Discrete Probability Distribution: Bernoulli, Binomial, and Multinomial Distribution

Probability distributions

A probability distribution describes the probabilities of each of the possible outcomes of a random trial. Some probability distributions can be described mathematically, while others are just a list of the possible outcomes and their probabilities. The precise meaning of a probability distribution depends on whether the variable is discrete or continuous.

A ***probability distribution*** is a list of the probabilities of all mutually exclusive outcomes of a random trial.

Discrete Probability Distribution

Random variable

Let us take an example of measuring the OD of a protein, then

$$\text{Measured OD} = \text{Real OD} + \text{Random Noise}$$

For example:

$$0.23 = 0.22 + (0.01)$$
$$0.20 = 0.22 + (-0.02)$$
$$0.21 = 0.22 + (-0.01)$$
$$0.24 = 0.22 + (0.02)$$

Thus, OD is a random variable.

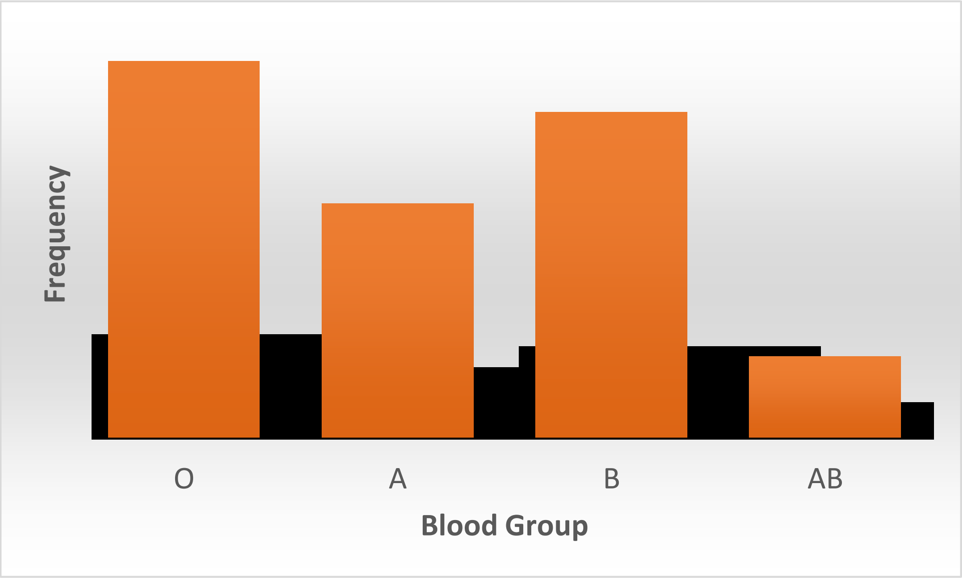
A **discrete random variable** takes discrete numerical values. For example
(1) Number of heads in 10 coin tosses
(2) Number of 6 in 10 die throws
(3) Number of people having a specific mutation
(4) Number of cells in an well of a 96 well plate.

These are few examples of discrete random variables.

Discrete Probability Distribution

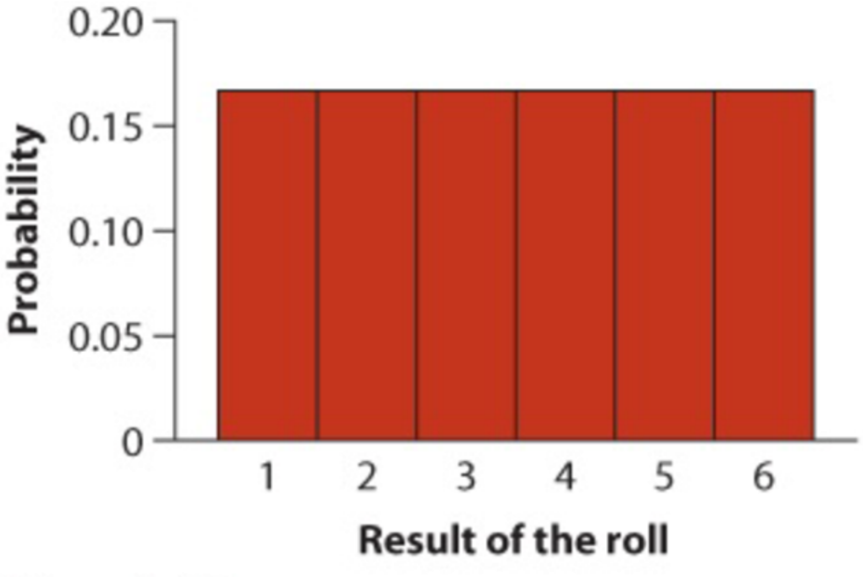
Frequency distribution

Blood Group	Population distribution (%)	Fraction
O	37	0.37
A	23	0.23
B	32	0.32
AB	8	0.08
Sum of fractions		1.00



