Random Variable and Introduction to Probability Distribution

- In a statistical experiment 3 electronic items are tested, D signifies the item is defective while N signifies the item is non-defective.
- $S = \{NNN, NND, NDN, DNN, NDD, DND, DDN, DDD\}$
- Anyone might be interested to know the number of defective items.
- With each sample point there is an associated number of defective items and it can be 0, 1, 2, 3.
- These values are, of course, random quantities *determined by the outcome of the experiment*.

- They may be viewed as values assumed by the *random variable X*, the number of defective items when three electronic components are tested.
- A **random variable** is a function that associates a real number with each element in the sample space.
- We shall use a capital letter, say *X*, to denote a random variable and its corresponding small letter, *x* in this case, for one of its values.

- In the electronic component testing illustration, we notice that the random variable X assumes the value 2 for all elements in the subset $E = \{DDN, DND, NDD\}$ of the sample space S. That is, each possible value of X represents an event that is a subset of the sample space for the given experiment.

- Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where Y is the **number of red balls**, are

Sample Point	y
RR	2
RB	1
BR	1
BB	0

- Statisticians use **sampling plans** to either accept or reject batches or lots of material. Suppose one of these sampling plans involves sampling independently 10 items from a lot of 100 items in which 12 are defective.

- Let *X* be the random variable defined as the number of items found defective in the sample of 10. In this case, the random variable takes on the values 0, 1, 2, . . . , 9, 10.

Random variables can be discrete or continuous

- **Discrete** random variables have a countable number of outcomes
  - Examples: Dead/alive, outcomes when a die is rolled, rain/not-rain

- **Continuous** random variables have an infinite continuum of possible values.
  - Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

- **Discrete Probability Distributions:** A discrete random variable assumes each of its values with a certain probability.
- In the case of tossing a coin three times, the variable *X*, representing the number of heads,

Sample Points	x
HHH	3
HHT	2
HTH	2
HTT	1
THH	2
THT	1
TTH	1
TTT	0

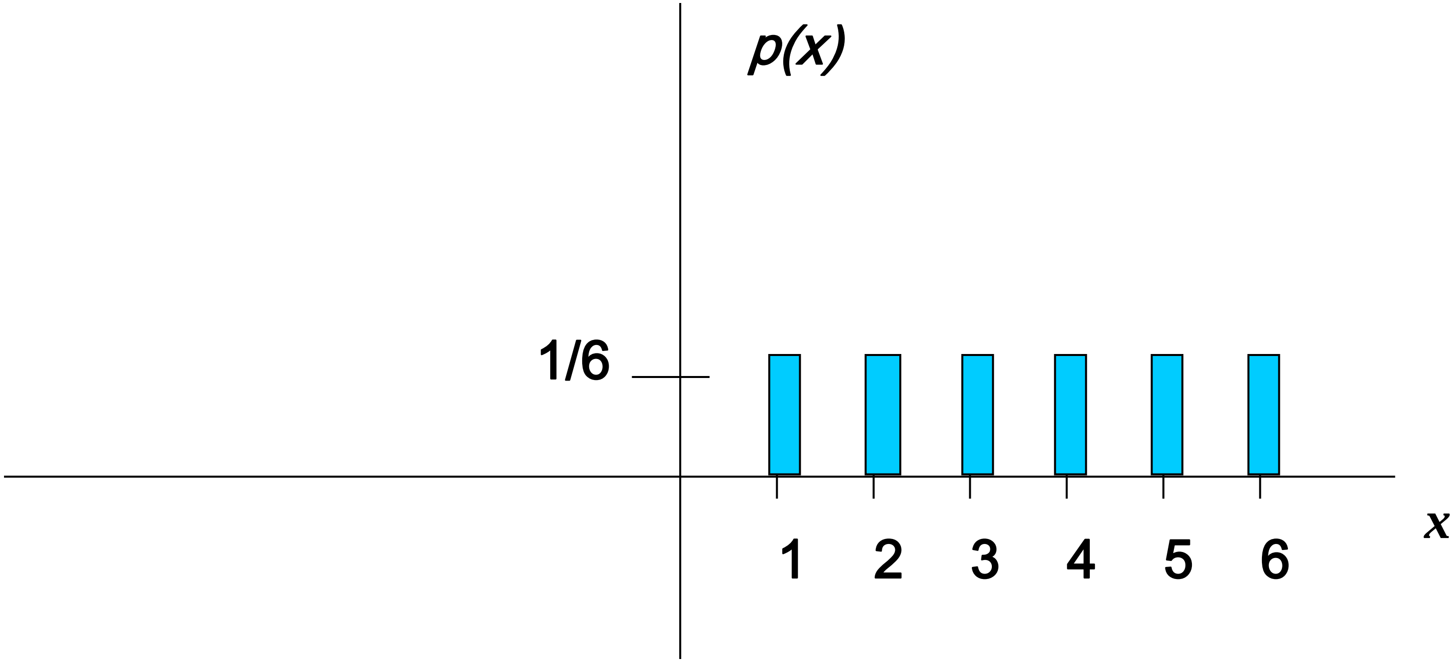
x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8

