# Biostatistics

## - Probability Review -

Sophia Daskalaki

Dept. of Electrical & Computer Engineering
✉: sdask@upatras.gr ☎: 2610-997810

# Summary

# Recommended Reading

- *Biostatistics*, A Foundation for Analysis in the Health Sciences, W.W. Daniel and C.L. Cross
- *Probability & Statistics for Engineers and Scientists* R.E.Walpole, R.H. Myers, S.L.Myers, K.Ye
- *Engineering Biostatistics*, An introduction using MATLAB and WINBUGS, B.Vidakovic

# Recommended Reading

- *Biostatistics*, A Foundation for Analysis in the Health Sciences, W.W. Daniel and C.L. Cross
- *Probability & Statistics for Engineers and Scientists* R.E.Walpole, R.H. Myers, S.L.Myers, K.Ye
- *Engineering Biostatistics*, An introduction using MATLAB and WINBUGS, B.Vidakovic

# Introduction to Data Analysis

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

- Descriptive statistics

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

- Descriptive statistics
- Inferential statistics

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

- Descriptive statistics
- Inferential statistics
- Experimental design

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

- Descriptive statistics
- Inferential statistics
- Experimental design
- Regression modeling and analysis

*Statistics* is a science that deals with designing experiments; collecting, summarizing, and processing data; but most importantly making inferences and aiding decisions in the presence of variability and uncertainty.

*Biostatistics* refers to the application of statistical tools to data derived from biological sciences and medicine

- Descriptive statistics
- Inferential statistics
- Experimental design
- Regression modeling and analysis
- Time Series analysis
- Survival analysis

*Data* is the raw material for our work

*Data* is the raw material for our work

Data can be *quantitative* (numbers) or *qualitative* (information)

*Data* is the raw material for our work

Data can be *quantitative* (numbers) or *qualitative* (information)

Sources of Data:

- records (routinely kept)
- surveys (organized by someone)
- experiments (set up to collect measurements)
- external sources (published reports, data banks, research literature, etc.)

*Variable* is a characteristic we are interested in studying

*Variable* is a characteristic we are interested in studying

A variable takes different values when observed in different subjects and we are interested in understanding its *variability*

*Variable* is a characteristic we are interested in studying

A variable takes different values when observed in different subjects and we are interested in understanding its *variability*

Types of Variables:

- *Quantitative*, when the characteristic can be measured
- *Qualitative*, when the characteristic categorize the subjects in different categories. Then we are interested in the frequency of each category.

*Population* is a collection of entities, such as people, animals, items, events, times, etc. for which we have interest, or a collection of all existing attribute values of some natural phenomenon, or the collection of all potential attribute values when a process is involved.

*Population* is a collection of entities, such as people, animals, items, events, times, etc. for which we have interest, or a collection of all existing attribute values of some natural phenomenon, or the collection of all potential attribute values when a process is involved.

A population can be *finite* or *infinite*

*Population* is a collection of entities, such as people, animals, items, events, times, etc. for which we have interest, or a collection of all existing attribute values of some natural phenomenon, or the collection of all potential attribute values when a process is involved.

A population can be *finite* or *infinite*

*Sample* is a subset (a portion) of a population that is selected with the help of a process that guarantees randomness.

Samples may be obtained through a retrospective study, an observational study, a survey using some questionnaire, or a designed study where variables are monitored and controlled to induce a cause/effect relationship.

*Descriptive Statistics* (or *Exploratory Data Analysis* ) refers to methodologies for approaching and summarizing experimental data. It is based solely on sample data and provides tools that calculate descriptive measures (e.g. sample mean and sample variance) and others that visualize data (e.g. pie charts, histograms, stem-and-leaf plots, box-and-whisker plots, scatter plots, etc.)

It is the first step prior to any further statistical analysis and contributes significantly to the goal of statistics:

INFERENCE ABOUT THE POPULATION

*Statistical Inference* refers to procedures by which we infer conclusions about a population on the basis of information contained in a sample drawn from that population.

*Statistical Inference* refers to procedures by which we infer conclusions about a population on the basis of information contained in a sample drawn from that population.

For valid inferences we need scientific sampling techniques and there are several ways of providing such samples. For example:

- Simple random sample: If all possible samples of size $n$ have the same chance of being selected. The sampling procedure can be done *with* or *without replacement*.
- Systematic sampling: If the subjects in the sample are collected from records in a systematic way. It is a common technique for medical research.
- Stratified random sampling: If the population is grouped in different groups (called *strata*) then it may be more useful to take random samples from each stratum and combine them to a single sample. In such a case the variability within each stratum should be less than the variability across strata.

*Data Types*: Data can be *numerical* or *categorical*

*Data Types*: Data can be *numerical* or *categorical*

A *measurement scale* refers to the assignment of numbers to objects or events according to a set of rules, so different rules give different scales.

*Data Types*: Data can be *numerical* or *categorical*

A *measurement scale* refers to the assignment of numbers to objects or events according to a set of rules, so different rules give different scales.

1. Nominal scale: Used when objects are categorized to different categories (e.g. male/female, Asian/Caucasian/Negroid/Australoid)

*Data Types*: Data can be *numerical* or *categorical*

A *measurement scale* refers to the assignment of numbers to objects or events according to a set of rules, so different rules give different scales.

1. Nominal scale: Used when objects are categorized to different categories (e.g. male/female, Asian/Caucasian/Negroid/Australoid)
2. Ordinal scale: Used when objects are categorized to different categories and categories can be ranked (e.g. child/adult/elderly, no fever/low fever/high fever/very high fever)

   **Note**: The distance between categories is not necessarily the same

*Data Types*: Data can be *numerical* or *categorical*

A *measurement scale* refers to the assignment of numbers to objects or events according to a set of rules, so different rules give different scales.

1. Nominal scale: Used when objects are categorized to different categories (e.g. male/female, Asian/Caucasian/Negroid/Australoid)
2. Ordinal scale: Used when objects are categorized to different categories and categories can be ranked (e.g. child/adult/elderly, no fever/low fever/high fever/very high fever)

   **Note**: The distance between categories is not necessarily the same
3. Interval scale: Used when the characteristic of interest can be measured with a scale where distance is well defined (e.g. temperature, student scores) but the origin has no natural meaning.

*Data Types*: Data can be *numerical* or *categorical*

A *measurement scale* refers to the assignment of numbers to objects or events according to a set of rules, so different rules give different scales.

1. Nominal scale: Used when objects are categorized to different categories (e.g. male/female, Asian/Caucasian/Negroid/Australoid)
2. Ordinal scale: Used when objects are categorized to different categories and categories can be ranked (e.g. child/adult/elderly, no fever/low fever/high fever/very high fever)

   **Note**: The distance between categories is not necessarily the same
3. Interval scale: Used when the characteristic of interest can be measured with a scale where distance is well defined (e.g. temperature, student scores) but the origin has no natural meaning.
4. Ratio scale: Used when zero point has the meaning of no existence (e.g. height, length, age)

The *scientific method* is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results. When applied then decisions and outcomes are based on data, therefore it is an *empirical approach*.

The *scientific method* is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results. When applied then decisions and outcomes are based on data, therefore it is an *empirical approach*.

A scientific method comprise four different stages:

1. Making observations: Observation leads to formulation of questions that need to be answered.

The *scientific method* is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results. When applied then decisions and outcomes are based on data, therefore it is an *empirical approach*.

A scientific method comprise four different stages:

1. Making observations: Observation leads to formulation of questions that need to be answered.
2. Formulating a Hypothesis: A hypothesis is formulated either due to observations made previously or due to background research and literature review on the field of interest.

