Solutions to Problem Set 2

Saeed Hedayatian
97100292
Probability and Applications

April 19, 2019

Problem 1

(a) For the random variable X,defined in the question, we can see that

$$supp(X) = \{1, 2, \cdots\}.$$

Also,

$$Pr(X = i) = (1-p)^{i-1}(1-q)^{i-1}(p+q-pq).$$

Let $\alpha = p + q - pq$. Now

$$Pr(X = i) = \alpha(1-\alpha)^{i-1}.$$

Now we can calculate the expected value of X.

$$\begin{aligned} E(X) &= \sum_{x \in supp(X)} xPr(x) = \sum_{i=1}^{\infty} i\alpha(1-\alpha)^{i-1} = \frac{1}{\alpha} \\ &= \frac{1}{p+q-pq}. \end{aligned}$$

The last equality holds because $|1-\alpha| < 1$. And this is true as $0 \leqslant p, q \leqslant 1$.

(b) We denote the event that both players die in the game by D. We can decompose D to $\{D_i | i \in \mathbb{N}\}$ where each D_i is the event that both players die in the ith round. Since for all i, j, $D_i \cap D_j = \emptyset$:

$$\begin{aligned} Pr(D) &= \sum_{i \in \mathbb{N}} Pr(D_i) = \sum_{i=1}^{\infty} pq(1-p)^{i-1}(i-q)^{i-1} = pq \sum_{i=0}^{\infty} (1-p-q+pq)^i \\ &= \frac{pq}{p+q-pq}. \end{aligned}$$

Problem 2

If We define X_i as a bernoulli random variable indicating the presence of a ball with ith color among the twenty selected balls, and X a random variable that shows the number of different colors seen among the twenty selected balls, it is easy to see that

$$E(X) = \sum_{i=1}^{7} E(X_i).$$

From this and the properties of a bernoulli random variable

$$E(X) = \sum_{i=1}^{7} Pr(X_i = 1) = 7 * (1 - \frac{\binom{60}{20}}{\binom{70}{20}})$$

Problem 3

First we prove that if $\forall m, n \in \mathbb{N} \quad Pr(X > n + m | X > m) = Pr(X > n) \implies \exists p \quad 0 \leqslant p \leqslant 1 : Pr(X = n) = p(1-p)^{n-1}$.

Consider a random variable X, $supp(X) = \mathbb{N}$, that satisfies the above condition(i.e is *memoryless*). We prove that X is a geometric random variable with parameter $p = Pr(X = 1)$. We use induction on n to prove that X is geometric(i.e $\forall n \in \mathbb{N}\ Pr(X = n) = p(1-p)^{n-1}$).

First, Note that

$$\forall m, n \in \mathbb{N} \quad Pr(X > n + m | X > m) = Pr(X > n). \tag{1}$$

Basis: For $n = 2$,

$$Pr(X > 2) = Pr(X > 1)Pr(X > 1) = (1-p)^2 \quad \text{where } p = Pr(X = 1).$$

Also,

$$Pr(X > 2) = 1 - p - Pr(X = 2).$$

Thus,

$$1 - p - Pr(X = 2) = (1-p)^2 \implies Pr(X = 2) = p(1-p).$$

Hypothesis: Assume that $\forall i \leqslant n\ \ Pr(X = i) = p(1-p)^{i-1}$.

We will prove that $Pr(X = n + 1) = p(1-p)^n$.

Equation (1) implies that

$$Pr(X > n + 1) = Pr(X > 1)Pr(X > n) = (1-p)(1 - \sum_{i=i}^{n} Pr(X = i))$$

$$= (1-p)(1 - \sum_{i=i}^{n} p(1-p)^{i-1}) = (1-p)^{n+1}.$$

Furthermore,

$$Pr(X > n + 1) = 1 - \sum_{i=1}^{n} Pr(X = i) - Pr(X = n + 1) = (1-p)^n - Pr(X = n + 1).$$

Combining these two results in

$$(1-p)^{n+1} = (1-p)^n - Pr(X = n+1) \implies Pr(X = n+1) = p(1-p)^n.$$

Now we show that a geometric random variable,X, is memoryless.

$$\begin{aligned}
Pr(X > n+m | X > m) &= \frac{Pr(X > n+m)}{Pr(X > m)} = \frac{1 - Pr(X \leqslant n+m)}{1 - Pr(X \leqslant m)} \\
&= \frac{1 - \sum_{i=1}^{m+n} p(1-p)^{i-1}}{1 - \sum_{i=1}^{m} p(1-p)^{i-1}} = \frac{(1-p)^{m+n}}{(1-p)^m} = (1-p)^n \\
&= Pr(X > n).
\end{aligned}$$

