CHAPTER 4

Probability Distributions

4.1. Probability Distributions of Discrete Random Variables:

Introductory Example:

Consider the following discrete random variable.

X = The number of times a Saudi person had a cold in January 2010.

x	freq. of x $n(X=x)$	$P(X=x) = \frac{n(X=x)}{16000000}$ (Relative frequency)
0	10000000	0.6250
1	3000000	0.1875
2	2000000	0.1250
3	1000000	0.0625
Total	16000000	1.0000

Note that the possible values of the random variable X are: x=0,1,2,3

The probability distribution of the discrete random variable X is given by the following table:

x	$P(X = x) = f(x)$
0	0.6250
1	0.1874
2	0.1250
3	0.0625
Total	1.0000

Notes:

The probability distribution of any discrete random variable X must satisfy the following two properties:

(1) $0 \leq P(X = x) \leq 1$

(2) $\sum_x P(X = x) = 1$

Considering the previous table, we can easily calculate:

$$(1)\ P(X \geq 2) = P(X = 2) + P(X = 3) = 0.1250 + 0.0625 = 0.1875$$

$$(2)\ P(X > 2) = P(X = 3) = 0.0625 \qquad [\text{note: } P(X > 2) \neq P(X \geq 2)]$$

$$(3)\ P(1 \leq X < 3) = P(X = 1) + P(X = 2) = 0.1875 + 0.1250 = 0.3125$$

$$(4)\ P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$= 0.6250 + 0.1875 + 0.1250 = 0.9375$$

another solution:

$$P(X \leq 2) = 1 - P((\overline{X \leq 2}))$$
$$= 1 - P(X > 2) = 1 - P(X = 3) = 1 - 0.625 = 0.9375$$

$$(5)\ P(-1 \leq X < 2) = P(X = 0) + P(X = 1)$$
$$= 0.6250 + 0.1875 = 0.8125$$

(6) $P(-1.5 \le X < 1.3) = P(X = 0) + P(X = 1)$

$$= 0.6250 + 0.1875 = 0.8125$$

(7) $P(X = 3.5) = P(\phi) = 0$

(8) $P(X \le 10) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) = P(\Omega) = 1$

(9) The probability that the selected person had at least 2 cold:

$$P(X \ge 2) = P(X = 2) + P(X = 3) = 0.1875$$

(10) The probability that the selected person had at most 2 colds:

$$P(X \le 2) = 0.9375$$

(11) The probability that the selected person had more than 2 colds:

$$P(X > 2) = P(X = 3) = 0.0625$$

(12) The probability that the selected person had less than 2 colds:

$$P(X < 2) = P(X = 0) + P(X = 1) = 0.8125$$

4.1.1 Mean and Variance of a Discrete random variable

Mean: The mean (or expected value) of a discrete random variable X is denoted by μ or μ_X. It is defined by:

$$\mu = \sum_x x\,P(X = x)$$

Variance: The variance of a discrete random variable X is denoted by σ^2 or σ_X^2. It is defined by:

$$\sigma^2 = \sum_x (x-\mu)^2 P(X = x)$$

Example:

We wish to calculate the mean μ and the variance of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X=x)$
0	0.05
1	0.25
2	0.45
3	0.25

Solution:

$$\mu = \sum_x x\ P(X=x) = (0)(0.05) + (1)(0.25) + (2)(0.45) + (3)(0.25) = 1.9$$

$$\sigma^2 = \sum_x (x-1.9)^2\ P(X=x)$$

$$= (0-1.9)^2(0.05) + (1-1.9)^2(0.25) + (2-1.9)^2(0.45) + (3-1.9)^2(0.25)$$

$$= 0.69$$

Theorem

The variance of the random variable X is given by:

$$\mathrm{Var}(X) = \sigma_X^2 = E(X^2) - \mu^2$$

$$\text{where } E(X^2) = \begin{cases} \sum_{all\ x} x^2 f(x)\,; & if\ X\ is\ discrete \\ \int_{-\infty}^{\infty} x^2 f(x)\,dx\,; & if\ X\ is\ continuous \end{cases}$$

Calculate the variance using this formula for the previous example

4.1.2 Cumulative distribution

The cumulative distribution function of a discrete r. v. X is defined by:

$$P(X \le x) = \sum_{a \le x} P(X = a)$$

Example:

Calculate the cumulative distribution of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X = x)$
0	0.05
1	0.25
2	0.45
3	0.25

