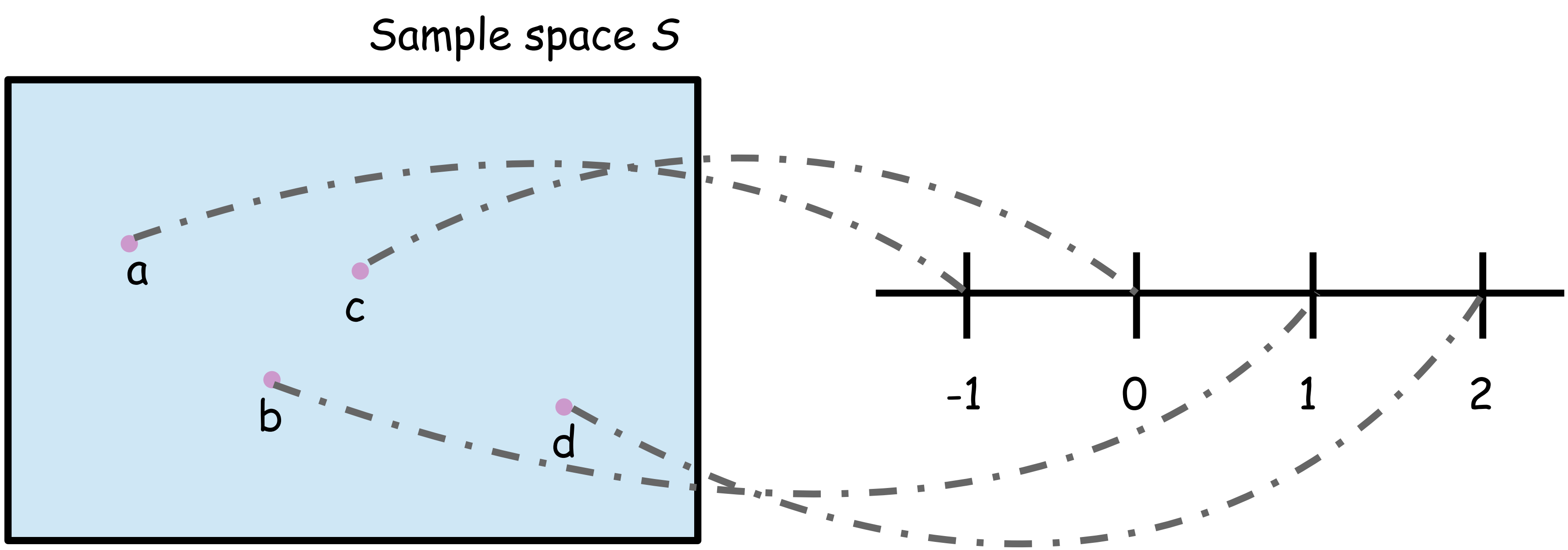
통계분석

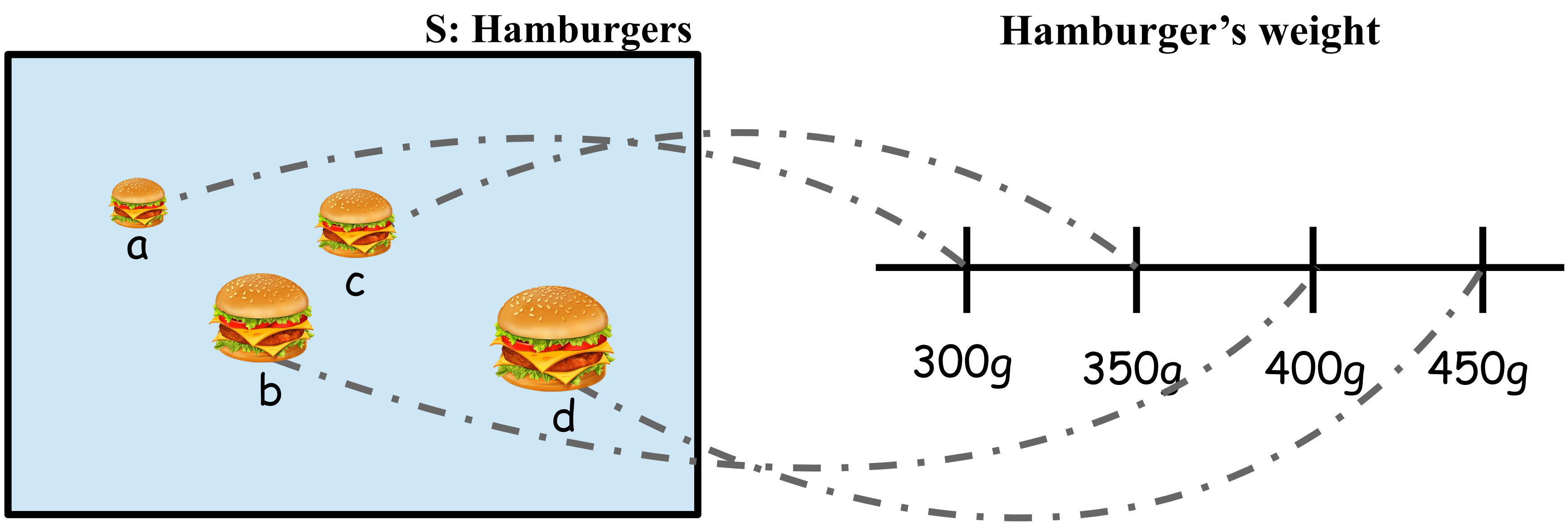
Statistical Analysis

Random Variables

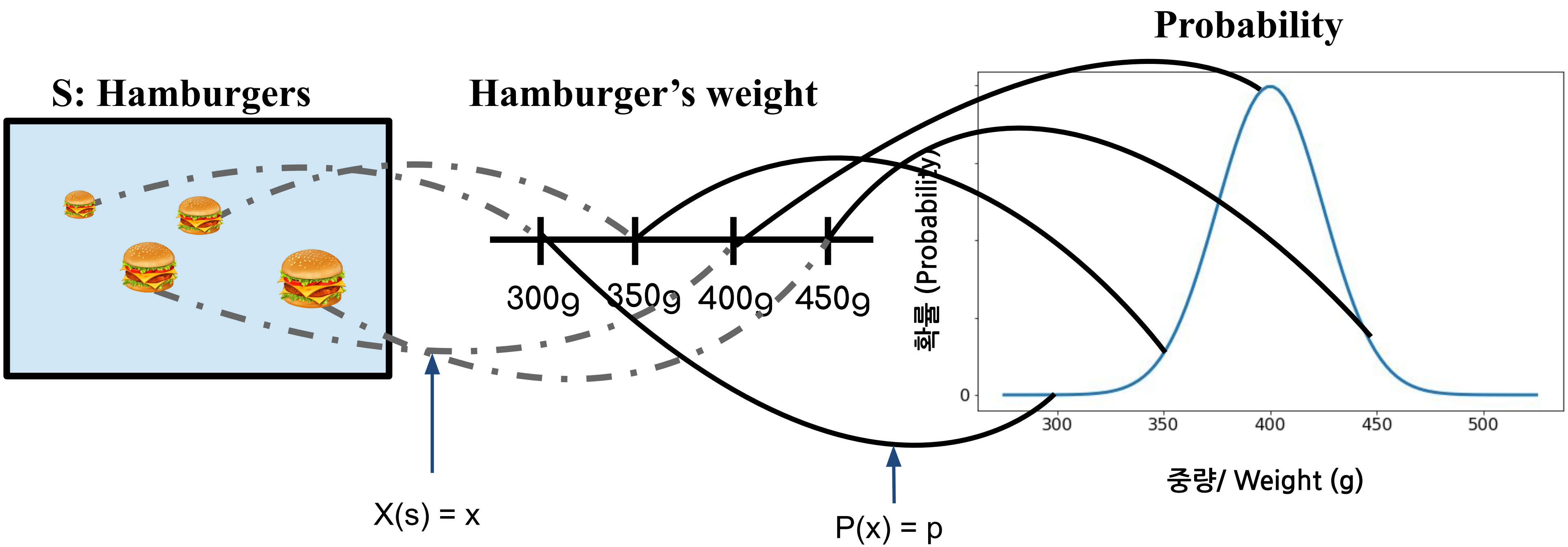


- Random Variable = mapping that relates **each outcome** in S with **a number**
- Mathematically, RV is a **function** from S (domain) to real numbers (range)
- Do not be confused with a probability that we define as a function in the previous lecture.

Random Variables: Example



Random Variables and Probability



- Elements of the sample space S (here hamburgers) are not numbers themselves.

