Solutions to Assignment 7

Yogeshwaran D.

December 20, 2021

1. Compute the variances of the uniform random variable on $[n]$, the Binomial and the hypergeometric random variable.

Proof: For the uniform random variable, let $\Omega = [n]$ with $p(\omega) = \frac{1}{n}$ for all $\omega \in \Omega$. Let $X(\omega) = \omega$. Then, it is clear that $P(X = \omega) = \frac{1}{n}$ for all $\omega \in \Omega$, and therefore

$$\mathbb{E}[X] = \sum_{i=1}^{n} i\mathbb{P}(X = i) = \sum_{i=1}^{n} \frac{i^2}{n} = \frac{n+1}{2}$$

and

$$\mathsf{VAR}[X] = \sum_{i=1}^{n} i^2\mathbb{P}(X = i) - \mathbb{E}[X]^2 = \frac{n^2-1}{12}$$

For $X \sim Bin(p)$, we have the sample space $\Omega = \{(a_1, ..., a_n) : a_i \in \{0, 1\}\}$ with $p(a_1, ..., a_n) = p^b(1-p)^c$ where b, c are the number of ones and zeros in the sequence $a_1, ..., a_n$ respectively. Now, with this, we define $X(a_1, ..., a_n)$ to be the number of 1s in the sequence a_i. Then $X \sim Bin(p)$. Consider the random variables $X_i = 1_{a_i=1}$. Then, note that X_i are independent, and $X = \sum_{i=1}^{n} X_i$. Therefore, we get :

$$\mathbb{E}[X_i] = \mathbb{P}[a_i = 1] = p \implies \mathbb{E}[X] = np$$

and because $X_i^2 = X_i$, we get

$$\mathsf{VAR}[X] = \sum_{i=1}^{n} \mathsf{VAR}[X_i] = \sum_{i=1}^{n} (\mathbb{E}[X_i] - \mathbb{E}[X_i]^2) = np(1-p)$$

For the hypergrometric distribution with parameters $n \geq m, r$, we consider the sample space of r-sized subsets of n with the uniform pmf, and let $X(S) = \#\{S \cap [m]\}$. Then X is hypergeometric with parameters n, m, r. We define the random variables $X_i(S) = 1_{i\in S}$ for $i \in [m]$. Now it is clear that

$$\mathbb{E}[X_i] = \frac{r}{n} \implies \mathbb{E}[X] = \frac{mr}{n}$$

However, the X_i are dependent, therefore we calculate

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} \mathbb{E}[X_i^2] + \sum_{i\neq j} \mathbb{E}[X_iX_j]$$

Note that $\mathbb{E}[X_iX_j] = \mathbb{P}[i,j \in S] = \frac{r(r-1)}{n(n-1)}$. Therefore, noting that $X_i^2 = X_i$, we get

$$\mathbb{E}[X^2] = \frac{mr}{n} + \binom{m}{2}\frac{r(r-1)}{n(n-1)}$$

whence $\mathsf{VAR}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ gives the answer. ■

2. Show that $\mathsf{VAR}[aX+b] = a^2\mathsf{VAR}[X]$.

Proof : We have :

$$\begin{aligned}\mathsf{VAR}[aX+b] &= \mathbb{E}[(aX+b)^2] - \mathbb{E}[(aX+b)]^2 \\ &= a^2\mathbb{E}[X^2] + 2ab\mathbb{E}[X] + b^2 - a^2\mathbb{E}[X]^2 - 2ab\mathbb{E}[X] - b^2 \\ &= a^2(\mathbb{E}[X^2] - \mathbb{E}[X]^2) = a^2\mathsf{VAR}[X]\end{aligned}$$

as desired. ■

3. Two fair dice are rolled independently. Find the pmf, mean and variance of the following random variables - (1) The sum of the two dice. (2) The maximum among the two dice.

Proof : By a case-by-case analysis, it is easy to see that if X is the sum of the dice, then $P(X=a) = \frac{6-|7-a|}{36}$ for $a \in \{2,3,...,12\}$. From this, it is easy to find , by definition just as in question 1, $\mathbb{E}[X] = 7$, $\mathsf{VAR}[X] = \frac{35}{6}$.

For the second, we note that if Y is the larger of the numbers, then $\{Y = a\} = \{(c,d) : c = a, d < a \text{OR} c < a, d = a\}$, so it has $2a-1$ elements. Thus, $\mathbb{P}(Y=a) = \frac{2a-1}{36}$ for $a \in [6]$. Using the definitions, one finds $\mathbb{E}[X] = \frac{161}{36}$ and $\mathsf{VAR}[X] = \frac{2555}{1296}$. ■

4. Balls are thrown one after another (uniformly at random) into two bins. Each throw is independent of the previous throw. The experiment stops when there is no empty bin. Let X be the total number of balls thrown. Find $\mathbb{P}(X \geq n)$ for all $n \geq 1$.

Proof: See that $\{X \geq n\}$ if and only if the first $n-1$ throws have landed in the same bin. That bin has two ways of being picked, and the probability of the ball going into the same bin each time is $\frac{1}{2}$, so we get $2(\frac{1}{2})^{n-1} = \frac{1}{2^{n-2}} = \mathbb{P}\{X \geq n\}$, for $n \geq 2$. ■

5. Let X be the number of empty cells corresponding to Maxwell-Boltzmann distribution. Compute the pmf, mean and variance of X.

