UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

DIPARTIMENTO DI MEDICINA E CHIURGIA

Random variable and Probability distributions

Random variable

A **random variable** is a variable (typically represented by X) that has a single numerical value, determined by chance, for each outcome of a procedure.

- **Discrete** (e.g. number of girls in two births):either a finite number of values or countable number of values, where "countable" refers to the fact that there might be infinitely many values, but they result from a counting process

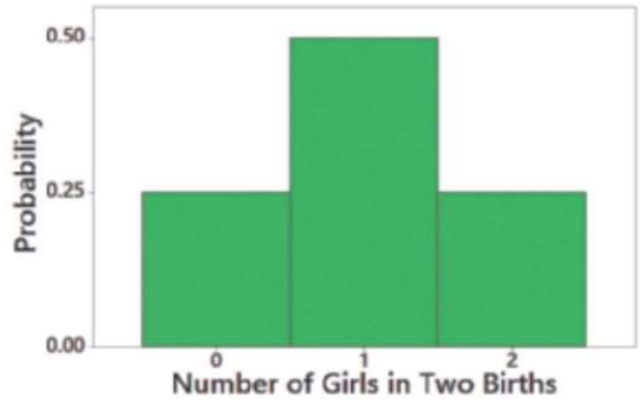
- **Continuous** (e.g. body temperature): infinitely many values, and those values can be associated with measurements on a continuous scale without gaps or interruptions

Probability distribution

is a description that gives the probability for each value of the random variable. It is often expressed in the format of a table, formula, or graph.

TABLE 5-1 Probability Distribution for the Number of Girls in Two Births

x: Number of Girls	$P(x)$
0	0.25
1	0.50
2	0.25



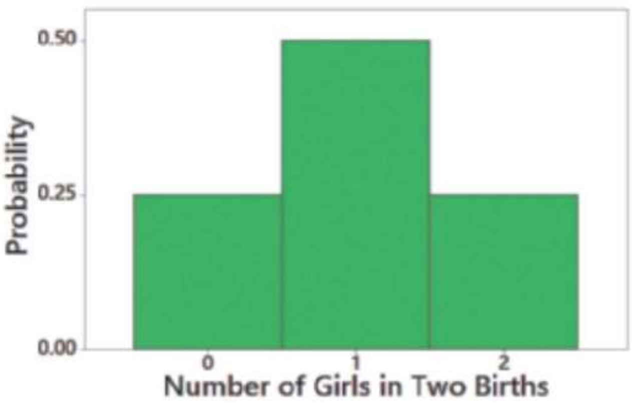
Probability distribution requirements

Every probability distribution must satisfy each of the following three requirements:

1. There is a numerical (not categorical) random variable X, and its number values are associated with corresponding probabilities.
2. ΣP(x)=1 where x assumes all possible values (The sum of all probabilities must be 1).
3. 0 ≤P(x) ≤ 1 for every individual value of the random variable x. (That is, each probability value must be between 0 and 1 inclusive.)

