WEEK 6

6. Discrete Probability Distributions

In this part of the class we are going to review e few special discrete probability distributions (Uniform, Bernoulli, Binomial, Geometric, Negative Binomial, Hypergeometric and Poisson). First we have to mention that there are so many discrete distributions in the literature. From any finite sum of a sequence you can obtain a discrete probability function. In this part of the class, we are going to give the probability function of the distribution and calculate the mean and the variance. We will also calculate some probabilities.

1. <u>Discrete Uniform Distribution</u>:

Consider a discrete random variable with the range $D_X = \{x_1, x_2, ..., x_n\}$. Consider a probability function

$$f(x) = \begin{cases} \frac{1}{n} & , \quad x \in D_X \\ 0 & , \quad elsewhere. \end{cases}$$

Notice that the function is a probability function because

$$\sum_{x \in D_X} f(x) = \sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} \frac{1}{n} = 1 .$$

We can also calculate the mean and the variance of the distribution as follows:

$$E(X) = \sum_{x \in D_X} x\, f(x) = \sum_{i=1}^{n} x_i\, P(X = x_i) = \sum_{i=1}^{n} x_i \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i = \overline{x}_n$$

and

$$E(X^2) = \sum_{x \in D_X} x^2\, f(x) = \sum_{i=1}^{n} x_i^2 P(X = x_i) = \sum_{i=1}^{n} x_i^2 \frac{1}{n} = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

and therefore the variance can be calculated as

$$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - (\overline{x}_n)^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \overline{x}_n^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x}_n)^2 .$$

<u>Example</u>: Consider a discrete random variable having the following probability function:

$$f(x) = P(X = x) = \begin{cases} c & , \quad x = 1, 2, ..., n \\ 0 & , \quad elsewhere. \end{cases}$$

The constant c is determined as

$$1 = \sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} c = n c \Rightarrow c = \frac{1}{n}$$

and thus the probability function can be written as

$$f(x)=\begin{cases} 1/n & , \quad x=1,2,\ldots,n \\ 0 & , \quad elsewhere. \end{cases}$$

Mean and the variance can be calculayed as

$$E(X)=\sum_{x\in D_X} x\,f(x)=\sum_{x=1}^{n} x\,P(X=x_i)=\frac{1}{n}\sum_{i=1}^{n} x=\frac{1}{n}\frac{n(n+1)}{2}=\frac{n+1}{2}$$

and

$$E(X^2)=\sum_{x\in D_X} x^2\,f(x)=\sum_{x=1}^{n} x^2 P(X=x_i)=\frac{1}{n}\sum_{i=1}^{n} x^2=\frac{1}{n}\frac{n(n+1)(2n+1)}{6}=\frac{(n+1)(2n+1)}{6}$$

and therefore the variance can be calculated as

$$Var(X)=E(X^2)-(E(X))^2=\frac{(n+1)(2n+1)}{6}-\frac{(n+1)^2}{4}=(n+1)\left[\frac{(2n+1)}{6}-\frac{n+1}{4}\right].$$

$$=(n+1)\left[\frac{4n+2-3n-3}{12}\right]=\frac{(n+1)(n-1)}{12}=\frac{n^2-1}{12}.$$

2. <u>Bernoulli Distribution</u>:

Consider an experiment that has only two possible outcomes like tossing a coin (there are only two possible outcomes head or tail). That is, we have only two possible outcomes, say success or failure, denoted by S and F. Therefore the sample space is $\Omega=\{S,F\}$. Such an experiment is called a Bernoulli experiment. Define a random variable X as

$$X:\Omega\to\mathbb{R}$$
$$w\to X(w)=\begin{cases} 0 & , \quad w=F \\ 1 & , \quad w=S \end{cases}$$