Use the cumulative distribution to find:

$P(X \le 2)$, $P(X < 2)$, $P(X \le 1.5)$, $P(X < 1.5)$, $P(X > 1)$, $P(X \ge 1)$

Solution:

The cumulative distribution of X is:

x	$P(X \le x)$
0	0.05
1	0.30
2	0.75
3	1.0000

$$P(X \le 0) = P(X = 0)$$
$$P(X \le 1) = P(X = 0) + P(X = 1)$$
$$P(X \le 2) = P(X = 0) + P(X = 1) + P(X = 2)$$
$$P(X \le 3) = P(X = 0) + \cdots + P(X = 3)$$

Using the cumulative distribution,

$$P(X \le 2) = 0.75$$
$$P(X < 2) = P(X \le 1) = 0.30$$
$$P(X \le 1.5) = P(X \le 1) = 0.30$$
$$P(X < 1.5) = P(X \le 1) = 0.30$$
$$P(X > 1) = 1 - P(\overline{(X > 1)}) = 1 - P(X \le 1) = 1 - 0.30 = 0.70$$
$$P(X \ge 1) = 1 - P(\overline{(X \ge 1)}) = 1 - P(X < 1) = 1 - P(X \le 0)$$
$$= 1 - 0.05 = 0.95$$

Combinations:

In many problems, we are interested in the number of ways of selecting r objects from n objects without regard to order. These selections are called combinations.

Combinations:

Notation: n factorial is denoted by $n!$ and is defined by:

$$n! = n(n-1)(n-2)\cdots(2)(1) \qquad for \quad n \geq 1$$

$$0! = 1$$

Example: $5! = (5)(4)(3)(2)(1) = 120$

Theorem:

The number of different ways for selecting r objects from n distinct objects is denoted by $\binom{n}{r}$ and is given by:

$$\binom{n}{r} = \frac{n!}{r!\ (n-r)!}; \qquad r = 0, 1, 2, \ldots, n$$

Notes:

$\binom{n}{r}$ is read as " n " choose " r ".

$$\binom{n}{n} = 1 \qquad \binom{n}{0} = 1 \qquad \binom{n}{r} = \binom{n}{n-r}$$

Example

If we have 10 equal–priority operations and only 4 operating rooms, in how many ways can we choose the 4 patients to be operated on first?

Answer:

$n = 10 \qquad r = 4$

The number of different ways for selecting 4 patients from 10 patients is

$$\binom{10}{4} = \frac{10!}{4!\,(10-4)!} = \frac{10!}{4!\,6!} = \frac{(10)(9)(8)\cdots(2)(1)}{(4)(3)(2)(1)\quad(6)(5)(4)(3)(2)(1)}$$

$$= 210 \qquad (\textit{different ways})$$

4.1.3 Binomial Distribution

- Bernoulli Trial: is an experiment with only two possible outputs: **S=success** and F=failure (Boy or girl, Saudi or non-Saudi)
- Binomial distribution: is used to model an experiment for which:

1. The experiment has a sequence of n Bernoulli trials.
2. The probability of success is $P(S) = p$, and the probability of failure is $P(F) = 1 - p = q$.
3. The probability of success $P(S) = p$ is constant for each trial.
4. The trials are independent; that is the outcome of one trial has no effect on the outcome of any other trial.

The possible values of X (number of success in n trails) are:

$$x = 0, 1, 3, \dots, n$$

The r.v. X has a binomial distribution with parameters *n* and p , and we write:

$$X \sim Binomial(n, p)$$

The probability distribution of *X* is given by:

$$P(X = x) = \begin{cases} {}_nC_x \; p^x \; q^{n-x} & for\ x = 0, 1, 2, \dots, n \\ 0 & otherwise \end{cases}$$

Where: $${}_nC_x = \frac{n!}{x!\,(n-x)!}$$

Mean and Variance of Binomial Distribution

If $X\sim$ Binomial(n, p) , then

- The mean: $\mu = np$ (expected value)
- The variance: $\sigma^2 = npq$

Example:

Suppose that the probability that a Saudi man has high blood pressure is 0.15. Suppose that we randomly select a sample of 6 Saudi men.

(1) Find the probability distribution of the random variable (X) representing the number of men with high blood pressure in the sample.

(2) Find the expected number of men with high blood pressure in the sample (mean of X).

(3) Find the variance X.

