Random variables

Discrete   Continuous

PDF → T

(Q) Approx 60% of full term newborn babies develop Jaundice. Suppose we randomly sample 2 full-term new born babies & let 'X' represent the number that develop jaundice. What is the PDF of x?

Ans) Possible values of x → 0, 1 & 2, sample space = [YY, YN, NY, NN]

~~P(0)~~ =

since probability of Jaundice is .6, $\bar{P}(J) = .4$

sample space = YY = 0.6×0.6 = 0.36 → P(2) = 0.36

YN = 0.6×0.4 = 0.24

NY = 0.4×0.6 = 0.24 } P(1) = 0.48

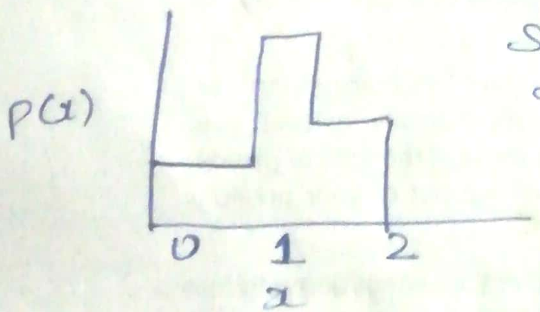
NN = 0.4×0.4 = 0.16

P(0) = 0.16

$P(X=0) = 0.16$, $P(X=1) = 0.48$, $P(X=2) = 0.36$

All discrete PP must satisfy → $0 \le p(x) \le 1$ for all x.

$\sum_{\text{all } x} p(x) = 1$



Some example of discrete probability distributions are → hypergeometric

→ geometric

→ binomial

→ Poisson

The Expected value of a random variable is the theoretical mean of the random variable denoted by $\boxed{E(x) = \mu}$

E(x) for random variable X :-

$$E(x) = \sum_{\text{all } x} x \cdot p(x)$$

var → $\sigma^2 = E[(X-\mu)^2]$

$$= \sum_{\text{all } x} (x-\mu)^2 \cdot p(x)$$

A useful relationship: $E[(X-\mu)^2] = E(X)^2 - [E(X)]^2$
$= E(x)^2 - \mu^2$

for the previous example $E(X) = 0 \times 0.16 + 1 \times 0.48 + 2 \times 0.36 = 1.2$

$\sigma^2 = (0-1.2)^2 \times 0.16 + (1-1.2)^2 \times 0.48 + (2-1.2)^2 \times 0.36$
$= 0.48$

Std. dev (σ) $= \sqrt{0.48}$

Bernoulli Distribution

Ex:- suppose we have a single trial. The trial can result in one of two possible outcomes, labelled success & failure

Two important conditions

$P(\text{success}) = p$ & $P(\text{failure}) = 1-p$

Also we def $X = 1$ if success & 0 if failure

Then X has a bernoulli distribution

$$\boxed{P(X = x) = p^x (1-p)^{1-x}}$$

Probability Mass function

This formula gives value for all x.

$P(X=0) = 1-p$

$P(X=1) = p$

The mean of bernoulli distribution is $\boxed{p}$ & its variance is $\boxed{p(1-p)}$

$E(X) = 1 \cdot p + 0 \cdot 1-p = p$

$E(x^2) = 1 \cdot p + 0 \cdot 1-p = p$

so $E[(x-\mu)^2] = E(x^2) - [E(x)]^2$
$= p - p^2 = \boxed{p(1-p)}$ Derived

Scanned by CamScanner

Binomial Distribution → The number of successes in 'n' independent bernoulli trials has a binomial distribution.

Let

X = no of success in n trials

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

Mean of binomial distribution = np

$$\sigma^2 = np(1-p)$$

(Q) A six-sided die is rolled 3 times. What is the probability a 5 comes up exactly twice?

Success → Rolling a 5
Failure → Rolling anything but a 5

X has a binomial distribution with $n = 3$ & $p = 1/6$

$$P(X=2) = {}^3C_2 \cdot \left(\frac{1}{6}\right)^2 \left(1 - \frac{1}{6}\right)^{3-2}$$

$$= 3 \cdot \frac{1}{36} \times \frac{5}{6}$$

$$= \boxed{0.0694}$$

Hypergeometric Distribution

- Trials are not independent.
- We are randomly sampling n objects without replacement from a series that contains 'a' success & 'N-a' failures

X represents the no of success in the sample.

