Index

Sept 8, 2021

~ PART 3 ~

RANDOM VARIABLES

Random Variables: numerical values computed from the outcomes of the experiment.

→ DEFINITION

A random variable is a function with domain as the sample space of an experiment and range as the real numbers, i.e., a function from the sample space to the real line.

- Toss a coin, Sample space = {H, T}
  → Random variable X : $X(H) = 0$, $X(T) = 1$
  → Random variable Y : $Y(H) = -10$, $Y(T) = 200$
  → Random variable Z : $Z(H) = \sqrt{2}$, $Z(T) = \pi$
- While all of the above are valid random variables, usually meaningful functions are considered!

→ EXAMPLE

(1) Throw A Die

Sample space = {1, 2, 3, 4, 5, 6}

A random variable X can be defined as follows:

- Pick any 6 real numbers $x_1, x_2, x_3, x_4, x_5, x_6$.
- $X(1) = x_1$, $X(2) = x_2$, $X(3) = x_3$, $X(4) = x_4$, $X(5) = x_5$, $X(6) = x_6$

What values for $x_1, x_2, x_3, x_4, x_5, x_6$?

→ If x_i's are distinct, random variable is a one-to-one function
- Essentially, same as sample space

→ However, x_i's need not be distinct
- Random variable E: $E(2) = E(4) = E(6) = 1$,
  $E(1) = E(3) = E(5) = 0$

$X<3$

Random Variable And Events

If X is a random variable,
$(X<x) = \{ s \in S : X(s) < x \}$ is an event for all real x.

So, $(X<x)$, $(X=x)$, $(X \le x)$, $(X \ge x)$ are all events.

Example : Throw a dice
$S = \{1, 2, 3, 4, 5, 6\}$

$X(1) = 1$, $X(2) = 2$, $X(3) = 3$, $X(4) = 4$, $X(5) = 5$, $X(6) = 6$

→ $X=1$: die shows 1, $X<4$: event $\{1, 2, 3\}$

→ Any event can be expressed in terms of X.
- Event $\{2, 5\}$: $(X=2) \cap (X=5)$

Why Random Variables?

- IPL powerplay over : Several random variables can be defined
  → No. of runs in the over
  → No. of actual deliveries
  → No. of wickets
  → No. of boundaries
  → No. of dot balls, etc.

- Instead of trying to assign probabilities to the entire outcome, we will usually study random variables and assign probabilities to events defined through them.

  → This reduces the detail in the outcome to something simpler
  → Using limited data, only such random variables can be studied

---

RANGE OF RANDOM VARIABLE

Range of random variable is the set of values taken by it. Range is a subset of the real line.

- Given the definition of a random variable, the first task is to find its range.

Examples : Throw a die

→ X = no. shown on a die, Range = $\{1, 2, 3, 4, 5, 6\}$

→ E = 0 if no. is odd, E = 1 if no. is even
Range = {0, 1}

~~DISCRETE~~

DISCRETE RANDOM VARIABLE & DISTRIBUTIONS

Range of a Random Variable

→ The range of a random variable is the set of values taken by it. Range is the subset of a real line.

* Given the definition of a random variable, the first task is to find its range.

Examples: (1) Throw a die, X = no. shown on a die
Range = {1, 2, 3, 4, 5, 6}

(2) Throw a die, E = 0 (odd), E = 1 (even)
Range = {0, 1}

Discrete Random Variable

→ A random variable is said to be discrete if its range is a discrete set.

Distribution of a Discrete Random Variable: Probability Mass Function (PMF)

→ PMF : The probability mass function (PMF) of a discrete random variable X with range set T is the function $f_X : T \to [0,1]$ defined as

$$f_X(t) = P(X=t) \quad \text{for } t \in T$$

* $(X=t)$ is an event
  $P(X=t) = P(\text{all outcomes that result in } X \text{ taking value } t)$

* Let B be a subset of T. Consider the event $(X \in B)$
  $P(X \in B) = P(\text{all outcomes that result in } X \text{ taking values in } B)$
  $= \sum_{t \in B} P(X=t) = \sum_{t \in B} f_X(t)$

EXAMPLE

(1) Suppose a fair coin is flipped 3 times.

(a) How many heads will appear?
(b) Which will be the first flip (if any) that shows heads?

Solution: Let X = no. of heads
Y = first flip that shows head

$S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$

outcome	p(outcome)	X	Y
HHH	1/8	3	1
HHT	1/8	2	1
HTH	1/8	2	1
HTT	1/8	1	1
THH	1/8	2	2
THT	1/8	1	2
TTH	1/8	1	3
TTT	1/8	0	none

$X \in \{0, 1, 2, 3\}$

$Y \in \{\overset{0}{\text{none}}, 1, 2, 3\}$

$f_X(0) = p(TTT) = 1/8$

$f_X(1) = p(HTT, THT, TTH) = 3/8$

$f_X(2) = p(HHT, HTH, THH) = 3/8$

$f_X(3) = p(HHH) = 1/8$

$f_Y(0) = 1/8$

$f_Y(1) = 4/8$

$f_Y(2) = 2/8$

$f_Y(3) = 1/8$

Sept 9, 2021

PROPERTIES OF P.M.F

Random Variable, Range and PMF

- Random variable X with range $T = \{t_1, t_2, \ldots t_k\}$, PMF f_X

t	t_1	t_2	t_3	....	t_k
$f_X(t)$	$f_X(t_1)$	$f_X(t_2)$	$f_X(t_3)$	....	$f_X(t_k)$

- PROPERTIES OF PMF

→ $0 \leq f_X(t) \leq 1$

→ $\Sigma_{t \in T} f_X(t) = 1$

- Often, random variables and their PMFs are studied without much mention of experiment or sample space

→ Events are defined using random variables.
→ probabilities evaluated using PMF.