The *scientific method* is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results. When applied then decisions and outcomes are based on data, therefore it is an *empirical approach*.

A scientific method comprise four different stages:

1. Making observations: Observation leads to formulation of questions that need to be answered.
2. Formulating a Hypothesis: A hypothesis is formulated either due to observations made previously or due to background research and literature review on the field of interest.
3. Designing an experiment: An experiment is designed to yield the data necessary to test the hypotheses set previously. The data should be consistent and reliable. The design of the experiment depends on the type of data that need to be collected.

The *scientific method* is a process by which scientific information is collected, analyzed, and reported in order to produce unbiased and replicable results. When applied then decisions and outcomes are based on data, therefore it is an *empirical approach*.

A scientific method comprise four different stages:

1. Making observations: Observation leads to formulation of questions that need to be answered.
2. Formulating a Hypothesis: A hypothesis is formulated either due to observations made previously or due to background research and literature review on the field of interest.
3. Designing an experiment: An experiment is designed to yield the data necessary to test the hypotheses set previously. The data should be consistent and reliable. The design of the experiment depends on the type of data that need to be collected.
4. Conclusion: At the end of any scientific method conclusions should be drawn regarding the hypotheses that were posed. However, such a research is never conclusive and often require several replications.

# Review of Basic Probability Concepts

*Theory of Probability provides the foundation for statistical inference*

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

**Basic terminology:**

- Random Experiment

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

**Basic terminology:**

- Random Experiment
- Sample Space ($\Omega$)

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

**Basic terminology:**

- Random Experiment
- Sample Space ($\Omega$)
- Event

# Definition of Probability

There have been several attempts in defining the notion of *probability*.

- **Subjective definition:** based on somebody's opinion
- **Classical definition:** $P(A) = \frac{\nu(A)}{\nu}$
- **Limit of relative frequency:** $P(A) = lim_{n \to \infty} \frac{n(A)}{n}$

None can satisfy all requirements of a mathematically proper definition.

**Basic terminology:**

- Random Experiment
- Sample Space ($\Omega$)
- Event
- Mutually Exclusive Events

# Notions from Set Theory

We can borrow notions from Set Theory to create new events.

- **Union:** $A \cup B$, when *at least one* of $A$ or $B$ occurs
- **Intersection:** $A \cap B$, when *both* $A$ and $B$ occur
- **Complement of** $E$**:** $E^c$, when *event* $E$ *is not realized* ( $E^c \cap E = \emptyset$ and with respect to the sample space $\Omega$: $E^c \cup E = \Omega$)
- **Difference:** $A - B$ or $A \cap B^c$, when event $A$ *is realized but* $B$ *is not*.
- **Symmetric Difference:** $A \oplus B$ or $(A \cap B^c) \cup (A^c \cap B)$, when *exactly one of* $A$ *or* $B$ occur.

# Axiomatic Definition of Probability

Let $\Omega$ be the sample space of an experiment. We call *probability* a function defined on the set of all subsets of $\Omega$, the events, and satisfy three axioms:

- For each event $A$ it is true that $P(A) \geq 0$
- $P(\Omega) = 1$
- For each infinite sequence of mutually exclusive events $A_1, A_2, \ldots$, it is true that

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

# Important Results

1. $P(\emptyset) = 0$

# Important Results

1. $P(\emptyset) = 0$
2. For each finite sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

# Important Results

1. $P(\emptyset) = 0$
2. For each finite sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

3. For every two subsets $A$ and $B$ where $B \subset A$ it is true that
   - $P(B) \leq P(A)$
   - $P(A - B) = P(A) - P(B)$

# Important Results

1. $P(\emptyset) = 0$
2. For each finite sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

3. For every two subsets $A$ and $B$ where $B \subset A$ it is true that
   - $P(B) \leq P(A)$
   - $P(A - B) = P(A) - P(B)$
4. For each event $A$: $P(A) \leq 1$

# Important Results

1. $P(\emptyset) = 0$
2. For each finite sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

3. For every two subsets $A$ and $B$ where $B \subset A$ it is true that
   - $P(B) \leq P(A)$
   - $P(A - B) = P(A) - P(B)$
4. For each event $A$: $P(A) \leq 1$
5. For each event $A$: $P(A^c) = 1 - P(A)$

# Important Results

1. $P(\emptyset) = 0$
2. For each finite sequence of mutually exclusive events $A_1, A_2, \ldots, A_n$

$$P(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

3. For every two subsets $A$ and $B$ where $B \subset A$ it is true that
   - $P(B) \leq P(A)$
   - $P(A - B) = P(A) - P(B)$
4. For each event $A$: $P(A) \leq 1$
5. For each event $A$: $P(A^c) = 1 - P(A)$
6. (Addition Rule) For any two events $A$ and $B$:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

# More Important Results

For any $n \geq 2$ events $E_1, E_2, \ldots, E_n$:

1. (Generalization of Addition Rule)

$$P(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} P(E_i) - \sum_{i} \underset{1 \leq i < j \leq n}{\sum_{j}} P(E_i \cap E_j)$$

$$+ \sum_{i} \underset{1 \leq i < j < k \leq n}{\sum_{j}} \sum_{k} P(E_i \cap E_j \cap E_k) + \ldots + (-1)^{n-1} P(\bigcap_{i=1}^{n} E_i)$$

2. (Alternatively)

$$P(\bigcup_{i=1}^{n} E_i) = 1 - P(\bigcap_{i=1}^{n} E_i^c)$$

# Odds of an event

## Definition

Let $A$ be an event and $P(A)$ the probability of $A$ taking place. Then *the odds of A* are:

$$odds(A) = \frac{P(A)}{P(A^c)}$$

# Odds of an event

## Definition

Let $A$ be an event and $P(A)$ the probability of $A$ taking place. Then *the odds of A* are:

$$odds(A) = \frac{P(A)}{P(A^c)}$$

# Odds of an event

## Definition

Let $A$ be an event and $P(A)$ the probability of $A$ taking place. Then *the odds of $A$* are:

$$odds(A) = \frac{P(A)}{P(A^c)}$$

Odds are alternative measures of the likelihood of events. Given the odds of $A$ we can write:

$$P(A) = \frac{odds(A)}{odds(A) + 1}$$

# Conditional Probability

## Definition

Let $A$ and $B$ be two random events of an experiment with $P(B) > 0$. Then the probability of $A$ occurring given that $B$ has already taken place is called *conditional probability* of $A$ given $B$, it is denoted as $P(A|B)$ and is calculated as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Conditional Probability

## Definition

Let $A$ and $B$ be two random events of an experiment with $P(B) > 0$. Then the probability of $A$ occurring given that $B$ has already taken place is called *conditional probability* of $A$ given $B$, it is denoted as $P(A|B)$ and is calculated as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

# Conditional Probability

## Definition

Let $A$ and $B$ be two random events of an experiment with $P(B) > 0$. Then the probability of $A$ occurring given that $B$ has already taken place is called *conditional probability* of $A$ given $B$, it is denoted as $P(A|B)$ and is calculated as:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

**Conditional Odds**

$$Odds(A|B) = \frac{P(A|B)}{P(A^c|B)} = \frac{P(A \cap B)}{P(A^c \cap B)}$$

# Product Rule and Independence

# Product Rule and Independence

**Product Rule**

- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(E_1 \cap E_2 \cap \ldots E_k) = P(E_1)P(E_2|E_1)\ldots P(E_k|E_1 \cap E_2 \cap \ldots E_{k-1})$

# Product Rule and Independence

**Product Rule**

- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(E_1 \cap E_2 \cap \ldots E_k) = P(E_1)P(E_2|E_1)\ldots P(E_k|E_1 \cap E_2 \cap \ldots E_{k-1})$