Discrete Probability Distribution

A discrete numerical variable is measured in indivisible units. Categorical variables are discrete, as are many numerical variables. A discrete probability distribution gives the probability of each possible value of a discrete variable. Categorical and discrete numerical variables have discrete probability distributions.



The probability distribution of outcomes resulting from the roll of a single six-sided fair die. The probability of each possible outcome is 1/6 = 0.167.

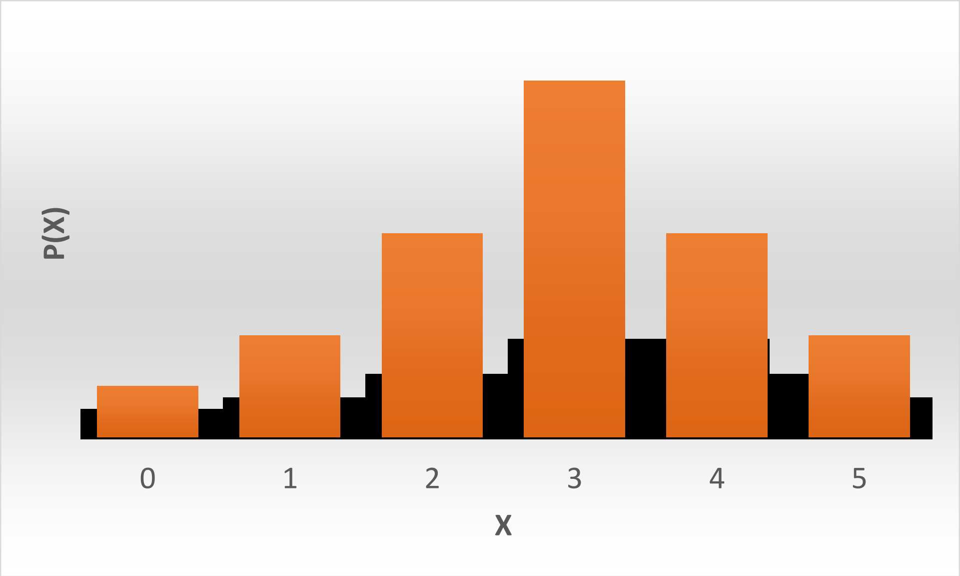
The probability distribution for the sum of the numbers resulting from rolling two 6-sided fair dice.

Discrete Probability Distribution

Probability distribution

Let X is a random variable

X	Probability P(X)
0	0.05
1	0.10
2	0.20
3	0.35
4	0.20
5	0.10
Sum	1.00



Probability distribution is a mathematical function that will give the probability of a specific value of a random variable.

Discrete Probability Distribution

Let us take an example of very common process of coin toss. We assume that pH = pT = 0.5

What is the probability of having 2 H in 5 toss ? i.e. P (H=2) = ?

To answer this question, we need to know that how many ways can one get 2 H in 5 toss?

$$ {}^{5}c_{2} = \frac{5!}{2!\,(5-2)!} = 10 $$

And those are: {HHTTT, HTHTT, HTTHT, HTTTH, THHTT, THTHT, THTTH, TTHHT, TTHTH, TTTHH}.

Note that these are mutually exclusive and independent.

Thus, using the product rule, P(HHTTT) = pH x pH x pT x pT x pT = $(½)^5$ = 1/32

P(HTHTT) = pH x pT x pH x pT x pT = $(½)^5$ = 1/32

and so on.

