Probability Distributions

Dr. Charu Arora

Assistant Professor

Applied Science

Some Discrete Probability Distributions:

Discrete Uniform Distribution:

If the discrete random variable X assumes the values $x_1, x_2, \ldots, x_k$ with equal probabilities, then X has the discrete uniform distribution given by:

$$f(x) = P(X = x) = f(x, k) = \begin{cases} \frac{1}{k}\,; & x = x_1, x_2, \cdots, x_k \\ 0\,; & elsewhere \end{cases}$$

Note:

- $f(x)=f(x;k)=P(X=x)$

k is called the parameter of the distribution.

Example 1 :

Experiment: tossing a balanced die.

- Sample space: S={1,2,3,4,5,6}
- Each sample point of S occurs with the same probability 1/6.
- Let X= the number observed when tossing a balanced die.
- The probability distribution of X is,

$$f(x) = P(X = x) = f(x;6) = \begin{cases} \frac{1}{6} \ ; & x = 1, 2, \cdots, 6 \\ 0 \ ; & elsewhere \end{cases}$$

Result:

If the discrete random variable X has a discrete uniform distribution with parameter k, then the mean and the variance of X are;

$$E(X) = \mu = \frac{\sum_{i=1}^{k} x_i}{k}$$

$$Var(X) = \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k}$$

Uniform Distribution

Also $E(X) = (k+1)/2$ (why?)

Hint: $(1/k).k(k+1)/2 = (k+1)/2$

Similarly, $Var(X) = (k+1)(k-1)/12$

Example 5.3:
Find E(X) and Var(X) in Example 1.

Solution:

$$E(X) = \mu = \frac{\sum_{i=1}^{k} x_i}{k} = \frac{1+2+3+4+5+6}{6} = 3.5$$

$$Var(X) = \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \mu)^2}{k} = \frac{\sum_{i=1}^{k} (x_i - 3.5)^2}{6}$$

$$= \frac{(1-3.5)^2 + (2-3.5)^2 + \cdots + (6-3.5)^2}{6} = \frac{35}{12}$$

Binomial Distribution:

Bernoulli Trial:

· Bernoulli trial is an experiment with only two possible outcomes.

· The two possible outcomes are labeled:

success (s) and failure (f)

· The probability of success is $P(s)=p$ and the probability of failure is $P(f)= q = 1-p$.

· Examples:

1. Tossing a coin (success=H, failure=T, and p=P(H))
2. Inspecting an item (success=defective, failure=non-defective, and p=P(defective))

Bernoulli Process:

Bernoulli process is an experiment that must satisfy the following properties:

1. The experiment consists of n repeated Bernoulli trials.
2. The probability of success, $P(s)=p$, remains constant from trial to trial.
3. The repeated trials are independent; that is the outcome of one trial has no effect on the outcome of any other trial.

Binomial Random Variable:

Consider the random variable :

