DSC 40A - Example Homework

not due (this is not a real homework assignment!)

This is not a real homework assignment — the two problems below are *not* required (but they are good practice!). Instead, this is an example document that uses the same LaTeX template that we'll provide you with in homework assignments. Make sure to watch the walkthrough videos linked on the course website, which not only talk about how to use this template and write LaTeX, but how to work on problems and begin proofs in DSC 40A in general!

Problem 1. When life gives you lemons...

Suppose you're operating a fruit stand near La Jolla Cove. Your stand only accepts cash, so you don't have access to digital receipts of all of your transactions.

You want to keep track of the mean transaction price so far during the day. Instead of writing down the price of each transaction and re-calculating the mean every time someone makes a purchase, you want to come up with a way to only keep track of the mean transaction price.

Let x_i represent the price of the ith transaction (e.g. the 5th transaction of the day is x_5). We define μ_n to be the mean of the first n transactions; that is, $\mu_n = \frac{1}{n}\sum_{i=1}^{n} x_i$.

a) Determine a formula for μ_{n+1} that only uses the variables μ_n, n, and x_{n+1}. (Hint: Start by writing out the definition of μ_{n+1}).

b) Why does the above result imply that you don't need to store the values of all transactions individually?

c) So far, you've sold 7 items and the mean transaction price (including the 7th item) is \$12. How much would your next transaction price have to be in order for your new mean transaction price to be \$14?

Problem 2. Big L

You may find the following properties of logarithms helpful in this question. Assume that all logarithms in this question are natural logarithms, i.e. of base e.

- $e^{\log(x)} = x$
- $\log(a) + \log(b) = \log(a \cdot b)$
- $\log(a) - \log(b) = \log\left(\frac{a}{b}\right)$
- $\log(a^c) = c\log(a)$
- $\frac{d}{dx}\log x = \frac{1}{x}$

Billy, the avocado-farmer-turned-waiter-turned-Instagram-influencer that you're all-too-familiar with, is trying his hand at coming up with loss functions. He comes up with the Billy loss, $L_B(h, y)$, defined as follows:

$$L_B(h, y) = \left[\log\left(\frac{y}{h}\right)\right]^2$$

Throughout this problem, assume that all ys are positive.

a) Show that

$$\frac{d}{dh}L_B(h, y) = -\frac{2}{h}\log\left(\frac{y}{h}\right)$$

b) Show that the constant prediction h^* that minimizes **empirical risk** for Billy loss is

$$h^* = (y_1 \cdot y_2 \cdot \ldots \cdot y_n)^{\frac{1}{n}}$$

You do not need to perform a second derivative test, but otherwise you must show your work.

Hint: To confirm that you're interpreting the result correctly, h^* *for the dataset 3, 5, 16 is* $(3 \cdot 5 \cdot 16)^{\frac{1}{3}} = 240^{\frac{1}{3}} \approx 6.214$.