Homework 5

Reading: Meester, Section 2.3, 2.4, 4.1. Notes: *expectation.pdf*.

Idea Journal: Remember to submit the idea journal entries after each class.

Problems/Exercises: All of these are due **Thursday, February 29 by 11:00pm**.

1. Meester Exercise 2.3.28.

2. Meester Exercise 2.7.15. Note: the procedure described in this problem is called pooled testing. During the pandemic, this procedure was used by Duke (and several other colleges and universities) to do large-scale COVID surveillance testing among students.

3. Suppose X is a discrete random variable.

   (i) Prove the following simple but very important fact: if $f(x) \geq g(x)$ for all x, then $\mathbb{E}[f(X)] \geq \mathbb{E}[g(X)]$, assuming these are well-defined.

   (ii) Suppose that $f(x) : \mathbb{R} \to \mathbb{R}$ is differentiable. Suppose $\mathbb{E}[X] = \mu$. Let

   $$\ell(x) = f(\mu) + f'(\mu)(x - \mu), \quad x \in \mathbb{R}$$

   be the line that is tangent to the graph of f at the point $(\mu, f(\mu))$. Suppose that the graph of f lies above the graph of ℓ everywhere (except that the graphs of f and ℓ touch at the point of tangency). Prove that $\mathbb{E}[f(X)] \geq f(\mathbb{E}[X])$. In particular, what does this imply about $\mathbb{E}[e^X]$ versus $e^{\mathbb{E}[X]}$? Which is bigger?

4. In a box there are n identical marbles, labeled $1, \ldots, n$. There are n people who take turns drawing a marble from the box, with replacement. So, possibly everyone draws the same marble, or maybe they draw distinct marbles. Let X_n be the number of marbles that were **not** drawn by anyone.

   (i) Compute $\mathbb{E}[\frac{1}{n}X_n]$, the expected fraction of marbles not chosen.

   (ii) What is $\lim_{n\to\infty} \mathbb{E}[\frac{1}{n}X_n]$?

   (iii) Compute $\mathrm{Var}(\frac{1}{n}X_n)$.

   Hint: try representing X_n as a sum of indicators. **Note:** this scenario corresponds to one generation of descent in an important model of population genetics known as the Wright-Fisher model. In that model, the "marbles" represent alleles in a population; each person in the next generation "chooses a parent" or allele at random from among those available. So, $\frac{1}{n}X_n > 0$ means that some fraction of alleles is not inherited by the next generation.

5. Suppose you are going to toss a fair coin 10,000 times. Use Chebychev's inequality (Corollary 2.3.27) to bound the probability of tossing more than 5,500 heads. How does this bound compare to the bound you get by applying the bound that we used in proving the law of large numbers for coin tossing? (here I am referring to Theorem 0.1 in the notes *largenumber1.pdf*) In particular, which bound is better here?

6. Suppose that every time you shop at a certain store, there is a small randomly selected prize that comes with your purchase. Suppose there are n different prizes that you could win, all equally likely. It is possible that you get the same prize multiple times (the prizes are chosen randomly by computer). How many visits to the store does it take to win all of the n prizes? Let X_n be the number of visits until you have won all n distinct prizes; this is a random variable. Compute $\mathbb{E}[X_n]$. Do this by answering the following:

   (i) How many visits (N_1) are needed to win one prize? (Yes, this is really easy.)

   (ii) Now that you have won one prize, let N_2 be the number of additional visits until you get a second prize that is different from the one you already have. What is the distribution of N_2? Note: you could continue to get the first prize many times in a row.

   (iii) Now suppose you have won already k distinct prizes (and $k < n$). There are $n-k$ prizes left to win. Let N_{k+1} be the number of additional visits until you get the $(k+1)^{th}$ prize (the next one that is different from the previous k you have already won). What is the distribution of N_{k+1}?

   (iv) How is X_n related to the random variables N_k?

   Now, having thought about the questions above, you should be able to compute a formula for $\mathbb{E}[X_n]$ in terms of n. How fast does $\mathbb{E}[X_n]$ grow with n? Obviously, $\mathbb{E}[X_n] \geq n$, but maybe it grows much faster as $n \to \infty$. How fast? Linearly? Exponentially? Is the answer surprising to you?