- Recall that we define any feature that elements in the population have as a variable.
- Here our variable is numeric (hamburger's weight).

- Elements (hamburgers) might have some value (weight) with a probability. In a sense that this can happen with a probability, the variable is called a ***random variable***.

Random Variables: Notations

$$\boxed{X(s) = x}$$

- ***X*** = Random Variable (RV)
- ***s*** = the outcome in the Sample Space ***S***
- ***x*** = a number associated with ***s*** by ***X***

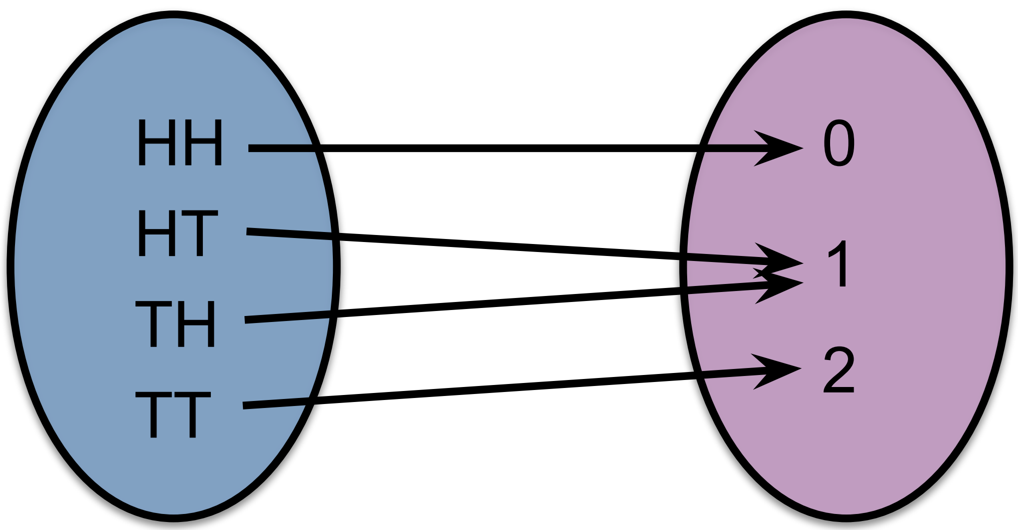
$X($ 🍔 $) = 400\ g$

Random Variables: Example

Tossing a fair coin twice

$$S = \{\mathrm{HH},\ \mathrm{HT},\ \mathrm{TH},\ \mathrm{TT}\}$$

X(s) = the number of tails in the outcome ***s***



Random Variables

$$X(s) = x$$

Discrete Random Variables

Continuous Random Variables

Discrete Random Variables

$$X(s) = x$$

$$x \in \{ \text{ a } \underline{\text{countable set}} \text{ of numbers } \}$$

Discrete Random Variables

$$X(s) = x$$

$$x \in \{ \text{ a } \underline{\text{countable set}} \text{ of numbers } \}$$

1. a finite set: {0,1,2}
2. a countably infinite set: a set of positive integers

cf. an uncountably infinite set : Real numbers

Probability Distribution

Probability Mass Function for Discrete RV

$$p(x) = P(X = x) = P(\text{all } s \in S : X(s) = x)$$

p(x) = the probability for all outcomes ***s*** such that ***X(s) = x***

Upper case : a function
Lower case : a single number

Probability Distribution