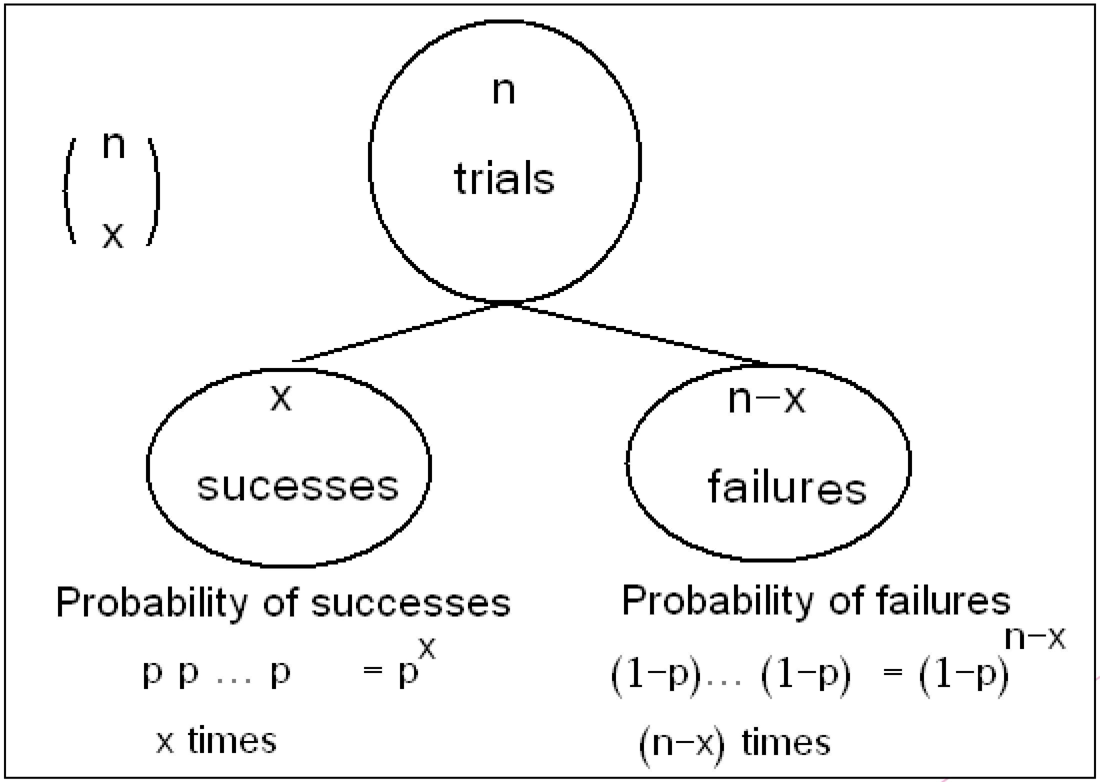
X = The number of successes in n trials in a Bernoulli process.

The random variable X has a binomial distribution with parameters n (number of trials) and p (probability of success), and we write:

$$X \sim \text{Binomial}(n,p) \quad \text{or} \quad X \sim b(x;n,p)$$

The probability distribution of X is given by:

$$f(x) = P(X = x) = b(x; n, p) = \begin{cases} \binom{n}{x} p^{x} (1 - p)^{n-x} ; & x = 0, 1, 2, \ldots, n \\ 0 ; & otherwise \end{cases}$$



We can write the probability distribution of *X* in table as follows.

x	f(x)=P(X=x)=b(x;n,p)
0	$\binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$
1	$\binom{n}{1} p^1 (1-p)^{n-1}$
2	$\binom{n}{2} p^2 (1-p)^{n-2}$
⋮	⋮
$n-1$	$\binom{n}{n-1} p^{n-1} (1-p)^1$
n	$\binom{n}{n} p^n (1-p)^0 = p^n$
Total	1.00

Example 2:

Suppose that 25% of the products of a manufacturing process are defective. Three items are selected at random, inspected, and classified as defective (D) or non-defective (N). Find the probability distribution of the number of defective items.

Solution:

- Experiment: selecting 3 items at random, inspected, and classified as (D) or (N).
- The sample space is

S={DDD,DDN,DND,DNN,NDD,NDN,NND,NNN}

- Let X = the number of defective items in the sample
- We need to find the probability distribution of X.