Such a random variable is called a Bernoulli random variable. The random variable has a range $D_X=\{0,1\}$ which is a countable subset of the real line so the random variable is discrete. Notice that $P(success)=P(\{S\})=P(\{w:X(w)=1\})=P(X=1)$. Denote the probability of observing a success as p. Therefore the probability of observing a failure will be $1-p$ say q. Thus, the probability function

$$f(x)=\begin{cases} p^x q^{1-x} & , \quad x=0,1 \\ 0 & , \quad elsewhere \end{cases}$$

or simply

$$P(X=x)=p^x q^{1-x}\ ,\ x=0,1$$

This is a probability function known as Bernoulli probability function. If a random variable has a probability function like this we say that X is distributed as Bernoulli and denoted by $X\sim Bern(p)$. Note that

$$\sum_{x=0}^{1} f(x)=\sum_{x=0}^{1} P(X=x)=\sum_{x=0}^{1} p^x q^{1-x}=q+p=1$$

and for any integer k the kth moment of the Bernoulli random variable is

$$E(X^k) = \sum_{x=0}^{1} x^k f(x) = \sum_{x=0}^{1} x^k p^x q^{1-x} = 0^k p^0 q^1 + 1^k p^1 q^{1-1} = p .$$

This means that any moment of a Bernoulli random variable is the probability of a success. That is, $E(X^k) = p$.

For $k=1$ $E(X) = p$ and for $k=2$ $E(X^2) = p$ then we can calculate the variance of Bernoulli distribution as $Var(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p) = pq$.

The moment generating function is

$$M_X(t) = E(e^{tX}) = \sum_{x=0}^{1} e^{tx} P(X=x) = \sum_{x=0}^{1} e^{tx} p^x q^{1-x} = q + pe^t .$$

3. <u>Binomial Distribution</u>:

Consider an experiment of repeating a Bernoulli experiment n time. At each trial the probability of observing a success is the same and the trials are independent of each other. Such an experiment is called a Binomial experiment.

When we repeat a Bernoulli experiment n times let a random variable X counts the number of successes. Therefore the random variable X has range $D_X = \{0,1,2,...,n\}$. Let S denotes a success and F denotes a failure. Then possible observations can be obtained as:

F S S F F . . . . F S S.

The probability of observing a success is p ($P(\{S\}) = p$ and therefore probability of observing a failure will be q namely $P(\{F\}) = q = 1-p$). We look at the probability observing x number of successes (each having probability p) and the rest will be a failure (each having the probability as q) and they are independent. Moreover, the x successes can be observed any time in n trials (x successes can be arranged into n cells). The probability function can be written as

$$P(X=x) = \binom{n}{x} p^x q^{m-x} \ , \ x = 0,1,2,...,n \ .$$

This is a probability function because

$$\sum_{x \in D_X} P(X=x) = \sum_{x=0}^{n} P(X=x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{m-x} = (p+q)^n = 1^n = 1.$$

If a random variable X has a probability function as given above, we say that X is distributed as Binomial with parameters n and p denoted by $X \sim Binom(n,p)$. The moment of Binomial distribution can be calculated directly but if we consider that a Binomial random variable is a sum of Bernoulli random variables, the moments can be calculated much easily. Let X_1 be a random variable which counts the number of successes at the first trial (it is either 0 or 1) and X_2 be a random variable which counts the number of successes at the second trial (it is again either 0 or 1) and continue until X_n which is a random variable which counts the number of successes at the n[th] trial (it is again either 0 or 1) then a Binomial random variable X is a sum of independent Bernoulli random variables. Therefore a Binomial random variable X which counts the number of successes in n independently repeated Bernoulli experiment can be considered as a sum of independent Bernoulli random variables as $X = X_1 + X_2 + ... + X_n$. Remember that if $X_1 \sim Ber(p)$ then $E(X_1^k) = p$ for all k and therefore $E(X_1) = p$ and $Var(X_1) = p\,q$. Therefore the expected value and the variance of Binomial distribution are calculated as

$$E(X) = E(X_1 + X_2 + ... + X_n) = E(X_1) + E(X_2) + ... + E(X_n) = p + p + ... + p = n\,p$$

and

$$Var(X) = Var(X_1 + X_2 + ... + X_n) = Var(X_1) + Var(X_2) + ... + Var(X_n) = pq + pq + ... + pq = n\,pq\,.$$

That is, if $X \sim Binom(n, p)$ then $E(X) = n\,p$ and $Var(X) = n\,pq$.

