Homework 1
(120 points)

1. (20 points) Upper-bound on Entropy. Let $\Omega = \{1, 2, \ldots, N\}$. Suppose $X$ is a random variable over the sample space $\Omega$. For shorthand, let us use $x_i = \mathbb{P}[X = i]$, for each $i \in \Omega$. The entropy of the random variable $X$ is defined to be the following function.

$$H(X) := \sum_{i \in \Omega} -p_i \log p_i$$

Use Jensen’s inequality on the function $f(x) = \log x$ to prove that the following inequality.

$$H(X) \leq \log N$$

Equality holds if and only if we have $p_1 = p_2 = \cdots = p_N$. 
2. **(20 points) Log-sum Inequality.** Let \( \{a_1, \ldots, a_N\} \) and \( \{b_1, \ldots, b_N\} \) be two sets of non-negative real numbers. Use Jensen’s inequality to prove the following inequality.

\[
\sum_{i=1}^{N} a_i \log \frac{a_i}{b_i} \geq a \log \frac{a}{b},
\]

where \( a = \sum_{i=1}^{n} a_i \) and \( b = \sum_{i=1}^{N} b_i \). Moreover, equality holds if and only if \( \frac{a_i}{b_i} \) is equal for all \( i \in \{1, \ldots, N\} \).
3. **(20 points) Approximating Square-root.** Our goal is to find a (meaningful and tight) upper-bound for $f(x) = (1 - x)^{1/2}$ using a quadratic function of the form

$$g(x) = 1 - \alpha x - \beta x^2$$

Use the Lagrange form of the Taylor’s remainder theorem on $f(x)$ around $x = 0$ to obtain the function $g(x)$. 

4. **(15+20+5 points) Coupon Collector Problem.** We shall show the following result. “The expected numbers of balls that we need to throw such that every bin has at least one ball is (roughly) \( n \log n \).”

(a) Let \( i \in \{1, \ldots, n-1\} \). Suppose a few balls have already been thrown and we have the guarantee that there are exactly \((i-1)\) bins with \( > 0 \) load. Let \( N_i \) be the random variable over the sample space \( \{1, 2, \ldots \} \) that represents the minimum number of additional balls thrown so that (a total of) \( i \) bins have \( > 0 \) load. For any \( j \in \{1, 2, \ldots \} \), prove that

\[
P[N_i = j] = \left( 1 - \frac{i-1}{n} \right) \left( \frac{i-1}{n} \right)^{j-1}
\]

(b) Prove that

\[
E[N_i] = \frac{n}{n-i+1}
\]

(c) Denote \( N = N_1 + N_2 + \cdots + N_n \). So, the random variable \( N \) represents the number of balls so that all bins have load \( > 0 \). Prove that

\[
E[N] = n \left( 1 + \frac{1}{2} + \cdots + \frac{1}{n} \right)
\]

This quantity is roughly \( n \log n \). So, we have proven the bound on the coupon collector problem, which was our goal!
5. (7+7+6 points) Poisson Distribution. We can also approach the Coupon Collector problem using the Poisson Approximation Theorem. To get us started towards that, let us consider the following problem.

Suppose we have \( m \) balls and \( n \) bins. Let \( X \) be the Poisson distribution with mean \( \mu = m/n \). That is, for all \( i \in \{0, 1, 2, \ldots\} \), we have

\[
P [ X = i ] = \exp(-\mu) \frac{\mu^i}{i!}
\]

(a) Find the following probability

\[ P [ X \geq 1 ] \]

(b) Suppose \((X^{(1)}, \ldots, X^{(n)})\) be \( n \) independent samples of the random variable \( X \). Find the probability

\[ P \left[ X^{(1)} \geq 1, \ldots, X^{(n)} \geq 1 \right] \]

(c) Substituting \( m = n \log n \), compute the probability

\[ P \left[ X^{(1)} \geq 1, \ldots, X^{(n)} \geq 1 \right] \]

It is left as an exercise to think on how to proceed from here to prove the bounds on the coupon collector problem.