PROBABILITY DISTRIBUTIONS

Sorana D. Bolboacă

OBJECTIVES

- Random variables
- Types of probability distributions (density function)
- Discrete probability distributions
- Continuous probability distributions

RANDOM VARIABLE

- A random variable
  - is a quantification of a probability model that allows to model random data
  - Is a quantity that may take any value of a given range that cannot be predicted exactly but can be described in terms of their probability
    - Discrete: generally assessed by counting
    - Continuous: generally assessed by measurements

RANDOM VARIABLE

- When throwing a die with 6 faces, let X be the random variable defined by:

  X = the square of the scores shown on the die

What is the expectation of X?

Solution:

- S = {1, 2^2, 3^2, 4^2, 5^2, 6^2} = {1, 4, 9, 16, 25, 36}
- Each face has a probability of 1/6 of occurring, so:

E(X) = 1*1/6+4*1/6+6*1/6+16*1/6+25*1/6+36*1/6 = 91*1/6

Random Variables

- Outcome of an experiment → a quantitative or a qualitative data
- A random variable is a function that assign a unique numerical value to each outcome of an experiment along with an associated probability.
  - Discrete random variable corresponds to an experiment that has a finite or countable number of outcomes.
  - A continuous random variable is a random variable for which the set of possible outcomes is continuous

Probability Distribution

Discrete

- The probabilities associated with each specific value

Continuous

- The probabilities associated with a range of values

DISCRETE PROBABILITY DISTRIBUTIONS

Event space

- Suppose that we toss 3 coins.
- Let X be the number of "heads" appearing
- X is a random variable taking one of the following values {0,1,2,3}

Event space

- Let us suppose that we have an urn with black and white balls. We win $1 for every white and lose $1 for every black. Let X = total winnings.
- X is a random variable that can take one of the following values {-2,0,2}

