Submitted By: Amit Kumar

AI Batch

CDAC- Mohali

Subject- Mathematical Foundation of Computer Science

Topics covered:

- Hypergeometric Distribution
- Geometric Distribution and Negative Binomial (Pascal) Distribution
- Gamma Distribution

Submitted By :- AmitKumar
AI Batch
CDAC Mohali

Date
8 Jan 2021

Subject :- Mathematical foundation of Comp. Science
sub code :- MTAI - 101 - 20

Topic :- Hypergeometric Distribution

Definition :- In probability theory and statistics, the hypergeometric distribution is a discrete probability distribution that describes the probability of successes in draws, without replacement.

Hypergeometric Distribution characteristics :-

- There are only 2 possible outcomes. - success or failure
- The probability of a success is not the same on each trial without replacement, thus events are not independent
- In which population is finite
- Trials are dependent.
- It is similar to the binomial distribution but the difference is the method of sampling.
- X is the no. of successes in the n trials
- It is discrete distribution, the no. of objects in the population, N, is finite & known.

- Probability function

N = population size
n = sample size
A = no. of successes in population
x = no of success in sample

$$\boxed{P(x) = \frac{({}_{A}C_{x})({}_{N-A}C_{n-x})}{{}_{N}C_{n}}}$$

mean value = $\boxed{\mu = \frac{A \cdot n}{N}}$

• Variance & Standard Deviation

$$\sigma^2 = \frac{A(N-A)n(N-n)}{N^2(N-1)}$$

$$\sigma = \sqrt{\sigma^2}$$

Ex1

3 light bulbs were selected from 10. ~~one~~ of the 10, there were 4 defective. What is the probability that 2 of the 3 selected are defective?

Sol
$$P(x) = \frac{({}^{A}C_x)({}^{N-A}C_{n-x})}{({}^{N}C_n)} = P(2) = \frac{({}^{4}C_2)({}^{6}C_1)}{{}^{10}C_3}$$

$$\boxed{P(x=2) = .30}$$

Ex2 Suppose that a shipment contains 5 defective items and 10 non defective items. If 7 items are selected at random without replacement, what is the probability that at least 3 defective items will be obtained?

Sol $N = 15$ (5 = defective, 10 = non defective) $n = 7$

$$P(x \geq 3) = 1 - P(x \leq 2)$$
$$= 1 - [P(0) + P(1) + P(2)] = 0.4267$$

$$P(0) = \frac{{}^{5}C_0 \cdot {}^{10}C_7}{{}^{15}C_7} = 0.0186$$

$$P(1) = \frac{{}^{5}C_1 \cdot {}^{10}C_6}{{}^{15}C_7} = 0.1631$$

$$P(2) = \frac{{}^{5}C_2 \cdot {}^{10}C_5}{{}^{15}C_7} = 0.3916$$

Ex3 If a random variable X has a hyper geometric Distribution with parameter $A = 8$, $B = 20$ & n, for what value of n will Var (X) be maximum?

$$Var(x) = \frac{n \cdot A \cdot B}{(A+B)^2} \cdot \frac{A+B-n}{A+B-1} = \frac{n \cdot 8 \cdot 20}{(8+20)^2} \cdot \frac{8+20-n}{8+20-1}$$

$$= \frac{160n}{28^2} \cdot \frac{(28-n)}{27} = 0$$

$n = 28$, or $n = 0$ for variance to be maximum

use case of Hypergeometric Distribution in real life

case → consider the situation in a factory where around 100 parts are made everyday. you supply these parts in boxes of 500 parts every week (so, lot size is 500).

→ It is known that 2% of parts produced are defective. your customer inspects each box by choosing 25 parts randomly one by one from a box

→ Then he records the no. of defective parts in his sample of 25 parts. The no. of defective parts out of 25 inspected parts can be analyzed through a Hypergeometric distribution

This distribution is used in all such cases where objects are choosen out of a finite lot size with a known no. of defects and the sample parts choosen without replacement.