Binomial Distribution

Binomial Distribution

$$P(HHTTT) = pH^2 \times pT^{(5-2)}$$
$$P(HTHTT) = pH^2 \times pT^{(5-2)}$$
$$P(HTTHT) = pH^2 \times pT^{(5-2)}$$
$$P(HTTTH) = pH^2 \times pT^{(5-2)}$$
$$P(THHTT) = pH^2 \times pT^{(5-2)}$$
$$P(THTHT) = pH^2 \times pT^{(5-2)}$$
$$P(THTTH) = pH^2 \times pT^{(5-2)}$$
$$P(TTHHT) = pH^2 \times pT^{(5-2)}$$
$$P(TTHTH) = pH^2 \times pT^{(5-2)}$$
$$P(TTTHH) = pH^2 \times pT^{(5-2)}$$

Since these events (outcomes) are **mutually exclusive**, by using **addition rule**,

$$P\ (H=2) = 10 \times pH^2 \times pT^{(5-2)}$$
$$P\ (H=2) = {}^5C_2 \times pH^2 \times pT^{(5-2)}$$

Binomial Distribution

Thus, it can be generalized for N coin tosses. Let us assume, pH = p and pT = 1-p, then for getting H = k (getting k number of times H),

$$p(H = k) = {}^{N}c_{k} p^{k} (1 - p)^{\mathrm{N}-k} \qquad\qquad 0 \leq k \leq N$$

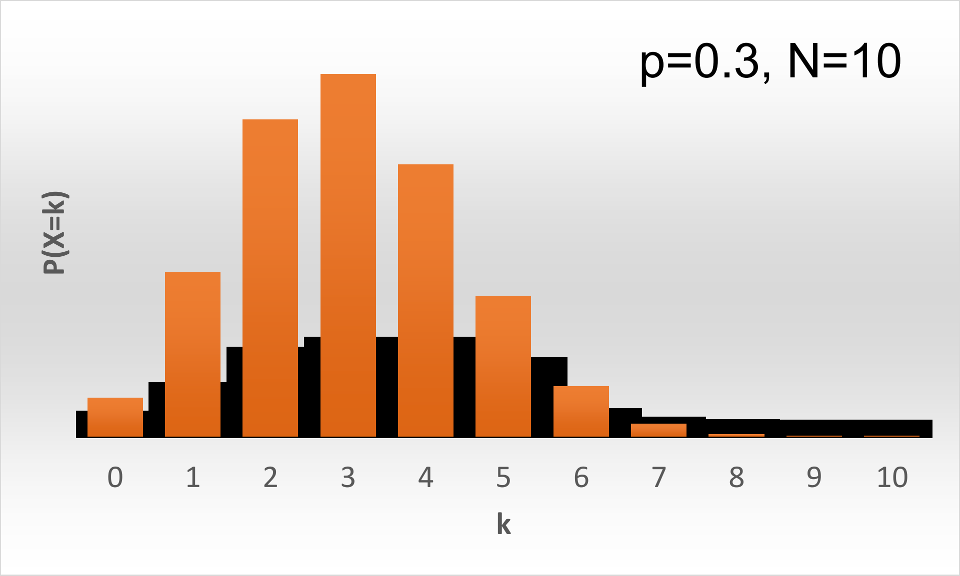
Binomial Distribution

Let us assume that we have N **independent** trials and each trial has two **mutually exclusive** outcomes, with P(success) = p, and X = total number of success in N trials (X is a discrete random variable 0 ≤ X ≤ N), then P (X=k) would be

$$p(\mathrm{X} = k) = {}^{N}c_{k} p^{k} (1 - p)^{\mathrm{N}-k}$$

This is called **Probability Mass Function (PMF)** of **Binomial Distribution**.

Discrete Probability Distribution



Binomial Distribution

Binomial probability distribution is a discrete probability distribution used for the events that give results in ‘Yes or No’ or ‘Success or Failure’. It’s particularly useful in scenarios where these outcomes are mutually exclusive, and the probability of success (usually denoted as “p”) and the probability of failure (usually denoted as “q”) is constant for each trial. This distribution helps calculate the probability of getting a specific number of successes in a fixed number of trials, making it valuable in fields such as statistics, economics, and quality control.

Binomial Distribution Definition

Binomial distribution for a random variable X = 0, 1, 2, …., n is defined as the probability distribution of two outcomes, success or failure, in a series of events. Binomial distribution in statistics uses one of the two independent variables in each trial where the outcome of each trial is independent of the outcome of other trials.

Each trial done to obtain the outcome in success or failure is called **Bernoulli Trial** and the probability distribution for each Bernoulli Trial is called the **Bernoulli Distribution**.

Binomial Distribution

Bernoulli Trial
It is a trial that gives results of dichotomous nature i.e. result in yes or no, head or tail, even or odd. It means it gives two types of outcomes out of which one favors the event while the other doesn't.

