Exercise: Variance and Standard Deviation

(7/7 points)

This exercise explores the important concept of *variance*, which measures how much a random variable deviates from its expectation. This can be thought of as a measure of uncertainty. Higher variance means more uncertainty.

The variance of a real-valued random variable X is defined as

$$\textbf{var}(X) \triangleq \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Note that as we saw previously, $\mathbb{E}[X]$ is just a single number. To keep the variance of X, what you could do is first compute the expectation of X.

For example, if X takes on each of the values 3, 5, and 10 with equal probability 1/3, then first we compute $\mathbb{E}[X]$ to get 6, and then we compute $\mathbb{E}[(X-6)^2]$, where we remember to use the result that for a function f, if $f(X)$ is a real-valued random variable, then $\mathbb{E}[f(X)] = \sum_x f(x)p_X(x)$. Here, f is given by $f(x) = (x-6)^2$. So

$$\mathbf{var}(X) = (3-6)^2 \cdot \frac{1}{3} + (5-6)^2 \cdot \frac{1}{3} + (10-6)^2 \cdot \frac{1}{3} = \frac{26}{3}.$$

Let's return to the three lotteries from earlier. Here, random variables L_1, L_2, and L_3 represent the amount won (accounting for having to pay $1):

		Probability
L_1	-1	$\frac{999999}{1000000}$
	-1+1000	$\frac{1}{1000000}$

		Probability
L_2	-1	$\frac{999999}{1000000}$
		1

-1+1000000	$\frac{1}{1000000}$

		Probability
L_3	-1	$\frac{9}{10}$
	-1+10	$\frac{1}{10}$

Compute the variance for each of these three random variables. (Please provide the **exact** answer for each of these.)

- $\mathbf{var}(L_1) =$ 0.999999 ✔
- $\mathbf{var}(L_2) =$ 999999.0 ✔
- $\mathbf{var}(L_3) =$ 9.0 ✔

What units is variance in? Notice that we started with dollars, and then variance is looking at the expectation of a dollar amount squared. Thus, specifically for the lottery example $\mathbf{var}(L_1)$, $\mathbf{var}(L_2)$, and $\mathbf{var}(L_3)$ are each in squared dollars.

Some times, people prefer keeping the units the same as the original units (i.e., without squaring), which you can get by computing what's called the *standard deviation* of a real-valued random variable X:

$$\mathbf{std}(X) \triangleq \sqrt{\mathbf{var}(X)}.$$

Compute the following standard deviations, which are in units of dollars. (Please be precise with at least 3 decimal places, unless of course the answer doesn't need that many decimal places. You could also put a fraction.)

- $\mathbf{std}(L_1) =$ 0.9999995
- $\mathbf{std}(L_2) =$ 999.9995
- $\mathbf{std}(L_3) =$ 3.0

Note that when we first introduced the three lotteries and computed average winnings, we didn't account for the uncertainty in the average winnings. Here, it's clear that the third lottery has far smaller standard deviation and variance than the second lottery.

As a remark, often in financial applications (e.g., choosing a portfolio of stocks to invest in), accounting for uncertainty is extremely important. For example, you may want to maximize profit while ensuring that the amount of uncertainty is not too high as to not be reckless in investing.

In the case of the three lotteries, to decide between them, you could for example use a score that is of the form

$$\mathbb{E}[L_i] - \lambda \cdot \text{std}(L_i) \qquad \text{for } i = 1, 2, 3,$$

where $\lambda \geq 0$ is some parameter that you choose for how much you want to penalize uncertainty in the lottery outcome. Then you could choose the lottery with the highest score.

Finally, a quick sanity check (this is more for you to think about the definition of variance rather than to compute anything out):

- Can variance be negative? If yes, give a specific distribution as a Python dictionary for which the variance is negative. If no, enter the text "no" (all lowercase, one word, no spaces).

no

You have used 3 of 5 submissions