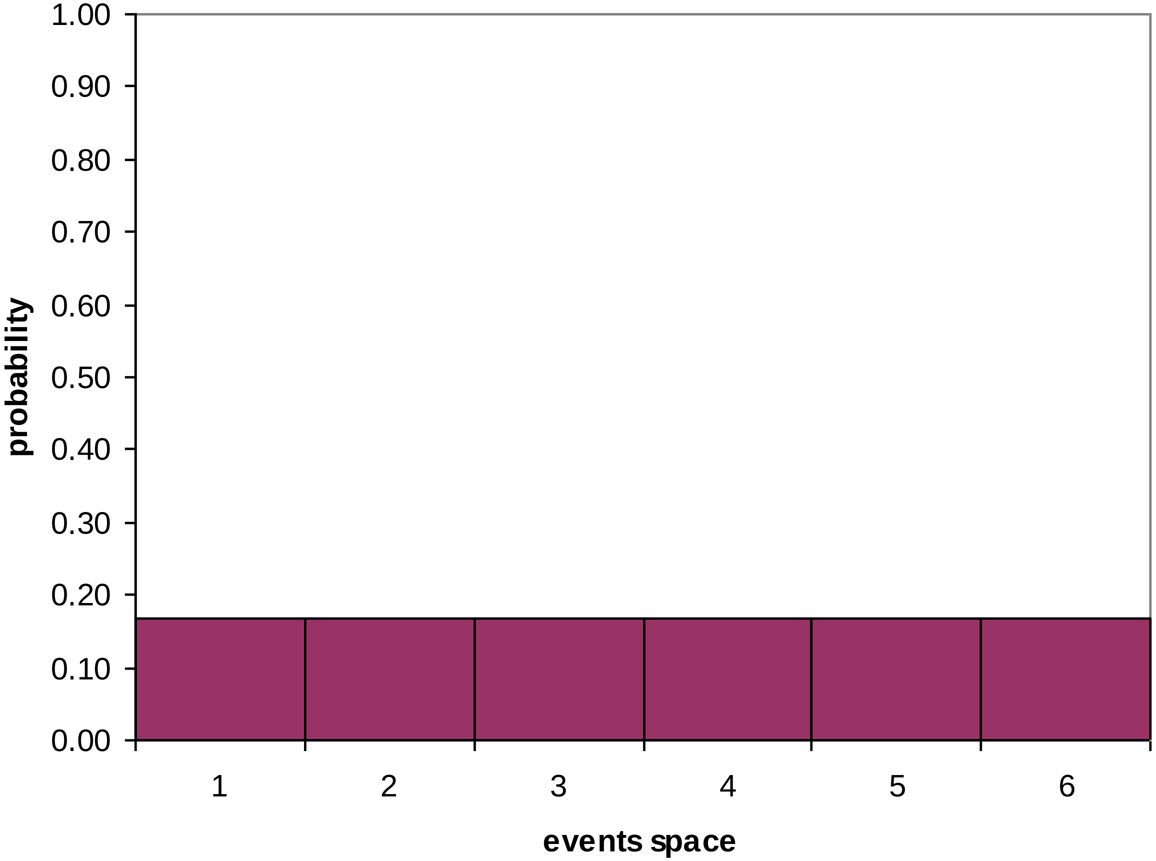
Discrete Probability Distributions

- The probability of X distribution: list of values from the events space and associated probabilities
  - Let X be the outcome of tossing a die
  - X is a random variable that can take one of the following values {1, 2, 3, 4, 5, 6}

X_i	$Pr(X_i)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Discrete Probability Distributions

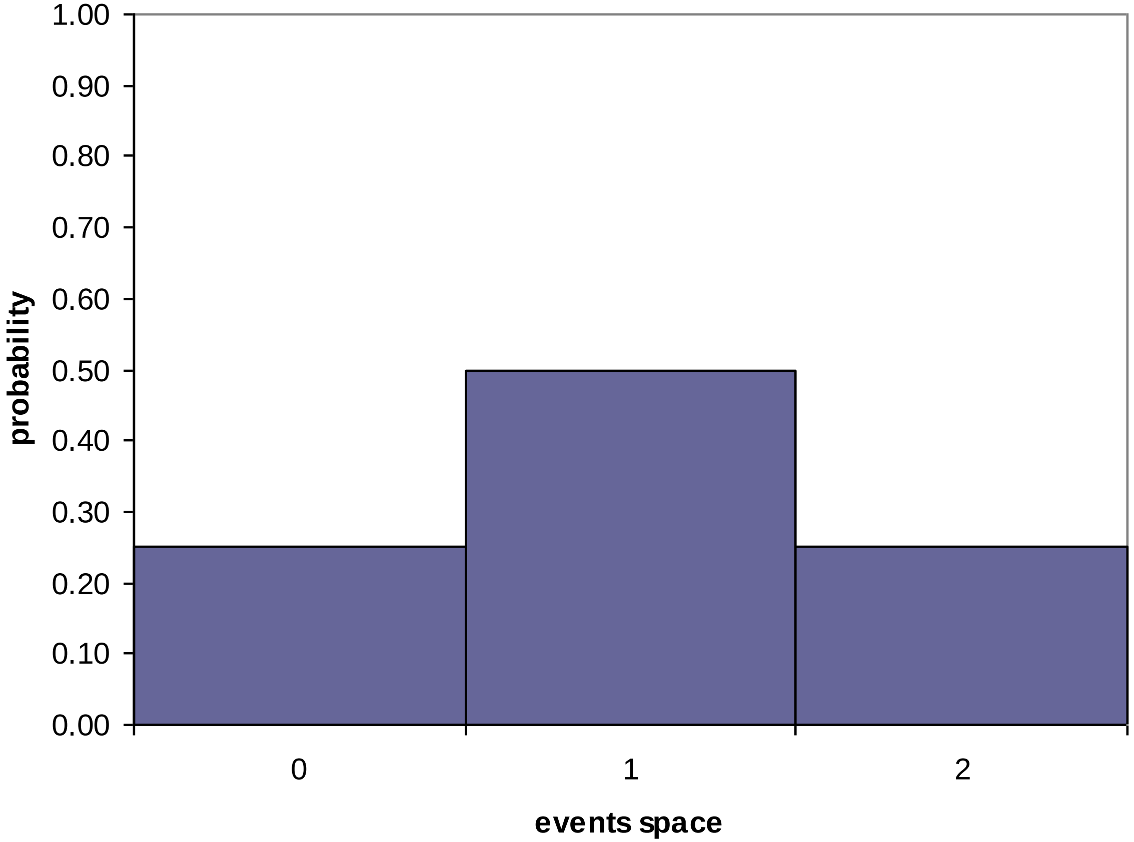
- The probability of X lists the values in the events space and their associated probabilities



X_i	Pr_i
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

DISCRETE PROBABILITY DISTRIBUTIONS

- Let X be the number of “head” results by throwing twice two coins. What is the probability distribution?



