MCA Sem. I Core

Mathematical Techniques for Computer Applications(MCAC 103) L 6

Discrete Distributions

10 Feb 2021

Vasudha Bhatnagar
Department of Computer Science,University of Delhi
Delhi, India.

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Outline

1. Bernoulli Distribution
2. Binomial Distribution
3. Geometric Distribution
4. Negative Binomial Distribution
5. Multinomial Distribution
6. Poisson Distribution

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Bernoulli Distribution

Boolean Random variable with $P(X = 1) = p$ and
$P(X = 0) = 1 - p$
$X \sim Bernoulli(p)$

Important Notes:

1. any experiment (random trial) that results in two outcomes is a Bernoulli experiment
2. the probabilities of the occurrence of these outcomes remain same throughout the trials
3. the trials are carried out independently

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Bernoulli Distribution

Boolean Random variable with $P(X = 1) = p$ and
$P(X = 0) = 1 - p$
$X \sim Bernoulli(p)$

Important Notes:

1. any experiment (random trial) that results in two outcomes is a Bernoulli experiment
2. the probabilities of the occurrence of these outcomes remain same throughout the trials
3. the trials are carried out independently

Use Cases

i. Good or defective components
ii. Transmitted or lost signals
iii. Working or malfunctioning hardware
iv. Benign or malicious attachments
v. Documents that contain or do not contain a keyword

Mean –> $E(X) = p$,

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Bernoulli Distribution

Boolean Random variable with $P(X=1)=p$ **and**
$P(X=0)=1-p$
$X \sim Bernoulli(p)$

Important Notes:

1. any experiment (random trial) that results in two outcomes is a Bernoulli experiment
2. the probabilities of the occurrence of these outcomes remain same throughout the trials
3. the trials are carried out independently

Use Cases

i. Good or defective components
ii. Transmitted or lost signals
iii. Working or malfunctioning hardware
iv. Benign or malicious attachments
v. Documents that contain or do not contain a keyword

Mean –> $E(X)=p$, Variance –> $Var(X)=p(1-p)$

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Binomial Distribution

1. $X \sim B(n, p)$, **X is a rv that denotes number of successes in** n **Bernoulli trials**

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Binomial Distribution

1. $X \sim B(n, p)$, **X is a rv that denotes number of successes in** n **Bernoulli trials**
2. **Effectively, X is sum of** n **Bernoulli rvs, each iid with probability** p **(sequence of** n **Bernoulli trials will yield** n **values of rv X)**

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Binomial Distribution

1. **$X \sim B(n, p)$, X is a rv that denotes number of successes in n Bernoulli trials**
2. **Effectively, X is sum of n Bernoulli rvs, each iid with probability p (sequence of n Bernoulli trials will yield n values of rv X)**
3. **Order in which successes occur is not important**

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Binomial Distribution

1. $X \sim B(n, p)$, **X is a rv that denotes number of successes in** n **Bernoulli trials**
2. **Effectively, X is sum of** n **Bernoulli rvs, each iid with probability** p **(sequence of** n **Bernoulli trials will yield** n **values of rv X)**
3. **Order in which successes occur is not important**
4. **PMF:** $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$

Mean: $E(X) = np$
Variance: $Var(X) = np(1-p)$

Practice Questions

1.
2.
3. **Random walk in 2-dimension - Probability of moving on X-axis is p andmoving on Y-axis is q. What is the joint probability distribution of the position after 5 steps?**

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane?

Practice Questions

1.
2.
3. **Random walk in 2-dimension - Probability of moving on X-axis is p andmoving on Y-axis is q. What is the joint probability distribution of the position after 5 steps?**

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane? ($p \geq 2/3$)

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Practice Questions

1.
2.
3. **Random walk in 2-dimension - Probability of moving on X-axis is p andmoving on Y-axis is q. What is the joint probability distribution of the position after 5 steps?**

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane? ($p \geq 2/3$)

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane?