Two events $A$ and $B$, with $P(A) > 0$ and $P(B) > 0$, are called *independent* iff

$$P(A \cap B) = P(A) \cdot P(B)$$

or equivalently

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

# Product Rule and Independence

**Product Rule**

- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(E_1 \cap E_2 \cap \ldots E_k) = P(E_1)P(E_2|E_1)\ldots P(E_k|E_1 \cap E_2 \cap \ldots E_{k-1})$

Two events $A$ and $B$, with $P(A) > 0$ and $P(B) > 0$, are called *independent* iff

$$P(A \cap B) = P(A) \cdot P(B)$$

or equivalently

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

# Product Rule and Independence

**Product Rule**

- $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$
- $P(E_1 \cap E_2 \cap \dots E_k) = P(E_1)P(E_2|E_1)\dots P(E_k|E_1 \cap E_2 \cap \dots E_{k-1})$

Two events $A$ and $B$, with $P(A) > 0$ and $P(B) > 0$, are called *independent* iff

$$P(A \cap B) = P(A) \cdot P(B)$$

or equivalently

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B)$$

**WATCH OUT:**

*Independent events* is different than *mutually exclusive events*.

# Pairwise and Global Independence

Three events $A$, $B$ and $C$ with $P(A) > 0$, $P(B) > 0$ and $P(C) > 0$, are called *pairwise independent* iff

$$P(A \cap B) = P(A) \cdot P(B), \quad P(A \cap C) = P(A) \cdot P(C)$$

$$\text{and} \quad P(B \cap C) = P(B) \cdot P(C)$$

# Pairwise and Global Independence

Three events $A$, $B$ and $C$ with $P(A) > 0$, $P(B) > 0$ and $P(C) > 0$, are called *pairwise independent* iff

$$P(A \cap B) = P(A) \cdot P(B), \quad P(A \cap C) = P(A) \cdot P(C)$$
$$\text{and} \quad P(B \cap C) = P(B) \cdot P(C)$$

# Pairwise and Global Independence

Three events $A$, $B$ and $C$ with $P(A) > 0$, $P(B) > 0$ and $P(C) > 0$, are called *pairwise independent* iff

$$P(A \cap B) = P(A) \cdot P(B), \quad P(A \cap C) = P(A) \cdot P(C)$$

$$\text{and} \quad P(B \cap C) = P(B) \cdot P(C)$$

**Note:**

- If three events are pairwise independent this does not imply that they are independent in totality.

# Pairwise and Global Independence

Three events $A$, $B$ and $C$ with $P(A) > 0$, $P(B) > 0$ and $P(C) > 0$, are called *pairwise independent* iff

$$P(A \cap B) = P(A) \cdot P(B), \quad P(A \cap C) = P(A) \cdot P(C)$$
$$\text{and} \quad P(B \cap C) = P(B) \cdot P(C)$$

**Note:**

- If three events are pairwise independent this does not imply that they are independent in totality.
- On the contrary three mutually exclusive events implies exclusiveness in totality.

# Total Probability

## Theorem

Let event $A \subset \Omega$ and $H_1, \dots H_n$ a sequence of $n$ mutually exclusive events that cover the sample space, i.e. $H_1 \cup H_2 \cup \dots H_n = \Omega$ and $P(H_i) > 0$. Then

$$P(A) = \sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)$$

**In other words:** The rule of total probability expresses the probability of $A$ as the weighted average of its conditional probabilities given the $H_i$'s, which are mutually exclusive events that partition $\Omega$.

# Bayes Theorem

### Theorem

Let event $A \subset \Omega$ and $H_1, \dots H_n$ a sequence of $n$ mutually exclusive events that cover the sample space. Suppose the conditional probabilites $P(A|H_i)$, as well as the (*prior*) probabilities $P(H_i)$ are known $\forall i$. Then we can write the (*posterior*) probabilities:

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{P(A)} = \frac{P(A|H_i)P(H_i)}{\sum_{i=1}^{n} P(A|H_i) \cdot P(H_i)}$$

Many applications in screening and diagnostic tests.

# Example

Suppose that a certain screening test, designed to identify subjects with a specific disease, is successful 99.5% of the times if the subject carries the disease. Moreover, let's assume that the test gives *falsely positive* result 1% of the times for the subjects that do not carry the disease. If it is known that only the 0.8% of the population carries the disease, find the probability that a subject examined by the screening test (a) carries the disease when the result is positive, and (b) does not carry the disease when the result is negative.

# Example

Suppose that a certain screening test, designed to identify subjects with a specific disease, is successful 99.5% of the times if the subject carries the disease. Moreover, let's assume that the test gives *falsely positive* result 1% of the times for the subjects that do not carry the disease. If it is known that only the 0.8% of the population carries the disease, find the probability that a subject examined by the screening test (a) carries the disease when the result is positive, and (b) does not carry the disease when the result is negative.

Define the events:

$D$: the subject carries the disease, and $D^c$: the subject doesn't carry the disease
$T^+$: the test is positive, and $T^-$: the test is negative

# Example

Suppose that a certain screening test, designed to identify subjects with a specific disease, is successful 99.5% of the times if the subject carries the disease. Moreover, let's assume that the test gives *falsely positive* result 1% of the times for the subjects that do not carry the disease. If it is known that only the 0.8% of the population carries the disease, find the probability that a subject examined by the screening test (a) carries the disease when the result is positive, and (b) does not carry the disease when the result is negative.

Define the events:

$D$: the subject carries the disease, and $D^c$: the subject doesn't carry the disease
$T^+$: the test is positive, and $T^-$: the test is negative

Then:

$P(T^+/D) = .995$ and $P(T^-/D^c) = 1 - P(T^+/D^c) = .99$

# Example

Suppose that a certain screening test, designed to identify subjects with a specific disease, is successful 99.5% of the times if the subject carries the disease. Moreover, let's assume that the test gives *falsely positive* result 1% of the times for the subjects that do not carry the disease. If it is known that only the 0.8% of the population carries the disease, find the probability that a subject examined by the screening test (a) carries the disease when the result is positive, and (b) does not carry the disease when the result is negative.

Define the events:
$D$: the subject carries the disease, and $D^c$: the subject doesn't carry the disease
$T^+$: the test is positive, and $T^-$: the test is negative

Then:
$P(T^+/D) = .995$ and $P(T^-/D^c) = 1 - P(T^+/D^c) = .99$

(a) $P(D/T^+) =??$

(b) $P(D^c/T^-) =??$

# Sensitivity and Specificity

# Sensitivity and Specificity

Suppose $n$ subjects are randomly selected from a population which is comprised by diseased $(D)$ and non-diseased $(D^c)$ persons. A diagnostic test is applied to all subjects in the sample and the result is either positive $(T^+)$ or negative $(T^-)$.