Topic :- Geometric and Negative Binomial Distribution (Pascal Distribution)

- The geometric distribution is a special case of the negative Binomial distribution

- A Negative Binomial Experiment has following properties
  - → the Experiment consist of x repeated trials
  - → each trial results in 2 possible outcomes < success / failure
  - → The probability of success, denoted by P, is the same on every trial
  - → The trials are independent, that is, the outcome on one trial does not affect the outcome on other trials
  - → the experiment continues until r successes are observed, where r is specified in advance

consider the following statistical experiment (use case)

→ you flip a coin repeatedly and the count the no. of times the coin lands on heads. You continue flipping the coin until it has landed 5 times on heads. This is a ~~go~~ negative binomial experiment because :-

→ The experiment consists of repeated trials we flip a coin repeatedly until it has landed 5 times on heads

→ each trial can result in just two possible outcomes $<$ Head / Tails

→ The probability of success is constant = 0.5 on every trial

→ The trials are independent that is, getting heads on one trial does not affect whether we get heads on other trials

→ The experiment continues until a fixed no. of success have occurred, in this case, 5 heads

Notation and formulas

A negative binomial random variable is the no. X of repeated trials to produce r success in a negative binomial experiment. The probability distribution of a negative binomial random variable is called a negative binomial Distribution, also known as pascal distribution

suppose a negative binomial experiment consists of x trials, result in r success. If the probability of success on an individual trial is P, then the negative binomial probability is

$$b^*(x; r, P) = ({}^{x-1}C_{r-1}) * P^r * (1-P)^{x-r}$$

x = no. of trials required to produce r success
r = no of successes
P = Prob. of success on an individual trial
q = The prob. of failure " " = (1-P)
$b^*(x; r, P)$ = Negative Binomial Probability - The probability that an x-trial negative binomial experiment results in the rth success on the xth trial, when success Prob is P.

nC_r :- The no. of combinations of n things, taken r at a time

The mean of Negative Binomial Distribution :-

If we define the mean of the negative binomial distribution as the average no of trials required to produce r successes, then the mean is equal to

$$\boxed{\mu = r/p}$$

μ = mean no. of trials,
r = no. of successes;
p = Prob. of a success on any given trial

→ if no of successes before the binomial experiment results in K failures, the mean of R is :-

$$\boxed{\mu_R = KP/Q}$$

→ The no. of failures before the binomial experiment results in r successes, the mean of K is

$$\boxed{\mu_K = rQ/P}$$

→ Geometric Distribution

The geometric distribution is a special case of the negative binomial distribution. It deals with the no. of trials required for a single success. Thus the geometric distribution is negative binomial distribution where the no. of success (r) is equal to 1.

Geometric Probability formula

suppose a negative binomial experiment consists of x trials and results in one success. If the probability of success on an individual trial is P, then the geometric probability is

$$\boxed{g(x ; P) = P * Q^{x-1}}$$

Ex

Bob is a high school basketball Player. He is a 70% free throw shooter. During the season, what is the Probability that Bob makes his third free throw on his 5th shot?

To solve this, — This is an example of a negative binomial experiment.

Solution

Success (P) = 0.7
no of trials (x) = 5
no of success (r) = 3

So, Acc to negative binomial formula

$$b^*(x; r, P) = {}^{x-1}C_{r-1} * P^r * q^{x-r}$$

$$b^*(5; 3, 0.7) = {}^4C_2 * (0.7)^3 * (0.3)^2$$

$$b^*(5; 3, 0.7) = 6 * .343 * 0.09$$

$$= 0.18522$$

thus, the probability that Bob will make his third successful free throw on his fifth shot is = 0.18522

Q2. Let's consider question I, slightly different question what is the probability that Bob makes his first free throw on his 5th shot?

→ This is an example of a geometric distribution which is an special case of negative binomial distribution

Prob. of success (P) = 0.70
no. of trials (x) = 5
no of success (r) = 1

Acc to geometric formula,

$$g(x; P) = P * q^{x-1}$$

$$g(5; 0.7) = 0.7 * (0.3)^{5-1}$$

$$= 0.7 * (0.3)^4$$

$$= 0.00567$$

Topic — Gamma Distribution

The gamma distribution is a family of right skewed, continous Probability distributions. These distributions are useful in real-life where something has a natural minimum of 0. for example, it is commonly used in finance, for elapsed times or during poisson process.

gamma Distribution PDf

if X is a continous random variable then the Probability distribution function is

$$P(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^{\alpha}\,\Gamma(\alpha)} \qquad x \geq 0$$

where

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t}\, dt \qquad \Gamma(x) = \text{the gamma function}$$