(1) First Solution:

Outcome	Probability	X
NNN	$\frac{3}{4}\times\frac{3}{4}\times\frac{3}{4}=\frac{27}{64}$	0
NND	$\frac{3}{4}\times\frac{3}{4}\times\frac{1}{4}=\frac{9}{64}$	1
NDN	$\frac{3}{4}\times\frac{1}{4}\times\frac{3}{4}=\frac{9}{64}$	1
NDD	$\frac{3}{4}\times\frac{1}{4}\times\frac{1}{4}=\frac{3}{64}$	2
DNN	$\frac{1}{4}\times\frac{3}{4}\times\frac{3}{4}=\frac{9}{64}$	1
DND	$\frac{1}{4}\times\frac{3}{4}\times\frac{1}{4}=\frac{3}{64}$	2
DDN	$\frac{1}{4}\times\frac{1}{4}\times\frac{3}{4}=\frac{3}{64}$	2
DDD	$\frac{1}{4}\times\frac{1}{4}\times\frac{1}{4}=\frac{1}{64}$	3

The probability distribution of X is,

x	f(x)=P(X=x)
0	$\frac{27}{64}$
1	$\frac{9}{64}+\frac{9}{64}+\frac{9}{64}=\frac{27}{64}$
2	$\frac{3}{64}+\frac{3}{64}+\frac{3}{64}=\frac{9}{64}$
3	$\frac{1}{64}$

(2) Second Solution:

Bernoulli trial is the process of inspecting the item. The results are success=D or failure=N, with probability of success $P(s)=25/100=1/4=0.25$.

The experiments is a Bernoulli process with:

- number of trials: $n=3$
- Probability of success: $p=1/4=0.25$
- $X \sim$ Binomial(n,p)=Binomial(3,1/4)
- The probability distribution of X is given by:

$$f(x) = P(X = x) = b(x;3,\frac{1}{4}) = \begin{cases} \binom{3}{x} (\frac{1}{4})^x (\frac{3}{4})^{3-x}; & x = 0, 1, 2, 3 \\ 0 \; ; & otherwise \end{cases}$$

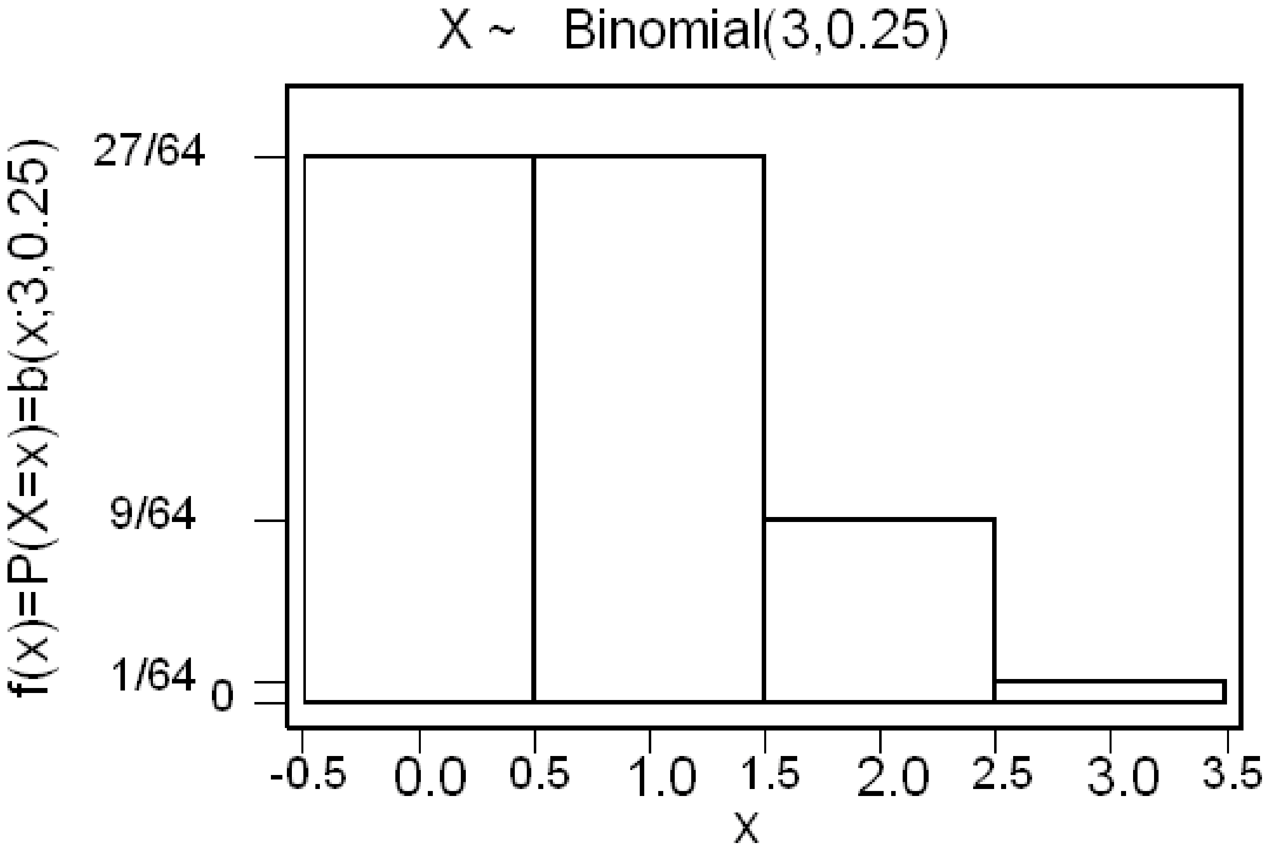
$$f(0) = P(X = 0) = b(0;3,\frac{1}{4}) = \binom{3}{0} (\frac{1}{4})^0 (\frac{3}{4})^3 = \frac{27}{64}$$