# Sensitivity and Specificity

Suppose $n$ subjects are randomly selected from a population which is comprised by diseased ($D$) and non-diseased ($D^c$) persons. A diagnostic test is applied to all subjects in the sample and the result is either positive ($T^+$) or negative ($T^-$).

| | Disease ($D$) | No Disease ($D^c$) | Total |
|---|---|---|---|
| Test Positive | $TP$ | $FP$ | $nT^+ = TP + FP$ |
| Test Negative | $FN$ | $TN$ | $nT^- = FN + TN$ |
| Total | $nD = TP + FN$ | $nD^c = FP + TN$ | $n = nD + nD^c = nT^+ + nT^-$ |

# Sensitivity and Specificity

Suppose $n$ subjects are randomly selected from a population which is comprised by diseased ($D$) and non-diseased ($D^c$) persons. A diagnostic test is applied to all subjects in the sample and the result is either positive ($T^+$) or negative ($T^-$).

| | Disease ($D$) | No Disease ($D^c$) | Total |
|---|---|---|---|
| Test Positive | $TP$ | $FP$ | $nT^+ = TP + FP$ |
| Test Negative | $FN$ | $TN$ | $nT^- = FN + TN$ |
| Total | $nD = TP + FN$ | $nD^c = FP + TN$ | $n = nD + nD^c = nT^+ + nT^-$ |

e.g. Disease: breast cancer, Test: mammogram

# Sensitivity and Specificity

Suppose $n$ subjects are randomly selected from a population which is comprised by diseased ($D$) and non-diseased ($D^c$) persons. A diagnostic test is applied to all subjects in the sample and the result is either positive ($T^+$) or negative ($T^-$).

| | Disease ($D$) | No Disease ($D^c$) | Total |
|---|---|---|---|
| Test Positive | $TP$ | $FP$ | $nT^+ = TP + FP$ |
| Test Negative | $FN$ | $TN$ | $nT^- = FN + TN$ |
| Total | $nD = TP + FN$ | $nD^c = FP + TN$ | $n = nD + nD^c = nT^+ + nT^-$ |

e.g. Disease: breast cancer, Test: mammogram

## Definitions

1. *Sensitivity* is the ratio of the true positives over the number of the diseased subjects: $Se = TP/(TP + FN) = TP/nD$
2. *Specificity* is the ratio of the true negatives over the number of the non diseased subjects: $Sp = TN/(TN + FP) = TN/nD^c$

# More measures in diagnostics

Besides *Se* and *Sp* a number of other measures are important and should be monitored when diagnostic tests are designed. These are:

# More measures in diagnostics

Besides $Se$ and $Sp$ a number of other measures are important and should be monitored when diagnostic tests are designed. These are:

1. *Positives Predictive Value:* $PPV = TP/(TP + FP) = TP/nT^+$

   Proportion of true positives among all subjects with positive test.

# More measures in diagnostics

Besides $Se$ and $Sp$ a number of other measures are important and should be monitored when diagnostic tests are designed. These are:

1. *Positives Predictive Value:* $PPV = TP/(TP + FP) = TP/nT^+$
   Proportion of true positives among all subjects with positive test.
2. *Negatives Predictive Value:* $NPV = TN/(TN + FN) = TN/nT^-$
   Proportion of true negatives among all subjects with negative test.

# More measures in diagnostics

Besides $Se$ and $Sp$ a number of other measures are important and should be monitored when diagnostic tests are designed. These are:

1. *Positives Predictive Value:* $PPV = TP/(TP + FP) = TP/nT^{+}$
   Proportion of true positives among all subjects with positive test.
2. *Negatives Predictive Value:* $NPV = TN/(TN + FN) = TN/nT^{-}$
   Proportion of true negatives among all subjects with negative test.
3. *Likelihood Ratio Positive:* $LRP = \frac{Se}{1-Sp}$.
   It represents the extent by which a positive test result would increase the likelihood of the disease

# More measures in diagnostics

Besides $Se$ and $Sp$ a number of other measures are important and should be monitored when diagnostic tests are designed. These are:

1. *Positives Predictive Value:* $PPV = TP/(TP + FP) = TP/nT^+$

   Proportion of true positives among all subjects with positive test.

2. *Negatives Predictive Value:* $NPV = TN/(TN + FN) = TN/nT^-$

   Proportion of true negatives among all subjects with negative test.

3. *Likelihood Ratio Positive:* $LRP = \frac{Se}{1-Sp}$.

   It represents the extent by which a positive test result would increase the likelihood of the disease

4. *Likelihood Ratio Negative:* $LRN = \frac{1-Se}{Sp}$.

   It represents the extent by which a negative test result would increase the likelihood of no disease or rather would reduce the likelihood of disease.

# Example:

# Example:

In a study about the diagnostic capability of a certain mammogram type there were 96420 subjects (women above 40 yrs old). Among them 5401 had positive mammogram and 91019 had negative. The women were further examined for breast cancer using routine screening and other diagnostic techniques. 655 were actually diagnosed with breast cancer out of which only 495 had positive mammogram. At the same time 4906 subjects that were not diagnosed with cancer had a positive mammogram. If we can assume that the sample used in the study was quite representative of all women above 40 yrs old calculate all relevant measures to evaluate the diagnostic capability of this mammogram.

# Example:

In a study about the diagnostic capability of a certain mammogram type there were 96420 subjects (women above 40 yrs old). Among them 5401 had positive mammogram and 91019 had negative. The women were further examined for breast cancer using routine screening and other diagnostic techniques. 655 were actually diagnosed with breast cancer out of which only 495 had positive mammogram. At the same time 4906 subjects that were not diagnosed with cancer had a positive mammogram. If we can assume that the sample used in the study was quite representative of all women above 40 yrs old calculate all relevant measures to evaluate the diagnostic capability of this mammogram.

| | Disease ($D$) | No Disease ($D^c$) | Total |
|---|---|---|---|
| Test + | 495 | 4906 | 5401 |
| Test − | | | 91019 |
| Total | 655 | | 96420 |

# Prevalence

If the sample in our study is representative of the population then all previous formulas are working fine and in fact we can see that:

$$Se = Pr(T^+|D), \;\; Sp = Pr(T^-|D^c), \;\; PPV = Pr(D|T^+), \;\; NPV = Pr(D^c|T^-)$$

$$\text{and} \quad LRP = \frac{Pr(T^+|D)}{Pr(T^+|D^c)}, \;\; LRN = \frac{Pr(T^-|D)}{Pr(T^-|D^c)}$$

If the sample is not representative then one needs to know the prevalence. *Prevalence of a disease*: probability that a randomly selected person from the population will have the disease (i.e. *prior probability*).

The value of prevalence ($Pre$) influences heavily $PPV$ and $NPV$, since by Bayes:

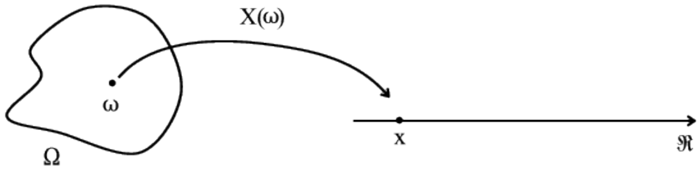
$$PPV = \frac{Se \times Pre}{Se \times Pre + (1 - Sp) \times (1 - Pre)}$$

and

$$NPV = \frac{Sp \times (1 - Pre)}{Sp \times (1 - Pre) + (1 - Se) \times Pre}$$

# Random Variables

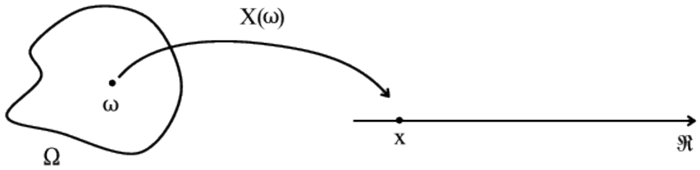
**Random Variable** is any function $X : \Omega \to D \subseteq \mathbb{R}$




(Assigns arithmetic values to the results of a random experiment)

# Random Variables

**Random Variable** is any function $X : \Omega \to D \subseteq \mathbb{R}$




(Assigns arithmetic values to the results of a random experiment)

**Types of random variables**

- Discrete
- Continuous

# Discrete Random Variables

## Probability Mass Function (PMF)

Let $X$ be a discrete random variable. The function $f(x)$:

$$f(x) = P(X = x), \ \forall x \in \mathbb{R}$$

defined on $\mathbb{R}$ and values in $[0, \ 1]$ is called *probability mass function (pmf)* of $X$.