- Frequently, it is convenient to represent all the probabilities of a random variable X by a formula. Therefore, we write $f(x) = P(X = x)$.

- The set of ordered pairs $(x, f(x))$ is called the **probability function**, **probability mass function**, or **probability distribution** of the discrete random variable X.

- The set of ordered pairs $(x, f(x))$ is a p**robability function**, **probability mass function**, or **probability distribution** of the discrete random variable X if, for each possible outcome x,
  1. $f(x) \geq 0$,
  2. $\sum f(x) = 1$
  3. $P(X = x) = f(x)$.

Discrete example: roll of a die



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

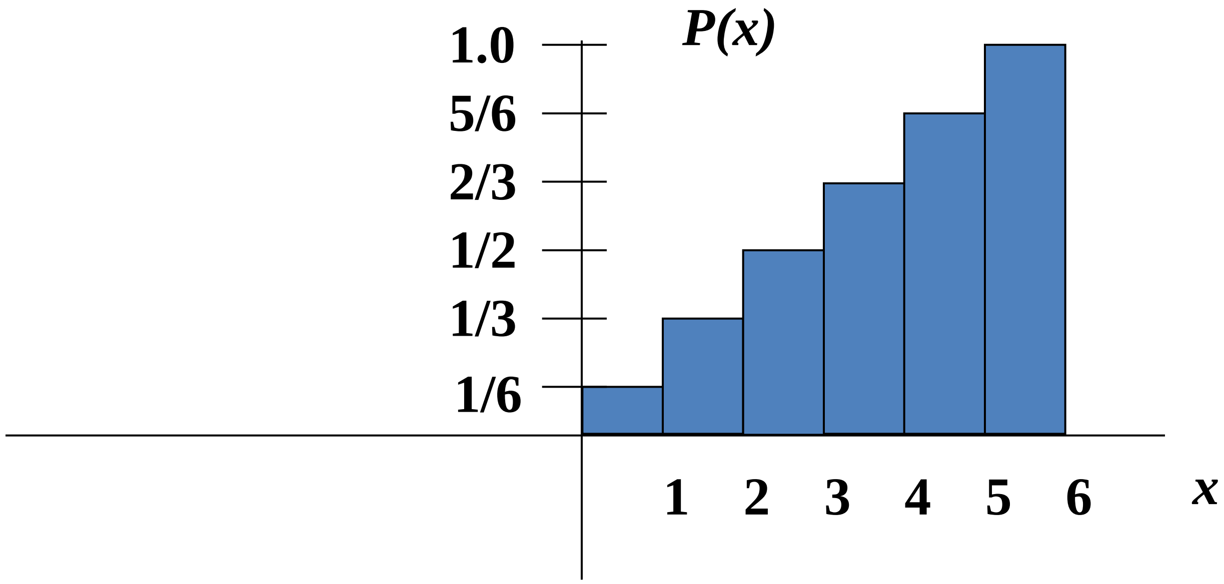
Probability mass function (pmf)

x	$p(x)$
1	$p(x=1)=1/6$
2	$p(x=2)=1/6$
3	$p(x=3)=1/6$
4	$p(x=4)=1/6$
5	$p(x=5)=1/6$
6	$p(x=6)=1/6$

- A shipment of 20 similar laptops to a retail outlet contains 3 laptops which are defective. If a school makes a random purchase of 2 of these laptops, find the probability distribution for the number of defectives laptops.
- Let X be a random variable whose values x are the possible numbers of defective laptops purchased by the school. Then x can only take the values 0, 1, and 2. Now
- $f(0) = P(X = 0) = {}^{3}C_{0} * {}^{17}C_{2} / {}^{20}C_{2} = 68/95$
- $f(1) = P(X = 1) = {}^{3}C_{1} * {}^{17}C_{1} / {}^{20}C_{2} = 51/190$
- $f(2) = P(X = 2) = {}^{3}C_{2} * {}^{17}C_{0} / {}^{20}C_{2} = 3/190$
- Thus, the probability distribution of X is

x	0	1	2
f(x)	68/95	51/190	3/190

Cumulative distribution function (CDF)



