

## Homework 2, Spring 2024

*"Facts are stubborn things, but statistics are pliable."*

MARK TWAIN

Homework 2 has questions 1 through 5 with a total of 5 points. Your recorded score will be scaled to twenty points. The point value for each question or part of a question is in the box following each question or part of a question. This work is due **Saturday 3 Feb** at 11:59 PM.

For this assignment, convert your Jupyter notebook (a IPYNB file) to HTML and submit the HTML file to Canvas. To convert your Jupyter notebook to HTML, do File → Download as → HTML. (For the File menu, look toward the upper right corner.)

The harmonic mean HM of positive numbers  $x_1, x_2, \dots, x_n$  is defined as

$$\text{HM}(x_1, x_2, \dots, x_n) = \frac{n}{\sum_{k=1}^n \frac{1}{x_k}}.$$

For example

$$\text{HM}(5, 45) = \frac{2}{\frac{1}{5} + \frac{1}{45}} = 9, \quad \text{and} \quad \text{HM}(5, 45, 200) = \frac{3}{\frac{1}{5} + \frac{1}{45} + \frac{1}{200}} = \frac{5400}{409}.$$

The harmonic mean has lots of applications to statistics, engineering, and finance. One simple example is that if you drive from Kearney to North Platte at 60 mph and turn around and drive back to Kearney at 70 mph, your average speed for the trip is the harmonic, not arithmetic mean, of 60 and 70; thus your average speed is

$$\frac{2}{\frac{1}{60} + \frac{1}{70}} = \frac{840}{13} \approx 64.6 \text{ mph}.$$

The harmonic mean has the property that for all  $\alpha \in \mathbf{R}_{\neq 0}$  and for all  $x_1, x_2, \dots, x_n \in \mathbf{R}_{>0}$ , the equation

$$\text{HM}(x_1, x_2, \dots, x_n) = \alpha \text{HM}(x_1/\alpha, x_2/\alpha, \dots, x_n/\alpha).$$

is an identity. Additionally, for all  $x \in \mathbf{R}_{>0}$ , the equation  $\text{HM}(x, x, \dots, x) = x$  is an identity.

The Julia package StatsBase has a function 'harmmean' that computes the harmonic mean of the members of a vector. You'll need to download and install this package. This is done using the package manager

```
julia> using Pkg
julia> Pkg.add("StatsBase")
Resolving package versions...
```

Once you have downloaded and installed the this package, you don't need to do these steps again. To make the package available enter

```
julia> using StatsBase;
```

```
julia> x = 2.0^-1074  
5.0e-324
```

```
julia> harmmean([x,x])  
0.0
```

Given the identity  $HM(x, x) = x$ , a better output is 5.0e-324. But it's not mysterious why the output is 0. The reciprocal of the denormalized number  $2.0^{-1074}$  overflows to Inf. But in Julia  $1/\text{Inf}$  is zero. So effectively Julia (miss) computes

$$\frac{2}{\frac{1}{5.0 \times 10^{-324}} + \frac{1}{5.0 \times 10^{-324}}} = \frac{2}{\text{Inf} + \text{Inf}} = \frac{2}{\text{Inf}} = 0. \quad (1)$$

Here is a proposed function 'harmonicMean' that will better handle denormalized inputs

```
function harmonicMean(L)  
    n = length(L)  
    m = max(map(abs, L)...) # maximum abs value  
    L = map(x -> m/x, L)  
    m*(n / sum(L))  
end
```

1. For  $x = 2.0^{-1074}$ , find the value of 'harmonicMean([x,x])' and 'harmonicMean([x,x,x])'. Does this version of the harmonic mean seem to resolve the issue for the function from the StatsBase package?
2. In Julia arithmetic  $1/\text{Inf} = 0$ . Using this standard, we should have

$$HM(\infty, 5, 45) = \frac{3}{\frac{1}{\infty} + \frac{1}{5} + \frac{1}{45}} = \frac{3}{\frac{1}{5} + \frac{1}{45}} = \frac{27}{2}.$$

Evaluate `harmonicMean([5,45,Inf])`. The value should be 13.5. Is it? Step through the Julia code and explain why `harmonicMean([5,45,Inf])` returns NaN. **Hint** In Julia arithmetic, we have  $\text{Inf}/\text{Inf} = \text{NaN}$ .

3. To fix the defect in harmonicMean when one or more input is Inf, one approach is to delete every member of the input that is Inf before computing  $m$ . To do this, use

```
filter!(x-> x != Inf,L) # remove Inf from L
```

The Julia function `filter!` modifies its argument. Similarly, remove every input of `-Inf`. Modify `harmonicMean` to do this.

- 1 4. Test your modified `harmonicMean` function with the input

```
harmonicMean[2.0-1074,2.0-1074,Inf])
```

 (2)

Is the output correct?

- 1 5. Using both `harmonicMean` and the function `harmmean` from `StatBase`, find the harmonic mean of the one thousand member vector `[10,10,10,...,10]`. To do this use

```
harmonicMean([10.0 for k=1:103])
```

and

```
harmmean([10.0 for k=1:103])
```

Which function is more accurate for this test? Why? Congratulations, you have built a better Julia function for the harmonic mean.