# Discrete Random Variables

## Probability Mass Function (PMF)

Let $X$ be a discrete random variable. The function $f(x)$:

$$f(x) = P(X = x),\ \forall x \in \mathbb{R}$$

defined on $\mathbb{R}$ and values in $[0,\ 1]$ is called *probability mass function (pmf)* of $X$.

# Discrete Random Variables

## Probability Mass Function (PMF)

Let $X$ be a discrete random variable. The function $f(x)$:

$$f(x) = P(X = x),\ \forall x \in \mathbb{R}$$

defined on $\mathbb{R}$ and values in $[0,\ 1]$ is called *probability mass function (pmf)* of $X$.

Any function $f(x)$ can be pmf iff:

- $f(x) \geq 0\ \forall x \in \mathbb{R}$, and
- $\sum_x f(x) = 1$

# Continuous Random Variables

## Probability Density Function (PDF)

Let $X$ be a continuous random variable. Then its *probability density function* $f(x)$ is a function defined on $\mathbb{R}$ with values also in $\mathbb{R}$ and for each interval $[a, b] \subset \mathbb{R}$ we can write:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Any function $f(x)$ can be pdf iff:

- $f(x) \geq 0, \ \forall x \in \mathbb{R}$, and
- $\int_{-\infty}^{\infty} f(x)dx = 1$

# Continuous Random Variables

## Probability Density Function (PDF)

Let $X$ be a continuous random variable. Then its *probability density function* $f(x)$ is a function defined on $\mathbb{R}$ with values also in $\mathbb{R}$ and for each interval $[a, b] \subset \mathbb{R}$ we can write:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Any function $f(x)$ can be pdf iff:

- $f(x) \geq 0, \ \forall x \in \mathbb{R}$, and
- $\int_{-\infty}^{\infty} f(x)dx = 1$

# Continuous Random Variables

## Probability Density Function (PDF)

Let $X$ be a continuous random variable. Then its *probability density function* $f(x)$ is a function defined on $\mathbb{R}$ with values also in $\mathbb{R}$ and for each interval $[a, b] \subset \mathbb{R}$ we can write:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Any function $f(x)$ can be pdf iff:

- $f(x) \geq 0, \; \forall x \in \mathbb{R}$, and
- $\int_{-\infty}^{\infty} f(x)dx = 1$

# Cumulative Distribution Function (CDF)

Let $X$ be a random variable. Then its *cumulative distribution function* $F(x)$ is a function defined on $\mathbb{R}$ with values in $[0, 1]$ so that

$$F(x) = P(X \leq x), \quad \forall x \in \mathbb{R}$$

# Cumulative Distribution Function (CDF)

Let $X$ be a random variable. Then its *cumulative distribution function* $F(x)$ is a function defined on $\mathbb{R}$ with values in $[0,1]$ so that

$$F(x) = P(X \leq x), \quad \forall x \in \mathbb{R}$$

# Cumulative Distribution Function (CDF)

Let $X$ be a random variable. Then its *cumulative distribution function* $F(x)$ is a function defined on $\mathbb{R}$ with values in $[0,1]$ so that

$$F(x) = P(X \leq x), \quad \forall x \in \mathbb{R}$$

Obviously:

- $P(a < X \leq b) = F(b) - F(a)$
- $P(X > x) = 1 - F(x)$

# Cumulative Distribution Function

## Properties

1. Non-negative and non-decreasing in $\mathbb{R}$
2. Right-continuous: $\lim_{x \to x_0^+} F(x) = F(x_0)$ *(for discrete r.v.)*,

   Continuous: $\lim_{x \to x_0} F(x) = F(x_0)$ *(for continuous r.v.)*
3. $\lim_{x \to \infty} F(x) = 1$ and $\lim_{x \to -\infty} F(x) = 0$
4. $F(x) = \sum_{u \leq x} f(u)$ *(for discrete r.v.)*,

   $F(x) = \int_{-\infty}^{x} f(u)du$ *(for continuous r.v.)*
5. $f(x) = F(x) - F(x^-)$ *(for discrete r.v.)*,

   $f(x) = \frac{d}{dx} F(x)$ *(for continuous r.v.)*

# Expected Value

- Let $X$ be a random variable. Then

$$E(X) = \mu_X = \begin{cases} \sum_x x\ f(x) & \text{if the r.v. } X \text{ is discrete} \\ \int_x x\ f(x)dx & \text{if the r.v. } X \text{ is continuous} \end{cases}$$

# Expected Value

- Let $X$ be a random variable. Then

$$E(X) = \mu_X = \begin{cases} \sum_x x\ f(x) & \text{if the r.v. } X \text{ is discrete} \\ \int_x x\ f(x)dx & \text{if the r.v. } X \text{ is continuous} \end{cases}$$

- Let $Y = g(X)$ a function of the r.v. $X$. Then

$$E(Y) = \mu_Y = E(g(X)) = \begin{cases} \sum_x g(x)f(x) & \text{if } X \text{ is discrete} \\ \int_x g(x)f(x)dx & \text{if } X \text{ is continuous} \end{cases}$$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$
- $E(X + b) = E(X) + b$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$
- $E(X + b) = E(X) + b$
- $E[X - \mu_X] = 0$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$
- $E(X + b) = E(X) + b$
- $E[X - \mu_X] = 0$
- $E(aX + b) = aE(X) + b$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$
- $E(X + b) = E(X) + b$
- $E[X - \mu_X] = 0$
- $E(aX + b) = aE(X) + b$
- If $X$ and $Y$ are r.v. then $E(X + Y) = E(X) + E(Y)$

# Expected Value (cont'd)

**Properties**

- $E(a) = a$
- $E(X + b) = E(X) + b$
- $E[X - \mu_X] = 0$
- $E(aX + b) = aE(X) + b$
- If $X$ and $Y$ are r.v. then $E(X + Y) = E(X) + E(Y)$
- If $X$ and $Y$ are independent r.v. then $E(XY) = E(X)E(Y)$

# More Centrality Measures

**Median** Let $X$ be a r.v. and $F(x)$ its cdf

*Median* of $X$ is called any real number $\delta$ for which

$$F(\delta) = P(X \leq \delta) \geq \frac{1}{2} \geq P(X < \delta) = F(\delta^{-})$$

# More Centrality Measures

**Median** Let $X$ be a r.v. and $F(x)$ its cdf

*Median* of $X$ is called any real number $\delta$ for which

$$F(\delta) = P(X \leq \delta) \geq \frac{1}{2} \geq P(X < \delta) = F(\delta^{-})$$

The median of a random variable always exists; this is not true for the mean.

# More Centrality Measures

## Percentiles

Let $X$ be a r.v., $F(x)$ its cdf and let $p$ a number in $(0,1)$. Then any real value $x_p$ for which

$$F(x_p^-) = P(X < x_p) \leq p \leq P(X \leq x_p) = F(x_p)$$

is called the *p–percentile* of $X$.

# More Centrality Measures

## Percentiles

Let $X$ be a r.v., $F(x)$ its cdf and let $p$ a number in $(0,1)$. Then any real value $x_p$ for which

$$F(x_p^-) = P(X < x_p) \leq p \leq P(X \leq x_p) = F(x_p)$$

is called the *p–percentile* of $X$.

- Obviously, when $X$ is continuous then a $p$–percentile is the number $x_p$ that satisfies the equation $F(x_p) = p$
- The percentile $x_{0.5}$ is the median of $X$
- The points $x_{0.25}$ and $x_{0.75}$ are known as first and third quartiles.