Cumulative distribution function

x	$P(x \leq A)$
1	$P(x \leq 1)$=1/6
2	$P(x \leq 2)$=2/6
3	$P(x \leq 3)$=3/6
4	$P(x \leq 4)$=4/6
5	$P(x \leq 5)$=5/6
6	$P(x \leq 6)$=6/6

Examples

1. What’s the probability that you roll a die and outcome is 3 or less?

$P(x \leq 3)$=1/2

2. What’s the probability that you roll a die and outcome is 5 or higher?

$P(x \geq 5) = 1 – P(x \leq 4) = 1\text{-}2/3 = 1/3$

Practice Problem

Which of the following are probability functions?

a. $f(x)=0.25$ for x=9,10,11,12

b. $f(x)= (3-x)/2$ for x=1,2,3,4

c. $f(x)= (x^2+x+1)/25$ for x=0,1,2,3

Answer (a)

a. *f(x)=0.25* for x=9,10,11,12

x	***f(x)***
9	**0.25**
10	**0.25**
11	**0.25**
12	**<u>0.25</u>**

1.0

Yes, probability function!

Answer (b)

b. $f(x)= (3-x)/2$ for x=1,2,3,4

x	*f(x)*
1	**(3-1)/2=1.0**
2	**(3-2)/2=.5**
3	**(3-3)/2=0**
4	**(3-4)/2=-.5**

Though this sums to 1, you can't have a negative probability; therefore, it's not a probability function.

Answer (c)

c. $f(x) = (x^2+x+1)/25$ for x=0,1,2,3

x	**f(x)**
0	**1/25**
1	**3/25**
2	**7/25**
3	**13/25**

24/25

Doesn't sum to 1. Thus, it's not a probability function.

Example

- The number of ships to arrive at a harbor on any given day is a random variable represented by *x*. The probability distribution for *x* is:

x	10	11	12	13	14
P(x)	0.4	0.2	0.2	0.1	0.1

Find the probability that on a given day:

a. exactly 14 ships arrive $p(x=14) = 0.1$

b. At least 12 ships arrive $p(x\geq 12) = (0.2 + 0.1 + 0.1) = 0.4$

c. At most 11 ships arrive $p(x\leq 11) = (0.4 + 0.2) = 0.6$

Example

You are lecturing to a group of 1000 students. You ask them to each randomly pick an integer between 1 and 10. Assuming, their picks are truly random:

- What's your best guess for how many students picked the number 9?

  Since $p(x=9)$ = 1/10, we'd expect about 1/10th of the 1000 students to pick 9. 100 students.

- What percentage of the students would you expect picked a number less than or equal to 6?

  Since $p(x \leq 6)$ = 1/10 + 1/10 + 1/10 + 1/10 + 1/10 + 1/10 =0.6= 60%

Joint Probability Distributions

- There are many situations, where we may find it desirable to record the simultaneous outcomes of several random variables.

- If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function with values $f(x, y)$ for any pair of values (x, y) within the range of the random variables X and Y. It is customary to refer to this function as the **joint probability distribution** of X and Y.

- Hence, in the discrete case $f(x, y) = P(X = x, Y = y)$;

- The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if
  1. $f(x, y) \geq 0$ for all (x, y),
  2. $\sum_x \sum_y f(x,y) = 1$
  3. $P(X = x, Y = y) = f(x, y)$.
- For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x,y)$

Example

- **Two ballpoint** pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find
  a) the joint probability function $f(x, y)$,
  b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) \mid x + y \leq 1\}$.