X_i	Pr_i
0	1/4
1	2/4
2	1/4

Discrete Probability Distributions

- Probability Distribution: symbols

$$X:\begin{pmatrix} X_1 & X_2 & \dots & X_n \\ Pr(x_1) & Pr(x_2) & \dots & Pr(x_n) \end{pmatrix}$$

- **Property:** the probabilities that appear in distribution of a finite random variable verify the formula:

$$\sum_{i=1}^{n} Pr(X_i) = 1$$

DISCRETE PROBABILITY DISTRIBUTIONS

- The mean of discrete probability distribution (called also expected value) is give by the formula:

$$M(X) = \sum_{i=1}^{n} X_i \cdot Pr(X_i)$$

- Represents the weighted average of possible values, each value being weighted by its probability of occurrence.

DISCRETE PROBABILITY DISTRIBUTIONS

Example:

- Let X be a random variable represented by the number of otitis episodes in the first two years of life in a community. This random variable has the following distribution:

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- What is the expected number (average) of episodes of otitis during the first two years of life?

Discrete Probability Distributions

- What is the expected number (average) of episodes of otitis during the first two years of life?

$$X:\begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 \\ 0.129 & 0.264 & 0.271 & 0.185 & 0.095 & 0.039 & 0.017 \end{pmatrix}$$

- $M(X) = 0 \cdot 0.129 + 1 \cdot 0.264 + 2 \cdot 0.271 + 3 \cdot 0.185 + 4 \cdot 0.095 + 5 \cdot 0.039 + 6 \cdot 0.017$
- $M(X) = 0 + 0.264 + 0.542 + 0.555 + 0.38 + 0.195 + 0.102$
- $M(X) = 2.038$

Discrete Probability Distributions

- Variance: is a weighted average of the squared deviations in X

$$V(X) = \sum_{i=1}^{n} (X_i - M(X))^2 \cdot Pr(X_i)$$

- Standard deviation:

$$\sigma(X) = \sqrt{V(X)} = \sqrt{\sum_{i=1}^{n} (X_i - M(X))^2 \cdot Pr(X_i)}$$

Discrete Probability Distributions

X_i	$Pr(X_i)$	$X_i*Pr(X_i)$	$X_i-M(X)$	$(X_i-M(X))^2$	$(X_i-M(X))^2*Pr(X_i)$
0	0.129	0	-2.038	4.153	0.536
1	0.264	0.264	-1.038	1.077	0.284
2	0.271	0.542	-0.038	0.001	0.000
3	0.185	0.555	0.962	0.925	0.171
4	0.095	0.38	1.962	3.849	0.366
5	0.039	0.195	2.962	8.773	0.342
6	0.017	0.102	3.962	15.697	0.267
		M(X)=**2.038**			V(X)=**1.967**
					σ(X)=**1.402**

Discrete Probability Distributions By Examples

- **Bernoulli:** head versus tail (two possible outcomes)
- **Binomial:** number of 'head' obtained by throwing a coin of n times
- **Poisson:** number of patients consulted in a emergency office in one day

BERNOULLI DISTRIBUTION

- A random variable whose response are dichotomous is called a Bernoulli random variable
- The experiment has just two possible outcomes (success and failure – dichotomial variable):
  - Gender: boy or girl
  - Results of a test: positive or negative
- Probability of success = p
- Probability of failure = 1-p

X	1	0
Pr(X=x)	p	1-p

BERNOULLI DISTRIBUTION

X	1	0
Pr(X=x)	p	1-p

- Mean of X:

$$M(X) = 1 \cdot p + 0 \cdot (1-p)$$

- Variance of X:

$$V(X) = p \cdot (1-p)$$

Binomial Distribution

- An experiment is given by repeating a test of n times (n = known natural number).
- The possible outcomes of each attempt are two events called success and failure
- Let noted with p the probability of success and with q the probability of failure ($q = 1 - p$)
- The n repeated tests are independent

BINOMIAL DISTRIBUTION

- In a binomial experiment:
  - It consists of a fixed number n of identical experiments
  - There are only two possible outcomes in each experiments, denoted by S (success) and F (failure)
  - The experiments are independent with the same probability of S (denoted p)
- ${}_4C_2$ = 4 choose 2 (combination choosing 2 from 4)