# Measures of Variability

## Variance

$$
\begin{aligned}
Var(X) &= \sigma^2 = E\left((X-\mu)^2\right) \\
&= \begin{cases} \sum_x (x-\mu)^2 \cdot f(x) & \text{if } X \text{ is a discrete r.v.} \\ \int_x (x-\mu)^2 \cdot f(x)dx & \text{if } X \text{ is a continuous r.v.} \end{cases}
\end{aligned}
$$

# Measures of Variability

**Variance**

$$
\begin{aligned}
Var(X) &= \sigma^2 = E\left((X-\mu)^2\right) \\
&= \begin{cases} \sum_x (x-\mu)^2 \cdot f(x) & \text{if } X \text{ is a discrete r.v.} \\ \int_x (x-\mu)^2 \cdot f(x)dx & \text{if } X \text{ is a continuous r.v.} \end{cases}
\end{aligned}
$$

**Standard Deviation**

$$\sigma = \sqrt{\sigma^2}$$

# Measures of Variability

## Variance

$$
\begin{aligned}
Var(X) &= \sigma^2 = E\left((X-\mu)^2\right) \\
&= \begin{cases} \sum_x (x-\mu)^2 \cdot f(x) & \text{if } X \text{ is a discrete r.v.} \\ \int_x (x-\mu)^2 \cdot f(x)dx & \text{if } X \text{ is a continuous r.v.} \end{cases}
\end{aligned}
$$

## Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

## Coefficient of Variation

$$CV = \frac{\sigma}{\mu} \quad \text{or} \quad \frac{\sigma}{\mu} \times 100\%$$

# Measures of Variability

## Variance

$$
\begin{aligned}
Var(X) &= \sigma^2 = E\left((X-\mu)^2\right) \\
&= \begin{cases} \sum_x (x-\mu)^2 \cdot f(x) & \text{if } X \text{ is a discrete r.v.} \\ \int_x (x-\mu)^2 \cdot f(x)dx & \text{if } X \text{ is a continuous r.v.} \end{cases}
\end{aligned}
$$

## Standard Deviation

$$\sigma = \sqrt{\sigma^2}$$

## Coefficient of Variation

$$CV = \frac{\sigma}{\mu} \quad \text{or} \quad \frac{\sigma}{\mu} \times 100\%$$

- CV is just a number (unit-less)
- it gives the standard deviation as a fraction of the mean.

# Functions of random variables

Let $X$ be a r.v. with known distribution and $Y = g(X)$ a function of $X$, which is also a r.v. We are looking for the distribution of $Y$.

- If $X$ is discrete then the probability of $Y$ getting the value $y$ can be found from the probability of $X$ getting all values $x$ for which $g(x) = y$.

# Functions of random variables

Let $X$ be a r.v. with known distribution and $Y = g(X)$ a function of $X$, which is also a r.v. We are looking for the distribution of $Y$.

- If $X$ is discrete then the probability of $Y$ getting the value $y$ can be found from the probability of $X$ getting all values $x$ for which $g(x) = y$.
  - Let $X$ be the number we get when a dice is rolled and $Y = X^2$. Find the probability mass function of $Y$.
  - Let $X \in \{-2, 0, 2\}$ and $f(-2) = 0.2,\ f(0) = 0.4, f(2) = 0.4$. Find the probability mass function of $Y = X^2$.

# Functions of random variables

Let $X$ be a r.v. with known distribution and $Y = g(X)$ a function of $X$, which is also a r.v. We are looking for the distribution of $Y$.

- If $X$ is discrete then the probability of $Y$ getting the value $y$ can be found from the probability of $X$ getting all values $x$ for which $g(x) = y$.
  - Let $X$ be the number we get when a dice is rolled and $Y = X^2$. Find the probability mass function of $Y$.
  - Let $X \in \{-2, 0, 2\}$ and $f(-2) = 0.2,\ f(0) = 0.4, f(2) = 0.4$. Find the probability mass function of $Y = X^2$.
- If $X$ is continuous, finding the distribution of $Y = g(X)$ is not as easy.
  - One approach is based on finding first the cdf of $Y$, using the cdf of $X$ and then, if needed, find the pdf by differentiation.
  - If $g(x)$ is monotone and differentiable then we can use the following theorem.

# Functions of random variables

## Change of Variable

Let $X$ be a continuous r.v. with pdf $f_X(x)$ and let $Y = g(X)$ a function of X. If the function $y = g(x)$ is monotone and differentiable in the area of all possible values of $X$, then its reverse function $x = g^{-1}(y)$ exists and

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}(y)}{dy} \right|$$

# Multivariate Random Variables

In many occasions we are interested in studying jointly the behavior of two or more random variables

# Multivariate Random Variables

In many occasions we are interested in studying jointly the behavior of two or more random variables

## Examples

- Suppose we study the cholesterol and sugar level in blood for a given population.
- Studying air pollution in the center of a city we need to take into account several variables, like carbon monoxide ($CO$), nitrogen oxides ($NO$, $NO_2$), Ozone ($O_3$), Ammonia ($NH_3$).

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i,\ i = 1, \ldots, n$
- Right continuous for each $x_i,\ i = 1, \ldots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \ldots, x_{i+1}, x_i, x_{i+1}, \ldots, x_n) = F(x_1, \ldots, x_{i+1}, x_{i0}, x_{i+1}, \ldots, x_n)$$

- $F(x_1, \ldots, x_n)$ satisfies the following properties

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \ i = 1, \ldots, n$
- Right continuous for each $x_i, \ i = 1, \ldots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \ldots, x_{i+1}, x_i, x_{i+1}, \ldots, x_n) = F(x_1, \ldots, x_{i+1}, x_{i0}, x_{i+1}, \ldots, x_n)$$

- $F(x_1, \ldots, x_n)$ satisfies the following properties
  - $F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = \lim_{x_i \to -\infty}, F(x_1, \ldots, x_n) =$

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \ i = 1, \ldots, n$
- Right continuous for each $x_i, \ i = 1, \ldots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \ldots, x_{i+1}, x_i, x_{i+1}, \ldots, x_n) = F(x_1, \ldots, x_{i+1}, x_{i0}, x_{i+1}, \ldots, x_n)$$

- $F(x_1, \ldots, x_n)$ satisfies the following properties
  - $F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \ldots, x_n) = 0$

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \; i = 1, \ldots, n$
- Right continuous for each $x_i, \; i = 1, \ldots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \ldots, x_{i+1}, x_i, x_{i+1}, \ldots, x_n) = F(x_1, \ldots, x_{i+1}, x_{i0}, x_{i+1}, \ldots, x_n)$$

- $F(x_1, \ldots, x_n)$ satisfies the following properties
  - $F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \ldots, x_n) = 0$
  - $F(\infty, \ldots, \infty) = \lim_{x_1 \to \infty, \ldots, x_n \to \infty} F(x_1, \ldots, x_n) =$

# Joint Distribution Function

If $X_1, \ldots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \ldots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \; i = 1, \ldots, n$
- Right continuous for each $x_i, \; i = 1, \ldots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \ldots, x_{i+1}, x_i, x_{i+1}, \ldots, x_n) = F(x_1, \ldots, x_{i+1}, x_{i0}, x_{i+1}, \ldots, x_n)$$

- $F(x_1, \ldots, x_n)$ satisfies the following properties
  - $F(x_1, \ldots, x_{i-1}, -\infty, x_{i+1}, \ldots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \ldots, x_n) = 0$
  - $F(\infty, \ldots, \infty) = \lim_{x_1 \to \infty, \ldots, x_n \to \infty} F(x_1, \ldots, x_n) = 1$

