PROBABILITY DISTRIBUTIONS
INTRODUCTION

DR. GARGI BASU
ASSISTANT PROFESSOR
DEPARTMENT OF ECONOMICS
S.A. JAIPURIA COLLEGE

FOR SEM IV
DSCC -7

What is a Random Variable?

A **variable** is any characteristic, observed or measured. A variable can be either **random** or **constant** in the population of interest.

Note this differs from common English usage where the word variable implies something that **varies** from individual to individual.

For a defined population, every **random variable** has an associated distribution that defines the **probability** of occurrence of each possible value of that variable (if there are a finitely countable number of unique values) or all possible sets of possible values (if the variable is defined on the real line).

Probability Distribution

A **probability distribution** (function) is a list of the probabilities of the values (simple outcomes) of a random variable.

Table: Number of heads in two tosses of a coin

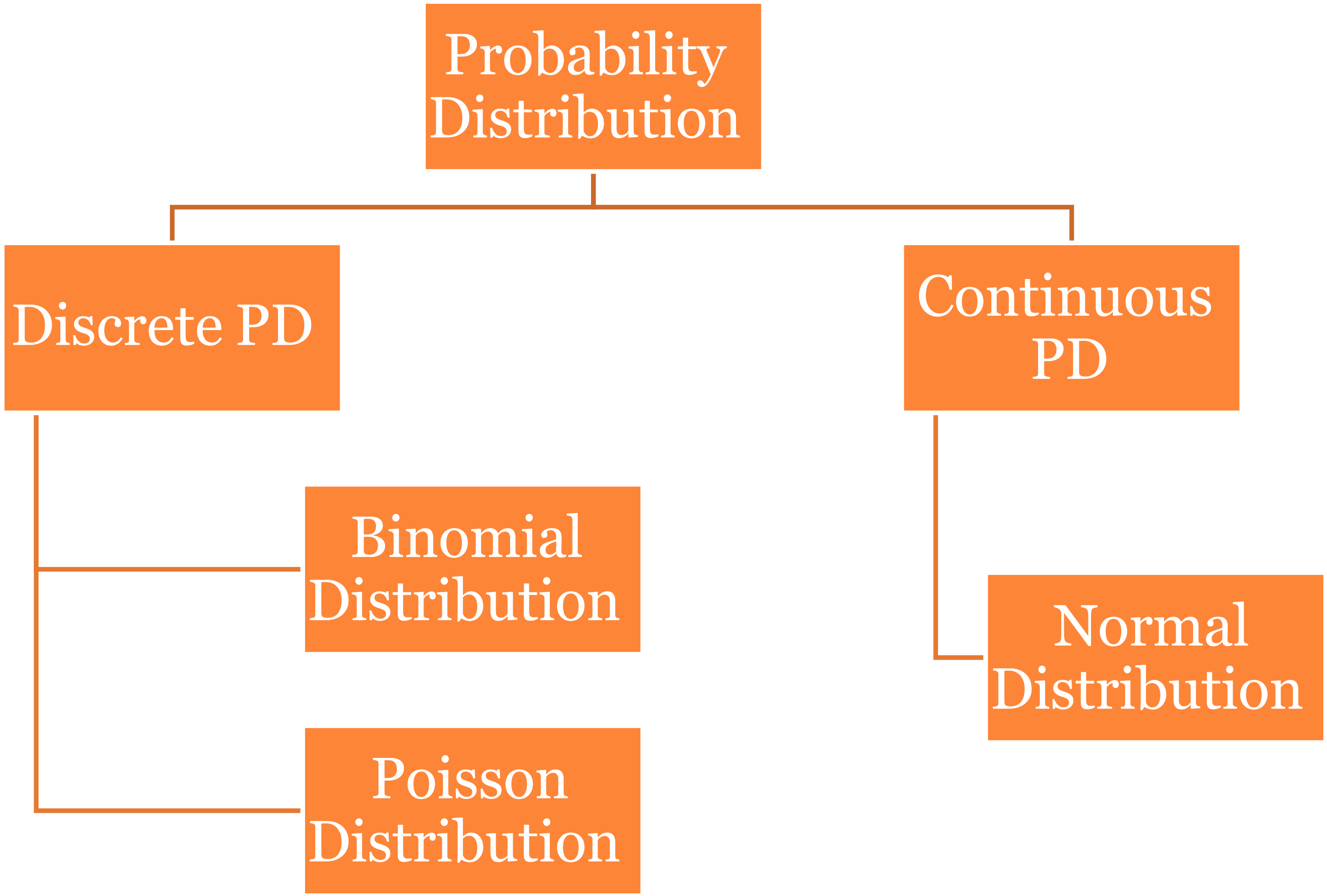
y	P(y)
outcome	*probability*
0	1/4
1	2/4
2	1/4