TABLE 5-1 Probability Distribution for the Number of Girls in Two Births

x: Number of Girls	*P*(*x*)
0	0.25
1	0.50
2	0.25



Parameters of a probability distribution

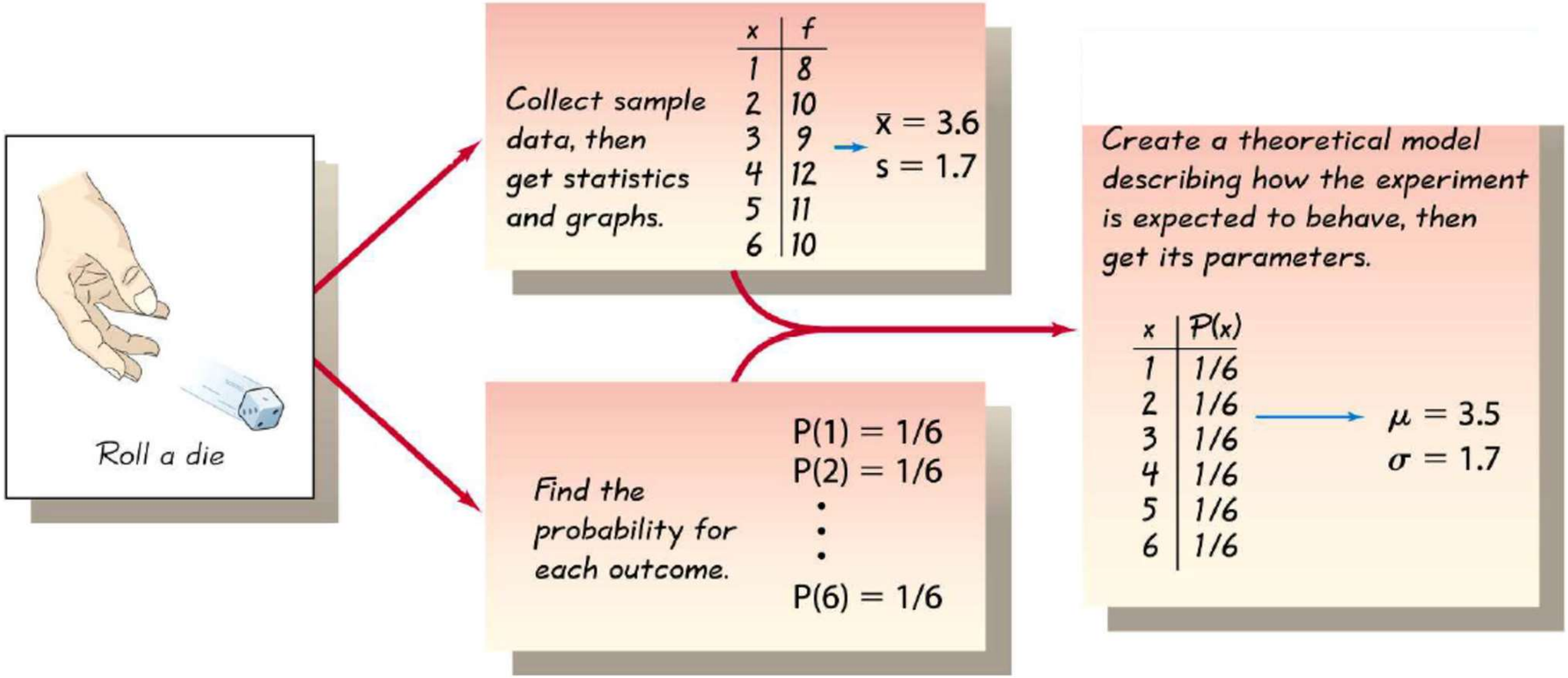
The expected value of a discrete random variable X is denoted by E, and it is the mean value of the outcomes, so E = μ

$$E(X) = \mu = \sum (x * P(x))$$

Variance (indicates how spread is the distribution):

$$\sigma^2 = \sum \left[(x - \mu)^2 P(x)\right]$$

Combining Descriptive Methods and Probabilities



DIPARTIMENTO DI MEDICINA E CHIURGIA

Discrete Probability distributions:
The binomial distribution

A discrete random variable: Binomial distribution

Binomial probability distributions allow us to deal with circumstances in which the outcomes belong to two categories, such as cured/not cured or acceptable/defective or survived/died.

A binomial probability distribution results from a procedure that meets these four requirements:

1. The procedure has a fixed number of trials. (A trial is a single observation.)
2. The trials must be independent, meaning that the outcome of any individual trial doesn’t affect the probabilities in the other trials.
3. Each trial must have all outcomes classified into exactly two categories, commonly referred to as success and failure.
4. The probability of a success remains the same in all trials.

Example

In a certain population (eg Italian) the probability of being born Male is ≈ 0.51 and it is constant (in recent years).

Which is the probability that on 3 children 2 are males?

Example

In a certain population (eg Italian) the probability of being born Male is ≈ 0.51 and it is constant (in recent years).

Which is the probability that on 3 children 2 are males?

M M F

$$P(M \cap M \cap F) = 0.51 * 0.51 * 0.49 = 0.127$$

$P(M \cap M \cap F)$ does not give the probability of getting exactly two boys out of three birth because it assumes a particular arrangement /order. Other arrangements are possible:

M F M

F M M

Thus the probability to get MMF or MFM or FMM is the sum of the probability of each sequence:

$$P([M \cap M \cap F] \cup [M \cap F \cap M] \cup [F \cap M \cap M]) =$$
$$= P(M \cap M \cap F) + P(M \cap F \cap M) + P(F \cap M \cap M) =$$
$$= 3[(0.51)^2 0.49] = 0.38$$

The Binomial Probability Formula

$X \sim Bi(\pi, n)$

$$P(x|n) = \frac{n!}{x! \cdot (n-x)!} \cdot \pi^x \cdot (1-\pi)^{n-x} =$$

$$= \binom{n}{x} \cdot \pi^x \cdot (1-\pi)^{n-x}$$ for x = 0,1,2,…,n

where:

n = number of trials

x = number of successes among n trials

π = probability of success in any one trial

1-π = probability of failure in any one trial

The factorial symbol ! denotes the product of decreasing factors. Two examples of factorials are 3! = 3 * 2 * 1 = 6 and 0! = 1 (by definition).