$$f(2) = P(X = 2) = b(2;3,\frac{1}{4}) = \binom{3}{2} (\frac{1}{4})^2 (\frac{3}{4})^1 = \frac{9}{64}$$

$$f(3) = P(X = 3) = b(3;3,\frac{1}{4}) = \binom{3}{3} (\frac{1}{4})^3 (\frac{3}{4})^0 = \frac{1}{64}$$

The probability distribution of X is,

x	f(x)=P(X=x) =b(x;3,1/4)
0	27/64
1	27/64
2	9/64
3	1/64



Result:

The mean and the variance of the binomial distribution b(x;n,p) are:

$$\mu = n\,p$$
$$\sigma^2 = n\,p\,(1-p)$$

Example:

In the previous example, find the expected value (mean) and the variance of the number of defective items.

Solution:

- X = number of defective items
- We need to find $E(X)=\mu$ and $Var(X)=\sigma^2$
- We found that $X \sim$ Binomial(n,p)=Binomial(3,1/4)
- .n=3 and p=1/4

The expected number of defective items is

$$E(X)=\mu = n\ p = (3)\ (1/4) = 3/4 = 0.75$$

The variance of the number of defective items is

$$Var(X)=\sigma^2 = n\ p\ (1 - p) = (3)\ (1/4)\ (3/4) = 9/16 = 0.5625$$

Example:

In the previous example, find the following probabilities:

(1) The probability of getting at least two defective items.

(2) The probability of getting at most two defective items.

Solution:

X ~ Binomial(3,1/4)

$$f(x) = P(X = x) = b(x;3,\frac{1}{4}) = \begin{cases} \binom{3}{x} (\frac{1}{4})^x (\frac{3}{4})^{3-x} & for\ x = 0, 1, 2, 3 \\ 0 & otherwise \end{cases}$$

x	.f(x)=P(X=x)=b(x;3,1/4)
0	27/64
1	27/64
2	9/64
3	1/64

(1) The probability of getting at least two defective items:

$$P(X\geq 2)=P(X=2)+P(X=3)= f(2)+f(3)= \frac{9}{64}+\frac{1}{64}=\frac{10}{64}$$

(2) The probability of getting at most two defective item:

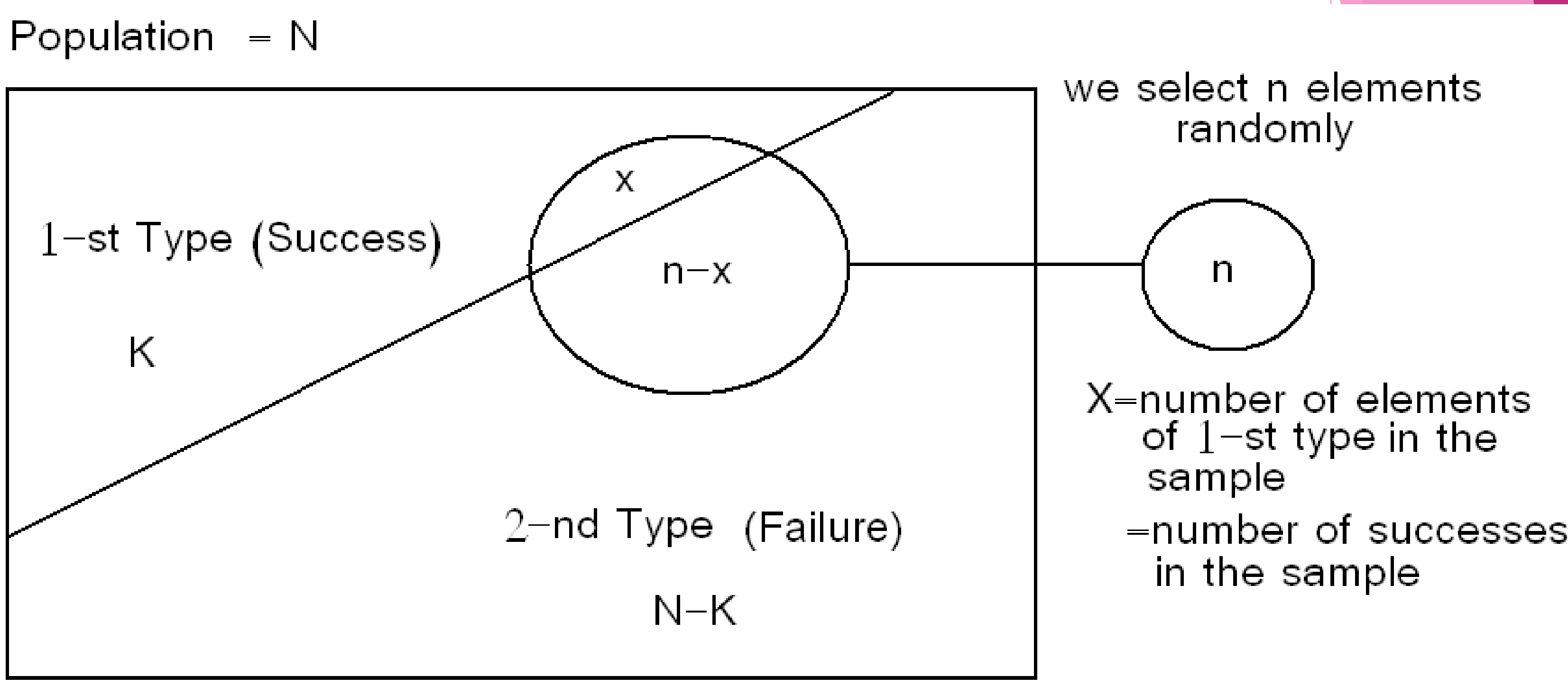
$$P(X\leq 2) = P(X=0)+P(X=1)+P(X=2)$$

$$= f(0)+f(1)+f(2) = \frac{27}{64}+\frac{27}{64}+\frac{9}{64}=\frac{63}{64}$$

or

$$P(X\leq 2)= 1-P(X>2) = 1-P(X=3) = 1- f(3) = 1- \frac{1}{64}=\frac{63}{64}$$

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.