(4) What is the probability that there will be exactly 2 men with high blood pressure?

(5) What is the probability that there will be at most 2 men with high blood pressure?

(6) What is the probability that there will be at lease 4 men with high blood pressure?

Solution:

We are interested in the following random variable:

X = The number of men with high blood pressure in the sample of 6 men.

Notes:

- Bernoulli trial: diagnosing whether a man has a high blood pressure or not. There are two outcomes for each trial:

S = Success: The man has high blood pressure

F = failure: The man does not have high blood pressure.

- Number of trials = 6 (we need to check 6 men)
- Probability of success: $P(S) = p = 0.15$
- Probability of failure: $P(F) = q = 1 - p = 0.85$
- Number of trials: $n = 6$
- The trials are independent because of the fact that the result of each man does not affect the result of any other man since the selection was made ate random.

The random variable X has a binomial distribution with parameters: n=6 and p=0.15, that is:

$$X \sim \text{Binomial (n, p)}$$

$$X \sim \text{Binomial (6, 0.15)}$$

The possible values of X are:

$$x = 0, 1, 3, 4, 5, 6$$

(1) The probability distribution of X is:

$$P(X=x)=\begin{cases} {}_6C_x \ (0.15)^x (0.85)^{6-x} & ; x=0,1,2,3,4,5,6 \\ 0 & ; \quad otherwise \end{cases}$$

The probabilities of all values of X are:

$$P(X=0)={}_6C_0\ (0.15)^0(0.85)^6=(1)(0.15)^0(0.85)^6=0.37715$$

$$P(X=1)={}_6C_1\ (0.15)^1(0.85)^5=(6)(0.15)(0.85)^5=0.39933$$

$$P(X=2)={}_6C_2\ (0.15)^2(0.85)^4=(15)(0.15)^2(0.85)^4=0.17618$$

$$P(X=3)={}_6C_3\ (0.15)^3(0.85)^3=(20)(0.15)^3(0.85)^3=0.04145$$

$$P(X=4)={}_6C_4\ (0.15)^4(0.85)^2=(15)(0.15)^4(0.85)^2=0.00549$$

$$P(X=5)={}_6C_5\ (0.15)^5(0.85)^1=(6)(0.15)^5(0.85)^1=0.00039$$

$$P(X=6)={}_6C_6\ (0.15)^6(0.85)^0=(1)(0.15)^6(1)=0.00001$$

(2) The mean of the distribution (the expected number of men out of 6 with high blood pressure) is:

$$\mu = np = (6)(0.15) = 0.9$$

(3) The variance is:

$$\sigma^2 = npq = (6)(0.15)(0.85) = 0.765$$

(4) The probability that there will be exactly 2 men with high blood pressure is:

$$P(X = 2) = 0.17618$$

(5) The probability that there will be at most 2 men with high blood pressure is:

$$\begin{aligned} P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= 0.37715 + 0.39933 + 0.17618 \\ &= 0.95266 \end{aligned}$$

(6) The probability that there will be at lease 4 men with high blood pressure is:

$$\begin{aligned} P(X \geq 4) &= P(X=4) + P(X=5) + P(X=6) \\ &= 0.00549 + 0.00039 + 0.00001 \\ &= 0.00589 \end{aligned}$$

4.1.4 Poisson Distribution

The discrete r. v. X is said to have a Poisson distribution with parameter (average or mean) **λ** if the probability distribution of X is given by

$$P(X=x)=\begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; \quad for \quad x=0,1,2,3,\ldots \\ 0 & ; \quad otherwise \end{cases}$$

where $e = 2.71828$ (the natural number), and we write

$$\boldsymbol{X \sim \text{Poisson}(\lambda)}$$

- The mean (average) of X is : $\mu = \lambda$ (Expected value)
- The variance of X is: $\sigma^2 = \lambda$

Some applications:

- No. of patients in a waiting room in an hours.
- No. of surgeries performed in a month.
- No. of rats in each house in a particular city.

Note:

If X = The number of patients seen in the emergency unit in a day, and if $X \sim Poisson\ (\lambda)$, then:

1. The average (mean) of patients seen every day in the emergency unit = λ.
2. The average (mean) of patients seen every month in the emergency unit =30λ.
3. The average (mean) of patients seen every year in the emergency unit = 365λ.
4. The average (mean) of patients seen every hour in the emergency unit = $\lambda/24$.