The Binomial Probability Formula

X~Bi(π,n)

$$P(X=x)=\underbrace{\frac{n!}{x!\,(n-x)!}}_{\text{Number of outcomes with exactly } x \text{ successes among } n \text{ trials}}\ \underbrace{\pi^{x}(1-\pi)^{n-x}}_{\text{The probability of } x \text{ successes among } n \text{ trials for any one particular order}}$$

where:

n = number of trials

x = number of successes among n trials

π = probability of success in any one trial

Example - method 2 (using Binomial)

In a certain population (eg Italian) the probability of being born Male is ≈ 0.51 and it is constant (in recent years).
What is the probability that on 3 children 2 are males?

$$P(X = x) = \boxed{\frac{n!}{x!\,(n-x)!}}\pi^x(1-\pi)^{n-x}$$

Binomial coefficient: denotes the number of outcomes with exactly x successes among n trials

The probability of x successes among n trials for any one particular order

x=2
n=3
π=0.51

$$P(X = 2) = \frac{3!}{2!\,(3-2)!}0.51^2 0.49 = 3 * 0.51^2 0.49 = 0.38$$

Exercise

When Gregor Mendel conducted his famous hybridization experiments, he used peas with green pods and peas with yellow pods. Because green is dominant and yellow is recessive, when crossing two parents with the green/yellow pair of genes, we expect that 3/4 of the offspring peas should have green pods. That is, P(green pod) = 3/4. Assume that all parents have the green/yellow combination of genes.

1. Find the probability that 5 out of ten offspring peas have green pods.
2. Find the probability that 9 out of ten offspring peas have green pods.
3. Find the probability that all of the ten offspring peas have green pods.

$$P(X = x) = \frac{n!}{x!\,(n-x)!}\pi^x(1-\pi)^{n-x}$$

Exercise – point 1

$$P(X=5)=\frac{10!}{5!\,(10-5)!}0.75^5(1-0.75)^{10-5}=$$

$$=\frac{10*9*8*7*6*\cancel{5!}}{\cancel{5!}\,5*4*3*2*1}0.75^5(0.25)^5=$$

$=252\ 0.75^5(0.25)^5$

$=0.0584$

Exercise - point 2

$$P(X = 9) = \frac{10!}{9!\,(10-9)!}\,0.75^9(1-0.75)^{10-9} =$$

$$= \frac{10 * \cancel{9!}}{\cancel{9!}(1)!}\,0.75^9(0.25)^1 = 0.1877$$

Exercise – point 3

$$P(X = 10) = \frac{\cancel{10!}}{\cancel{10!}\,(10-10)!} 0.75^{10}(1-0.75)^{10-10} =$$
$$= 1 * 0.75^{10} = 0.0563$$

Exercise

By using the distribution probability below:

1. Compute the probability to have at least 4 green pods.
2. Compute the probability to have not more then 7 green pods.

x	f(x)	F(x)
0	0.0000	0.0000
1	0.0000	0.0000
2	0.0004	0.0004
3	0.0031	0.0035
4	0.0162	0.0197
5	0.0584	0.0781
6	0.1460	0.2241
7	0.2503	0.4744
8	0.2816	0.7560
9	0.1877	0.9437
10	0.0563	1.0000

Exercise

By using the distribution probability below:

1. Compute the probability to have at least 4 green pods.
2. Compute the probability to have not more then 7 green pods.

x	f(x)	F(x)
0	0.0000	0.0000
1	0.0000	0.0000
2	0.0004	0.0004
3	0.0031	0.0035
4	0.0162	0.0197
5	0.0584	0.0781
6	0.1460	0.2241
7	0.2503	0.4744
8	0.2816	0.7560
9	0.1877	0.9437
10	0.0563	1.0000

1. $P(X \geq 4) = 1 - 0.0035 = 0.9965$

2. $P(X \leq 7) = 0.4744$

Binomial distribution:

special formula for expected value and variance

$X \sim Bi(\pi=0.75, n=10)$

- How many green pods are expected?

