Homework 3 (due Friday, May 31st at 11:59pm)

1. (10 pts) Let P be a probability distribution on the sample space Ω and $B \subset \Omega$ an event such that $P(B) > 0$. Show that the conditional probability given B is a probability distribution on Ω.
   Hint: use Bayes theorem.

   Solution:
   We need to show that the conditional probability satisfies the three properties: non-negativity, norming, and countably additivity.
   Let $\omega_1, \ldots, \omega_n$ enumerate the different outcomes of Ω.
   Let $A \subset \Omega$ also denote any event.

   - $P(A|B) = \frac{P(A \cap B}{P(B)} \geq 0$ since the numerator is a probability (between 0 and 1) and the denominator is > 0.

   - $P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

   - Let $A_1, \ldots$ denote disjoint events. (Up to countably infinitely many; all sums in the following proof can be replaced with a finite sum)

   $$\begin{aligned} P(\bigcup_{i=1}^{\infty} A_i|B) &= \frac{P((\bigcup_{i=1}^{\infty} A_i) \cap B)}{P(B)} \\ &= \frac{P(\bigcup_{i=1}^{\infty}(A_i \cap B))}{P(B)} && \text{(Distributive Property of Sets)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap B)}{P(B)} && \text{(Disjoint events)} \\ &= \sum_{i=1}^{\infty} \frac{P(A_i \cap B)}{P(B)} \\ &= \sum_{i=1}^{\infty} P(A_i|B) \end{aligned}$$

   The three facts combined mean that the conditional probability is a probability distribution on Ω.

2. (10 pts) In answering a question on multiple choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer, and assume that each question has m multiple–choice alternatives (only one of which is correct). Find an expression (involving p and m) for the conditional probability that a student actually *knew* the answer to a question, given that he or she answered it correctly.

   Solution:
   Define the following events:
   A = student answered the question correctly.
   B = student actually knows the answer.
   We are given: $P(B) = p$, and $P(A|B^c) = \frac{1}{m}$. Also, we can presume that $P(A|B) = 1$.
   We need to find: $P(B|A)$. Using Bayes Rule:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)}$$
$$= \frac{1 \cdot p}{1 \cdot p + \frac{1}{m}(1-p)} = \frac{mp}{mp + (1-p)} = \frac{mp}{1 + (m-1)p}$$

3. (10 pts) A production facility employs 14 workers on the day shift, 15 workers on the swing shift, and 16 workers on the graveyard shift. A quality control consultant is to randomly select 6 of these workers for in-depth interviews. What is the probability that all 6 selected workers will be from the day shift? What is the probability that all 6 selected workers will be from the same shift? What is the probability that at least two different shifts will be represented among the selected workers?

   Solution:

   (a) What is the probability that all 6 selected workers will be from the day shift? Since the selection of workers is made at random, all samples of 6 are equally likely. Therefore, the probability that all 6 selected workers will be from the day shift is equal to the fraction

$$\frac{\text{\# of possible samples of size 6 with all workers from the day shift}}{\text{total \# of possible samples of size 6}}.$$

   There are a total of $14 + 15 + 16 = 45$ workers, so there are a total of $\binom{45}{6}$ possible samples of size 6. Since there are 14 workers in the day shift, the number of samples of size 6 so that all 6 workers are from the day shift is $\binom{14}{6}$. Thus, the probability that all 6 selected workers will be from the day shift is

$$\frac{\binom{14}{6}}{\binom{45}{6}} = 0.0003687.$$

(b) What is the probability that all 6 selected workers will be from the same shift?
The event that all 6 come from the same shift is the union of three events: 1) they all come from the day shift, 2) they all come from the swing shift, and 3) they all come from the graveyard shift. These are mutually exclusive events, so the desired probability is the sum of the probabilities of each of these 3. In part (a), we calculated the probability that they all come from the day shift. Following the same argument, we can obtain the probabilities of the other two events. The probability that they all come from the same shift is therefore

$$\frac{\binom{14}{6} + \binom{15}{6} + \binom{16}{6}}{\binom{45}{6}} = 0.001966.$$

(c) What is the probability that at least two different shifts will be represented among the selected workers?
The complement of {at least two different shifts will be represented among the selected workers} is {only one shift will be represented among the selected workers}, which is the event in part b. Thus, the probability that at least two different shifts will be represented among the selected workers is

$$1 - 0.001966 = 0.998.$$

4. (5 pts) For some r.v. X, prove that $V(X) = E[X^2] - E[X]^2$.

Solution:

$$\begin{aligned} V(X) &= E[(X - E[X])^2] = E[X^2 - 2XE[X] + E[X]^2] \\ &= E[X^2] - 2E[X]^2 + E[X]^2 = E[X^2] - E[X]^2 \end{aligned}$$

by linearity of the expectation operator.

5. (5 pts) Prove that $V(a+bX) = b^2V(X)$.

   Solution:

$$\begin{aligned} V(a+bX) &= E[(a+bX-E[a+bX])^2] \\ &= E[(a+bX-a-bE[X])^2] \\ &= E[b^2(X-E[X])^2] \\ &= b^2E[(X-E[X])^2] = b^2V(X) \end{aligned}$$

   by linearity of the expectation operator.