Remember that if $X_1 \sim Ber(p)$ then the moment generating function of X_1 is $M_{X_1}(t) = q + pe^t$ and the moment generating function of the sum is the product of the moment generating functions and therefore the moment generating function of a Binomial random variable X can be obtained as

$$M_X(t) = E\left(e^{t(X_1 + X_2 + ... + X_n)}\right) = \prod_{i=1}^{n} M_{X_i}(t) = \prod_{i=1}^{n} (q + pe^t) = (q + pe^t)^n\,.$$

That is, if $X \sim Binom(n, p)$ then the moment generating function of X is $M_X(t) = (q + pe^t)^n$.

Example: Consider an experiment of tossing a coin three times. Then all possible outcomes of this experiment is $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$. Define a random variable as

$$X : \Omega \to \mathbb{R}$$

$$w \to X(w) = \begin{cases} 0 \;, & w = HHH \\ 1 \;, & w = HHT, HTH, THH \\ 2 \;, & w = TTH, THT, HTT \\ 3 \;, & w = TTT. \end{cases}$$

Notice that the range of the random variable is $\{0,1,2,3\}$ which counts the number of tails in the experiment. Define a probability measure for any subset of Ω as $P(A) = n(A)/8$ where $n(A)$ represents the number of elements in A. The related probabilities and probability function of the random variable is given below:

$$P(X = 0) = P(\{HHH\}) = 1/8$$

$$P(X = 1) = P(\{HHT, HTH, THH\}) = 3/8$$

$X = x$	0	1	2	3
$P(X = x)$	$1/8$	$3/8$	$3/8$	$1/8$

$P(X=2)=P(\{TTH, THT, HTT\})=3/8$

$P(X=3)=P(\{TTT\})=1/8$

$$P(X=x)=\binom{3}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{3-x} \quad , \; x=0,1,2,3.$$

Using this probability function, we can calculate the probabilities as follows:

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)=1-(1/8)=7/8$$

$$P(1\leq X<3)=P(X=1)+P(X=2)=(3/8)+(3/8)=6/8.$$

4. <u>Geometric Distribution</u> :

Consider an experiment of repeating a Bernoulli experiment until we observe the first success. Such an experiment is called a geometric experiment. Let X be a random variable which counts the number of trials until we observe the first success. Such a random variable is called a Geometric random variable. Again, at each trial the probability of observing a success is the same. That is, $P(\{S\})=p$ and $P(\{F\})=q$. That is we may have a sequence of trials as

F F F . . . . . F S

The last observation will be a success (because we will stop when we observe a success). Since the probability of observing a success or a failure are the same at each trial and having independent trials we can write the probability function of a Geometric random variable as

$$P(X=x)=p\,q^{x-1} \; , \; x=1,2,...$$

If a random variable X has a probability function given above, we say that X distributed as geometric and denoted by $X\sim Geo(p)$. The probability function can also be given by

$$f(x)=\begin{cases} p\,q^{x-1} & , \quad x=1,2,... \\ 0 & , \quad elsewhere \end{cases}$$

This function is actually a probability function because

$$\sum_{x=1}^{\infty} f(x)=\sum_{x=1}^{\infty} p\,q^{x-1}=p\sum_{y=0}^{\infty} q^{y}=p\frac{1}{1-p}=\frac{p}{p}=1 .$$

The mean and the variance of a geometric random variable (can be calculated by using the property of the interchangeability of the derivation and infinite sum) are found to be $E(X)=1/p$ and $Var(X)=q/p^2$.

Let $X\sim Geo(p)$ and try to calculate $P(X>n)$. Note that

$$P(X>n)=1-P(X\leq n)=1-\sum_{x=1}^{n} P(X=x)=1-\sum_{x=1}^{n} p\,q^{x-1}=1-\sum_{y=0}^{n-1} p\,q^{y}=1-p\left(\frac{1-q^n}{1-q}\right)=q^n .$$

Let s and t be two integers such that $s>t$ and let us try to calculate the conditional probability $P(X>s\,|\,X>t)$.

