MATH 323 - EXERCISES 5

NOT FOR ASSESSMENT.

DISCRETE PROBABILITY DISTRIBUTIONS

1. Show that the function
$$p(y) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^y$$
for parameter $\lambda > 0$ is a valid probability mass function (pmf) for a discrete random variable Y taking values on $\{0, 1, 2, \ldots\}$, and find the corresponding cumulative distribution function (cdf), $F(.)$ defined by $F(y) = P(Y \leq y)$.

2. Discrete random variable Y has a *Negative Binomial* distribution, that is, the pmf of Y is given by
$$p(y) = \binom{y-1}{n-1} p^n (1-p)^{y-n} \qquad y = n, n+1, \ldots$$
and zero otherwise, for $n \geq 1$ and $0 < p \leq 1$. Find the probability mass function of discrete random variable X defined by $X = Y - n$. *Hint: $X = x$ if and only if $Y = x + n$.*

3. If a finite population of size N comprises R Type I objects and $N - R$ Type II objects, and a sample of size $n \leq N$ is obtained from the population, then the probability that the sample contains r Type I objects and $n - r$ Type II objects is computed using one of two methods:
$$\frac{\binom{n}{r}\binom{N-n}{R-r}}{\binom{N}{R}} \qquad \text{or} \qquad \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}$$

   - **Method A:** Choose items in the population to label Type I and Type II, broken down by whether they are IN the sample or OUT of the sample. The denominator is the total number of ways of choosing R from N items to label Type I. The numerator is the number of ways of choosing r from n items IN the sample to be Type I items, multiplied by the number of ways of choosing the $R - r$ from $N - n$ items OUT of the sample to be Type II items.
   - **Method B:** Consider items to be IN the sample, broken down by Type. The denominator is the total number of ways of choosing n items from N to be IN the sample. The numerator is the number of ways of choosing r Type I items from R to be IN the sample, multiplied by the number of ways of choosing the $n - r$ Type II items from $N - R$ to be IN the sample.

   We say that discrete random variable Y has a *Hypergeometric distribution* if its pmf takes the form
$$p(y) = \frac{\binom{n}{y}\binom{N-n}{R-y}}{\binom{N}{R}} = \frac{\binom{R}{y}\binom{N-R}{n-y}}{\binom{N}{n}} \qquad y \in \{\max\{0, n+R-N\}, \ldots, \min\{n, R\}\}$$
and zero otherwise: Y is the count of the number of Type I items in a sample of size n.

   If N and R are very large compared to n ($N, R >> n$), then it can be shown that
$$p(y) \approx \binom{n}{y} p^y (1-p)^{n-y} \qquad y = 0, 1, \ldots, n$$
and zero otherwise, where $p = R/N$, so that Y has an approximate Binomial distribution, with parameters n and p. Explain this result by considering sampling without replacement from a finite but large population.

4. A fair coin is tossed n times. Let Y be the discrete random variable corresponding to the difference between the number of heads and the number of tails observed. List the possible values that Y can take, and find the probability mass function (pmf) of Y.

5. Suppose that discrete random variable Y has a *Geometric* distribution with parameter p, so that

$$p(y) = (1-p)^{y-1}p \qquad y = 1, 2, 3, \ldots$$

and zero otherwise. Using the definition of conditional probability, show that for $n, k \geq 1$,

$$P(Y = n + k \mid Y > n) = P(Y = k).$$

This result is known as the *Lack of Memory* property.

6. For which values of k and α are the following functions valid probability mass functions on the ranges given.

   (a) For $y = 1, 2, 3, \ldots$,
   $$p(y) = \frac{k}{y(y+1)}$$
   and zero otherwise.

   (b) For $y = 1, 2, 3, \ldots$,
   $$p(y) = ky^{\alpha}$$
   and zero otherwise.

7. Consider a sequence of *Bernoulli* random variables $Y_1, \ldots, Y_n$ each with parameter p resulting from **independent and identical** binary trials, with $P(Y_i = 0) = 1 - p, P(Y_i = 1) = p$ for $i = 1, \ldots, n$. Find the probability distributions of the discrete random variables

$$Y = \min\{Y_1, \ldots, Y_n\} \qquad Z = \max\{Y_1, \ldots, Y_n\}$$

Hint: identify the possible values that Y and Z can take, and consider $P(Y = 1)$, $P(Z = 0)$

8. Consider event A in sample space S, and consider the *indicator function*, $\mathbb{1}_A$, defined for $s \in S$ by

$$\mathbb{1}_A(s) = \begin{cases} 1 & s \in A \\ 0 & s \notin A \end{cases}.$$

Show that $\mathbb{1}_A$ defines a *Bernoulli* random variable. Show furthermore that **any** discrete random variable can be expressed as a linear combination of indicator random variables.