For some experiments, the probability of a simple outcome can be easily calculated using a specific **probability function.** If y is a simple outcome and p(y) is its probability.

$$0 \leq p(y) \leq 1$$

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

TYPES OF PROBABILITY DISTRIBUTION



Probability distributions can be discrete or continuous

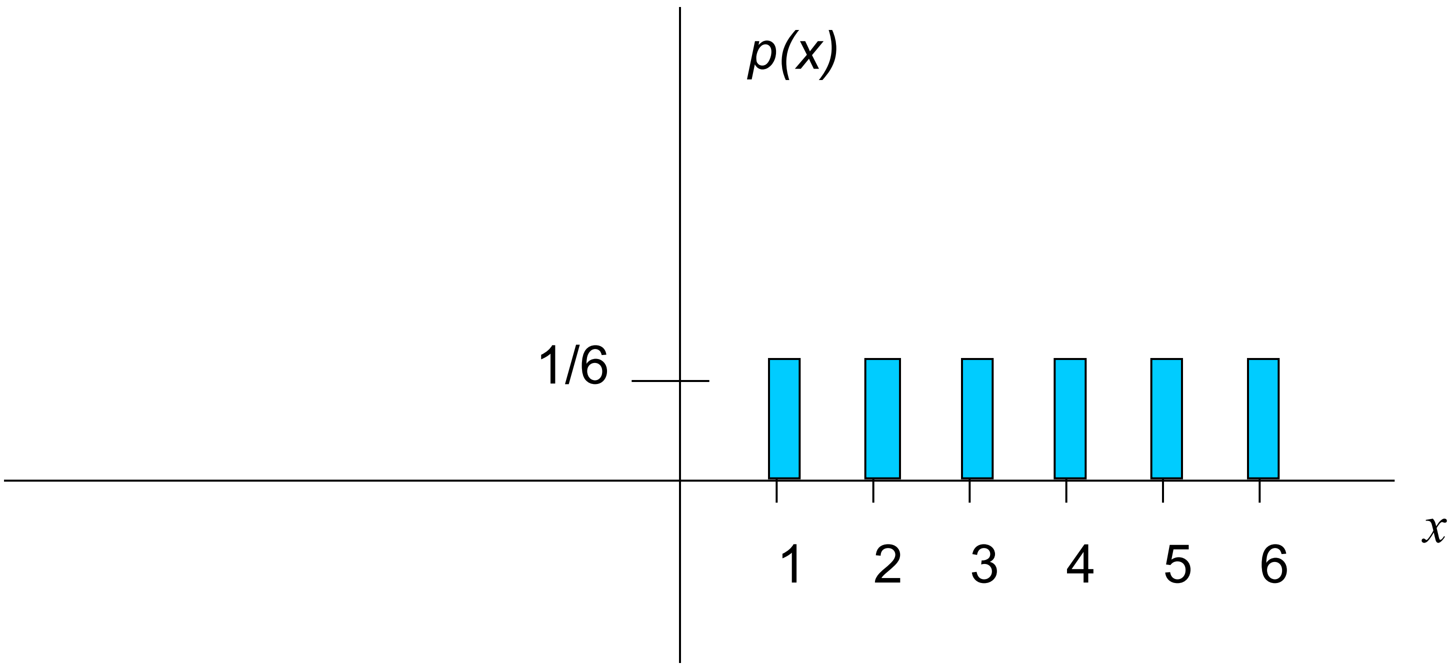
Discrete: has a countable number of outcomes

Examples: Dead/alive, treatment/placebo, dice, counts, etc.