Note that if $s > t$ then $\{X > s\} \subset \{X > t\}$ which implies that $\{X > s\} \cap \{X > t\} = \{X > s\}$. Therefore from the definition of the conditional probability we calculate $P(X > s \mid X > t)$ as

$$P(X > s \mid X > t) = \frac{P(X > s, X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} = \frac{q^s}{q^t} = q^{s-t} = P(X > s-t).$$

That is, $P(X > s \mid X > t) = P(X > s-t)$ which is known as the memoryless property of the random variable X. Among all the discrete probability distributions the Geometric distribution is the one that satisfies the memoryless property.

Example: Suppose two friends are playing a coin tossing game. The first one who obtains head the game will be over. That is, if player I get a head the game will be over otherwise the second player will toss the coin. If the second player (at his/her first trial) gets a head he/she will win the game and the game will be over again. That is the game will continue until one gets a head. Let us try to calculate the probability of the first player wins the game.

Note that the player wins the game at his/her first trial or second trial or so on. Therefore the probability can be calculated as

$$P(player\ wins) = \frac{1}{2} + \frac{1}{2^3} + \frac{1}{2^5} + \frac{1}{2^7} \ldots$$
$$= \frac{1}{2}\left[1 + \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^6} + \ldots\right] = \frac{1}{2}\sum_{x=0}^{\infty}\left(\frac{1}{2}\right)^{2x} = \frac{1}{2}\left(\frac{1}{1-(1/2^2)}\right) = \frac{1}{2}\left(\frac{1}{3/4}\right) = \frac{1}{2}\frac{4}{3} = \frac{2}{3}.$$

5. <u>Negative Binomial Distribution</u> :

When we repeat a Bernoulli experiment until we observe the first success the number is a geometric experiment and the random variable which counts the number of trials is a geometric random variable. Now, consider an experiment of Bernoulli trials until we observe the r successes. Now we have r independent geometric experiments. Let X be a random variable which counts the number of trials until we observe the rth success. Such an experiment is called a Negative Binomial experiment and the random variable X is called a Negative Binomial random variable. Notice that at each trial the probability of observing a success is the same and the trials are independent of each other.

As we have reviewed in Binomial distribution, let X_1 be a random variable which counts the number of trials until we observe the first success and after we observe the first success let X_2 be a random variable which counts the number of successes until we observe the first success (in total the second success) and continue defining the random variable X_r be a random variable which counts the number of trials until we observe the first success after we observe the (r-1)st success. Therefore, the random variables $X_1, X_2, \ldots, X_r$ independently distributed geometric random variables.

Therefore, the random variable $X = X_1 + X_2 + ... + X_r$ is a random variable which counts the number of trials until we observe the rth success. The probability function of X is given by,

$$P(X = x) = \binom{x-1}{r-1} p^r q^{x-r} \ , \ x = r, r+1, r+2, ...$$

A random variable having the above probability function is said to be distributed as Negative Binomial and denoted by $X \sim NB(r, p)$. The mean and the variance of a Negative Binomial random variable can be calculated by using the mean and the variance of Geometric distribution. Since a negative Binomial random variable is a sum of independent Geometric random variable, the mean and the variance $X = X_1 + X_2 + ... + X_r$ is calculated as

$$E(X) = E(X_1 + X_2 + ... + X_r) = E(X_1) + E(X_2) + ... + E(X_r) = \frac{1}{p} + \frac{1}{p} + ... + \frac{1}{p} = \frac{r}{p}$$

and

$$Var(X) = Var(X_1 + X_2 + ... + X_r) = Var(X_1) + Var(X_2) + ... + Var(X_r) = \frac{q}{p^2} + \frac{q}{p^2} + ... + \frac{q}{p^2} = \frac{rq}{p^2}.$$

That is, if $X \sim NB(r, p)$ then $E(X) = r / p$ and $Var(X) = rq / p^2$.

6. <u>Hypergeometric Distribution</u> :

Assume that there are N balls in a box and a of those balls are red, $N - a$ are blue. We take n number of balls at the same time from the box and count the number of red balls from those n balls. Let X be a random variable which counts the number of red balls in the experiment. The probability function of the random variable X is given by

$$P(X = x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}} \ , \ x = 0, 1, 2, ..., n \ .$$

Here,

$$\binom{N}{n} = \frac{N!}{n!(N-n)!},$$

the mean and variance of a hypergeometric random variable have been calculated as

$$E(X) = \frac{na}{N} \quad \text{and} \quad Var(X) = \frac{N-n}{N-1} \frac{an}{N} \left(1 - \frac{a}{N}\right).$$

<u>Example</u>: In a government office 3 Statistician, 4 Mathematician and 3 Econometrician are working. From the office a committee of 3 people will be selected randomly. Calculate the probability that there will be at least one statistician in this committee.