Problem 4

If X is a hyper geometric random variables with parameters n, D, N, we define its probability mass function as

$$f(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \quad \forall x \in \{0, 1, \cdots, n\}$$

To calculate the variance of X, we first determine $E[X^2]$.

$$\begin{aligned}
E[X^2] &= \sum_{x=0}^{n} x^2 f(x) = \sum_{x=0}^{n} \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} x^2 = \sum_{x=1}^{n} \frac{D\binom{D-1}{x-1}\binom{N-D}{n-x}}{\frac{N}{n}\binom{N-1}{n-1}} x \\
&= \frac{nD}{N} \sum_{x=1}^{n} \frac{\binom{D-1}{x-1}\binom{N-D}{n-x}}{\binom{N-1}{n-1}} x.
\end{aligned}$$

Let $N' = N-1, D' = D-1, x' = x-1, n' = n-1$. Using variable substitution we can see that

$$\begin{aligned}
E[X^2] &= \frac{nD}{N} \sum_{x'=0}^{n'} \frac{\binom{D'}{x'}\binom{N'-D'}{n'-x'}}{\binom{N'}{n'}} (x'+1) \\
&= \frac{nD}{N} \left[\sum_{x'=0}^{n'} \frac{\binom{D'}{x'}\binom{N'-D'}{n'-x'}}{\binom{N'}{n'}} x' + \sum_{x'=0}^{n'} \frac{\binom{D'}{x'}\binom{N'-D'}{n'-x'}}{\binom{N'}{n'}} \right].
\end{aligned}$$

The first sum in the bracket is the expected value of a hyper geometric random variable with parameters n', D', N'. The second sum is the sum of all values for this new variables probability mass function and is thus, equal to 1.

$$E[X^2] = \frac{nD}{N}\left(\frac{n'D'}{N'} + 1\right) = \frac{nD}{N}\left(\frac{(n-1)(D-1) + (N-1)}{N-1}\right).$$

Now we can use the formula for variance.

$$
\begin{aligned}
Var(X) = E[X^2] - E[X]^2 &= \frac{nD}{N}\left(\frac{(n-1)(D-1)+(N-1)}{N-1}\right) - \left(\frac{nD}{N}\right)^2 \\
&= \frac{nD(N-n)(N-D)}{N^2(N-1)}.
\end{aligned}
$$

The above formula for $Var(X)$ is equal to the one mentioned in the question.

Problem 5

Define random variables X, the number of cards drawn with a number more than ten on them, Y, the number of cards drawn, and X_i, the number of desired cards drawn, if we draw exactly i cards.$\{A_i\}$, the event in which exactly i cards are drawn, is a partition of sample space. From the law of total expectation, it can be derived that

$$
\begin{aligned}
\forall i \in \{1, \cdots, 6\} \quad Pr(A_i) = 1/6, \\
E[X] = \sum E[X|A_i]Pr(A_i).
\end{aligned}
$$

Thus,

$$
E[X] = \frac{1}{6}\sum_{i=1}^{6} E[X_i].
$$

So we must calculate $E[X_i]$. To do so note that X_i is a hypergeometric random variable with parameters $N = 52$, $D = 12$, $n = i$. Hence,

$$
E[X_i] = \frac{nD}{N} = \frac{12i}{52} = \frac{3i}{13}.
$$

Therefore,

$$
E[X] = 1/6\sum_{i=1}^{6}\frac{3i}{13} = \frac{21}{26}.
$$

Problem 6

(a) if $\alpha(\pi)$ is the number of inversion in permutation π and X is the number of inversions in a permutation of numbers $1, \cdots, 2n$,

$$
E[X] = \sum_{\pi} \alpha(\pi)Pr(\pi).
$$

So we should calculate total number of inversions among all permutations of numbers $1, \cdots, 2n$. To do so, note that for any $i, j \in \{1, \cdots, 2n\} \quad i \neq j$, these two numbers are inversions in exactly half of the permutations. Hence, $\sum \alpha(\pi) = \frac{\binom{2n}{2}(2n)!}{2}$. Now it is easy to see that

$$
E[X] = \frac{\binom{2n}{2}}{2}
$$

Problem 7

If we consider each selection from *tired mans* queue a success, then random variable X, as described in the question, is a negative binomial random variable with parameters $p = 1/3$ and $r = 10$ (number of distinct bernoulli experiments with success probability p, until rth success happens). For this kind of random variable we know that

$$E[X] = \frac{r}{p} = \frac{10}{\frac{1}{3}} = 30,$$

$$Var(X) = \frac{r(1-p)}{p^2} = \frac{10(\frac{2}{3})}{\frac{1}{9}} = 60.$$

Problem 8

If we denote the number of eggs layed in the upcoming year by X, then X is a random variable and question asks for its expectation.