A random experiment is called Bernoulli Trial if it satisfies the following conditions:

- Trials are finite in number.
- Trials are independent of each other.
- Each trial has only two possible outcomes.
- The probability of success and failure in each trial is the same.

Binomial Distribution

Binomial Distribution Calculation
Binomial Distribution in statistics is used to compute the probability of likelihood of an event using the above formula. To calculate the probability using binomial distribution we need to follow the following steps:

Step 1: Find the number of trials and assign it as ‘n’
Step 2: Find the probability of success in each trial and assign it as ‘p’
Step 3: Find the probability of failure and assign it as q where q = 1-p
Step 4: Find the random variable X = r for which we have to calculate the binomial distribution
Step 5: Calculate the probability of Binomial Distribution for X = r using the Binomial Distribution Formula.

Binomial Distribution

Binomial Distribution Examples

Let's say we toss a coin twice, and getting head is a success. We have to calculate the probability of success and failure then, in this case, we will calculate the probability distribution as follows:

In each trial getting a head that is a success, its probability is given as p =1/2 and n = 2 as we throw a coin twice

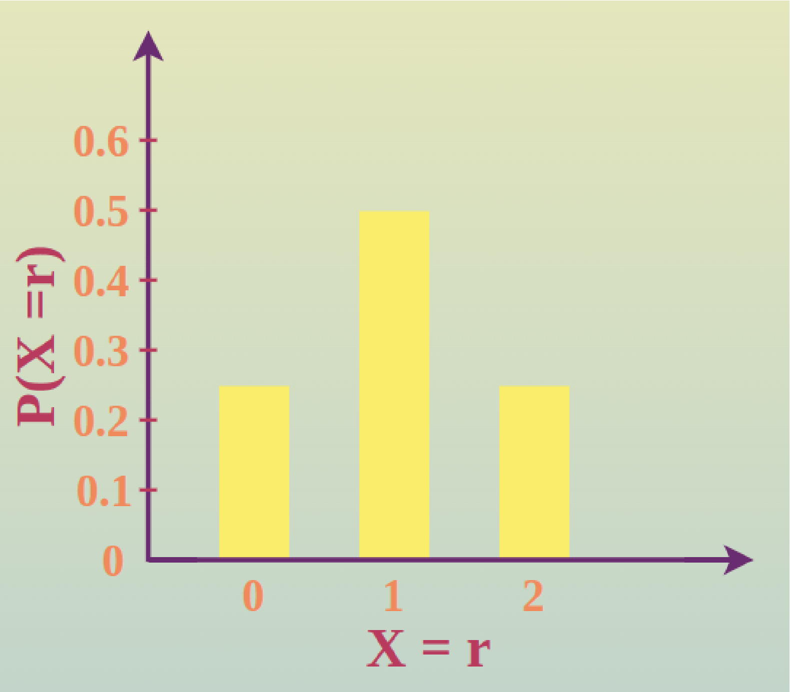
r = 0 for no success, r = 1 for getting head once, and r = 2 for getting head twice

P(Getting 0 heads) = P(X = 0) = ${}^{n}c_{r}p^{r}q^{n-r} = {}^{2}c_{0}(1/2)^{0}(1/2)^{2} = 1/4$

P(Getting 1 head) = P(X = 1) = ${}^{n}c_{r}p^{r}q^{n-r} = {}^{2}c_{1}(1/2)^{1}(1/2)^{1} = 2 \times 1/2 \times 1/2 = 1/2$

P(Getting 2 heads) = P(X = 2) = ${}^{n}c_{r}p^{r}q^{n-r} = {}^{2}c_{2}(1/2)^{2}(1/2)^{0} = 1/4$

Random Variable X = r	P(X = r)
X = 0 (Getting 0 Head)	1/4
X = 1 (Getting 1 Head)	1/2
X = 2 (Getting 2 Head)	1/4



Binomial Distribution

Binomial Distribution Examples

The binomial distribution for a situation when getting 6 is a success on throwing two dies. First of all, we see that it is a Bernoulli Trial as getting 6 is the only success, and getting any different is a failure. Now we can get six on both die in a trial or six on only one of the die in a trial and getting no six on both die. Hence, the random variable for which we have to find the probability takes the value X = r = 0, 1, 2.

Random Variable X = r	P(X = r)
X = 0 (Getting no 6)	25/36
X = 1 (Getting one 6)	10/36
X = 2 (Getting two 6)	1/36

We see that sum of all the probabilities 25/36 + 10/36 + 1/36 = 1.