EXAMPLE : Working With PMF 1

(1) → PMF of a random variable X is given partially in the foll. table :

t	-1	1	2	4
$f_X(t)$	0.5	0.25	0.125	

Find $f_X(4)$. Find the range of X. Find $P(X>3)$. Find $P(X<\frac{3}{2})$.

Solution: We know,

$\sum f_X(t) = 1$

$\Rightarrow f_X(-1) + f_X(1) + f_X(2) + f_X(4) = 1$

$\Rightarrow 0.5 + 0.25 + 0.125 + f_X(4) = 1$

$\Rightarrow 0.875 + f_X(4) = 1$

$\Rightarrow f_X(4) = 1 - 0.875 = 0.125$

(b) Range of X $\rightarrow$ $X = \{-1, 1, 2, 4\}$

(c) $P(X>3) = P(X=4) = f_X(4) = 0.125$

(d) $P(X<\frac{3}{2}) = P(X<1.5) = P(X=1) + P(X=-1) = 0.25 + 0.5 = 0.75$

---

12) The PMF of a random variable X is given below:

$f_X(k) = \frac{c}{3^k}$, for $k = 1, 2, 3, \ldots$

Find c. Find $P(X>10)$. Find $P(X>10 \mid X>5)$.

Sol.

k	1	2	3	....	n
$f_X(k)$	$c/3$	$c/9$	$c/27$	....	$c/3^n$

$\sum f_X(k) = 1 \Rightarrow \frac{c}{3} + \frac{c}{9} + \frac{c}{27} + \ldots + \frac{c}{3^\infty} = 1$

Formula of Geometric Progression

$$a + a^2 + a^3 + \dots\dots = \frac{a}{1-a} \quad [\text{if } a < 1]$$

$$\Rightarrow c\left[\frac{1}{3} + \frac{1}{9} + \frac{1}{27} + \dots + \frac{1}{3^{\infty}}\right] = 1$$

$$\Rightarrow c\left[\frac{1/3}{1 - 1/3}\right] = 1$$

$$\Rightarrow c\left[\frac{1}{2}\right] = 1$$

$$\Rightarrow \boxed{c = 2}$$

Now, we need to find $P(X > 10)$, which means

(b) $$P(X > 10) = P(X=11) + P(X=12) + \dots\dots$$

$$= \frac{2}{3^{11}} + \frac{2}{3^{12}} + \dots\dots = 2\left[\frac{1}{3^{11}} + \frac{1}{3^{12}} + \dots\right]$$

$$\Rightarrow P(X > 10) = 2\left[\frac{1/3^{11}}{1 - 1/3}\right] = \cancel{2}\left[\frac{1}{3^{11}} \times \frac{3}{\cancel{2}}\right] = \frac{1}{3^{10}}$$

(c) $$P(X > 10 \mid X > 5) = \frac{P([X > 10] \cap [X > 5])}{P(X > 5)} = \frac{P(X > 10)}{P(X > 5)}$$

$$\therefore P(X > 10 \mid X > 5) = \frac{1/3^{10}}{1/3^{5}} = \frac{1}{3^{5}}$$

'$\sim$' — is distributed as

COMMON DISTRIBUTIONS

(1) Uniform Random Variable

$X \sim \text{Uniform}(T)$, where T is some finite set

is distributed as

Range : Finite set T

PMF : $f_X(T) = \frac{1}{|T|}$, for all $t \in T$

size of T

Example : (1) Toss a fair coin,

$X \sim \text{Uniform}(\{0,1\})$, where 0 - head, 1 - Tail

(2) Throw a fair die,

$X \sim \text{Uniform}(\{\cancel{0}\,1,2,3,4,5,6\})$

---

(2) Bernoulli Random Variable

$X \sim \text{Bernoulli}(p)$, where $0 \le p \le 1$

RANGE : $\{0,1\}$

PMF : $f_X(0) = 1-p$, $f_X(1) = p$

Example : Bernoulli trial,

p = probability of success, $X \sim \text{Bernoulli}(p)$

---

(3) Binomial Random Variable

$X \sim$ Binomial (n, p), where n: +ve integer, $0 \leq p \leq 1$

RANGE : $\{0, 1, 2, \ldots, n\}$
PMF : $f_X(k) = {}^nC_k (p)^k (1-p)^{n-k}$

EXAMPLE :