Continuous: has an infinite continuum of possible values.

Examples: blood pressure, weight, the speed of a car, the real numbers from 1 to 6.

Discrete example: roll of a die

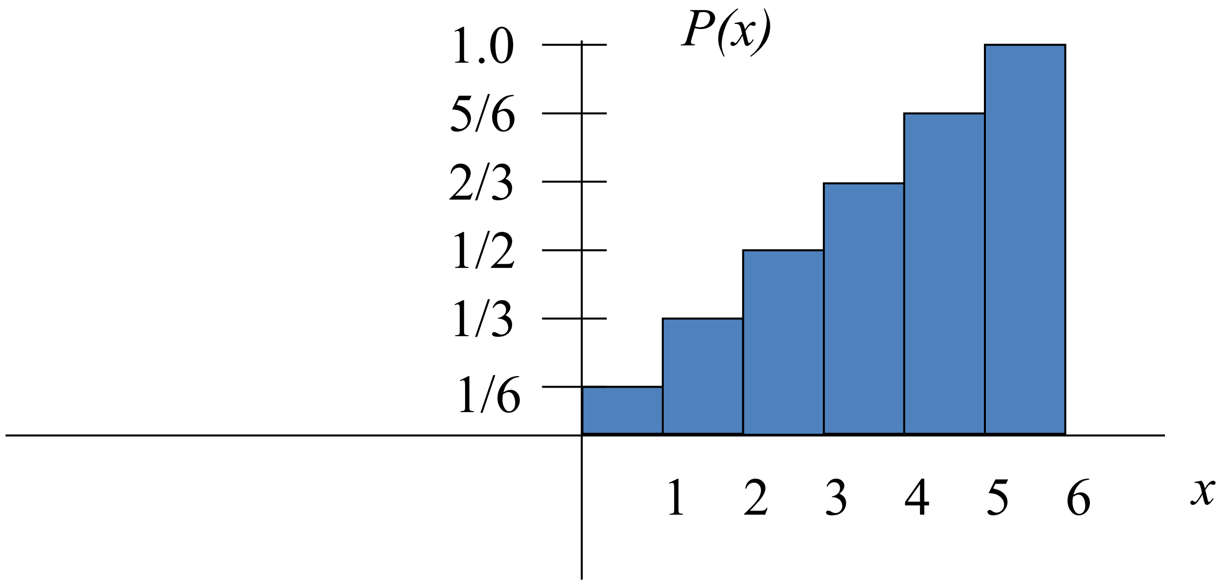


$$\sum_{\text{all } x} P(x) = 1$$

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	$\underline{p(x=6)=1/6}$
	1.0

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	$P(x \leq 6) = 6/6$

Practice Problem:

The number of patients seen in the ER in any given hour is a random variable represented by *x*. The probability distribution for *x* is:

x	10	11	12	13	14
P(x)	.4	.2	.2	.1	.1

Find the probability that in a given hour:

a. exactly 14 patients arrive
b. At least 12 patients arrive
c. At most 11 patients arrive

$p(x=14) = .1$

$p(x \geq 12) = (.2 + .1 + .1) = .4$

$p(x \leq 11) = (.4 + .2) = .6$

Review Question 1

If you toss a die, what's the probability that you roll a 3 or less?

a. 1/6
b. 1/3
c. 1/2
d. 5/6
e. 1.0

Review Question 2

Two dice are rolled and the sum of the face values is six. What is the probability that at least one of the dice came up a 3?