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Practice Questions

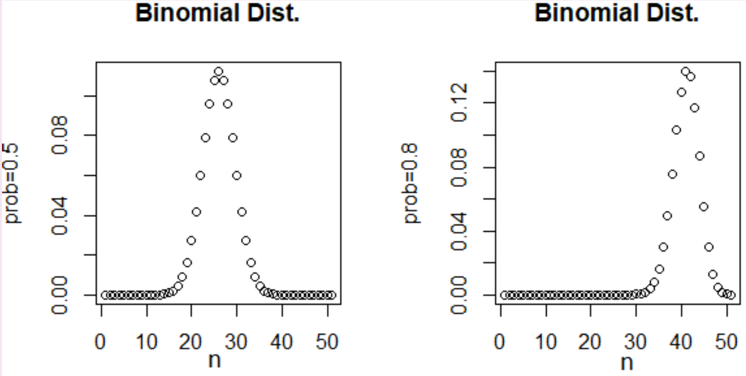
1.
2.
3. **Random walk in 2-dimension - Probability of moving on X-axis is p andmoving on Y-axis is q. What is the joint probability distribution of the position after 5 steps?**

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane? ($p \geq 2/3$)

Problem: Suppose that an airplane engine will fail, when in flight, with probability(1-p) independently from engine to engine; Suppose that the airplane will make a successful flight if atleast 50% of its engines remain operative. For what values of p is a four-engine plane is safer than a two-engine plane? ($p \geq 2/3$)

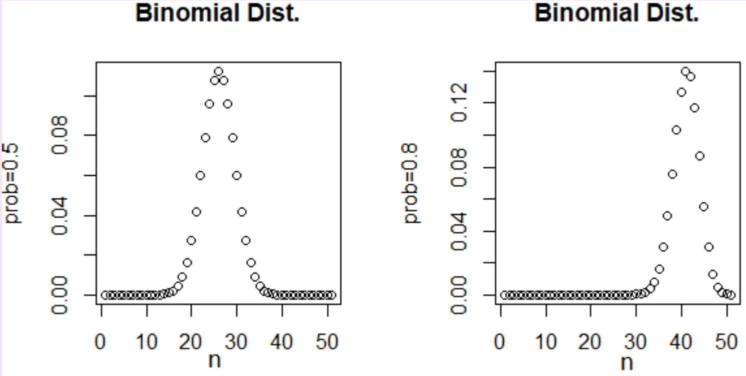
Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

The shape of Binomial distribution is determined by the values assigned to its two parameters, n and p.



Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

The shape of Binomial distribution is determined by the values assigned to its two parameters, n and p.



1. Symmetric around mean
2. Shape change when p increases
3. For large n, Stirling's approximtion is used $(n! \approx \sqrt{2n\pi}(\frac{n}{e})^n)$

Geometric Distribution

1. **Random variable X denotes number of trials to (and including) the first occurrence of success.**
2. $1 \leq X \leq \infty$
3. **PMF: P(X=k) = P(k) =** $q^{k-1}p,\ k = 1, 2, \ldots$

Mean: $1/p$
Variance: $(1-p)/p^2$
DF: $F(k) = P(X \leq k) = (1 - q^k)$

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Geometric Distribution

1. **Random variable X denotes number of trials to (and including) the first occurrence of success.**
2. $1 \leq X \leq \infty$
3. **PMF: P(X=k) = P(k) =** $q^{k-1}p,\ k = 1, 2, \ldots$

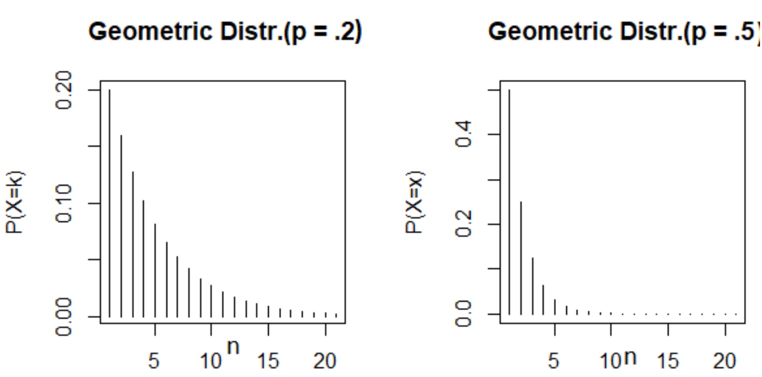
Mean: $1/p$
Variance: $(1-p)/p^2$
DF: $F(k) = P(X \leq k) = (1 - q^k)$

Use Case

1. Suppose a retail store want to know how many customers enter a store until that customer makes a purchase
2. A search engine goes through a list of sites looking for a given key phrase. Suppose the search terminates as soon as the key phrase is found. The number of sites visited is Geometric
3. The literacy rate for women in country AAA is 12%. The number of women you ask until one says that she is literate is Geometric rv.