Binomial Distribution

- The number of successes X obtained by performing the test n times is a random variable of n and p parameters and is noted as Bi(n,p)
- The random variable X can take the following values: 0, 1, 2,...n
- Probability that X to be equal with a value k is given by the formula:

$$\Pr(X = k) = C_n^k p^k q^{n-k}$$

where:

$$C_n^k = \frac{n!}{k! \cdot (n-k)!}$$

BINOMIAL DISTRIBUTION

Mean	**$M(X) = n \cdot p$**
Variance	$V(X) = n \cdot p \cdot q$
Standard deviation	$\sigma(X) = \sqrt{(n \cdot p \cdot q)}$

Binomial Distribution

- Suppose that 90% of adults with joint pain report symptomatic relief with a specific medication. If the medication is given to 10 new patients with joint pain, what is the probability that it is effective in exactly 7 subjects?
- Assumptions of the binomial distribution model?
  - The outcome is pain relief (yes or no), and we will consider that the pain relief is a success
  - The probability of success for each subject is 0.9 (p=0.9)
  - The final assumption is that the subjects are independent, and it is reasonable to assume that this is true.
- $Pr(X=7) = {}_{10}C_7 \cdot 0.9^7 \cdot (1-0.9)^{10-7} = 0.0574$ → there is a 5.74% chance that exactly 7 of 10 patients will report pain relief when the probability that any one reports relief is 90%.

BINOMIAL DISTRIBUTION

- What is the probability to have 2 boys in a family with 5 children if the probability to have a baby-boy for each birth is equal to 0.47 and if the sex of children successive born in the family is considered an independent random variable?

- p=0.47
- q=1-0.47=0.53
- n=5
- k=2
- **Pr(X=2)=10·0.47²·0.53³**
- **Pr(X=2) = 0.33**

$$\Pr(X = k) = C_n^k p^k q^{n-k}$$

$$C_5^2 = \frac{5!}{2! \cdot (5-2)!} = \frac{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (3 \cdot 2 \cdot 1)} = \frac{120}{12} = 10$$

Poisson Distribution

- Random Poisson variable take a countable infinity of values (0,1,2,...,k,...) that is the number of achievements of an event within a given range of time or place
  - number of entries per year in a given hospital
  - white blood cells on smear
  - number of decays of a radioactive substance in a given time T

POISSON DISTRIBUTION

- POISSON random variable:
  - Is characterized by theoretical parameter θ (expected average number of achievement for a given event in a given range)
- Symbol: Po(θ)
- Poisson Distribution:

$$X:\begin{pmatrix} k \\ e^{-\theta}\cdot\frac{\theta^k}{k!} \end{pmatrix}$$

$$\Pr(X=k)=\frac{e^{-\theta}\cdot\theta^k}{k!}$$

- Mean of expected values: M(X) = θ
- Variance: V(X) = θ

POISSON DISTRIBUTION

- The mortality rate for specific viral pathology is 7 per 1000 cases. What is the probability that in a group of 400 people this pathology to cause 5 deaths?
- n=400
- p=7/1000=0.007
- θ=n·p=400·0.007=2.8
- e=2.718281828=2.72

$Pr(X=5) = (2.72^{-2.8} \cdot 2.8^{5})/(5 \cdot 4 \cdot 3 \cdot 2 \cdot 1) = 10.45/120$