Binomial Distribution

Binomial Distribution Mean
The Mean of Binomial Distribution is the measurement of average success that would be obtained in ‘n’ number of trials. The Mean of Binomial distribution is also called Binomial distribution expectation.

$$\mu = np$$

Where, **μ** is the Mean or Expectation, **n** is the total number of trials, and **p** is the probability of success in each trial.

Example: If we toss a coin 20 times and getting head is the success then what is the mean of success?
Total Number of Trials n = 20
Probability of getting head in each trial, p = 1/2 = 0.5
Mean = n.p = 20 × 0.5
It means on average we would head 10 times on tossing a coin 20 times.

Binomial Distribution

Binomial Distribution Variance

The **variance** of Binomial Distribution tells about the dispersion or spread of the distribution. It is given by the product of the number of trials, probability of success, and probability of failure.

$$\boldsymbol{\sigma^2 = npq}$$

Where $\boldsymbol{\sigma^2}$ is the variance, **n** is the total number of trials, **p** is the probability of success in each trial, and **q** is the probability of failure in each trial.

Example: If we toss a coin 20 times and getting head is the success then what is the variance of the distribution?

We have, n = 20

Probability of Success in each trial (p) = 0.5

Probability of Failure in each trial (q) = 0.5

Variance of the Binomial Distribution, $\sigma = n.p.q = (20 \times 0.5 \times 0.5) = 5$

Binomial Distribution

Example 1: A die is thrown 6 times and if getting an even number is a success what is the probability of getting (i) 4 Successes (ii) No success

Given: $n = 6$, $p = 3/6 = 1/2$, and $q = 1 – 1/2 = 1/2$

$P(X = r) = {}^nC_r p^r q^{n-r}$

(i) $P(X = 4) = {}^6C_4(1/2)^4(1/2)^2 = 15/64$

(ii) $P(X = 0) = {}^6C_0(1/2)^0(1/2)^6 = 1/64$

Binomial Distribution

Example 2: A coin is tossed 4 times what is the probability of getting at least 2 heads?

Given: $n = 4$
Probability of getting head in each trial, $p = 1/2 \Rightarrow q = 1 - 1/2 = 1/2$

$P(X = r) = {}^4C_r(1/2)^r(1/2)^{4-r} \Rightarrow P(X = r) = {}^4C_r(1/2)^4$

And we know, Probability of getting at least 2 heads = $P(X \geq 2)$
$\Rightarrow$ Probability of getting at least 2 heads = $P(X = 2) + P(X = 3) + P(X = 4)$
$\Rightarrow$ Probability of getting at least 2 heads = ${}^4C_2(1/2)^4 + {}^4C_3(1/2)^4 + {}^4C_4(1/2)^4$
$\Rightarrow$ Probability of getting at least 2 heads = $({}^4C_2 + {}^4C_3 + {}^4C_4)(1/2)^4$
$\Rightarrow$ Probability of getting at least 2 heads = $11(1/2)^4 = 11/16$

Binomial Distribution

Example 3: A pair of die is thrown 6 times and getting sum 5 is a success then what is the probability of getting (i) no success (ii) two success (iii) at most two success

Given: $n = 6$

5 can be obtained in 4 ways (1,4) (4,1) (2,3) (3,2)

Probability of getting the sum 5 in each trial, $p = 4/36 = 1/9$

Probability of not getting sum 5 $= 1 – 1/9 = 8/9$

(i) Probability of getting no success, $P(X = 0) = {}^6C_0(1/9)^0(8/9)^6 = (8/9)^6$

(ii) Probability of getting two success, $P(X = 2) = {}^6C_2(1/9)^2(8/9)^4 = 15(8^4/9^6)$

(iii) Probability of getting at most two successes, $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$\Rightarrow P(X \leq 2) = (8/9)^6 + 6(8^5/9^6) + 15(8^4/9^6)$

Multinomial Distribution

Multinomial Distribution

Multinomial distributions specifically deal with events that have **multiple discrete outcomes**. The Binomial distribution is a specific subset of multinomial distributions in which there are only two possible outcomes to an event.

Multinomial distributions are not limited to events only having discrete outcomes. It is possible to categorize outcomes with continuous distributions to different degrees (high, medium, low). For instance, the water level - a continuous entity - in a storage tank can be made discrete by categorizing them into either "desirable" or "not desirable." Multinomial distributions, therefore, have expansive applications in process control.