Probability Mass Function for Discrete RV

$$p(x) = P(X = x) = P(\text{all } s \in S : X(s) = x)$$

p(x) = the probability for all outcomes ***s*** such that ***X(s) = x***

Tossing a fair coin twice

$$S = \{\text{HH, HT, TH, TT}\}$$

$$P(HH) = P(HT) = P(TH) = P(TT) = 0.25$$

X(s) = the number of tails in the outcome ***s***

Probability Distribution

Probability Mass Function for Discrete RV

$$p(x) = P(X = x) = P(\text{all } s \in S : X(s) = x)$$

p(x) = the probability for all outcomes ***s*** such that ***X(s) = x***

Tossing a coin twice

$$S = \{\text{HH, HT, TH, TT}\}$$

$$P(HH) = P(HT) = P(TH) = P(TT) = 0.25$$

For ***X(s)*** = the number of tails,

$$P(X = 1) = P(HT) + P(TH) = 0.5$$

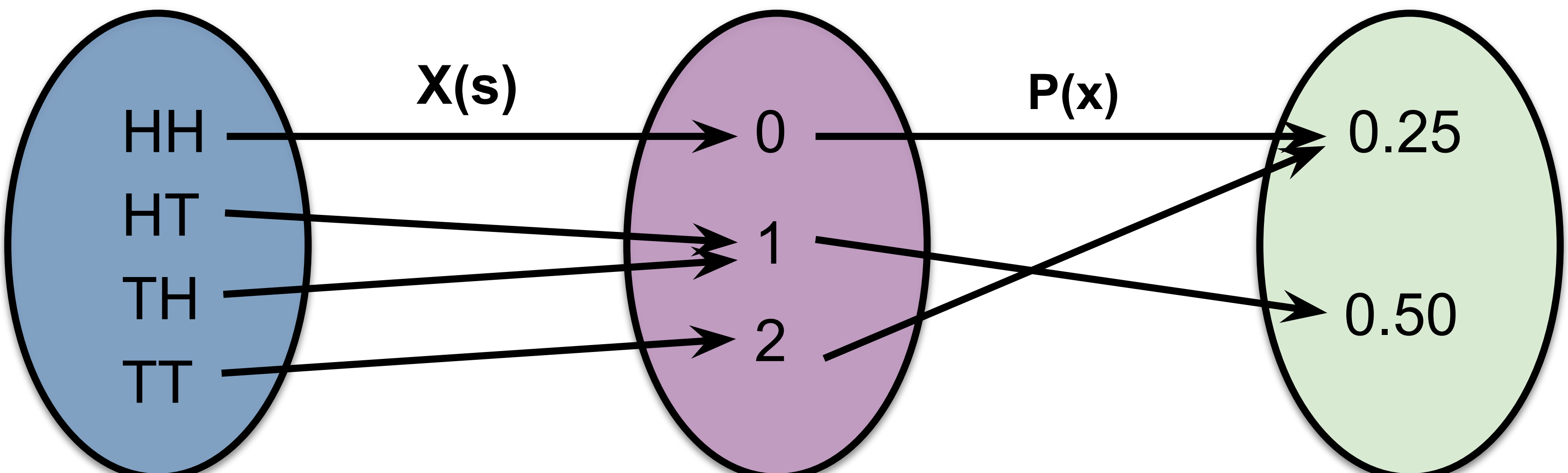
Probability Distribution

Probability Mass Function for Discrete RV

Tossing a coin twice

$$S = \{\text{HH, HT, TH, TT}\}$$

$X(s)$ = the number of tails in the outcome s



S

a finite set: $\{0,1,2\}$

Probability Space, [0,1]

Probability Distribution

Probability Mass Function for Discrete RV

$$0 \leq p(x) \leq 1 \text{ for any x}$$

$$p(x) = 0 \text{ if there is no s such that X(s)} = \text{x}$$

$$\sum_x p(x) = 1$$

Cumulative Distribution Function

$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

F(x) is the probability that the value of RV ***X*** is lesser than or equal to ***x***.

Cumulative Distribution Function

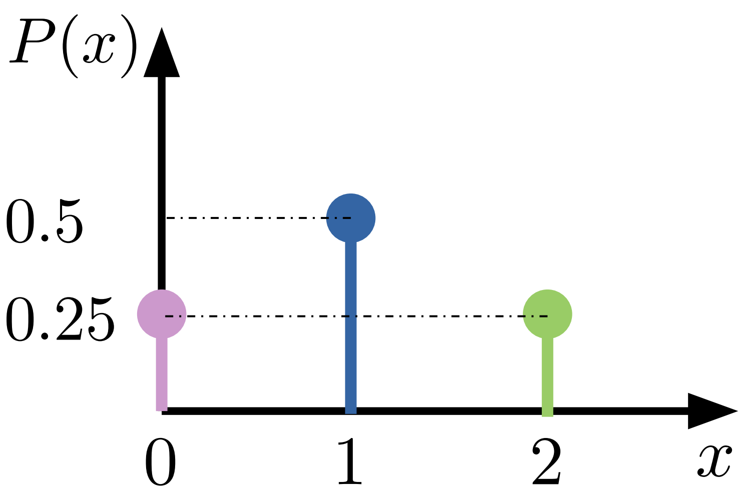
$$F(x) = P(X \leq x) = \sum_{y \leq x} p(y)$$