There are to two methods of selection:

1. selection with replacement
2. selection without replacement

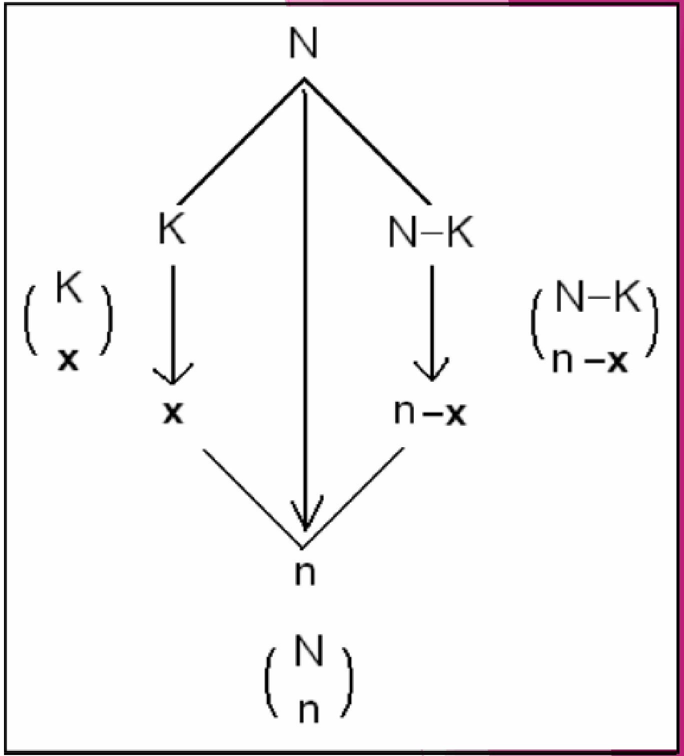
(1) If we select the elements of the sample at random and with replacement, then

X ~ Binomial(n,p); where $p = \dfrac{K}{N}$

(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 hyper geometric distribution with parameters N, n, and K. and we write X~h(x;N,n,K).

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

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



Note that the values of X must satisfy:

$0 \le x \le K$ and $0 \le n - x \le N - K$

$\Leftrightarrow$

$0 \le x \le K$ and $n - N + K \le x \le n$

Hypergeometric Distribution

Also $E(X)= (nK)/N$

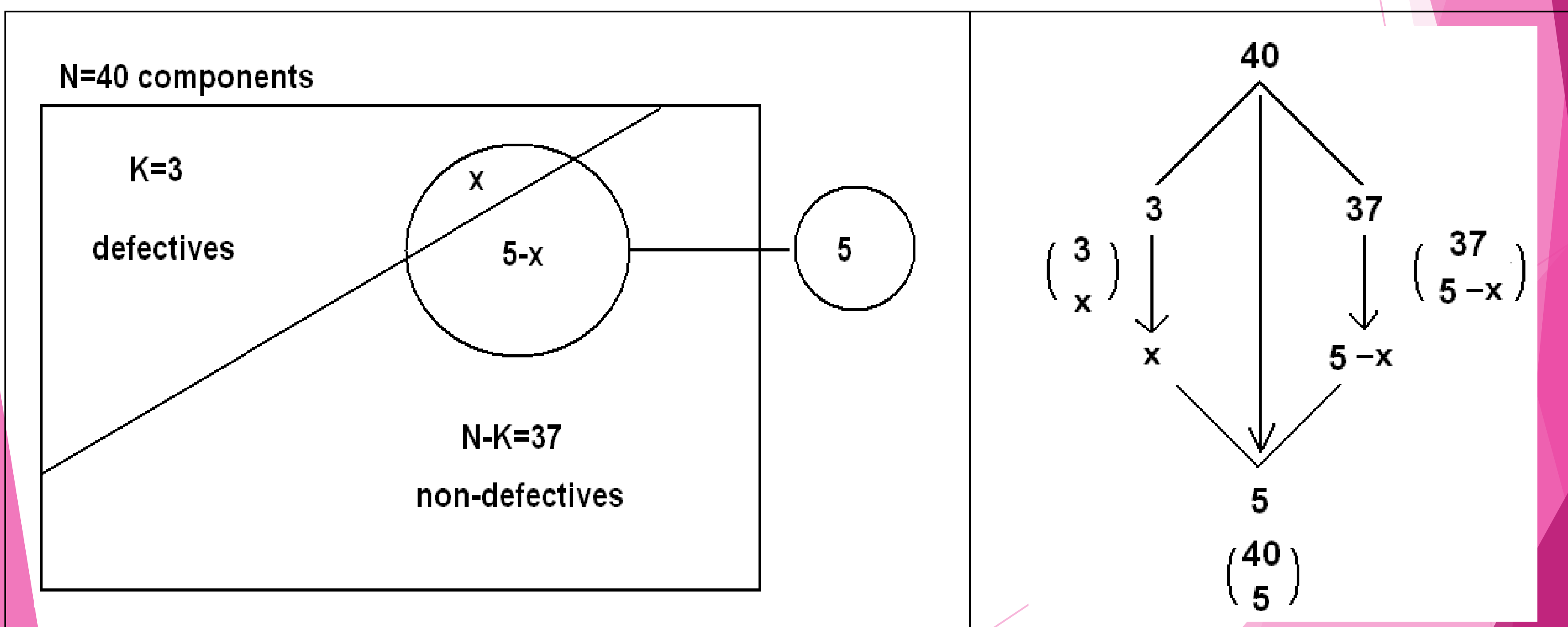
Hint: $(1/k).k(k+1)/2 = (k+1)/2$

Similarly, $Var (X)=NK(N-K)(N-n)/N^2(N-1)$

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~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:

$0 \le x \le K$ and $n-N+K \le x \le n \Leftrightarrow 0 \le x \le 3$ and $-42 \le x \le 5$

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

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

Question?

- A bag contains 50 light bulbs of which 5 are defective and 45 are not. A Quality Control Inspector randomly samples 4 bulbs without replacement. Let X = the number of defective bulbs selected. Find the probability mass function, $f(x)$, of the discrete random variable X.

Result.

- Note: One of the key features of the hypergeometric distribution is that it is associated with sampling without replacement. When the samples are drawn with replacement, the discrete random variable follows what is called the **binomial distribution**.
- Note: In cases where the sample size is relatively large compared to the population, a discrete distribution called **hypergeometric** may be useful.

Question 1. A lake contains 600 fish, eighty (80) of which have been tagged by scientists. A researcher randomly catches 15 fish from the lake. Find a formula for the probability mass function of the number of fish in the researcher's sample which are tagged.

Question 2. Let the random variable X denote the number of aces in a five-card hand dealt from a standard 52-card deck. Find a formula for the probability mass function of X.

Question 3: Determine whether the given scenario describes a binomial setting. Justify your answer.

(a) Genetics says that the genes children receive from their parents are independent from one child to another. Each child of a particular set of parents has probability 0.25 of having type O blood. Suppose these parents have 5 children. Count the number of children with type O blood.

(b). Shuffle a standard deck of 52 playing cards. Turn over the first 10 cards, one at a time. Record the number of aces you observe.

Question 4: Each child of a particular set of parents has probability 0.25 of having type O blood. Suppose these parents have 5 children. What’s the probability that exactly one of the five children has type O blood?

Solution 3 (a)

- **Binary?** “Success” 5 has type O blood. “Failure” 5 doesn’t have type O blood.
- **Independent?** Knowing one child’s blood type tells you nothing about another child’s because they inherit genes independently from their parents.
- **Number?** n = 5
- **Same probability?** p = 0.25

Solution 3 (b)

- **Binary?** "Success" get an ace. "Failure" don't get an ace.
- **Independent?** No. If the first card you turn over is an ace, then the next card is less likely to be an ace because you're not replacing the top card in the deck. If the first card isn't an ace, the second card is more likely to be an ace.