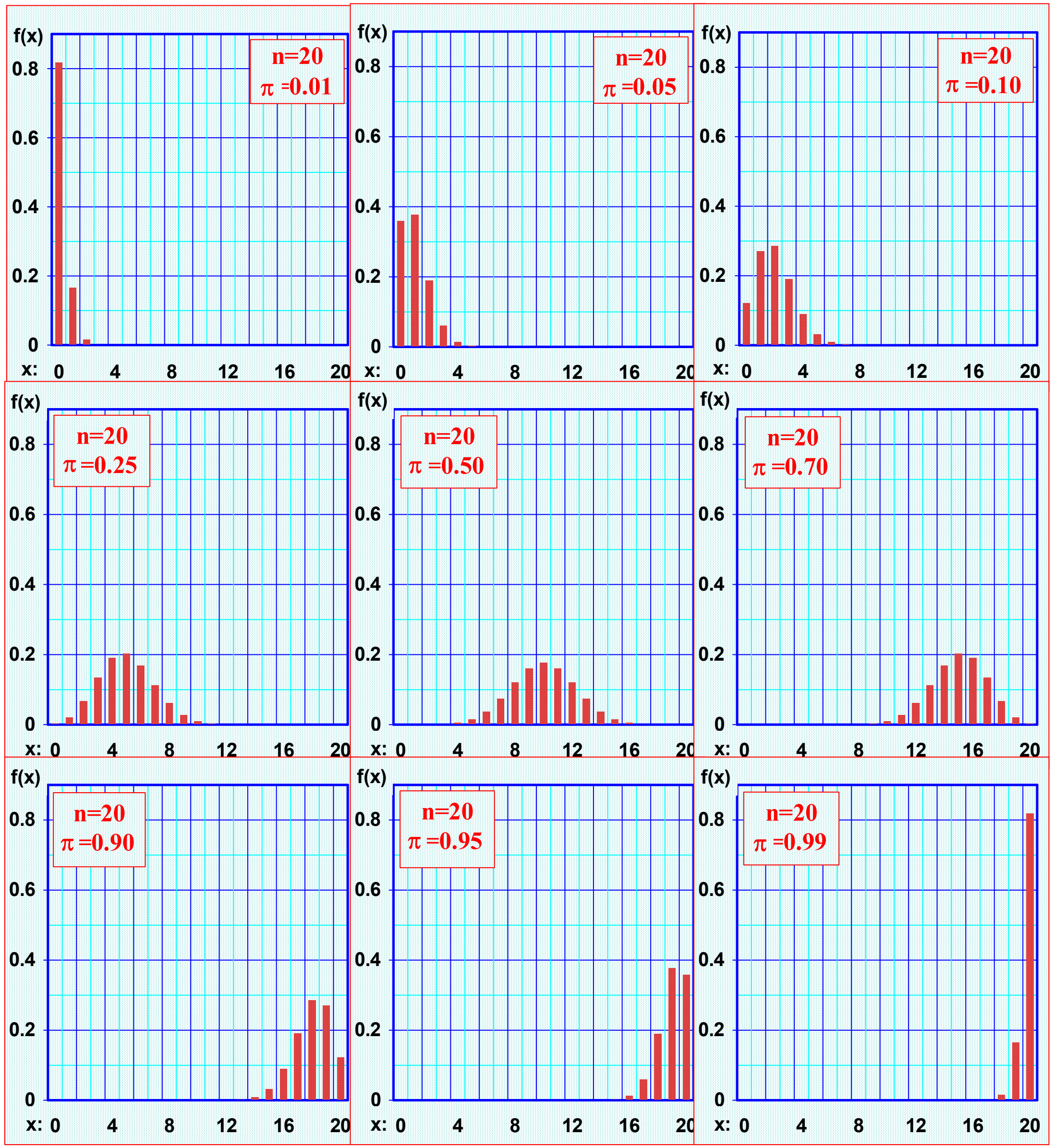
$$E(X) = n*\pi = 10*0.75 = 7.5$$

- Which is the variability of green pods?

$$VAR(X) = n*\pi*(1-\pi) = 10*0.75*0.25 = 1.875$$

Shape of the binomial function

Binomial random variables of size n with varying π (π=0.01, 0.05, … , 0.99).



The distribution is symmetrical:

π=0.50

Positive asymmetric

π→0

Negativ asymmetric

π→1

Exercise

A pharmaceutical product causes a serious side effect in 3 out of 100 patients. A pharmaceutical company wants to test the medicine.

1) What is the probability that the side effect will occur in at most 1 subject in a random sample of 20 patients taking the medicine?

2) What is the probability that all 20 subjects will experience the side effect?

3) How many patients with the side effect do I expect out of the 20 subjects?

1) P(X≤1)=P(X=1)+P(X=0)

$$P(X=0)=\binom{20}{0}0{,}03^{0}0{,}97^{20}=0{,}5438$$

$$P(X=1)=\binom{20}{1}0{,}03^{1}0{,}97^{19}=\frac{20*19!}{19!\,1!}0{,}03^{1}0{,}97^{19}=20*0{,}03^{1}0{,}97^{19}$$
$$=0{,}3364$$

$$P(X\leq 1)=P(X=0)+P(X=1)=0{,}5438+0{,}3364=0{,}8802$$

2) $P(X=20)=\binom{20}{20}0{,}03^{20}0{,}97^{0}=0{,}03^{20}=3.486784*10^{-31}\approx 0$

3) E(X)=n*π=20*0,03=0,6