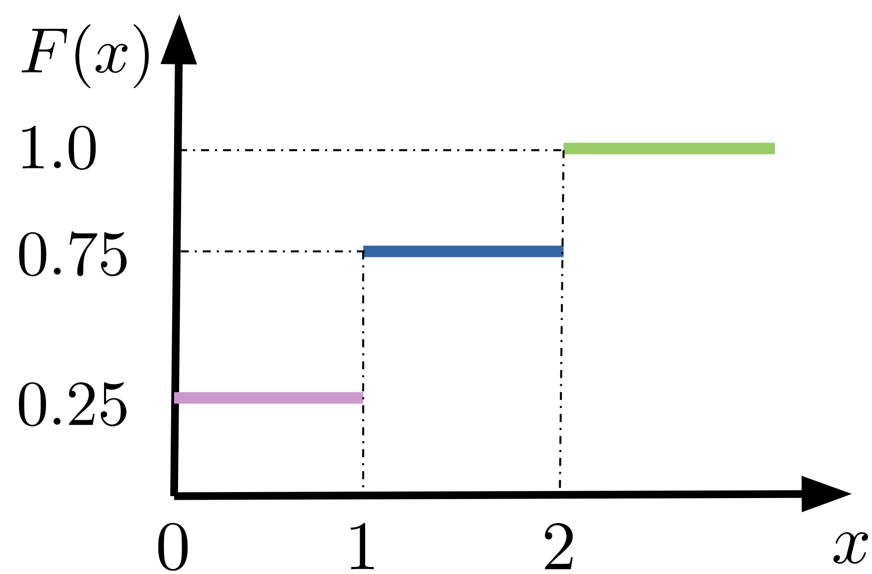
F(x) is the probability that the value of RV ***X*** is lesser than or equal to ***x***.

Tossing a fair coin twice

$$S = \{\text{HH, HT, TH, TT}\}$$

X(s) = the number of tails





Discrete Probability Distributions

- *Bernoulli Distribution*
- *Binomial Distribution*
- *Geometric Distribution*
- *Poisson Distribution*

