Problem Set 5 Solutions

Note: Thanks to Kevin Lee for some of the solutions.

1. In order for more than x tests to occur, he has to fail the first x. Therefore, we see that $\mathbb{P}(X > x) = (1-p)^x$ for $x < n$. Applying the tail-sum theorem on X, we get that $\mathbb{E}(X) = \sum_{x=0}^{\infty} \mathbb{P}(X > x) = \sum_{x=0}^{n-1} (1-p)^x = \frac{1-(1-p)^n}{p}$.

2. We have that $f_N(x) = a(1-a)^{x-1}$. Thus, $f_{N,K}(x,k) = a(1-a)^{x-1}\binom{x}{k}p^k(1-p)^{x-k}$. Note that

$$\begin{aligned}
f_K(k) &= \sum_x f_{N,K}(x,k) \\
&= a(1-a)^{-1} p^k \sum_{x=k}^{\infty} (1-a)^x (1-p)^{x-k} \binom{x}{k} \\
&= \frac{a(1-a)^{k-1} p^k (a+p-ap)^{-k}}{a+p-ap} \\
&= a(1-a)^{k-1} p^k (a+p-ap)^{-k-1}.
\end{aligned}$$

The formula for conditional expectation then gives us that

$$\mathbb{E}(N|K) = \sum_x xf(x|K) = \left(\sum_x xf_{N,K}(x,k)\right) / \left(a(1-a)^{k-1} p^k (a+p-ap)^{-k-1}\right).$$

Evaluating the top sum gives

$$\begin{aligned}
\sum_x xf_{N,K}(x,k) &= \sum_{x=k}^{\infty} xa(1-a)^{x-1}\binom{x}{k}p^k(1-p)^{x-k} \\
&= a(1-a)^{-1} p^k \sum_{x=k}^{\infty} x(1-a)^x(1-p)^{x-k}\binom{x}{k} \\
&= \frac{a(1-a)^{k-1} p^k (a+p-ap)^{-k}}{(a+p-ap)^2}(1-a+k-p+ap) \\
&= a(1-a)^{k-1} p^k (a+p-ap)^{-k-2}(1-a+k-p+ap).
\end{aligned}$$

We can then divide to obtain $\mathbb{E}(N|K) = \frac{1-a+k-p+ap}{a+p-ap}$.
Note: An alternate approach is to recognize this as an instance of the negative binomial distribution.

3. Let S_k be the set of consecutive pearls starting from pearl k and going clockwise. Let V_k be the total value of these pearls. Note that each pearl is counted exactly ten times, so $\sum_{x=1}^{23} V_k = \$7000 \cdot 10$. If we choose a sequence at random, then $\mathbb{E}(V_k) = \sum_{x=1}^{23} \frac{1}{23} V_k = \frac{70000}{23} > 3000$. This means that $\mathbb{E}(V_k) = \sum_{x=0}^{7} 1000x \cdot \mathbb{P}(V_k = 1000x) > 3000$, so we must have that $\mathbb{P}(V_k > 3000) > 0$, as the expected value would otherwise be at most $3000 < \frac{70000}{23}$. Therefore, at least one sequence has value greater than \$3000, but since the values must be integer multiples of \$1000, it must in fact be at least \$4000, as desired.

Consider the necklace with the sequence of pearls `pffpfffpffpffpfffpffpff`. Then we can see that every set of consecutive ten pearls contains at least three real pearls, so we cannot guarantee that there is a sequence worth \$2000 or less.
Note: It is important here to give an example of an instance where the statement is not true, as it is not enough to just show that a probabilistic argument does not work.

4. Since X and Y are independent random variables with the same distribution, it helps to first note that $\mathbb{P}(X=x) = \binom{6}{x}\frac{5^{6-x}}{6^6}$. Thus, $\mathbb{P}(A) = 1 - \mathbb{P}(X=0) = \frac{31031}{6^6} \approx 0.665$. And $\mathbb{P}(B) = \mathbb{P}(X=0)\mathbb{P}(Y \geq 2) + \mathbb{P}(X=1)\mathbb{P}(Y \geq 1) + \mathbb{P}(X \geq 2)\mathbb{P}(Y \geq 0) = \frac{1346704211}{6^{12}} \approx 0.619$. Since we have $\mathbb{P}(A \cap B) = \mathbb{P}(X=1)\mathbb{P}(Y \geq 1) + \mathbb{P}(X \geq 2)\mathbb{P}(Y \geq 0) = \frac{1154813586}{6^{12}}$, we have that $\mathbb{P}(B|A) = \frac{1154813586}{1447782336}$. Similarly, we can compute that $\mathbb{P}(B^c \cap A) = \mathbb{P}(X=1)\mathbb{P}(Y=0) = \frac{292968750}{6^{12}}$, that $\mathbb{P}(B \cap A^c) = \mathbb{P}(X=0)\mathbb{P}(Y \geq 2) = \frac{191890625}{6^{12}}$, and that $\mathbb{P}(B^c \cap A^c) = \mathbb{P}(X=0)\mathbb{P}(Y < 2) = \frac{537109375}{6^{12}}$. This gives us that $\mathbb{P}(B^c|A) = \frac{292968750}{1447782336}$, $\mathbb{P}(B|A^c) = \frac{191890625}{729000000}$, and $\mathbb{P}(B^c|A^c) = \frac{537109375}{729000000}$. Simplifying these fractions gives us