Multinomial Distribution

Multinomial Distribution
N independent trials. Each trial has k mutually exclusive outcomes. Let the probability of i^{th} outcome is p_i such that sum of p_i over all k is 1.

Let us take an example of a dice throwing 10 times. Then, what is the probability of getting 1 on 2 times, 2 on 1 time, 3 on 2 times, 4 on 2 times, 5 on 1 time and 6 on 2 times. That is P(1:2, 2:1, 3:2, 4:2, 5:1, 6:2).

This type of probability can be obtained by PMF of multinomial distribution, which is given as

$$p(x_1, x_2, x_3, \cdots, x_k) = \frac{N!}{x_1!x_2!x_3!\cdots x_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \cdots p_k^{x_k}$$

By using the above PMF, we can calculate

$$p(1\colon 2, 2\colon 1, 3\colon 2, 4\colon 2, 5\colon 1, 6\colon 2) = \frac{10!}{2!\,1!\,2!\,2!\,1!\,2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^1\left(\frac{1}{6}\right)^2$$

Multinomial Distribution

Multinomial Distribution Mean and Variance

$$E(X_i) = np_i$$

$$Var(X_i) = np_i(1 - p_i)$$

Multinomial Distribution

Example 1. If a fair die is tossed 12 times. What is the probability that each face value (1-6) will occur exactly twice?

The probability can be determined using a multinomial distribution in which 6 outcomes are possible. The individual probabilities are all equal given that it is a fair die, p = 1/6. The total number of trials N is 12, and the individual number of occurrences in each category n is 2.

$$P(2,2,2,2,2,2) = \frac{12!}{2!\,2!\,21.2!\,2!\,2!}\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2\left(\frac{1}{6}\right)^2 = 0.003488$$

Multinomial Distribution

Example 2. A bowl has 2 maize marbles, 3 blue marbles and 5 white marbles. A marble is randomly selected and then placed back in the bowl. You do this 5 times. What is the probability of choosing 1 maize marble, 1 blue marble and 3 white marbles?

N the number of trials = 5, k the number of possible outcomes = 3, n_i is the number of occurrences of outcome i, p_i is the probability of seeing outcome i

The three possible outcomes are choosing a maize marble, a blue marble or a white marble. We must determine n_i and p_i to solve the multinomial distribution.

The number of occurrences of the outcome are the number of times we wish to see each outcome. These are given in the problem statement.

$n_{maize} = 1$, $n_{blue} = 1$, $n_{white} = 3$

The probability of seeing each outcome is easy to find. For example, there are two maize marbles in the bowl of 10, so the probability of choosing a maize marble is 2/10. i.e.

$p_{maize} = 2/10$, $p_{blue} = 3/10$, $p_{white} = 5/10$

$$P(1,1,3) = \frac{5!}{1!\,1!\,3!} \times \left(\frac{2}{10}\right)^1 \left(\frac{3}{10}\right)^1 \left(\frac{5}{10}\right)^3 = 0 \cdot 15$$

Multinomial Distribution

Example 3. In India, 30% of the population has a blood type of O+, 33% has A+, 12% has B+, 6% has AB+, 7% has O-, 8% has A-, 3% has B- and 1% has AB-. If 15 Indian citizens are chosen at random, what is the probability that 3 have a blood type of O+, 2 have A+, 3 have B+, 2 have AB+, 1 has O-, 2 have A-, 1 has B- and 1 has AB-?

Given that N = 15 (trials), p1 = 0.30 (O+), p2 = 0.33 (A+), p3 = 0.12 (B+), p4 = 0.06 (AB+), p5 = 0.07 (O-), p6 = 0.08 (A-), p7 = 0.03 (B-), p8 = 0.01 (AB-).

Also, n1 = 3 (O+), n2 = 2 (A+), n3 = 3 (B+), n4 = 2 (AB+), n5 = 1 (O-), n6 = 2 (A-), n7 = 1 (B-), n8 = 1 (AB-). Thus, k = 8.

$$P(3,2,3,2,1,2,1,1) = \frac{15!}{3!\,2!\,3!\,2!\,1!\,2!\,1!\,1!} \times (0.30)^3(0 \cdot 33)^2(0.12)^3(0.06)^2(0.07)^1(0.08)^2(0.03)^1(0.01)^1 = 0.000011$$

Thank You