- No. of successes in n independent Bernoulli (p) trials
- Check the above PMF is valid
  - $\Rightarrow {}^nC_k (p)^k (1-p)^{n-k} \geq 0$
  - $\Rightarrow \sum_k {}^nC_k (p)^k (1-p)^{n-k} = 1$

---

(4) Geometric Random Variable

$X \sim$ Geometric (p), where $0 \leq p \leq 1$

RANGE : $\{1, 2, 3, \ldots\}$
PMF : $f_X(k) = (1-p)^k p$

EXAMPLE :

- No. of trials for first success in repeated independent Bernoulli (p) trials.
- Check that the above PMF is valid
  - $\Rightarrow$ Sum of G.P. formula
    - $a + ar + ar^2 + \ldots = \dfrac{a}{1-r}$
    - $a + a^2 + a^3 + \ldots = \dfrac{a}{1-a}$ (If $a < 1$)

Generalisation of Geometric R.V.

(5) Negative Binomial Random Variable

$X \sim$ Negative Binomial (r, p), where r: +ve integer, $0 < p \leq 1$

RANGE : $\{r, r+1, r+2, \dots\}$

PMF : $f_X(k) = {}^{k-1}C_{r-1} (1-p)^{k-r} p^r$

EXAMPLE :

→ No. of trials for r success in repeated independent Bernoulli (p) trials

→ $r = 1$: geometric random variable

→ Check that the above PMF is valid

(6) Poisson Random Variable

$X \sim$ Poisson (λ), where $\lambda > 0$ → non-negative real parameter (real no.)

RANGE : $\{0, 1, 2, 3, \dots\}$

PMF : $f_X(k) = \dfrac{e^{-\lambda} \cdot \lambda^k}{k!}$

k	0	1	2	3	4	....
$f_X(k)$	$e^{-\lambda}$	$e^{-\lambda} \cdot \lambda$	$\dfrac{e^{-\lambda}\lambda^2}{2}$	$\dfrac{e^{-\lambda}\lambda^3}{6}$	$\dfrac{e^{-\lambda}\lambda^4}{24}$	....

- Check PMF is valid : use expansion for $e^{\lambda} = 1 + \lambda + \dfrac{\lambda^2}{2!} + \dfrac{\lambda^3}{3!} + \dots$

(7) Hypergeometric Random Variable

$X \sim$ HyperGeo (N, r, m), where N, r, m : the integers

- Consider a population of 'N' persons with 'r' of Type 1 and 'N-r' of Type 2
- Select 'm' persons uniformly at random without replacement
- X = no. of persons of Type 1 selected

<u>Range of X</u>

- $N = 100,\ r = 50,\ m = 20 \Rightarrow X \in 0, 1, 2, \ldots, 20$
- $N = 100,\ r = 10,\ m = 20 \Rightarrow X \in 0, 1, 2, \ldots, 10$
- $N = 100,\ r = 90,\ m = 20 \Rightarrow X \in 10, 11, 12, \ldots 20.$

Range

$X \in \max(0, m-(N-r)), \ldots, \min(r, m)$

<u>PMF of HyperGeo R.V</u>

$$\text{PMF}: f_X(k) = \frac{{}^{r}C_{k}\ {}^{N-r}C_{m-k}}{{}^{N}C_{m}}$$

Application Of Poisson Random Var.

\# Events occuring over a period of time

EXAMPLES:
- Arrival of a person to a booking counter queue
- Arrival of a visitor to a website
- Emission of a particle by a radioactive decay
- A meteorite entering earth's atmosphere.

What is common to above examples?

- Arrival Rate : can be assumed to be constant
- Given 1 arrival, the time for next arrival can be assumed to be independent.

★ Under above assumptions, the no. of arrivals in a fixed period of time becomes a Poisson R.V.

\# EXAMPLE 1 : Radioactive decay

In 2608 time intervals of 7.5 seconds each, emission of alpha particles is as follows:

#particles →	0	1	2	3	4	5	6	7	8
#times →	57	203	383	525	532	408	273	139	45
fraction →	0.022	0.078	0.147	0.201	0.204	0.156	0.105	0.053	0.017

$$\text{Emmission Rate} = \frac{\text{total no. of particles}}{2608} = 3.8673\ (\lambda)$$

Time of next emission : independent of past

Poisson Model : $P(\text{no. of particle} = k) = \frac{e^{-\lambda} \lambda^k}{k!}$

The fraction given above should be ~~closely~~ close to probability under Poisson model.

FUNCTIONS OF ONE RANDOM VAR.

X : a random variable with range T and PMF $f_X(t)$

X : function from sample space to T

T : subset of real line

Let $f(x)$ be a function from real line to real line.

Then, $f(X)$ can be seen as a composition of two functions

So, $f(X)$ is also a random variable in the same probability space.

General Case : X : random variable with PMF $f_X(t)$

$f(X)$: random variable whose PMF is given as follows :

$$f_{f(X)}(a) = P(f(X) = a) = P(X \in \{t : f(t) = a\})$$

$$= \sum_{t : f(t) = a} f_X(t)$$