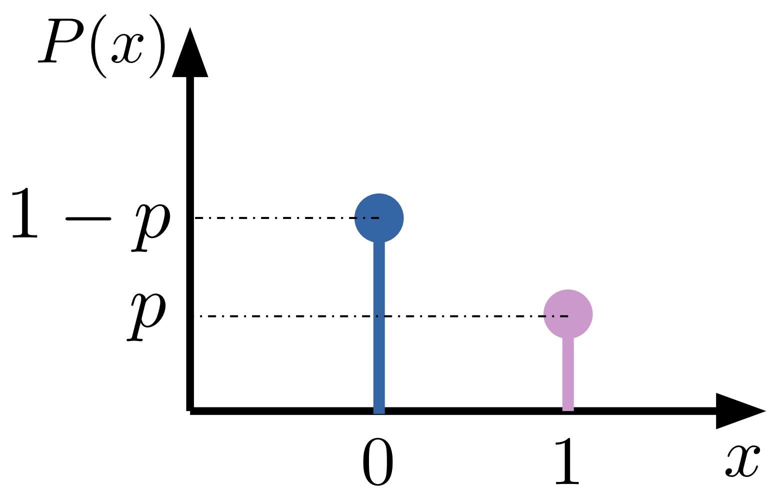
Bernoulli Distribution

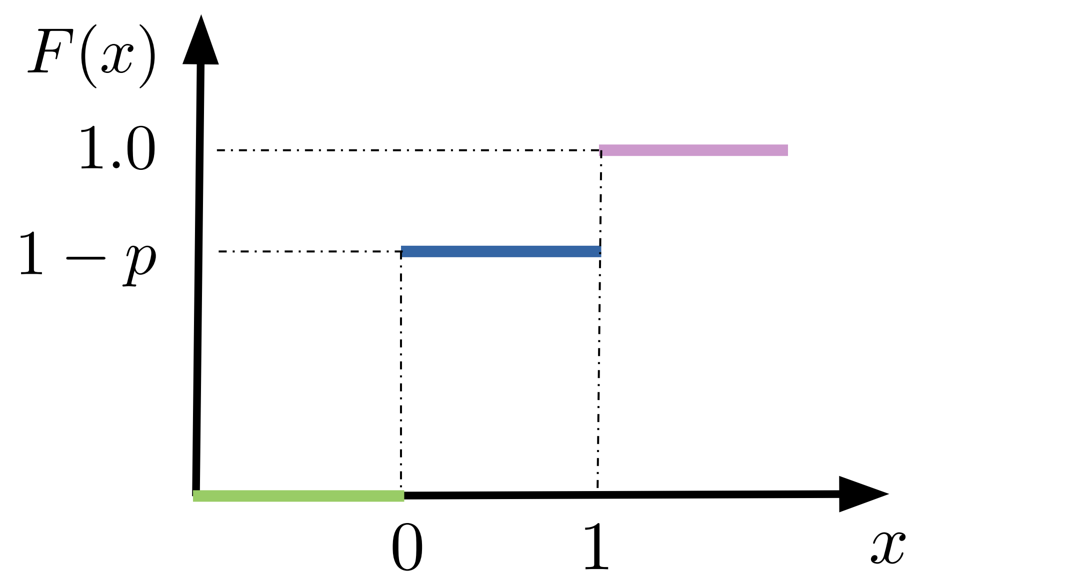
Probability distribution for one trial experiment resulting in two outcomes, ***success*** *(head)* or ***failure (tail)***

Tossing an unfair coin

Head (success) $P(X=1)=p$

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





Parameters

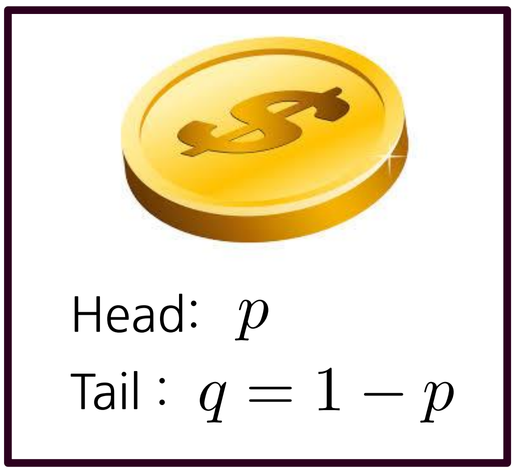
$$P(X) = P(X, \underbrace{p}_{\textbf{Parameter}})$$

Suppose the probability P(x) depends on a quantity that can be assigned any one of a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a family of probability distributions.

Binomial Distribution

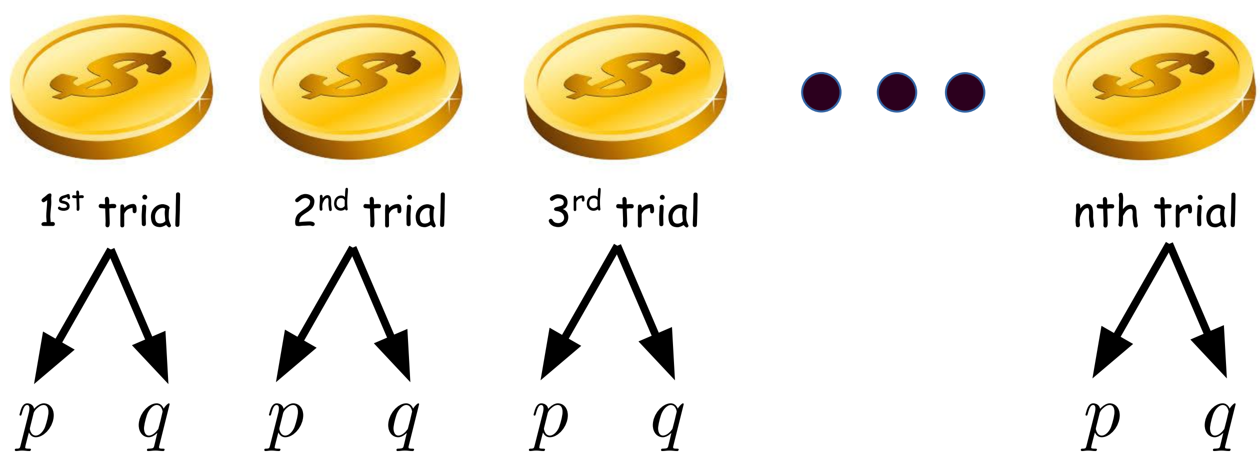
Probability distribution for experiment of **n successive trials**, each of which results in two outcomes, ***success*** or ***failure (Bernoulli)***

- Bernoulli distribution is the special case of Binomial distribution (n=1).
- Binomial distribution with n trials consists of a sequence of n Bernoulli trials.