$$\mathbb{P}(B|A) = \frac{192468931}{241297056} \approx 0.798, \mathbb{P}(B^c|A) = \frac{48828125}{241297056} \approx 0.202, \mathbb{P}(B|A^c) = \frac{12281}{46656} \approx 0.263, \mathbb{P}(B^c|A^c) = \frac{34375}{46656} \approx 0.737$$

Note that this is the case of Pepys' problem with $n=1$. The conclusion is then equivalent to $\mathbb{P}(A) > \mathbb{P}(B)$.

5. The last animal captured needs to be marked, as well as $m-1$ of the previously captured animals. Thus, if we capture n animals, the probability of capturing $m-1$ marked animals, then capturing a marked animal last is

$$\begin{aligned}
\mathbb{P}(X=n) &= \binom{n-1}{m-1}\frac{a(a-1)\cdots(a-m+1)(b-a)(b-a-1)\cdots(b-a-n+m+1)}{b(b-1)\cdots(b-n+1)} \\
&= \frac{(n-1)!}{(m-1)!(n-m)!}\frac{a!}{(a-m)!}\frac{(b-a)!}{(b-a-n+m)!}\frac{(b-n)!}{b!} \\
&= \frac{a}{b}\frac{(a-1)!}{(m-1)!(a-m)!}\frac{(b-a)!}{(n-m)!(b-a-n+m)}\frac{(n-1)!(b-n)!}{(b-1)!} \\
&= \frac{a}{b}\binom{a-1}{m-1}\binom{b-a}{n-m}\binom{b-1}{n-1}^{-1}.
\end{aligned}$$

Now, randomly order all the animals. The desired expected value is the expected location of the mth marked animal. The probability that the nth animal is the mth marked animal is $\binom{n-1}{m-1}\binom{b-n}{a-m}\binom{b}{a}^{-1}$ since we need to choose $m-1$ marked animals from $n-1$ before, and $a-m$ from $b-n$ animals after. The expected value is hence

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{n=m}^{b-a+m} n\binom{n-1}{m-1}\binom{b-n}{a-m}\binom{b}{a}^{-1} \\
&= \sum_{n=0}^{b-a}(n+m)\binom{n+m-1}{m-1}\binom{b-n-m}{a-m}\binom{b}{a}^{-1} \\
&= \sum_{n=0}^{b-a}(b-a-n+m)\binom{b-a-n+m-1}{m-1}\binom{a-m+n}{a-m}\binom{b}{a}^{-1} \\
&= b-a+m-\left(\frac{(b-a)(a-m+1)}{a+1}\right) \\
&= \frac{m(b+1)}{a+1}.
\end{aligned}$$

Here, the third equality follows by making the change of variable $n \mapsto b-a-n$. For the fourth, interpret $\binom{b-a-n+m-1}{m-1}\binom{a-m+n}{a-m}\binom{b}{a}^{-1}$ as the probability that the mth marked animal has $b-a-n$ unmarked animals before and $a+n-m$ after. Thus the $b-a+m$ term in the summation is summed over a weight of 1, while the n term gives the expected number of unmarked animals before the mth marked animal, which by symmetry is $\frac{a-m+1}{a+1}(b-a)$.

6. There is a probability 1 that he will get a picture he doesn't have initially. Afterwards, there is a probability of $\frac{2}{3}$. After getting two different pictures, there is a probability of $\frac{1}{3}$. Thus, $\mathbb{E}(Y) = \mathbb{E}(Y_1) + \mathbb{E}(Y_2) + \mathbb{E}(Y_2)$ where Y_k is the number of times he has donate to get the kth new picture. Since we know that the expected value of a geometric distribution is p^{-1}, we have that $\mathbb{E}(Y) = 1 + \frac{3}{2} + 3 = \frac{11}{2}$.