3-**blue(X)**, 2-**red(Y)**, 3-green

- The possible pairs of values (x, y) are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), and (2, 0).
- Now, $f(x, y)$, for example, represents the probability that x blue and y red pens are selected.
- The total number of equally likely ways of selecting any 2 pens from the 8 is $^{8}C_{2}$= 28.
- f(0,0)= $^{3}C_{2}/\ ^{8}C_{2}$ =3/28 [choosing 2 green pens out of 3 green pens / all possible ways]
- f(0,1)= $^{3}C_{1} * {}^{2}C_{1} / {}^{8}C_{2}$ =6/28 [choosing 1 green pen out of 3 green pens and 1 red pen out of 2 red pens / all possible ways]
- f(1,0)= $^{3}C_{1} * {}^{3}C_{1} / {}^{8}C_{2}$ =9/28 [choosing 1 blue pen out of 3 blue pens and choosing 1 green pen out of 3 green pens/ all possible ways]
- f(1,1)= $^{3}C_{1} * {}^{2}C_{1} / {}^{8}C_{2}$ =6/28 [choosing 1 blue pen out of 3 blue pens and choosing 1 red pen out of 2 red pens/ all possible ways]
- f(0,2)= $^{2}C_{2} / {}^{8}C_{2}$ =1/28 [choosing 2 red pens out of 2 red pens/ all possible ways]
- f(2,0)= $^{3}C_{2} / {}^{8}C_{2}$ =3/28 [choosing 2 blue pens out of 3 blue pens/ all possible ways]
- $f(x,y)=\frac{^{3}C_{x} * {}^{2}C_{y} * {}^{3}C_{2-x-y}}{^{8}C_{2}}$ for x = 0, 1, 2; y = 0, 1, 2; and $0 \le x + y \le 2$.

- The probability that (X, Y) fall in the region A is
- $P[(X, Y) \in A] = P(X + Y \leq 1) = f(0, 0) + f(0, 1) + f(1, 0) =$ 3/28+6/28+9/28=18/28=9/14

- The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

  1. $f(x, y) \geq 0$, for all (x, y),
  2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
  3. $P[(X, Y) \in A] = \iint f(x, y)\, dx\, dy$ for any region A in the xy plane.

A

- A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let *X* and *Y* , respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Verify condition 2 of Definition.

- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy =$
  $\int_0^1\int_0^1 \frac{2}{5}(2x+3y)dxdy$
  $=\int_0^1 \left(\frac{2x^2}{5}+\frac{6xy}{5}\right)\Big|_{X=0}^{X=1} dy$
  $=\int_0^1 (\frac{2}{5}+\frac{6y}{5})\text{dy} = \frac{2}{5}y\Big|_{y=0}^{y=1} + \frac{6}{5}\frac{y^2}{2}\Big|_{y=0}^{y=1} = \frac{2}{5}+\frac{3}{5} = 1$

Marginal Distribution

- Given the joint probability distribution $f(x, y)$ of the discrete random variables X and Y, the probability distribution $g(x)$ of X alone is obtained by summing $f(x, y)$ over the values of Y. Similarly, the probability distribution $h(y)$ of Y alone is obtained by summing $f(x, y)$ over the values of X. We define $g(x)$ and $h(y)$ to be the **marginal distributions** of X and Y, respectively.

- The **marginal distributions** of *X* alone and of *Y* alone are $g(x) = \sum_y f(x, y)$ and $h(y) = \sum_x f(x, y)$ for the discrete case, and
- $g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$ and $h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$ for the continuous case.
- The term *marginal* is used here because, in the discrete case, the values of $g(x)$ and $h(y)$ are just the marginal totals of the respective columns and rows when the values of $f(x, y)$ are displayed in a rectangular table.

Example

- Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected. Give the marginal distribution of X alone and of Y alone.
- f(0,0)=3/28, f(0,1)=6/28 , f(0,2)=1/28 , f(1,0)=9/28 , f(1,1)=6/28, f(2,0)=3/28
- For the random variable X,
- $g(0) = f(0, 0) + f(0, 1) + f(0, 2)$ =10/28=5/14
- $g(1) = f(1, 0) + f(1, 1) + f(1, 2)$ =15/28
- $g(2) = f(2, 0) + f(2, 1) + f(2, 2)$=3/28
- For random variable Y
- h(0)=f(0,0)+f(1,0)+f(2,0)=15/28
- h(1)=f(0,1)+f(1,1)+f(2,1)=12/28=3/7
- h(2)=f(0,2)+f(1,2)+f(2,2)=1/28

Example

- The joint density for the random variables (*X*, *Y*), where *X* is the unit temperature change and *Y* is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Find the marginal densities $g(x)$, $h(y)$

- $g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{x}^{1} 10xy^2 dy$

$$= \frac{10}{3} xy^3 \Big|_{y=x}^{y=1} = \frac{10}{3} x(1 - x^3),\ 0 < x < 1,$$

- $h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{0}^{y} 10xy^2 dx$

$$= 5x^2y^2 \Big|_{x=0}^{x=y} = 5y^4,\ 0 < y < 1.$$