STAT 215A Problem Set #4, Spring 2022

1. Let $F(x)$ be a strictly increasing cdf for a continuous random variable X.

   (a) Show that $Y = F(X)$ is a continuous random variable. In particular, check that Y has a continuous uniform distribution on the interval $(0, 1)$.

   (b) Verify the statement in part (a) for the cdf of exponential distribution. In other words, when $X \tilde{} Exp(\lambda)$ and $F(x) = 1 - e^{-\lambda x}$, for $x > 0$ (and $F(x) = 0$, for $x \leq 0$), check that $1 - e^{-\lambda X} \sim Unif(0, 1)$.

2. *Discrete uniform distribution.* For $n > 1$, let V be a discrete random variable with the pmf $P(V = k) = 1/n$, for all $1 \leq k \leq n$. Determine the mean and variance of V.

3. Let X have a normal distribution with parameters μ and σ^2. In other words, the pdf of X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

   for $x \in \mathbb{R}$. Let $Y = e^X$. Determine the pdf of Y (this distribution is called the *lognormal* distribution).

4. Poisson approximation to binomial. Let $X \sim Binomial(n, p)$ and $Y \sim Poisson(\lambda)$. When n goes to infinity and p goes to zero (for large n and small p) in such a way that np approaches λ, we can show that the probability distributions of X and Y are close to each other. Show that

   (a) the recursion formulas below are valid for X and Y, respectively:

$$P(X = x) = \frac{n - x + 1}{x} \frac{p}{1 - p} P(X = x - 1),$$

   for $x = 1, 2, ..., n$.
   $P(Y = y) = \frac{\lambda}{y} P(Y = y - 1)$, for $y = 1, 2, ...$

   (b) if p is small and $\lambda = np$, then $\frac{n-x+1}{x} \frac{p}{1-p} \simeq \frac{\lambda}{x}$, for $x = 1, 2, ..., n$.

   (c) if p is small, n is large and $\lambda = np$, then $P(X = 0) \simeq P(Y = 0)$.

   (d) $P(X = x) \simeq P(Y = x)$, for all $x = 0, 1, 2, ..., n$ (so their distributions are approximately identical).

5. The probability of twin birth in a large city is given to be about 0.015. On a randomly selected day, 120 births were reported. Assuming that these births were independent of each other, what is the probability that at most two of them were twin births? Explain which probability distribution you are using.

6. Let X and Y be two independent discrete random variables. If g and h are some real valued functions defined on the ranges of X and Y, respectively, then show that the random variables $g(X)$ and $h(Y)$ are also independent.

7. *Hypergeometric distribution.* Consider a finite population of N items of binary form (e.g. each item can be classified as either a success or a failure). Assume that there are R successes and $N - R$ failures in total. We select n of these items without replacement, where the sample size n is less than N. Let X denote the number of successes in the sample.

   (a) Explain why the pmf of X is given by this formula: $P(X = k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}$, for $\max\{0, R + n - N\} \leq k \leq \min\{R, n\}$.

   (b) Let E_k be the event that the k^{th} item selected is a success. Show that $\mathbb{P}(E_k) = R/N$, for each $k, 1 \leq k \leq n$.

   (c) Show that $E[X] = nR/N$. Hint: First verify that $X = \sum\limits_{k=1}^{n} I_{E_k}$.

8. Consider again the setup of this problem from Problem Set 2: We are given two urns, each containing a collection of colored balls. Urn I contains two white and three blue balls, while urn II contains three white and four blue balls. Draw a ball at random from urn I, put it into urn II and then pick a ball from urn II. For $i = 1, 2$, let W_i denote the number of white balls picked from urn i.

   (a) Construct the joint pmf of W_1 and W_2.

   (b) Obtain the marginal pmf of each W_i.

   (c) Are W_1 and W_2 independent? Justify your answer.