Example:

Suppose that the number of snake bites cases seen at KKUH in a year has a Poisson distribution with average 6 bite cases.

(1) What is the probability that in a year:

(i) The no. of snake bite cases will be 7?

(ii) The no. of snake bite cases will be less than 2?

(2) What is the probability that there will be 10 snake bite cases in 2 years?

(3) What is the probability that there will be no snake bite cases in a month?

Solution:

(1) $X =$ no. of snake bite cases in a year.

$$X \sim \text{Poisson}(6) \qquad (\lambda=6)$$

$$P(X=x)=\frac{e^{-6}6^{x}}{x!}\ ; \quad x=0,1,2,\ldots$$

(i) $$P(X=7)=\frac{e^{-6}6^{7}}{7!}=0.13768$$

(ii) $$P(X<2)=P(X=0)+P(X=1)$$

$$=\frac{e^{-6}6^{0}}{0!}+\frac{e^{-6}6^{1}}{1!}=0.00248+0.01487=0.01735$$

(2) Y = no of snake bite cases in 2 years

$$Y \sim \text{Poisson}(12) \qquad (\lambda^{*}=2\lambda=(2)(6)=12)$$

$$P(Y=y)=\frac{e^{-12}12^{y}}{y!}: \quad y=0\ ,1\ ,2\ldots$$

$$\therefore P(Y=10)=\frac{e^{-12}12^{10}}{10!}=0.1048$$

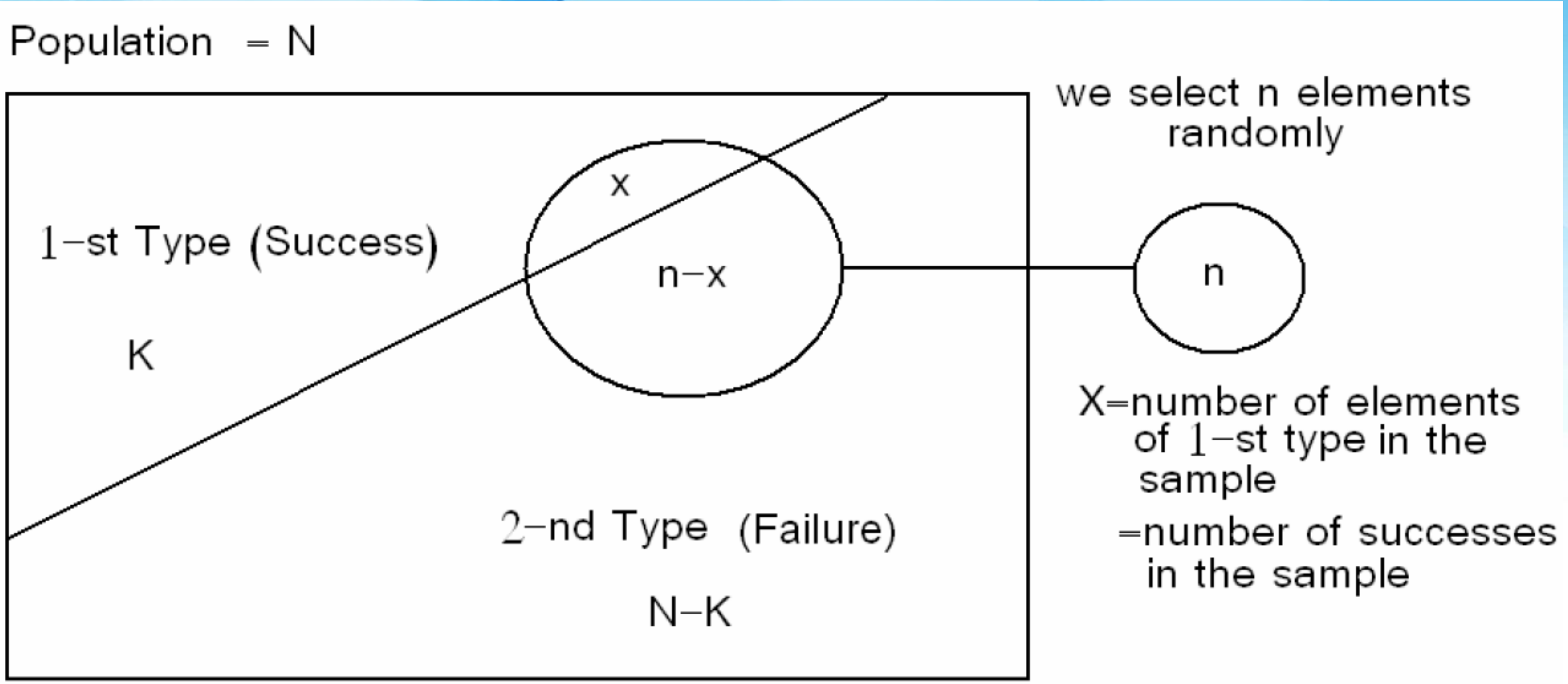
(3) W = no. of snake bite cases in a month.