Bernoulli trial

Tossing an unfair coin n times



Binomial Distribution

Binomial distribution for n trials: $b(x; n, p)$

parameter

variable
(observable)

x : random variable = the number of successes
n : the number of trials
p : probability of success for Bernoulli trial

$$b(x; n, p) = \underbrace{\left(\begin{array}{c} n \\ x \end{array} \right)} \underbrace{p^x (1-p)^{(n-x)}}$$

Probability for any case where x trials succeed
Probability for any case where x heads are observed

$$\left(\begin{array}{c} n \\ x \end{array} \right) = \frac{n!}{x!(n-x)!}$$: # of cases where x trials succeed among n total trials

n flips, observed x heads

Binomial Distribution: Example

$n = 4, x = 2$

1^{st} trial 2^{nd} trial 3^{rd} trial 4^{th} trial

Binomial Distribution: Example

$n = 4, x = 2$

1st trial	2nd trial	3rd trial	4th trial
H	**H**	**T**	**T**

Binomial Distribution: Example

$n = 4, x = 2$

1st trial | 2nd trial | 3rd trial | 4th trial

H **H** **T** **T**

Probability : $p \times p \times (1-p) \times (1-p) = p^2(1-p)^{4-2}$

Binomial Distribution: Example

$n = 4, x = 2$

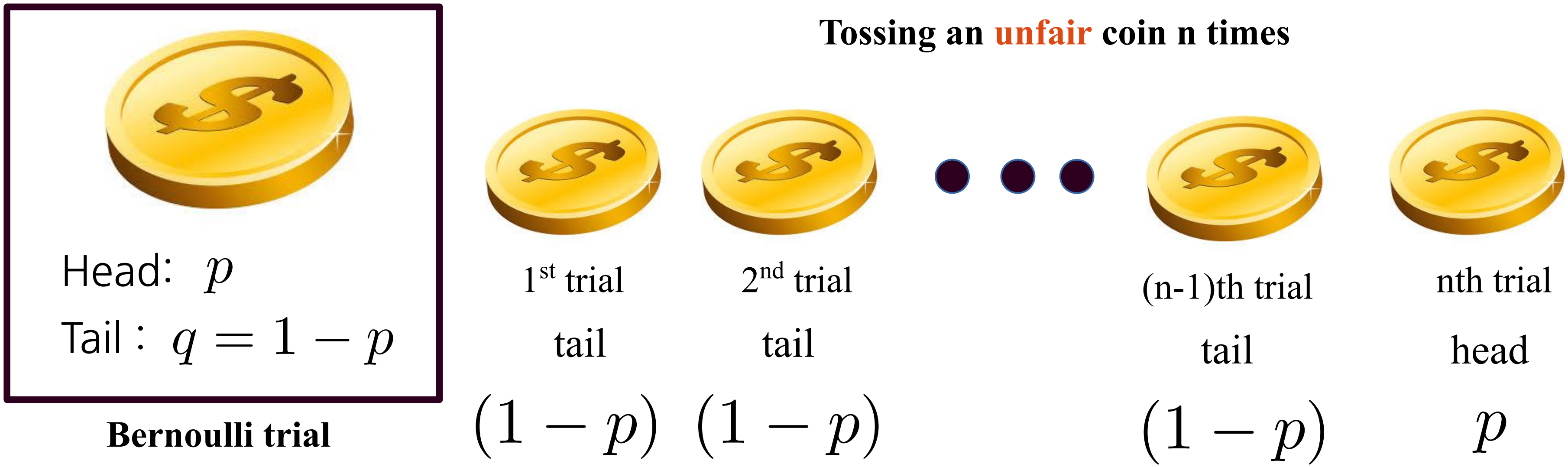
1st trial	2nd trial	3rd trial	4th trial
H	**H**	**T**	**T**
H	**T**	**H**	**T**
H	**T**	**T**	**H**
T	**H**	**H**	**T**
T	**H**	**T**	**H**
T	**T**	**H**	**H**

$$\binom{4}{2} = \frac{4!}{2!2!} = 6$$

6 p^2 (1-p)^2

Geometric Distribution

Let us consider a coin-tossing game. What is the probability distribution of observing a head at the first time at the *n*th tossing? This discrete distribution is called the **geometric distribution**.



Geometric Distribution : $P(x = n, p) = (1-p)^{n-1} p$