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

The shape of Geometric distribution is determined by the value assigned to p



1. Positive Skew
2. How does shape change when p increases?

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Negative Binomial Distribution

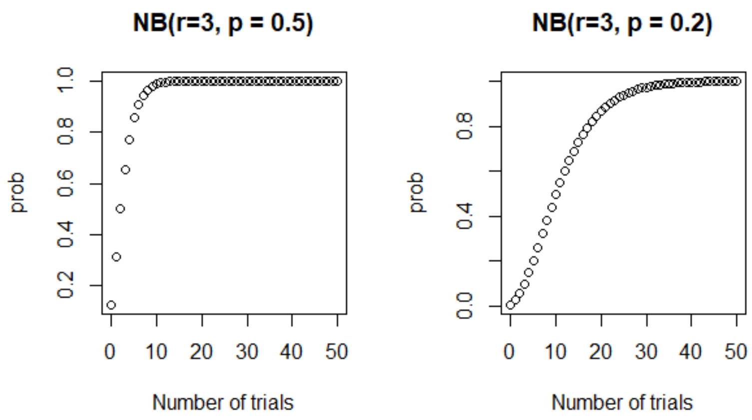
1. **Generalization of Geometric rv - X counts number of trials for r successes in a sequence of independent Bernoulli trials**
2. **Generalization of Binomial rv - X counts number of trials instead of successes in a sequence of independent Bernoulli trial**
3. $X \sim NB(r, p), X = r, r + 1, \ldots \infty$
4. **PMF:** $\binom{(x-1)}{(r-1)} p^r q^{(x-r)}$

Real world Use Cases:

1. Number of attempts to clear 4 papers with p as probability of clearing a paper
2. Number of years before flood reoccurs, with p as probability of flood
3. Number of inspections before 5 defective components are found, with p probabily of defective component (success)

$$\boxed{E(X) = r/p,\ Var(X) = rq/p^2}$$

Shape variation with p



ernoulli Distribution
nomial Distribution
eometric Distribution
egative Binomial
stribution
ultinomial
stribution
oisson Distribution

Multinomial Distribution

1. **Generalization of Bernoulli in another dimension - relax the requirement that there be only two possible outcomes**
2. **Let there be r possible outcomes for each trial, denoted by $E_1, E_2, ..., E_r$, with respective probabilities as $p_1, p_2, ..., p_r$. Let rv $X_1, X_2, ..., X_r$ denote number of E_is in n trials. Then the joint pmf of $X_1, X_2, ..., X_r$ is given as**

$$p_{X_1 X_2 \dots X_r}(k_1, k_2, \dots, k_r) = \frac{n!}{k_1! k_2! \dots k_r!} p_1^{k_1} p_2^{k_2} \dots p_r^{k_r},$$

Since each rv X_i is itself a Binomial rv with parameters n and p_i,

$\mu_{X_i} = np_i$ and $Var(X_i) = np_i(1 - p_i)$

Show that $Cov(X_i, X_j) = -np_i p_j$

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Real World Use cases

1. Text Analysis - Language Modeling
2. Number of faces in Dice experiment
3. Useful in multiclass classification

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution

Poisson Distribution

1. **Related to a concept of rare events (two such events are extremely unlikely to occur simultaneously or within a very short period of time)**
2. **A Poisson random variable can take any nonnegative integer value**
3. **Approximation of Binomial distribution (p is very small, n is lavery large, np = λ is finite**
4. **PMF:** $P(X = x) = e^{-\lambda}\lambda^x/x!, x = 0, 1, \ldots, \infty$
5. **E(X) = Var(X) =** λ

Real World Use Cases: The Poisson distribution is useful in determining the probability that a certain number of events occur over a given time period.

1. Arrivals of customers at service counter, telephone calls, e-mail messages in a given period
2. Number of work-related accidents over a given time
3. Number of network blackouts, virus attacks, errors in software in a given period

Bernoulli Distribution
Binomial Distribution
Geometric Distribution
Negative Binomial Distribution
Multinomial Distribution
Poisson Distribution