- This is not a binomial setting because the independent condition is not met.

Poisson Distribution:

- It is discrete distribution.
- The discrete r. v. X is said to have a Poisson distribution with parameter (average) 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.

We write:

X ~ Poisson (λ)

- The mean (average) of Poisson (λ) is

$\mu = \lambda$ $\qquad\qquad \sigma^2 = \lambda$

- The variance is:

- The Poisson distribution is used to model a discrete r. v. which is a <u>count</u> of how many times a specified random event occurred in an interval of time or space.

<u>Example:</u>

- No. of patients in a waiting room in an hour.
- No. of serious injuries in a particular factory in a month.
- No. of calls received by a telephone operator in a day.
- No. of rates in each house in a particular city.

<u>Note:</u>

- λ is the average (mean) of the distribution.

If X = The number of calls received in a month and

$X \sim$ Poisson (λ)

then:

(i) Y = The no. calls received in a year.

$Y \sim$ Poisson (λ^{*}), where $\lambda^{*}=12\lambda$

$Y \sim$ Poisson (12λ)

(ii) W = The no. calls received in a day.

$W \sim$ Poisson (λ^{*}), where $\lambda^{*}=\lambda/30$

$W \sim$ Poisson $(\lambda/30)$

Example

Suppose that the number of snake bites cases seen 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 <u>in</u> <u>2</u> <u>years</u> there will be 10 bite cases?

3- What is the probability that <u>in</u> <u>a</u> <u>month</u> there will be no snake bite cases?

<u>Solution:</u>

(1) X = no. of snake bite cases in <u>a year.</u>

X ~ Poisson (6) (λ=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)

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

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

Y = no of snake bite cases in <u>2 years</u>

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

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

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

3- *W* = no. of snake bite cases in a <u>month.</u>

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

$$P(W = w) = \frac{e^{-.5}.5^{w}}{w!} : \qquad w = 0 , 1 , 2 ...$$

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

Note: Poisson distribution is for counts—if events happen at a constant rate over time, the Poisson distribution gives the probability of X number of events occurring in time T.

$$Result:\ p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!} \text{ (for } k \in \mathbb{N})$$

$$\sum_{k=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!}$$

$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$$

$$= e^{-\lambda} e^{\lambda} = e^0 = 1$$

(Why? Taylor Series for e^x

$$\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x)$$

$X \sim \text{Unif}(a, b)$

$$p_X(k) = \frac{1}{b - a + 1}$$

$$\mathbb{E}[X] = \frac{a + b}{2}$$

$$\text{Var}(X) = \frac{(b - a)(b - a + 2)}{12}$$

$X \sim \text{Ber}(p)$

$$p_X(0) = 1 - p;\quad p_X(1) = p$$

$$\mathbb{E}[X] = p$$

$$\text{Var}(X) = p(1 - p)$$

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

$$p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

$$\mathbb{E}[X] = np$$

$$\text{Var}(X) = np(1 - p)$$

$X \sim \text{Geo}(p)$

$$p_X(k) = (1 - p)^{k-1} p$$

$$\mathbb{E}[X] = \frac{1}{p}$$

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

$X \sim \text{NegBin}(r, p)$

$$p_X(k) = \binom{k - 1}{r - 1} p^r (1 - p)^{k-r}$$

$$\mathbb{E}[X] = \frac{r}{p}$$

$$\text{Var}(X) = \frac{r(1 - p)}{p^2}$$

$X \sim \text{HypGeo}(N, K, n)$

$$p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$$

$$\mathbb{E}[X] = n\frac{K}{N}$$

$$\text{Var}(X) = \frac{K(N - K)(N - n)}{N^2(N - 1)}$$

$X \sim \text{Poi}(\lambda)$

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$\mathbb{E}[X] = \lambda$$

$$\text{Var}(X) = \lambda$$