6. (5 pts) Prove that for the discrete r.v.'s X_1, X_2 with $X_1 \perp\!\!\!\perp X_2$ (X_1 is independent of X_2), $E[X_1X_2] = E[X_1]E[X_2]$. **Solution:**
   Let p_{X_1,X_2} be joint probability mass function. Then

$$\begin{aligned} E[X_1X_2] &= \sum_{x_1 \in supp(X_1), x_2 \in supp(X_2)} x_1x_2p_{X_1,X_2}(x_1,x_2) \\ &= \sum_{x_1 \in supp(X_1), x_2 \in supp(X_2)} x_1x_2p_{X_1}(x_1)p_{X_2}(x_2) \\ &= \left(\sum_{x_1 \in supp(X_1)} x_1p_{X_1}(x_1)\right)\left(\sum_{x_2 \in supp(X_2)} x_2p_{X_2}(x_2)\right) \\ &= E[X_1]E[X_2]. \end{aligned}$$

7. (10 pts) Suppose that when in flight, airplane engines will fail with probability $1-p$ independently from engine to engine. If an airplane needs a majority of its engines operative to make a successful flight, for what values of p is a 5-engine plane preferable to a 3-engine plane?
   Solution:
   The probability that a 5-engine plane is operative is

$$P_5 = \binom{5}{3}p^3(1-p)^2 + \binom{5}{4}p^4(1-p) + p^5$$

   and the probability that a 3-engine plane is operative is

$$P_3 = \binom{3}{2}p^2(1-p) + p^3.$$

Therefore we need to find the values of p for which P_5 is no smaller than P_3. After some algebra this is equivalent of finding the values of p for which $6p^3 - 15p^2 + 12p3 \geq 0$, which simplifies to

$$6(p - 1/2)(p - 1)^2 \geq 0,$$

which can hold only if $p \geq 1/2$. Therefore, for any $p \geq 1/2$, a 5-engine plane is preferable to a 3-engine plane.

8. (10 pts) You roll a die 20 times. What is the probability of

- observing a sequence of 6.
- observing the first 6 in the 20-th throw.
- observing the first 6 in the 10-th throw.
- observing the fifth 6 in the 20-th throw.
- observing the following sequence: $1, 2, 3, 4, 5, 6, 1, 2, 3, \ldots, 6, 1, 2$.
- the 4-th throw being a 4.
- observing half of the times odd numbers.

Solution: Let A be the event of interest in each case.

- $$P(A) = 6^{-20}$$
- $$P(A) = \left(\frac{5}{6}\right)^{19} \frac{1}{6}$$
- $$P(A) = \left(\frac{5}{6}\right)^{9} \frac{1}{6}$$
- $$P(A) = \frac{1}{6}\binom{19}{4}\left(\frac{1}{6}\right)^{4}\left(\frac{5}{6}\right)^{15}$$
- $$P(A) = 6^{-20}$$
- $$P(A) = 6^{-1}$$
- $$P(A) = \binom{20}{10}\left(\frac{1}{2}\right)^{20}$$

9. (15 pts) Let $X \sim$Binom(n, p). Compute E$[X]$.
Note: you need to prove it via the definition of the expectation, not as done in class.
Hint: use the fact that $\sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = 1$.

Solution:

$$\begin{aligned}
E[X] &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n(n-1)!}{x(x-1)!(n-x)!} p^x (1-p)^{n-x} \\
&= \sum_{x=1}^{n} x \frac{n}{x} \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} p (1-p)^{n-x} \\
&= \sum_{x=1}^{n} np \binom{n-1}{x-1} p^{x-1} (1-p)^{n-x} \\
&= np \sum_{x=0}^{n-1} \binom{n-1}{x} p^x (1-p)^{n-x-1} \\
&= np \sum_{x=0}^{m} \binom{m}{x} p^x (1-p)^{m-x} = np
\end{aligned}$$

choosing $m = n - 1$.

10. (20 pts) Imagine that you are collecting n different coupons in order to win a prize. Every time you buy your favourite mozzarella, you get a new coupon. What's the expected number of mozzarellas that you should buy in order to get all n coupons? Show that this is equal to $n \cdot H_n$ where H_n is the n-th armonic number, eg $1+2^{-1}+3^{-1}+\cdots+n^{-1}$.
Hint: think about the case of $n = 2$. Let T_i be the expected number of mozzarellas needed to find the $i - th$ coupon. That is, T_1 will be the number of mozzarellas necessary to find the first coupon (this is obvious), T_2 the number to find the second one, and so on... How is T_2 distributed, given that there is only 1 coupon left to find? Generalize to the case of n coupons.

Solution:
It's easy to see that T_i is distributed as a geometric r.v. with mean

$n/(n-i+1)$. Let $T := \sum_{i=1}^{n} T_i$. Then

$$
\begin{aligned}
E[T] &= E[\sum_{i=1}^{n} T_i] = \sum_{i=1}^{n} E[T_i] \\
&= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-3} + \cdots + \frac{n}{2} + \frac{n}{1} \\
&= nH_n
\end{aligned}
$$