a. 1/5
b. 2/3
c. 1/2
d. 5/6
e. 1.0

Continuous case

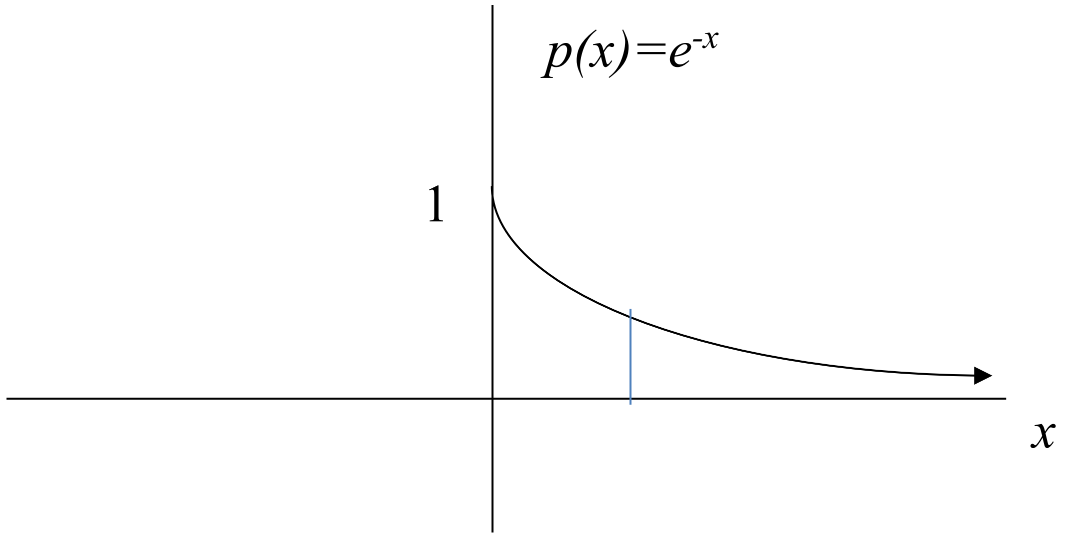
- The probability function that accompanies a continuous random variable is a continuous mathematical function that integrates to 1.
  - For example, recall the negative exponential function (in probability, this is called an "exponential distribution"):

$$f(x) = e^{-x}$$

- This function integrates to 1:

$$\int_0^{+\infty} e^{-x} = -e^{-x} \Big|_0^{+\infty} = 0 + 1 = 1$$

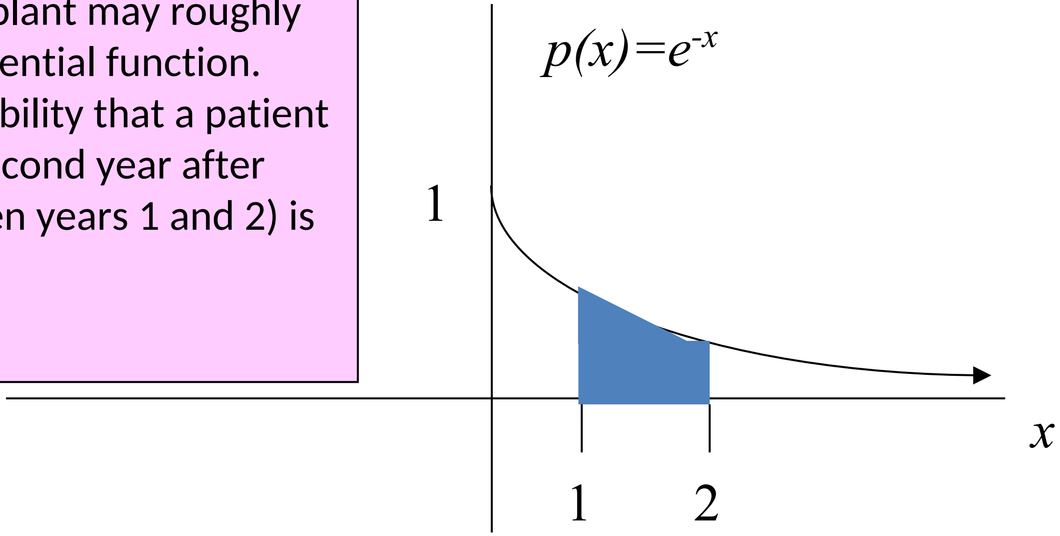
Continuous case: “probability density function” (pdf)