$W \sim$ Poisson (0.5) $\qquad \left(\lambda^{**} = \frac{\lambda}{12} = \frac{6}{12} = 0.5\right)$

$$P(W = w) = \frac{e^{-0.5}\ 0.5^{w}}{w!} : \qquad w = 0, 1, 2\ldots$$

$$P(W = 0) = \frac{e^{-0.5}(0.5)^{0}}{0!} = 0.6065$$

Hypergeometric Distribution



Suppose there is a population with 2 types of elements: 1-st Type = success 2-nd Type = failure

- N= population size • K= number of elements of the 1-st type
- $N-K$ = number of elements of the 2-nd type
- We select a sample of n elements at random from the population
- Let X = number of elements of 1-st type (number of successes) in the sample
- We need to find the probability distribution of X.

<u>There are two methods of selection:</u>

1. selection with replacement
2. selection without replacement

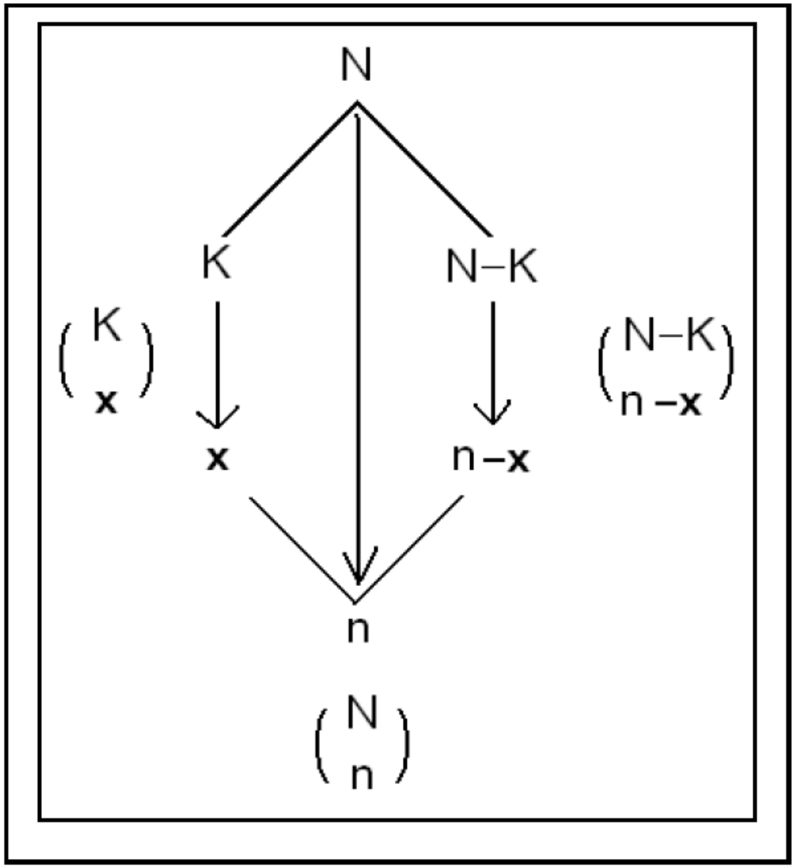
(1) If we select the elements of the sample at random and with replacement, then

$X \sim Binomial(n, p)$; where $p = \frac{k}{N}$

(2) Now, suppose we select the elements of the sample at random and without replacement. When the selection is made without replacement, the random variable X has a **<u>hypergeometric distribution</u>** with parameters *N*, *n*, and *K*. and we write $X \sim h(x; N, n, K)$.

$$f(x) = P(X = x) = h(x; N, n, K)$$

$$= \begin{cases} \dfrac{\dbinom{K}{x} \times \dbinom{N-K}{n-x}}{\dbinom{N}{n}}; & x = 0, 1, 2, \cdots, n \\ 0; \; otherwise \end{cases}$$



Example

Lots of 40 components each are called acceptable if they contain no more than 3 defectives. The procedure for sampling the lot is to select 5 components at random (without replacement) and to reject the lot if a defective is found. What is the probability that exactly one defective is found in the sample if there are 3 defectives in the entire lot.