Proof: The sample space is $\Omega = \{(r_1,...,r_n) : r_i = 0, \sum r_i = r\}$ with the probability distribution $p(r_1,...,r_n) = \frac{1}{n^r}\binom{r}{r_1,r_2,...,r_n}$. Let $E_i = \mathbb{P}(r_i = 0)$. Then, we note that

$$\begin{aligned}P(E_{i_1} \cap \ldots \cap E_{i_j}) &= \sum_{r_i \geq 0, \sum r_i = n, r_{i_l} = 0 \forall l} \frac{1}{n^r}\binom{r}{r_1,r_2,...,r_n} \\ &= \sum_{r_i \geq 0, \sum r_i = n, i = 1,...,n-j} \frac{1}{n^r}\binom{r}{r_1,r_2,...,r_{n-j}} \\ &= \frac{(n-j)^r}{n^r}\end{aligned}$$

Therefore, using the generalized IEP,

$$\mathbb{P}(X = k) = \sum_{j=k}^{m} (-1)^{j-k} \binom{j}{k} \binom{n}{j} \frac{(n-j)^r}{n^r}$$

Let $X_i = 1_{E_i}$. Then $X = \sum_{i=1}^{n} X_i$. Note that for $i \neq j$,

$$\mathbb{E}[X_i] = \mathbb{E}[X_i^2] = \frac{(n-1)^r}{n^r} \quad ; \quad \mathbb{E}[X_i X_j] = \mathbb{P}[E_i \cap E_j] = \frac{(n-2)^r}{n^r}$$

Therefore

$$\mathbb{E}[X] = n\frac{(n-1)^r}{n^r}$$

and

$$\mathbb{E}[X^2] = \sum_{i=1}^{n} E[X_i^2] + \sum_{i \neq j} E[X_i X_j] = n\frac{(n-1)^r}{n^r} + \binom{n}{2} \frac{(n-2)^r}{n^r}$$

and one can finish from here. ■

6. Let X be the number of empty cells corresponding to Fermi-Dirac distribution. Compute the pmf, mean and variance of X.

Proof: In the Fermi-Dirac distribution, it is always true that $n-r$ cells are empty. Therefore, $\mathbb{P}(X = n - r) = 1$ and $\mathbb{P}(X = t) = 0$ for $t \neq n - r$. It is easily seen that $\mathbb{E}[X] = n - r$ and $X - \mathbb{E}[X] = 0$, so $\mathsf{VAR}[X] = 0$. ■

7. Prove Markov's inequality and Chebyshev's inequality just using pmf of a random variable. Recall that for a finite random variable with pmf p_X, $\mathbb{P}(X \in A) = \sum_{x \in A} p_X(x)$ for any subset $A \subset \mathbb{R}$.

Proof: Suppose that X takes values $\{a_1, ..., a_n\}$ with $0 \leq a_i < a_j$ for $i < j$. Let $t > 0$. If $t \leq a_1$ or $t > a_n$, the Markov inequality is obvious. If not, then there exists k such that $a_k < t \leq a_{k+1}$, following which we do

$$t\mathbb{P}(X \geq t) \leq \sum_{i=k+1}^{n} i\mathbb{P}(X = a_i) \leq \mathbb{E}[X]$$

and the inequality follows. Chebyshev follows from

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) = \mathbb{P}((X - \mathbb{E}[X])^2 \geq t^2) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^2]}{t^2} = \frac{\mathsf{VAR}[X]}{t^2}$$

whence, we are done. ■

8. In a population of size N, m people prefer candidate A. In an opinion poll, n people are chosen at random and asked their preferences. Let Y denote the proportion of people who prefer candidate A among the n randomly chosen people. Find t (depending on N, n, m) such that $\mathbb{P}(|Y - \frac{m}{N}| \geq t) \leq 10^{-2}$. [1]

[1] EXTRA : Can you find n such that $t \leq 10^{-4}$.

Proof: Note that $nY \sim Bin(n, \frac{m}{N})$. Therefore, $\mathbb{E}[Y] = \frac{m}{N}$ and $\mathsf{VAR}[Y] = \frac{m(N-m)}{nN^2}$. By Chebyshev's inequality,

$$\mathbb{P}\left[|Y - \frac{m}{N}\| \geq t\right] \leq \frac{m(N-m)}{nN^2t^2}$$

and therefore, $t^2 \geq \frac{100m(N-m)}{nN^2}$ is sufficient. ■

9. Suppose two permutations of $[n]$ are chosen at random and independently. Let X denote the number of matches between the two random permutations i.e., the number of co-ordinates at which the permutations match. Compute the pmf and mean of X.

Proof: To match at exactly m places, we choose a subset of size m in $\binom{n}{m}$ ways, and ensure the rest of the indices are deranged, which has probability $\frac{D_{n-m}}{(n-m)!}$.Thus, we get $\mathbb{P}(X = m) = \binom{n}{m}\frac{D_{n-m}}{(n-m)!}$.

The expectation is much easier : If X_i denotes the random variable that there is a match at the ith position, then $\mathbb{E}[X_i] = \frac{1}{n}$ and $X = \sum_{i=1}^{n} X_i$, so $\mathbb{E}[X] = 1$. ■

10. Let a standard fair die be rolled. Suppose the die shows the number i, then we choose a coin with probability of heads being $i/6$. Now this coin is tossed independently and repeatedly until we get a heads. Let the random variable X be the number of tosses. Find the probability $\mathbb{P}(X = n)$ for all $n \geq 1$.

Proof: Note that $\mathbb{P}(X = n) = \sum_{i=1}^{6} \mathbb{P}(X = n|D = i)\mathbb{P}(D = i)$ by the law of total probability, where D denotes the value of the dice roll. However, $\mathbb{P}(X = n|D = i) = (1 - \frac{i}{6})^{n-1}\frac{i}{6}$, so we simply get

$$\mathbb{P}(X = n) = \frac{1}{6}\sum_{i=1}^{6}(1 - \frac{i}{6})^{n-1}\frac{i}{6}$$

which is the final answer. ■