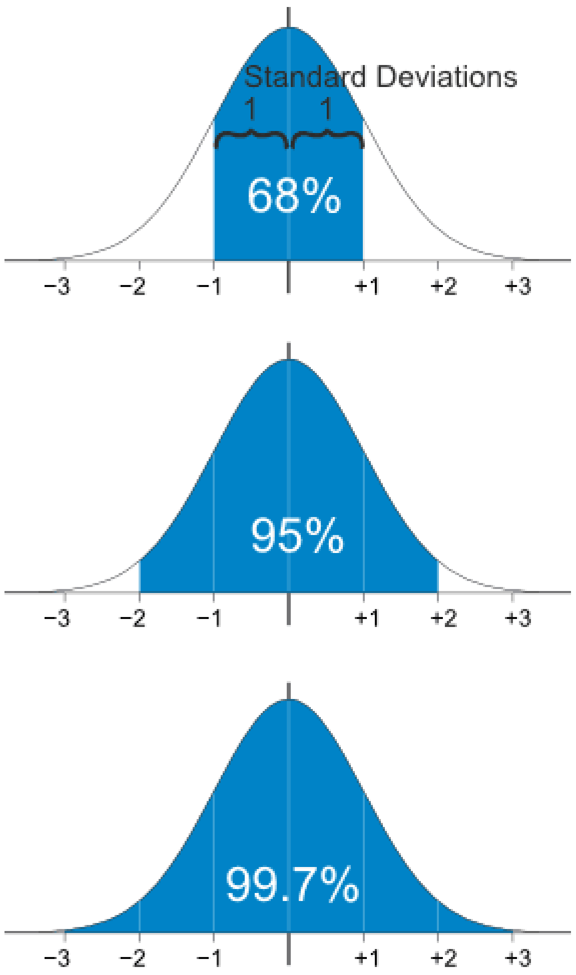
$Pr(X=5) = 0.09$

Continuous Probability Distributions

- Uniform distribution and its standard form
- Normal distribution and its standard form
  - Does your data follow a "bell shaped" pattern? (mean ~ median ~ mode)

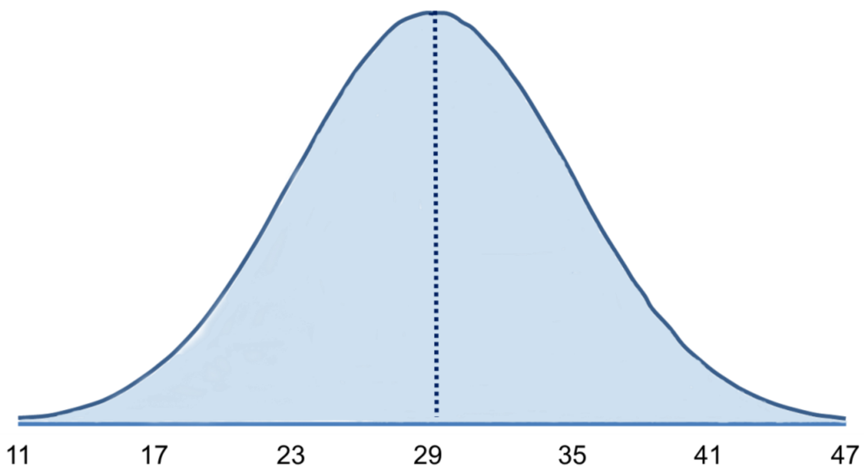
NORMAL DISTRIBUTION

- Also known Gaussian distribution
- Characteristics of normal distribution:
  - ~ 68% of values fall between mean and one standard deviation (in either direction)
  - ~ 95% of values fall between mean and two standard deviations (in either direction)
  - ~ 99.9% of values fall between mean and three standard deviations (in either direction)



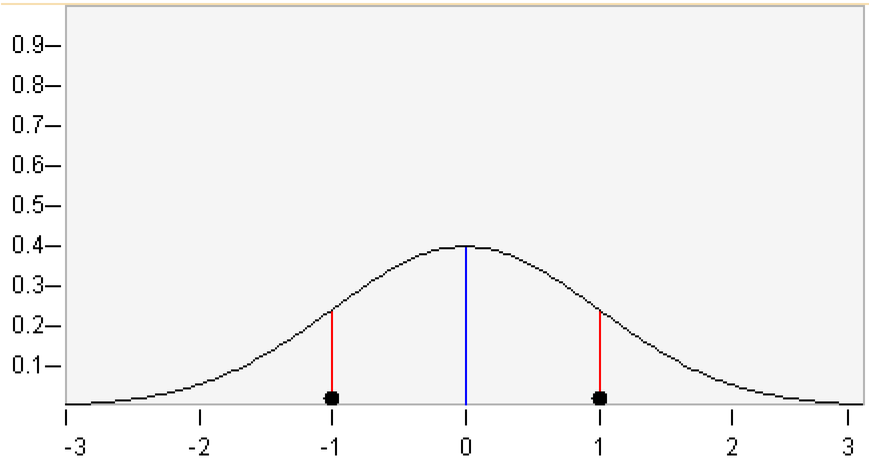
NORMAL DISTRIBUTION

- When we have a normal distributed variable and we know the population mean (μ) and standard deviation (σ), we can compute the probability of particular values using the following formula:
- $Pr(X) = 1/\sigma\sqrt{2\pi} \cdot e\hat{}(-(X-\mu)\hat{}2/(2\sigma\hat{}2))$