The probability that *x* is any exact particular value (such as 1.9976) is 0; we can only assign probabilities to possible ranges of x.

For example, the probability of *x* falling within 1 to 2:

Clinical example: Survival times after lung transplant may roughly follow an exponential function. Then, the probability that a patient will die in the second year after surgery (between years 1 and 2) is 23%.

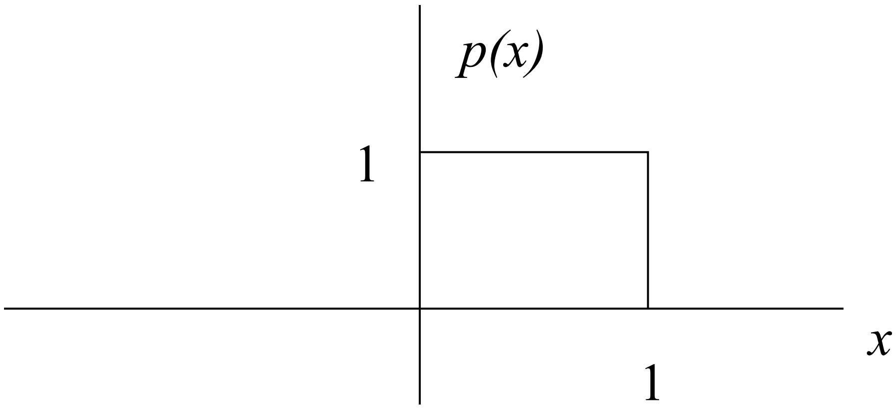


$$P(1 \le x \le 2) = \int_1^2 e^{-x} = -e^{-x} \Big|_1^2 = -e^{-2} - -e^{-1} = -.135 + .368 = .23$$

Example 2: Uniform distribution

The uniform distribution: all values are equally likely.

$f(x) = 1$, for $1 \geq x \geq 0$

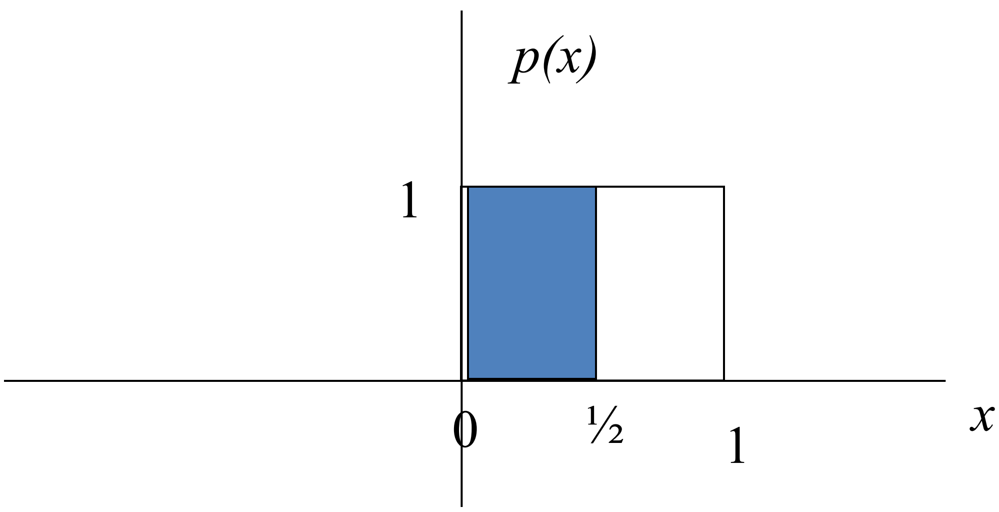


We can see it's a probability distribution because it integrates to 1 (the area under the curve is 1):

$$\int_0^1 1 = x \Big|_0^1 = 1 - 0 = 1$$

Example: Uniform distribution

What's the probability that *x* is between 0 and ½?



