDS('test.1.rds')
> dim(P)
[1] 10 10
> stationaryDistribution(P)
 [1] 0.07795888 0.07348301 0.09982255 0.07332319 0.17685264 0.09218236
 [7] 0.11057127 0.09921174 0.11857071 0.07802365
```

There are specific requirements in the implementation:

- You are NOT allowed to use any functions outside the `base` package in your implementation. Use `help(...)` to check whether a function you want to use belongs to the `base` package or not.
- Your answer should be accurate up to 8 significant digits (not in the digits under the decimal point). Maintaining high accuracy of your answer is therefore very important in this problem.
- Each test case must finish within 2 seconds. The size of the largest input matrix is 1000×1000.

Please note that the official solution finishes within 0.3 seconds, so this should be a reasonable time limit.

You do not need to implement error handling for malformed arguments in this function.