Solution: Let X be a random variable which counts the number of statisticians in that committee. Note that the random variable is distributed as hypergeometric with $n=3$, $a=3$ and the rest is 7 (not statisticians). We want to calculate $P(X \geq 1)$. Since,

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0)$$

it is enough to calculate $P(X = 0)$. Note that

$$\binom{10}{3} = \frac{10!}{3!7!} = \frac{7!8.9.10}{3!7!} = \frac{8.9.10}{6} = \frac{720}{6} = 120, \ \binom{3}{0} = 1 \text{ and } \binom{7}{3} = \frac{7!}{3!4!} = \frac{4!5.6.7}{3!4!} = \frac{5.6.7}{6} = 35$$

and therefore, we calculate the required probability as

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{\binom{3}{0}\binom{7}{3}}{\binom{10}{3}} = 1 - \frac{1(35)}{120} = 1 - \frac{35}{120} = \frac{85}{120} = \frac{17}{24}.$$

7. Poisson Distribution :

So far, we obtained the probability function of a discrete random variable after we defined the experiment. In those experiments we were repeating a Bernoulli experiment and counting the number of successes or the number of trials. Now, consider an experiment of counting the number of successes in a continuous period (like time). For example, we count the number of customers enter to a store between 12:00 to 13:00 (count the number of telephone calls receive between 20:00 to 21:00). Obviously, here the number of customers who enter to a store (or telephone calls received by the operator) are discrete and the period of time is continuous. Such an experiment is called a Poisson experiment. Let λ be the expected number of successes ($\lambda > 0$) in this experiment then the probability to observe x number of successes in the experiment is given by

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}, \ x = 0,1,2,\ldots .$$

This is a probability functions because (Taylor series expansion of e^{λ})

$$\sum_{x=0}^{\infty} P(X = x) = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\lambda^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

If a random variable X has a probability function given above, we say that X is distributed as Poisson with parameter λ and denoted by $X \sim Poisson(\lambda)$. Again using the Taylor series expansion we can find the moment generating function of Poisson random variable. That is, the moment generating function of the Poisson distribution is

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \sum_{x=0}^{\infty} e^{tx} P(X = x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^{x}}{x!} = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^{x}}{x!} \\ &= \sum_{x=0}^{\infty} \frac{e^{-\lambda}(\lambda e^{t})^{x}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^{t})^{x}}{x!} = e^{-\lambda} e^{t\lambda} = e^{\lambda(e^{t}-1)} \end{aligned}.$$

Simply, the moment generating function of the Poisson distribution is $M_X(t) = e^{\lambda(e^t - 1)}$. Notice that we can use the moment generating function of a random variable X to calculate the moments. Note that,

$$\left.\frac{d^k M_X(t)}{dt^k}\right|_{t=0} = \left.\frac{d^k}{dt^k}\Big(E(e^{tX})\Big)\right|_{t=0} = E\Big(X^k e^{tX}\Big)\Big|_{t=0} = E(X^k)$$

The first two derivatives of the moment generating function are

$M'_X(t) = \lambda e^t e^{\lambda(e^t-1)}$ and $M''_X(t) = \lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)}$

and therefore the first two moments of the Poisson distribution are

$$E(X) = M'_X(t)\big|_{t=0} = \left.\lambda e^t e^{\lambda(e^t-1)}\right|_{t=0} = \lambda$$

$$E(X^2) = M''_X(t)\big|_{t=0} = \left.\left(\lambda e^t e^{\lambda(e^t-1)} + (\lambda e^t)^2 e^{\lambda(e^t-1)}\right)\right|_{t=0} = \lambda + \lambda^2 .$$

Thus the variance of Poisson distribution is calculated as

$$Var(X) = E(X^2) - (E(X))^2 = (\lambda + \lambda^2) - \lambda^2 = \lambda .$$

Example: A restaurant at Kızılay Square opens the doors at 7:00 o'clock in the morning. They plan to change their opening hour from 7:00 to 8:00. Before they made the final decision they conduct 30 days experiment and obtain that 5 customers in the average come to the restaurant (30 days-average is 5 customers). For a certain day at the same time period find the probability that less than 3 customers will come to restaurant between 7:00 to 8:00 o'clock in the morning.

