Midterm Solution

1. **Cheap Watches** (35 points) Two factories A and B manufacture watches. Factory A produces on average one defective item out of 100 and defects happen independently from one watch to the next in the production line, so states of any two watches produced by the factory are independent. Similarly, Factory B produces on average one bad watch out of 200 and the states of any two watches produced by the factory are independent. A retailer receives a container of watches that is equally likely to belong to any of the two factories.

   (a) (10 points) He checks the first watch. What is the probability that it works?

   (b) (15 points) Assume the first watch works! What is the probability that the second watch he will check also works?

   (c) (10 points) Are the states of the first 2 watches independent?

**Solution.**

(a) Denote the following events:

\[ F_A : \text{ the container is from factory A} \]
\[ F_B : \text{ the container is from factory B} \]
\[ W_1 : \text{ the first watch works} \]

By the law of total probability:

\[
P(W_1) = P(F_A)P(W_1|F_A) + P(F_B)P(W_1|F_B)
\]
\[
= \frac{1}{2} \left( 1 - \frac{1}{100} \right) + \frac{1}{2} \left( 1 - \frac{1}{200} \right)
\]
\[
= \frac{397}{400} = 99.25\%.
\]

(b) Denote by \( W_2 \) the event that the second watch works. The probability that the second watch will work given that the first watch works:

\[
P(W_2|W_1) = \frac{P(W_2 \cap W_1)}{P(W_1)}.
\]

The probability of the intersection can be computed similarly to part (a):

\[
P(W_2 \cap W_1) = P(F_A)P(W_2 \cap W_1|F_A) + P(F_B)P(W_2 \cap W_1|F_B)
\]

Since the defects occur independently:

\[
P(W_2 \cap W_1) = P(F_A)P(W_1|F_A)P(W_2|F_A) + P(F_B)P(W_1|F_B)P(W_2|F_B)
\]
\[
= \frac{1}{2} \left( 1 - \frac{1}{100} \right)^2 + \frac{1}{2} \left( 1 - \frac{1}{200} \right)^2 = \frac{15761}{16000}
\]

Combining with the result of part (a):

\[
P(W_2|W_1) = \frac{15761/16000}{397/400} = \frac{15761}{15880} \approx 99.25063\%
\]

(c) By symmetry,

\[
P(W_2) = P(W_1) = \frac{397}{400} \neq P(W_2|W_1),
\]

so the states of the first two watches are not independent.
2. **The Entomologist.** (65 points) Each individual of a species of insects is a male with probability \( p \), independently of the rest of individuals. An entomologist seeks to collect exactly \( M \) males and \( N \) females, and therefore stops hunting as soon as she captures \( M \) males and \( N \) females. She has to capture an insect in order to determine its gender. Let \( X \) be the total number of insects she catches to collect the \( M \) male and \( N \) female insects she needs. (You can use the below table listing the mean and variance of famous distributions if you find it useful.)

(a) (10 points) Assume \( M = 1 \) and \( N = 0 \), i.e she needs only a single male. What is the distribution of \( X \)? How many insects does she need to catch on average?

(b) (10 points) Assume \( M = 1 \) and \( N = 1 \), i.e. she needs one male and one female insect. What is the distribution of \( X \)?

(c) (10 points) How many more insects does she need to catch on average in part (b) as compared to part (a)?

(d) (10 point) Assume she needs to catch \( M > 1 \) male insects and no females \( (N = 0) \). How many insects does she need to catch on average?

(e) (10 points) Can you give an upper bound on \( P(X \geq k) \)?

(f) (15 points) What is the distribution of \( X \) in part (d)?

**Solution.**

(a) This is an experiment of repeated trials with probability of success \( p \), so \( X \) is Geometrically distributed with parameter \( p \): \( X \sim \text{Geom}(p) \). The number of insects needed to catch on average is the expected value of \( X \):

\[
E[X] = \frac{1}{p}
\]

(b) First, observe that \( X \) must be at least 2. When the entomologist is done collecting insects, there are two options: either the last one is male, which means that all the previous insects were female, or vice versa: the last one is female and the rest are male. Therefore, the distribution of \( X \) is

\[
P(X = k) = \begin{cases} 
  p^{k-1}(1-p) + (1-p)^{k-1}p & \text{if } k \geq 2, \\
  0 & \text{otherwise}.
\end{cases}
\]

(c) The expected value of \( X \) is

\[
E[X] = \sum_{k=2}^{\infty} k[p^{k-1}(1-p) + (1-p)^{k-1}p]
\]

\[
= \sum_{k=1}^{\infty} kp^{k-1}(1-p) + \sum_{k=1}^{\infty} k(1-p)^{k-1}p - 1
\]

\[
= \frac{1}{1-p} \frac{1}{p} - 1 = \frac{1}{p} + \frac{p}{1-p}.
\]

The first term is the expectation of a \( \text{Geom}(1-p) \) random variable, and the second term is the expectation of a \( \text{Geom}(p) \) random variable. Hence the expectation of \( X \) is

\[
E[X] = \frac{1}{p} + \frac{1}{1-p} - 1 = \frac{1}{p} + \frac{p}{1-p}.
\]

The entomologist needs to catch \( \frac{p}{1-p} \) more insects on average as compared to part (a).

(d) We write \( X \) as a sum of \( M \) random variables:

\[
X = \sum_{i=1}^{M} Y_i,
\]
where $Y_1$ is the number of insects caught until the first male is collected, and $Y_i$, $i > 1$, is the number of insects caught after collecting the $(i - 1)$-th male and until the $i$-th male is collected. They are all identically distributed: $Y_i \sim Geom(p)$. By linearity of expectation:

$$
\mathbb{E}[X] = \sum_{i=1}^{M} \mathbb{E}[Y_i] = \frac{M}{p}
$$

(e) We apply Markov’s inequality, using the result of part (d):

$$
\mathbf{P}(X \geq k) \leq \frac{\mathbb{E}[X]}{k} = \frac{M}{kp}
$$

(f) If the entomologist caught the last male insect after catching a total of $k$ insects, then there must be exactly $M - 1$ other males and the rest are females. The probability of catching exactly $M - 1$ males out of $k - 1$ insects is given by a Binomial distribution:

$$
\binom{k - 1}{M - 1} p^{M-1} (1 - p)^{k-M}.
$$

We need to multiply this by $p$, the probability that the last insect is male. Therefore, the distribution of $X$ is

$$
\mathbf{P}(X = k) = \begin{cases} 
(k-1) \binom{M}{M-1} p^{M}(1 - p)^{k-M}, & \text{if } k \geq M \\
0, & \text{otherwise}.
\end{cases}
$$