Then X has the hypergeometric distribution

$$P(X=x) = \frac{\binom{a}{x}\binom{N-a}{n-x}}{\binom{N}{n}}$$

(Q) Suppose a large high school has 1100 female students & 900 male students. A random sample of 10 students is drawn. What is the prob exactly 7 of them are female.

$$P(X=7) = \frac{({}^{1100}C_7)({}^{900}C_3)}{{}^{2000}C_{10}}$$

Geometric Distribution

The Geometric distribution is the dist of the number of trials needed to get the first success in repeated independent bernoulli trials.

Let X represents the no of trials needed to get the first success

For the first success to occur on the x th trial

(1) The first x-1 trials must be failures

(2) The x th trial must be a success.

$(1-p)\cdot(x-1)\cdot p$

$$P(X=x) = (1-p)^{x-1} p$$

$$\mu = \frac{1}{p} \ \& \ \sigma^2 = \frac{1-p}{p^2}$$

Poisson Distribution

Suppose we are counting the number of occurrences of an event in a given unit of time, distance, area or volume.

Ex:- no of car accidents in a day

If Events are occurring independently & prob of event occurring in a given length of time does not change then X has a poisson distribution

$$P(X=x) = \frac{\lambda^x e^{-\lambda}}{\lfloor x}, \ \mu = \lambda \ \& \ \sigma^2 = \lambda$$

Negative binomial Distribution

Q A coin is tossed repeatedly until heads comes up for the sixth time. What is the prob this happens on the 15th toss?

The negative binomial distribution is the distribution of the number of trials needed to get the r^{th} success.

It is diff from binomial dis as in binomial we try to find the no of success from given trials, here it is inverted

For the rth success to occur on the xth trial
The first $(x-1)$ trials must result in $(r-1)$ successes
$\binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)}$ using binomial

& the xth trial must be a success which has a probability of p

$\therefore$

$$PMF = p \cdot \binom{x-1}{r-1} p^{r-1} (1-p)^{(x-1)-(r-1)}$$

$$= \binom{x-1}{r-1} p^{r} (1-p)^{x-r} \quad \text{for } x = r, r+1 \ldots$$

$\mu = \frac{r}{p}$ & $\sigma^2 = \frac{r \cdot (1-p)}{p^2}$

Q) A person conducting telephonic surveys must get 3 more completed surveys before their job is finished. On each randomly dialed number, there is a 9% chance of reaching an adult who will complete the survey.
What is the prob the 3rd completed survey occurs on the 10th call

On 9 calls, $X = 2$ → Binomial

$$P(X=2) = {}^9C_2 \times (.09)^2 \times (1-.09)^7$$

$$= 36 \times .09 \times .09 \times .9 \times (.91)^7 = 0.1$$

$$P(X=3) = 0.14 \times .09 = \boxed{0.013}$$

<u>Multinomial distribution</u>

Suppose

→ There are n independent trials

→ Each trial results in one of k mutually exclusive & exhaustive outcomes

→ On any single trial these k outcomes occurs with probabilities $p_1 \dots p_k$.

$$\sum_{i=1}^{k} p_i = 1.$$

Let the random variable X_i represent the no of occurrences of outcome i

Then

$$P(X_1 = x_1, \dots X_k = x_k) = \frac{\lfloor n}{\lfloor x_1 \dots \lfloor x_k} p_1^{x_1} \cdots p_k^{x_k}$$

(Q) In a random sample of 10 american cars what is the prob 6 have O, 2 have A, 1 has B & 1 has AB

given

$P(O) = 0.44$

$P(A) = 0.42$

$P(B) = 0.10$

$P(AB) = 0.04$

→

$$P(O=6, A=2, B=1, AB=1)$$

$$= \frac{\lfloor 10}{\lfloor 6 \lfloor 2 \lfloor 1 \lfloor 1} \times 0.44^6 \times 0.42^2 \times 0.10 \times 0.04$$

$=$

Central Limit theorem :) :)

The sample mean will be approximately normaly distributed for large sample size, which regardless of the distribution from which we are sampling.

Mean of samples = Mean of whole distribution

$\sigma_{sample} = \frac{\sigma}{\sqrt{n}}$