The mean (μ = 29) is in the center of the distribution, and the horizontal axis is scaled in increments of the standard deviation (σ = 6)

NORMAL DISTRIBUTION



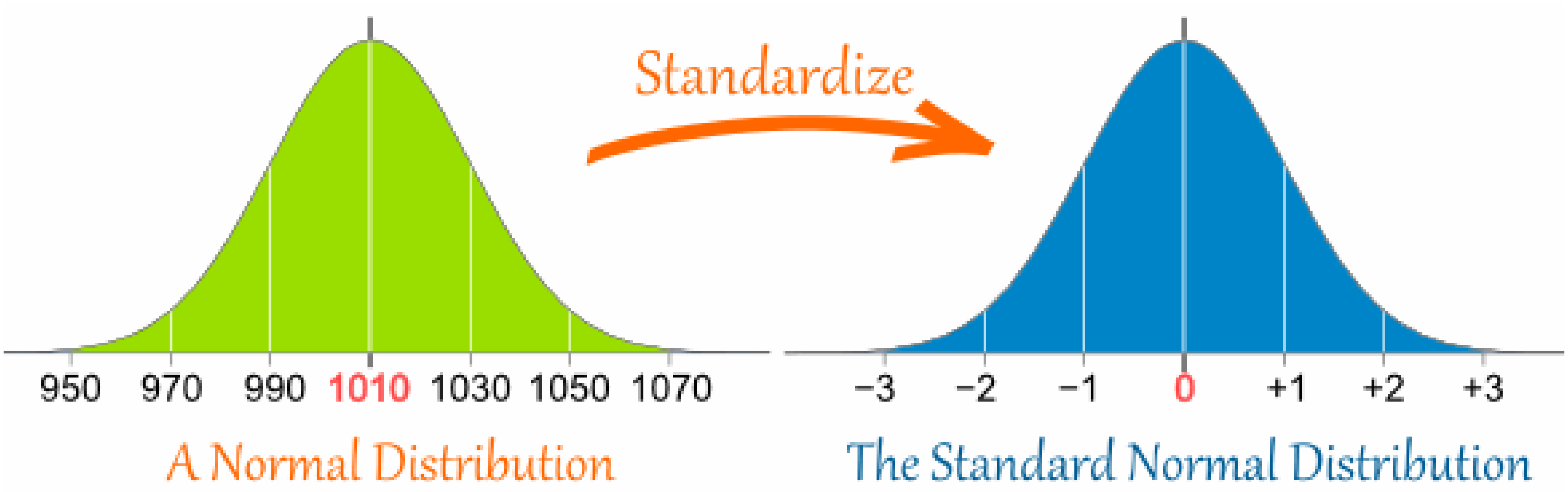
- Normal distribution:
  - Gaussian distribution
  - Symmetric
  - Not skewed
  - Unimodal
  - Described by two parameters:
    - Probability density function:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

  - μ & σ are parameters
    - μ = mean
    - σ = standard deviation
    - π, e = constants

STANDARD NORMAL DISTRIBUTION

- The standard normal distribution is a normal distribution with a mean of zero and standard deviation of 1.



SUMMARY

- Random variables could be discrete or continuous.
- For random variables we have:
  - Discrete probability distributions
  - Continuous probability distribution
- It is possible to compute the expected values (means), variations and standard deviation for discrete and random variables.

SUMMARY

- Normal distribution
  - can be used to describe a variety of variables
  - Is bell-shaped, unimodal, symmetric, and continuous; its mean, median, and mode are equal
  - Its standard form has a mean of 0 and a standard deviation of 1
  - Can be used to approximate other distributions to simplify the analysis of data

PROBLEMS

1. If X is binomially distributed with 6 experiments and a probability of success equal to ¼ at each experiment, what is the probability of:
   - Exactly 4 successes
   - At least on success
2. When an unbiased coin is tossed eight times what is the probability of obtaining:
   - Less than 4 heads
   - More than 5 heads

Problems

3. The serum level of 1,25 dihydroxyvitamin D in adolescent girls is believed to be normally distribution with mean 65 pg/ml and standard deviation 12.5 pg/ml.
   - What percent of adolescent girls will have a level higher than 65 pg/ml?
   - What percent are lower than 65 pg/ml?
   - What percent are between 40 pg/ml and 90 pg/ml?