# Joint Distribution Function

If $X_1, \dots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \dots, x_n) = P(X_1 \le x_1, \dots, X_n \le x_n)$$

**Properties**

- Non negative function $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \;\; i = 1, \dots, n$
- Right continuous for each $x_i, \;\; i = 1, \dots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \dots, x_{i+1}, x_i, x_{i+1}, \dots, x_n) = F(x_1, \dots, x_{i+1}, x_{i0}, x_{i+1}, \dots, x_n)$$

- $F(x_1, \dots, x_n)$ satisfies the following properties
  - $F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \dots, x_n) = 0$
  - $F(\infty, \dots, \infty) = \lim_{x_1 \to \infty, \dots, x_n \to \infty} F(x_1, \dots, x_n) = 1$
  - $F(\infty, \dots, \infty, x_i, \infty, \dots \infty) =$ $\lim_{x_1 \to \infty, \dots, x_{i-1} \to \infty, x_{i+1} \to \infty, \dots x_n \to \infty} F(x_1, \dots, x_n) =$

# Joint Distribution Function

If $X_1, \dots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \;\; i = 1, \dots, n$
- Right continuous for each $x_i, \;\; i = 1, \dots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \dots, x_{i+1}, x_i, x_{i+1}, \dots, x_n) = F(x_1, \dots, x_{i+1}, x_{i0}, x_{i+1}, \dots, x_n)$$

- $F(x_1, \dots, x_n)$ satisfies the following properties
  - $F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \dots, x_n) = 0$
  - $F(\infty, \dots, \infty) = \lim_{x_1 \to \infty, \dots, x_n \to \infty} F(x_1, \dots, x_n) = 1$
  - $F(\infty, \dots, \infty, x_i, \infty, \dots \infty) = \lim_{x_1 \to \infty, \dots, x_{i-1} \to \infty, x_{i+1} \to \infty, \dots x_n \to \infty} F(x_1, \dots, x_n) =$

# Joint Distribution Function

If $X_1, \dots X_n$ are random variables then we define the *joint cumulative distribution function* as:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$$

**Properties**

- Non negative function $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$
- Non decreasing function for each $x_i, \;\; i = 1, \dots, n$
- Right continuous for each $x_i, \;\; i = 1, \dots, n$, i.e.

$$\lim_{x_i \to x_{i0}^+} F(x_1, \dots, x_{i+1}, x_i, x_{i+1}, \dots, x_n) = F(x_1, \dots, x_{i+1}, x_{i0}, x_{i+1}, \dots, x_n)$$

- $F(x_1, \dots, x_n)$ satisfies the following properties
  - $F(x_1, \dots, x_{i-1}, -\infty, x_{i+1}, \dots, x_n) = \lim_{x_i \to -\infty,} F(x_1, \dots, x_n) = 0$
  - $F(\infty, \dots, \infty) = \lim_{x_1 \to \infty, \dots, x_n \to \infty} F(x_1, \dots, x_n) = 1$
  - $F(\infty, \dots, \infty, x_i, \infty, \dots \infty) = \lim_{x_1 \to \infty, \dots, x_{i-1} \to \infty, x_{i+1} \to \infty, \dots x_n \to \infty} F(x_1, \dots, x_n) = F_{X_i}(x_i) = P(X_i \leq x_i)$

(marginal cumulative distribution function of $X_i$).

# Multivariate Discrete Random Variables

We define the *joint probability mass function* of r.v. $X_1, X_2, \ldots, X_n$ as:

$$f(x_1, x_2, \ldots, x_n) \equiv P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), \ \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

# Multivariate Discrete Random Variables

We define the *joint probability mass function* of r.v. $X_1, X_2, \ldots, X_n$ as:

$$f(x_1, x_2, \ldots, x_n) \equiv P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n), \ \forall (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

We define as *marginal probability mass function* of $X_i$ the probability mass functions of $X_i$ by itself, i.e.

$$f_{X_i}(x) = \sum_{x_1} \sum_{x_2} \ldots \sum_{x_{i-1}} \sum_{x_{i+1}} \ldots \sum_{x_n} f(x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) \text{ for } i = 1, \ldots, n$$

# Bivariate Discrete Random Variables

- When the random variables are only two, $X$ and $Y$, we then define the *Joint probability mass function* as:

$$f(x, y) \equiv P(X = x, Y = y), \ \forall (x, y) \in \mathbb{R}^2$$

- If the joint probability mass function of $X, Y$ is known, then the *marginal probability mass functions* of $X$ and $Y$, respectively can be calculated as:

$$f_X(x) \equiv P(X = x) = \sum_y f(x, y) \ \text{ and } \ f_Y(y) \equiv P(Y = y) = \sum_x f(x, y)$$

- Moreover, the *conditional probability mass function* of $X$ given that $Y = y$ can be calculated as:

$$f_{X|Y=y}(x) = \frac{f(x, y)}{f_Y(y)}$$

## Multivariate continuous random variables

The *joint probability density function* for r.v. $X_1, X_2, \ldots, X_n$ is a function defined in $\mathbb{R}^n$ for which:

- $f(x_1, x_2, \ldots, x_n) \geq 0$
- $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)dx_1 dx_2 \ldots dx_n = 1$
- For each area $E$ of $\mathbb{R}^n$ the probability of observing the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ in $E$ can be calculated as:

$$P(\mathbf{X} \in E) = \int_{\mathbf{x} \in E} f(\mathbf{x})d\mathbf{x}$$

# Multivariate continuous random variables

The *joint probability density function* for r.v. $X_1, X_2, \ldots, X_n$ is a function defined in $\mathbb{R}^n$ for which:

- $f(x_1, x_2, \ldots, x_n) \geq 0$
- $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \ldots dx_n = 1$
- For each area $E$ of $\mathbb{R}^n$ the probability of observing the random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ in $E$ can be calculated as:

$$P(\mathbf{X} \in E) = \int_{\mathbf{x} \in E} f(\mathbf{x}) d\mathbf{x}$$

Marginal probability density function

$$f_{x_i}(x) = \int_{x_1} \int_{x_2} \ldots \int_{x_{i-1}} \int_{x_{i+1}} \ldots \int_{x_n} f(x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

# Independence of random variables

The r.v. $X, Y$ are called *independent* if one of the following equivalent relations is true.

1. $P(X \in E_1, Y \in E_2) = P(X \in E_1)P(Y \in E_2)$
2. $f(x, y) = f(x)f(y)$
3. $F(x, y) = F(x)F(y)$

# Independence of random variables

The r.v. $X, Y$ are called *independent* if one of the following equivalent relations is true.