Solution:

Let X be a random variable which counts the number of customers between 7:00 to 8:00 o'clock in the morning. Here, $\lambda = 5$ which is the average number of customers come to the restaurant between 7:00 to 8:00 o'clock in the morning. We want to calculate $P(X > 3)$ when X is distributed as Poisson with $\lambda = 5$. The probability function of X will be

$$P(X = x) = \frac{e^{-5} 5^x}{x!} \, , \, x = 0,1,2,...$$

The probability can be calculated directly as

$$P(X > 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \left(\frac{e^{-5}5^0}{0!}\right) + \left(\frac{e^{-5}5^1}{1!}\right) + \left(\frac{e^{-5}5^2}{2!}\right) = e^{-5}\left(1 + 5 + \frac{25}{2}\right) = \frac{37}{2} e^{-5} \cong 0.12465.$$

8. **Multinomial Distribution** :

We have studied a few univariate special discrete distributions. The multinomial distribution (or multivariate binomial) is a well known multivariate discrete probability distribution that we are going to summarize below.

In Bernoulli experiments, there are are only two possible outcomes (succress or failure). That is, there is only one succes at each trial. If there are more possibe success in the experiment it is not a bernoulli experiment any more.

Let $E_1, E_2, ..., E_k$ be the outputs of an experiment (for example, in a box there are k different colors of a ball and we randomly select n balls by replacement. E_1 is the output of first color, E_2 is the second color etc.) And let X_i counts the number of the event E_i occurs. Assume that the probability of the event occurs is p_i, that is $P(E_i) = p_i$. Then the joint probability function of a random vector $(X_1, X_2, ..., X_k)'$ is

$$P(X_1 = x_1, X_2 = x_2, ..., X_k = x_k) = \frac{n!}{x_1! x_2! ... x_k!} p_1^{x_1} p_2^{x_2} ... p_k^{x_k}$$

where $p_1 + p_2 + ... + p_k = 1$, $x_1 + x_2 + ... + x_k = n$ such that $x_i = 0, 1, 2, ..., n$.

Example: Assume that there are 5 black, 4 red and 3 yellow balls in a box. By replacement we randomly select 6 balls from the box. Suppose we want to find the probability that we observe 3 black, 2 red and 1 yellow balls.

Note that, $E_1 = \{black\, ball\}$, $E_2 = \{red\, ball\}$ and $E_3 = \{yellow ball\}$ and note that these events are disjoint and $P(E_1) = p_1 = 5/12$, $P(E_2) = p_2 = 4/12$ and $P(E_3) = p_3 = 3/12$. We randomly select 6 balls from the box and therefore $n = 6$. Let X_1 be the random variable which counts the number of black balls among all these 6 balls, X_2 be the random variable which counts the number of red balls among all these 6 balls and X_3 be the random variable which counts the number of yellow balls among all these 6 balls. Therefore we need to calculate $P(X_1 = 3, X_2 = 2, X_3 = 1)$. This probability can be calculated by using

$$P(X_1 = x_1, X_2 = x_2, X_3 = x_3) = \frac{n!}{x_1! x_2! x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

and putting $n = 6$, $p_1 = 5/12$, $p_2 = 4/12$ and $p_3 = 3/12$ we calculate the required probability as

$$\begin{aligned} P(X_1 = 3, X_2 = 2, X_3 = 1) &= \frac{6!}{(3!)(2!)(1!)} \left(\frac{5}{12}\right)^3 \left(\frac{4}{12}\right)^2 \left(\frac{3}{12}\right)^1 = \frac{6!\,(5)^3 (4)^2 (3)^1}{(3!)(2!)(1!)(12)^6} . \\ &= \frac{4320000}{35831808} = \frac{625}{5184} \cong 0.12056 \end{aligned}$$