$$E[X] = \sum_{i \in \mathbb{N}} iPr(X = i) = \sum_{i=1}^{\infty} \frac{i}{2^i}.$$

Using the ratio test, it can be understood that this series converges. If it converges to S,

$$\begin{aligned}
&S = 1/2 + 2/4 + 3/8 + \cdots \\
&\implies S/2 = 1/4 + 2/8 + 3/16 + \cdots \\
&\implies S - S/2 = S/2 = 1/2 + 1/4 + 1/8 + \cdots = 1 \\
&\implies S = 2.
\end{aligned}$$

Finally,

$$E[X] = 2.$$

Problem 9

(a) The city is divided into $\frac{1000}{50}^2 = 400$ regions. If X is the number of fires that start in the city every month and Y_i is the number of fires that start in the ith region every month, then,

  i) Due to symmetry, $\forall i, j \;\; Y_i = Y_j = Y$.
  ii) $X = \sum_{i=1}^{400} Y_i$.

  Thus,

  $$X = 400Y \implies E[X] = 400E[Y] \implies E[Y] = 1.25.$$

  So the number of fires in a region per month is a poisson random variable(Y), with $\lambda = 1.25$. Using poissons mass distribution function we can calculate the probability of a fire starting in a region is

  $$Pr(Y > 0) = 1 - Pr(Y = 0) = 1 - e^{-1.25} \approx 0.71.$$

(b) We define random variable Y, as we did in the previous part except that now, $E[Y] = 500/100 = 5$.
If T_i is the number of days until ith fire in a region, T_i a random variable. Also $E[T_5] = 30$(a month). Now (because of Y being poisson) it can be seen that $T_n = nT_1$. Therefore, $T_1 = T_5/5$. Hence,

$$E[T_1] = \frac{E[T_5]}{5} = 6.$$

Which means that on average, a fire starts in a region every 6 days.

Problem 11

First, note that $510510 = 2 \times 3 \times 5 \times 7 \times 11 \times 13 \times 17$. We can rephrase the question as follows:
In every step, and for any of the prime factors, we either keep it, or remove it. We define X_i, a random variable that shows the value of ith prime factor of 510510(Obviously this is either 1 or the prime itself). It is easy to see that X_is are independent of each other. Hence, if we denote the remaining number, after the tenth step, by X,

$$X = \prod X_i.$$

And,

$$E[X] = \prod E[X_i].$$

Now from the question, it can be understood that in every step, there is a 0.5 probability that X_is value stay unchanged and a 0.5 probability that it changes to 1. Knowing this we can calculate the probability mass function of a prime factor after i steps. We can then calculate $E[X_i]$s which is the expected value of ith factor after ten steps.
If ith prime factor is d_i,

$$\begin{aligned} E[X_i] &= \sum_{x\in\{1,d_i\}} xPr(X_i = x) = Pr(X_i = 1) + d_iPr(X_i = d_i) = (1 - \frac{1}{2^{10}}) + \frac{d_i}{2^{10}} \\ &= 1 + \frac{d_i - 1}{2^{10}}. \end{aligned}$$

So,

$$E[X] = \prod_{d\in\{2,3,5,7,11,13,17\}} (1 + \frac{d-1}{2^{10}}) \approx 1.05.$$

Problem 12

(a) If X is the number of edges and X_i is a bernoulli random variable indicating the presence of the ith edge in the graph, it can be seen that,

$$X = \sum_{i=1}^{\binom{100}{2}} X_i.$$

Using independence of X_is,

$$E[X] = \sum_{i=1}^{\binom{100}{2}} E[X_i] = p\binom{100}{2},$$

$$Var(X) = \sum_{i=1}^{\binom{100}{2}} Var(X_i) = p(1-p)\binom{100}{2}.$$

(b) The desired probability is

$$Pr(X = 4851) = \frac{\binom{\binom{100}{2}}{4851}}{2^{\binom{100}{2}}}$$