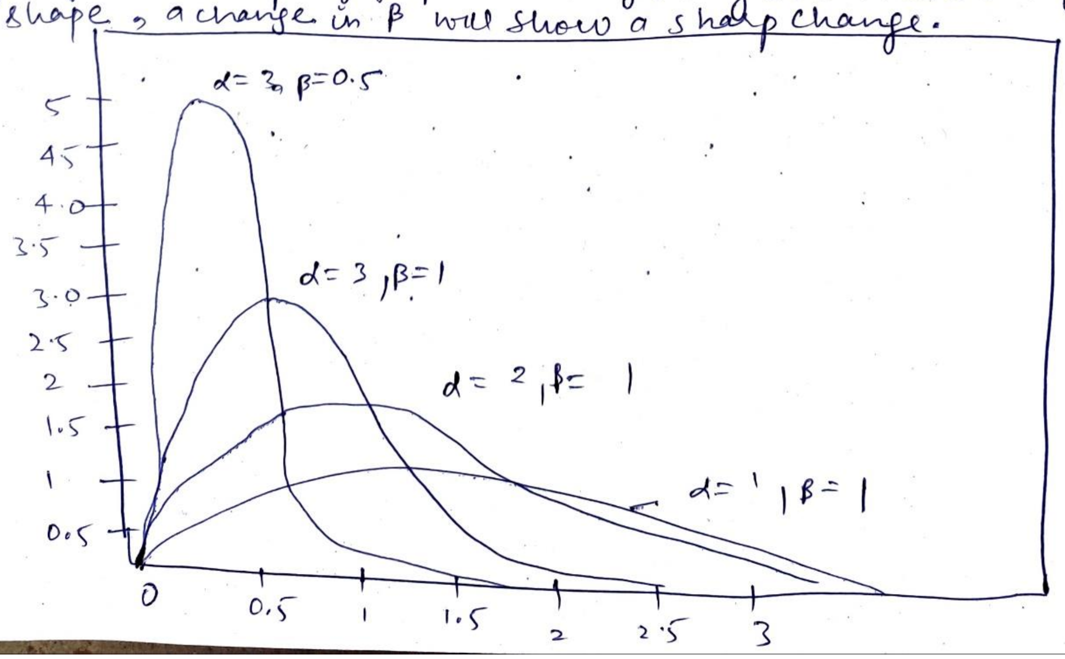
α = the shape parameter

β (sometimes used θ is used instead) = The rate parameter (the reciprocal of the scale parameter)

When, α & β are both greater than 1

when $\alpha = 1$, this becomes the exponential distribution
$\beta = 1$, this becomes the standard gamma distribution

- α & β define the shape of graph, although they both have an effect on the shape, a change in β will show a sharp change.



mean = $E(X) = \alpha\beta$

variance = $Var(X) = \alpha\beta^2$

Moment generating function $= M_X(t) = \frac{1}{(1-\beta t)^{\alpha}}$

- Important properties of gamma function
  - from defenitation of the integral, $\Gamma(1) = 1$
  - for any $\alpha > 0$, $\Gamma(\alpha) = (\alpha-1)\Gamma(\alpha-1)$
  - for any positive integers n, $\Gamma(n) = (n-1)!$ for $n = 1, 2, 3 \ldots$
  - $\Gamma(\frac{1}{2}) = \sqrt{\pi}$
  - $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
  - $\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^{\alpha}}$, for $\lambda > 0$;

Q claim amounts on a certain type of policy are modelled as following a gamma distribution with parameters $\alpha = 120$ & $\lambda = 1.2$ calc, an approximate value for the prob. that an individual claim amt exceeds 120, giving a reason for the approach you use?

$E(X) = \frac{\alpha}{\lambda} = \frac{120}{1.2} = 100$, variance $= Var(X) = \frac{\alpha}{\lambda^2} = \frac{120}{(1.2)^2} = 83.33$

$P(X > 120) = P(Z > [\frac{120 - 100}{9.129}])$ $\Big\{ X \approx N(100, 9.128^2)$

$= P(Z > 2.191)$

$= 1 - 0.986 = 0.014$

Example — Image you are solving difficult maths theorems & you expect to solve one every $\frac{1}{2}$ hr, compute the probability that you will have to wait b/w 2 to 4 hrs before you solve four of them.

Sol. One theorem every $\frac{1}{2}$ hr means we would suppose to get $\theta = \frac{1}{0.5} = 2$ theorem every hour on avg using $\theta = 2$ & $K = 4$,

$$P(2 \le x \le 4) = \sum_{x=2}^{4} \frac{x^4 e^{-x/2}}{\Gamma(4)\, 2^4} = 0.12388$$