Clinical Research Example: When randomizing patients in an RCT, we often use a random number generator on the computer. These programs work by randomly generating a number between 0 and 1 (with equal probability of every number in between). Then a subject who gets X<.5 is control and a subject who gets X>.5 is treatment.

$$P(\tfrac{1}{2} \geq x \geq 0) = \tfrac{1}{2}$$

Expected Value and Variance

All probability distributions are characterized by an expected value (mean) and a variance (standard deviation squared).

Expected value of a random variable

Expected value is just the average or mean (μ) of random variable *x*.

It's sometimes called a "weighted average" because more frequent values of X are weighted more highly in the average.

It's also how we expect X to behave on-average over the long run ("frequentist" view again).

Expected value, formally

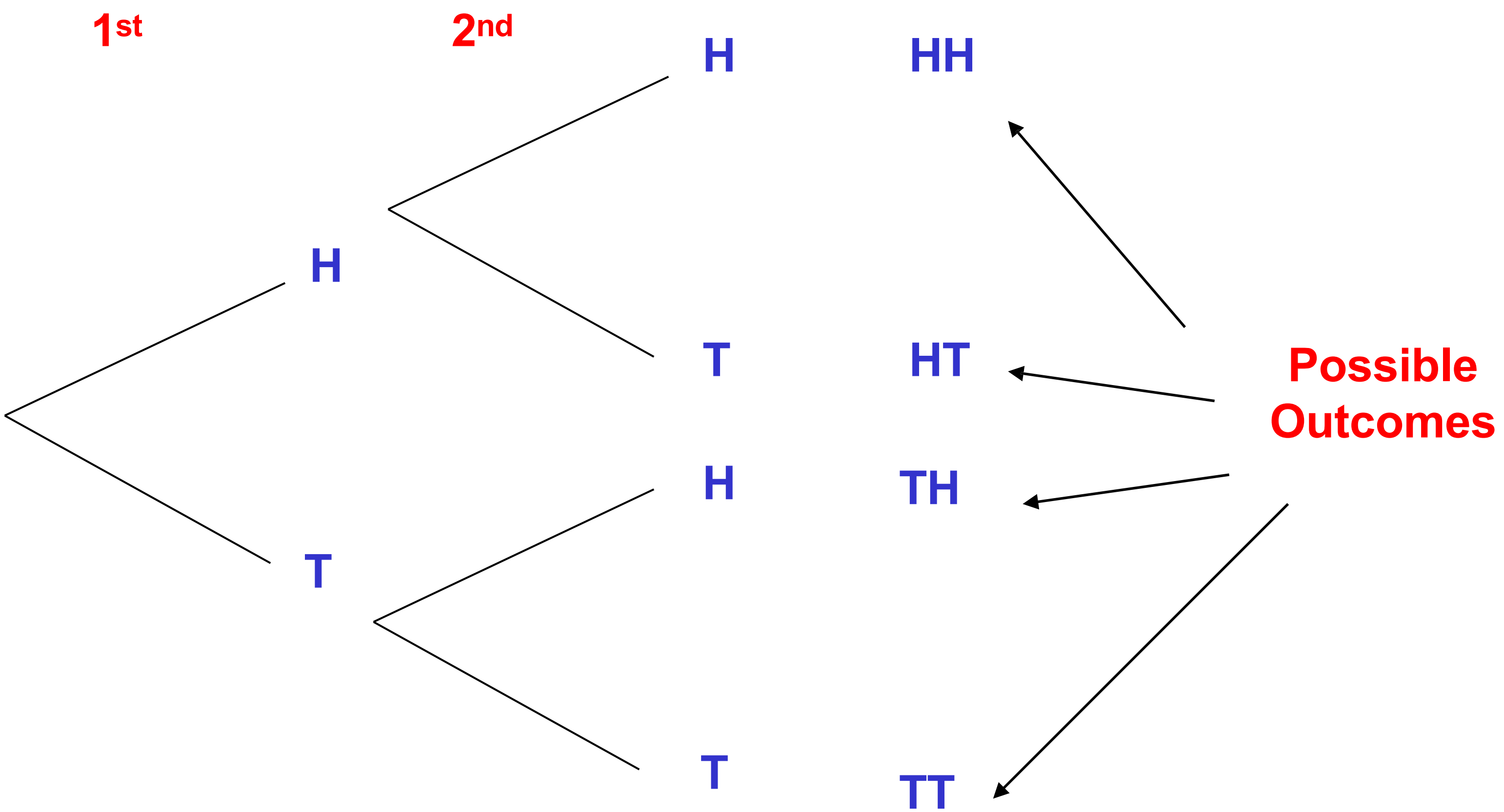
Discrete case:

$$E(X) = \sum_{\text{all } x} x_i \, p(x_i)$$

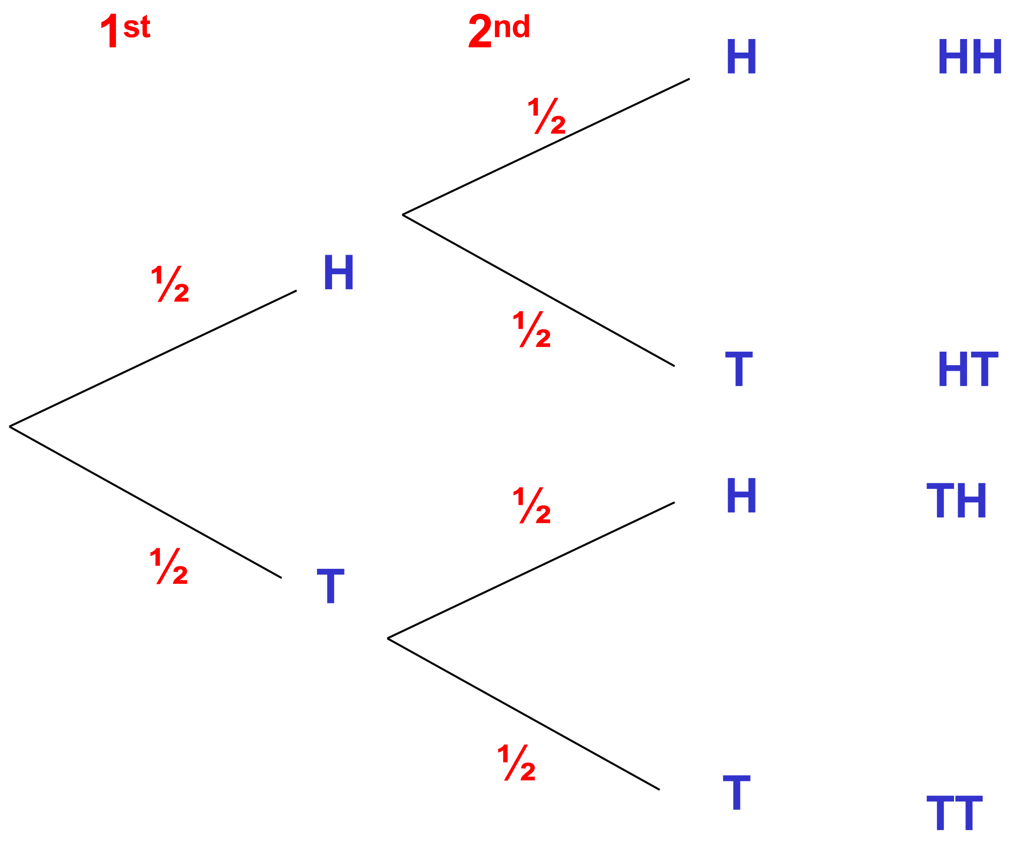
Continuous case:

$$E(X) = \int_{\text{all } x} x_i \, p(x_i) \, dx$$

Tree Diagram –
A fair coin is tossed twice



Attach probabilities



$P(H,H) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

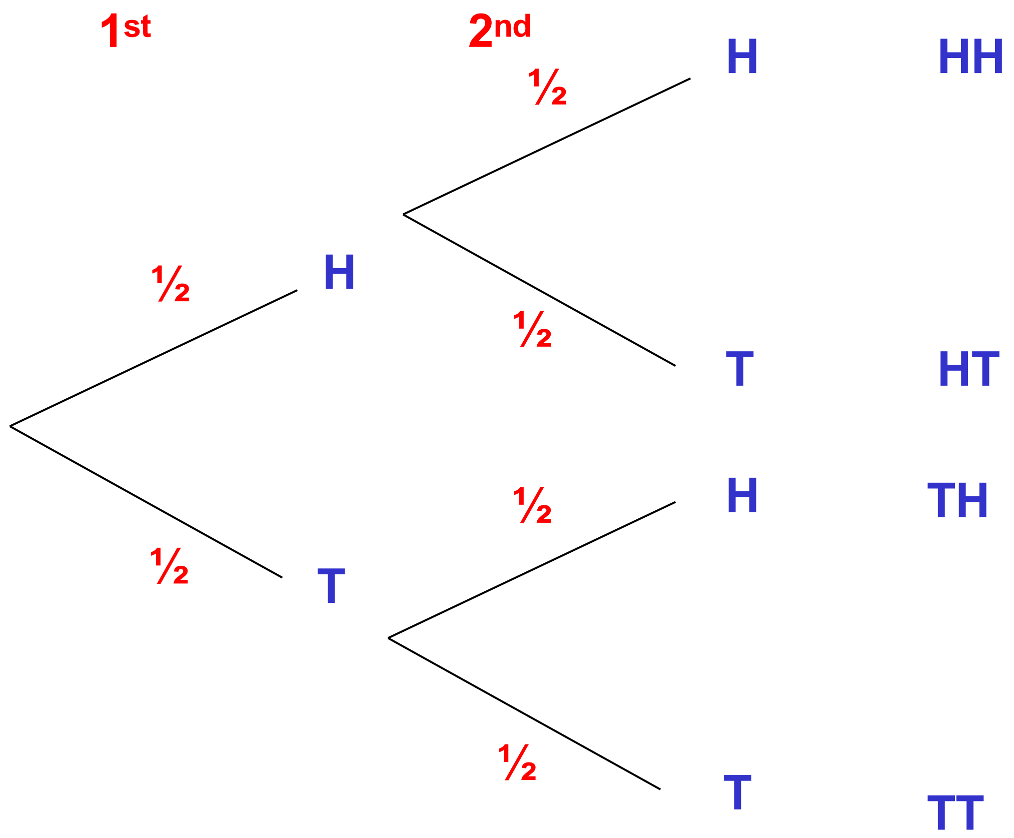
$P(H,T) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$P(T,H) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

$P(T,T) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

INDEPENDENT EVENTS – 1^{st} spin has no effect on the 2^{nd} spin

Calculate probabilities



$P(H,H)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$ *

$P(H,T)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$ *

$P(T,H)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$ *

$P(T,T)=\frac{1}{2}\times\frac{1}{2}=\frac{1}{4}$

Probability of *at least* one Head?

Ans: $\frac{1}{4}+\frac{1}{4}+\frac{1}{4}=\frac{3}{4}$

DISCRETE PD – EXAMPLE (TABLE)

- Tossing a coin three times:
  S = $\{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
  - Let X represents "No. of heads"

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