Questions for Discussion

1.1

> Write down the formula for grades that corresponds to the sentence above.

Let H be the homework grade, M_1 and M_2 be the midterm exam grades, and F be the final grade. The overall grade G would be

$$G = 0.2H + 0.2M_1 + 0.2M_2 + 0.4F = \boxed{0.2(H + M_1 + M_2 + 2F)}.$$

2.1

> Explain, in your own words, the difference between the three values computed in the previous cell.

The first value is somewhat unrelated to the second and third values; it represents the average number of values in each bin. The second and third values, however, are closely related: both are a measure of the average value generated by each "trial" of the random number generator. The second value was calculated on the raw data generated by the random number generator and thus represents the true mean. The third value was calculated on a quantized version of the raw data—where some information was lost during the quantization process—and thus represents an approximation of the true mean.

2.2

> How would these numbers change if you changed the bin size when histograming the data?

As the size of the bins tends towards zero and the number of bins tends towards infinity, the first number would get smaller, the second number would stay the same, and the third number would converge to the second number.

3.1

> In many cases the data might be presented already summarized, or binned into a histogram. Can you think of some examples in real-world data when

> this might be the case? List a few.

One example is U.S. census data where the number of people each age is reported. These represented pre-summarized data because it quantizes peoples' ages at the granularity of the year and we lose finer-grained information (e.g., how old they are in months, days, hours, or seconds). Another example is the homework/exam statistics as reported by Gradescope. As students, we can see aggregated figures on number of people that got scores within certain ranges but not the individual scores themselves.

3.2

> Oftentimes the way we collect data involves some averaging or sampling, so that we are effectively making a histogram as we actually collect the data. An example of this might be an X-ray detector that counts how many X-rays it detects per second for 6 seconds, then only sends the total number of X-rays it detected each second (i.e., it sends out 6 numbers). Explain how this corresponds to the example above. What does the total number of X-rays seen correspond to from our earlier example? How about the rate of X-rays?

The total number of X-rays seen corresponds to the total number of dice rolls from our earlier example and the rate of X-rays correspond to the average value of dice rolls computed from the histogramed statistics (in that both compute on summative results as opposed to raw values).