Probability and Statistics
Tsinghua Math Camp Prof. Paul Horn
Homework 4: Due Friday 7/29/2016

Note: You should get to know your coaches, Gavin St. John (`gavin.stjohn@du.edu`) and Zhe Liu (刘喆) (`z-liu16@mails.tsinghua.edu.cn`). If you need to contact me, my email address is `paul.horn@du.edu` or you may ask via WeChat.

I want very much to learn all of your names, but please be patient with me as I am not so good at pronouncing Chinese and I am very bad at remembering names (even non-Chinese names). So please help me by telling me your name very slowly and being patient with me as I try to learn and pronounce it.

Homework problems: To be turned in.

1. Suppose X is a non-negative continuous random variable (that is $\mathbb{P}(X \geq 0) = 1$). Let $F_X(x)$ denote the cdf of X. Show that
$$\mathbb{E}[X] = \int_0^\infty \mathbb{P}(X \geq x)dx = \int_0^\infty (1 - F_X(x))dx.$$
Too easy? Can you come up with a corresponding statement for random variables X that aren't assumed to be non-negative? Note that a similar statement also holds for discrete random variables.

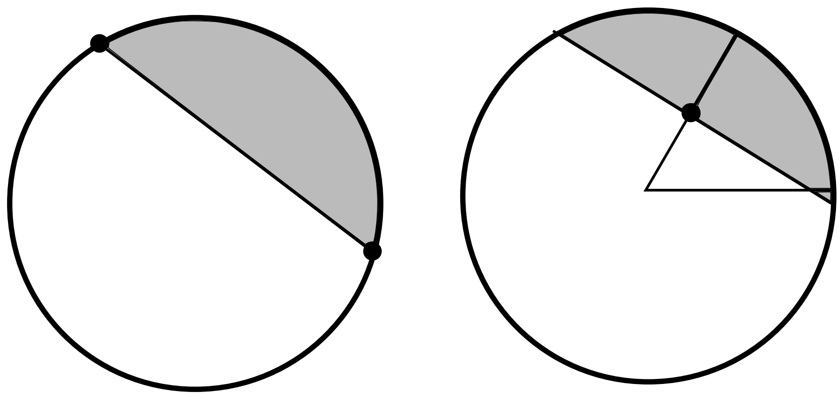
2. Find the variance of a hypergeometric random variable Recall, a hypergeometric random variable with parameters N, n, t has
$$f_X(k) = \frac{\binom{t}{k}\binom{N-t}{n-k}}{\binom{N}{n}}$$
for $0 \leq k \leq \max\{t, n\}$. (Note/hint: You can think of $X = X_1 + \cdots + X_n$ where X_i is the event that the ith pick is a head. These are *not* independent, but they are identically distributed and it's easier to compute $\mathbb{E}[X_iX_j] = \mathbb{E}[X_1X_2]$ to find $\mathbb{E}[X^2]$ than to directly find the whole thing.)

3. Find an example of two random variables X and Y that are dependent but satisfy $\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$, or explain why no such random variables can exist.

4. Find an example of a random variable distribution which proves that Chebyschev's inequality
$$\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sigma) \leq \frac{1}{k^2}$$
cannot be improved in general.

5. Consider a circle. From this circle we choose a random chord (line connecting two points on the circle) which cuts the circle into two points.

- Suppose both points are chosen uniformly at random from the points on the circle. Find the expected length of the chord.
- Suppose the experiment is now changed: first a (uniformly) random angle $0 \leq \theta \leq 2\pi$ is chosen. Then a number $0 \leq r \leq 1$ is chosen. A chord is formed by taking the chord perpendicular to the line at angle θ from the origin. (See picture). Is this the same experiment before in disguise? If so, explain. If not, find the expected length of the chord.



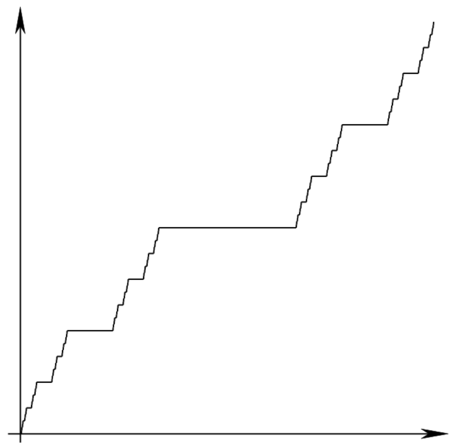
Method one: Pick endpoints uniformly at random. **Method two:** First pick angle, then pick point on line from center of circle to outside along angle.

Hint: It looks like there are two random choices in each method, but in terms of finding the area you can reduce it to one. How?
Too easy? Are there any other methods of finding a 'random chord' you can think of? How do they compare?

6. **WARNING: CHALLENGE!!** The c-function $c(x)$ is a function defined as follows: $c(x) = 0$ for $x \leq 0$ and $c(x) = 1$ for $x \geq 1$. For any other $0 < x < 1$, c is defined as follows:

   - Step 1: Write
   $$x = \sum_{i=1}^{\infty} a_i \left(\frac{1}{3}\right)^i,$$
   where $a_i \in \{0, 1, 2\}$. This is the base-3 expansion of x.
   - Step 2: Create a new sequence $\tilde{a}_i \in \{0, 1\}$ If j is the smallest value (if any) so that $a_j = 1$, let $\tilde{a}_i = 0$ for $i \geq j$. For other values, $\tilde{a}_i = 1$ if $a_i \in \{1, 2\}$, and $\tilde{a}_i = 0$ otherwise. (So make every two into a one, and cut off the sequence after the first one in a_i appears.)
   - Let $c(x) = \sum_{i=1}^{\infty} \tilde{a}_i 2^{-i}$.

This sounds horrible, but it's not actually so bad. Here's another way to build the same function. Make $c(0) = 0, c(1) = 1$. Then $c(x) = \frac{1}{2}$ for $\frac{1}{3} \leq x \leq \frac{2}{3}$. So we took the middle of the empty interval between one third and two thirds, and made the value there the average of those endpoings (ie. 1/2). Next we make $c(x) = \frac{1}{4}$ for $\frac{1}{9} \leq x \leq \frac{2}{9}$ and $c(x) = \frac{3}{4}$ for $\frac{7}{9} \leq x \leq \frac{8}{9}$. Again, we took the endpoints we've already defined, took the middle third between them and defined the c-function to be constant between them. Now we iterate. Here's a picture.



Let X be a random variable with CDF $F_X(x) = c(x)$.

(a) Find $\mathbb{E}[X]$. (**Hint:** Use problem 1.)

(b) **Hard!** Find Var(X).

Note: Although I think this is quite a hard problem – (a) isn't so bad, but (b) is quite difficult – there's a 'cheap' way of completing (b) which feels like magic and actually barely requires a knowledge of how to construct $c(x)$ but rather some very basic properties it has.