Solution:

Let X= number of defectives in the sample

- $N = 40, K = 3, and\ n = 5$
- X has a hypergeometric distribution with parameters $N = 40, n = 5, and\ K = 3.$
- $X \sim h(x; N, n, K) = h(x; 40,5,3).$
- The probability distribution of X is given by:

$$f(x) = P(X = x) = h(x;40,5,3) = \begin{cases} \dfrac{\binom{3}{x} \times \binom{37}{5-x}}{\binom{40}{5}}; & x = 0, 1, 2, \cdots, 5 \\ 0; \ otherwise \end{cases}$$

But the values of X must satisfy:

$$\mathbf{0 \le x \le K \ and \ n - N + K \le \ x \le \ n \ \Leftrightarrow \ 0 \le x \le 3 \ and \ -32 \le \ x \le \ 5}$$

Therefore, the probability distribution of X is given by:

$$f(x) = P(X = x) = h(x;40,5,3) = \begin{cases} \dfrac{\binom{3}{x} \times \binom{37}{5-x}}{\binom{40}{5}}; & x = 0, 1, 2, 3 \\ 0; & otherwise \end{cases}$$

Now, the probability that exactly one defective is found in the sample is

$$\text{f(1)=P(X=1)=h(1;40,5,3)=} \frac{\binom{3}{1} \times \binom{37}{5-1}}{\binom{40}{5}} = \frac{\binom{3}{1} \times \binom{37}{4}}{\binom{40}{5}} = 0.3011$$

Theorem

The mean and the variance of the hypergeometric distribution $h(x; N, n, K)$ are:

$$\mu = n\frac{k}{N}$$

$$\sigma^2 = n\frac{k}{N}\left(1 - \frac{k}{N}\right)\frac{N-n}{N-1}$$

Example

In the previous example, find the expected value (mean) and the variance of the number of defectives in the sample.

Solution:

- X = number of defectives in the sample
- We need to find $E(X)=\mu$ and $Var(X)=\sigma^2$
- We found that $X \sim h(x; 40,5,3)$
- $N = 40, n = 5, and\ K = 3$

The expected number of defective items is

$$E(X)=\mu=n\frac{K}{N}=5\times\frac{3}{40}=0.375$$

The variance of the number of defective items is

$$Var(X)=\sigma^2=n\frac{K}{N}\left(1-\frac{K}{N}\right)\frac{N-n}{N-1}=5\times\frac{3}{40}\left(1-\frac{3}{40}\right)\frac{40-5}{40-1}=0.311298$$

Relationship to the binomial distribution

Binomial distribution	Hypergeometric distribution
$b(x;n,p)=\binom{n}{x}p^x(1-p)^{n-x};\ x=0,1,\ldots,n$	$h(x;N,n,K)=\dfrac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}};x=0,1,\cdots,n$

If n is small compared to N and K, then the hypergeometric distribution $h(x;N,n,K)$ can be approximated by the binomial distribution $b(x;n,p)$, where $p={}^{K}/_{N}$; i.e., for large N and K and small n, we have:

$$\mathrm{h}(\mathrm{x};N,n,K) \approx \mathrm{b}(\mathrm{x};n,\frac{K}{N})$$

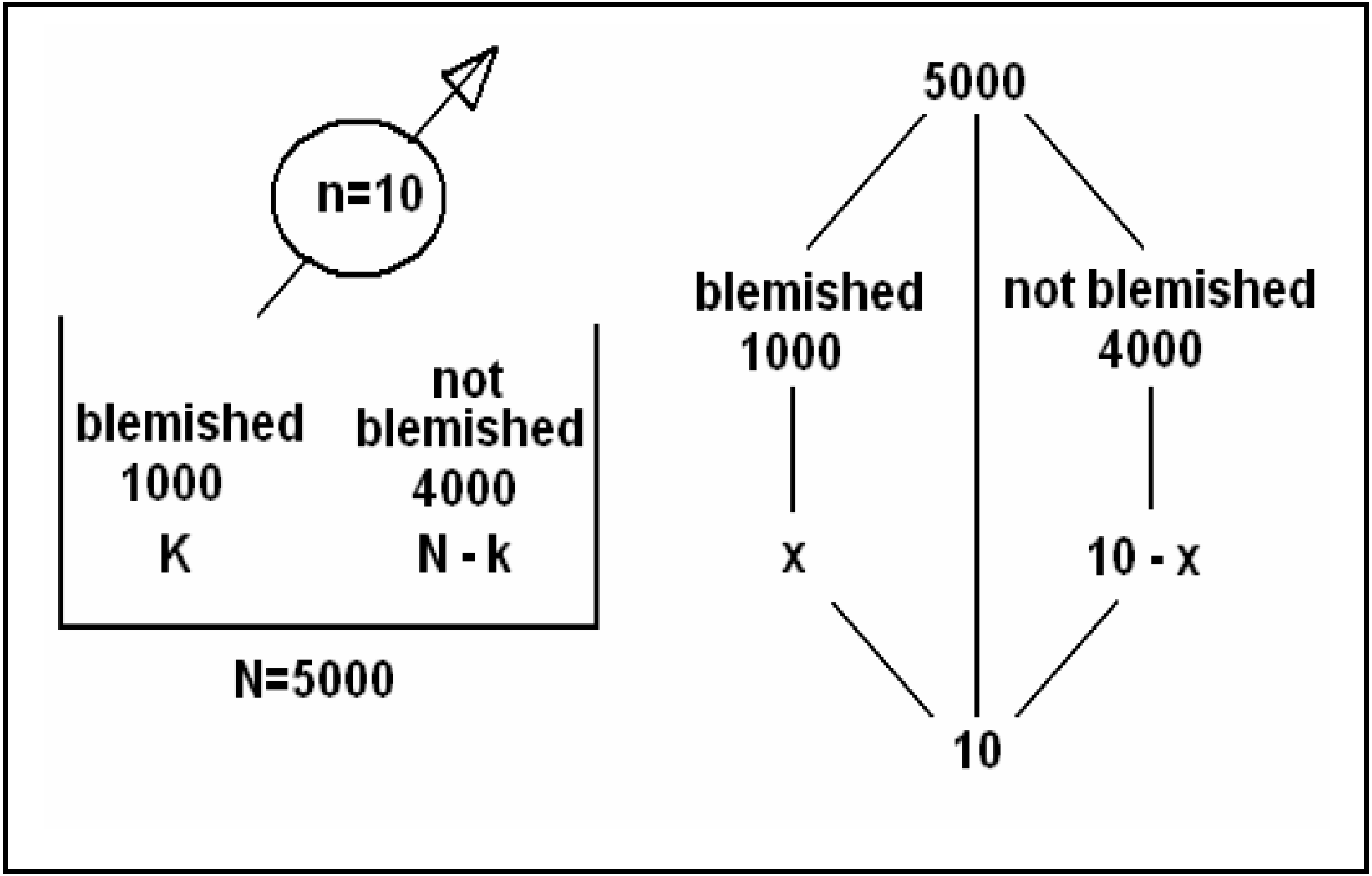
$$\frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}} \approx \binom{n}{x}\left(\frac{K}{N}\right)^{x}\left(1-\frac{K}{N}\right)^{n-x} ;x=0,1,\cdots,n$$

Note:

If n is small compared to N and K, then there will be almost no difference between selection without replacement and selection with replacement

$$\left(\frac{K}{N} \approx \frac{K-1}{N-1} \approx \cdots \approx \frac{K-n+1}{N-n+1}\right).$$

Example: Blemished Tires اطارات معيبة



Solution: (1)

N=5000   K=1000   n=10

X=Number of blemished tires in the Sample

X~h(x;5000,10,1000)

The exact probability is

$$P(X=3) = \binom{1000}{3}\binom{4000}{7}\Big/\binom{5000}{10}$$

$$= \underline{0.201}477715$$

$$\approx \underline{0.201}$$

Solution: (2)

Since n=10 is small relative to N=5000 and K=4000, we can approximate the hypergeometric probabilities using binomial probabilities as follows:

.n=10 (no. of trials)

.p=K/N=1000/5000=0.2 (probability of success)

$$X \sim h(x;5000,10,1000) \approx b(x;10,0.2)$$

$$P(X=3) \approx \binom{10}{3}(0.2)^3(0.8)^7 = \underline{0.201}326592$$

$$\approx \underline{0.201}$$