1. $P(X \in E_1, Y \in E_2) = P(X \in E_1)P(Y \in E_2)$
2. $f(x, y) = f(x)f(y)$
3. $F(x, y) = F(x)F(y)$

In general,
the r.v. $X_1, X_2, \dots, X_n$ are independent iff

$$f(x_1, x_2, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n)$$

# Expected Value

## Expected value of any function of bivariate r.v.

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)f(x,y) & \text{if } (X,Y) \text{ are discrete} \\ \int_x \int_y h(x,y)f(x,y)dxdy & \text{if } (X,Y) \text{ are continuous} \end{cases}$$

# Expected Value

## Expected value of any function of bivariate r.v.

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)f(x,y) & \text{if } (X,Y) \text{ are discrete} \\ \int_x \int_y h(x,y)f(x,y)dxdy & \text{if } (X,Y) \text{ are continuous} \end{cases}$$

# Expected Value

## Expected value of any function of bivariate r.v.

$$E[h(X,Y)] = \begin{cases} \sum_x \sum_y h(x,y)f(x,y) & \text{if } (X,Y) \text{ are discrete} \\ \int_x \int_y h(x,y)f(x,y)dxdy & \text{if } (X,Y) \text{ are continuous} \end{cases}$$

## Conditional expected value of $X$ given that $Y = y$,

$$E(X|Y=y) = \sum_x x \cdot f_X(x|y) = \sum_x x \cdot \frac{f(x,y)}{f_Y(y)} \quad \text{if X is discrete}$$

$$E(X|Y=y) = \int_x x \cdot f_X(x|y) = \int_x x \cdot \frac{f(x,y)}{f_Y(y)} \quad \text{if X is continuous}$$

# Conditional Variance

## Conditional Variance of $X$ given $Y = y$

$$Var(X|Y = y) = E(X^2|Y = y) - [E(X|Y = y)]^2$$

where

$$E(X^2|Y = y) = \sum_x x^2 \cdot f_X(x|y) = \sum_x x^2 \cdot \frac{f(x, y)}{f_Y(y)}$$

# Conditional Variance

## Conditional Variance of $X$ given $Y = y$

$$Var(X|Y = y) = E(X^2|Y = y) - [E(X|Y = y)]^2$$

where

$$E(X^2|Y = y) = \sum_x x^2 \cdot f_X(x|y) = \sum_x x^2 \cdot \frac{f(x,y)}{f_Y(y)}$$

# Conditional Variance

## Conditional Variance of $X$ given $Y = y$

$$Var(X|Y=y) = E(X^2|Y=y) - [E(X|Y=y)]^2$$

where

$$E(X^2|Y=y) = \sum_x x^2 \cdot f_X(x|y) = \sum_x x^2 \cdot \frac{f(x,y)}{f_Y(y)}$$

# Covariance

*Covariance* is a parameter that measures the co-variability of two r.v. $X$ and $Y$.

$$Cov(X,Y) = E\left((X - E(X))(Y - E(Y))\right)$$

# Covariance

*Covariance* is a parameter that measures the co-variability of two r.v. $X$ and $Y$.

$$Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$$

# Covariance

*Covariance* is a parameter that measures the co-variability of two r.v. $X$ and $Y$.

$$Cov(X, Y) = E\left((X - E(X))(Y - E(Y))\right) = E(XY) - E(X)E(Y)$$

If $X$ and $Y$ are independent r.v. then

$$Cov(X, Y) = 0$$

# Covariance

*Covariance* is a parameter that measures the co-variability of two r.v. $X$ and $Y$.

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))) = E(XY) - E(X)E(Y)$$

If $X$ and $Y$ are independent r.v. then

$$Cov(X, Y) = 0$$

The opposite is not true!

# Properties of Variance

- If $\alpha, \beta \in \mathbb{R}$ then

$$Var(\alpha X + \beta) = \alpha^2 Var(X)$$

# Properties of Variance

- If $\alpha, \beta \in \mathbb{R}$ then

$$Var(\alpha X + \beta) = \alpha^2 Var(X)$$

- If $\alpha, \beta \in \mathbb{R}$ and $(X, Y)$ a vector of r.v. jointly distributed then

$$Var(\alpha X \pm \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y) \pm 2\alpha\beta Cov(X, Y)$$

# Properties of Variance

- If $\alpha, \beta \in \mathbb{R}$ then

$$Var(\alpha X + \beta) = \alpha^2 Var(X)$$

- If $\alpha, \beta \in \mathbb{R}$ and $(X, Y)$ a vector of r.v. jointly distributed then

$$Var(\alpha X \pm \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y) \pm 2\alpha\beta Cov(X, Y)$$

- However, if the r.v. $X$ and $Y$ are independent, then

$$Var(\alpha X \pm \beta Y) = \alpha^2 Var(X) + \beta^2 Var(Y)$$

# Correlation Coefficient

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

*Correlation coefficient* is a parameter that measures the strength of the linear relation that may exist between two r.v. $X$ and $Y$

- $\rho$ is a pure number and $-1 \leq \rho \leq 1$
- If $\rho = 1$ there is perfect positive correlation between $X$ and $Y$
- If $\rho = -1$ there is perfect negative correlation between $X$ and $Y$
- If $\rho = 0$ then $X$ and $Y$ are uncorrelated

# Correlation Coefficient

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$$

*Correlation coefficient* is a parameter that measures the strength of the linear relation that may exist between two r.v. $X$ and $Y$

- $\rho$ is a pure number and $-1 \leq \rho \leq 1$
- If $\rho = 1$ there is perfect positive correlation between $X$ and $Y$
- If $\rho = -1$ there is perfect negative correlation between $X$ and $Y$
- If $\rho = 0$ then $X$ and $Y$ are uncorrelated

- If the r.v. $X$ and $Y$ are independent then $\rho = 0$

# Correlation Coefficient

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

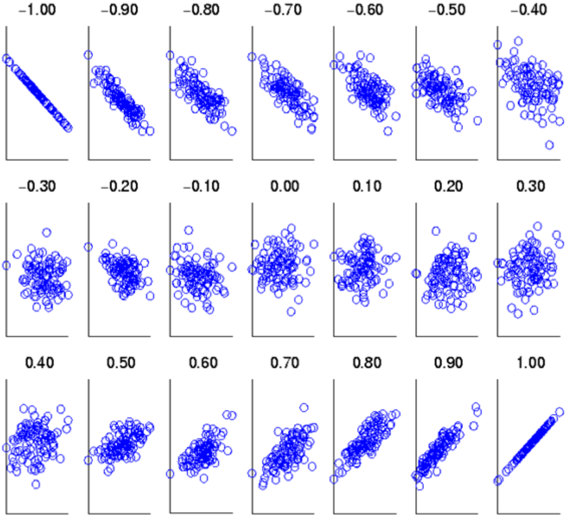
*Correlation coefficient* is a parameter that measures the strength of the linear relation that may exist between two r.v. $X$ and $Y$

- $\rho$ is a pure number and $-1 \leq \rho \leq 1$
- If $\rho = 1$ there is perfect positive correlation between $X$ and $Y$
- If $\rho = -1$ there is perfect negative correlation between $X$ and $Y$
- If $\rho = 0$ then $X$ and $Y$ are uncorrelated

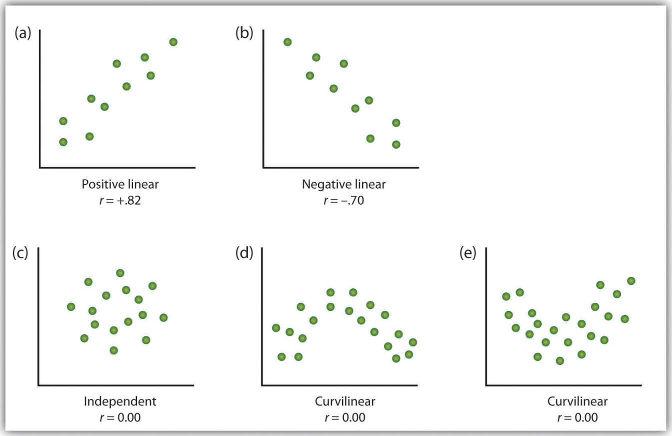
- If the r.v. $X$ and $Y$ are independent then $\rho = 0$

The opposite is not true!

# Examples of Correlation Coefficient



# More Examples of Correlation Coefficient



# Binomial Distribution

A r.v. $X$ follows a *binomial distribution* with parameters $n$ and $p$ if $X$ counts the number of successes in $n$ independent Bernoulli trials, i.e. experiments that result to two possible outcomes: success or failure, with $p$ being the probability of success at each trial.

## Probability mass function:

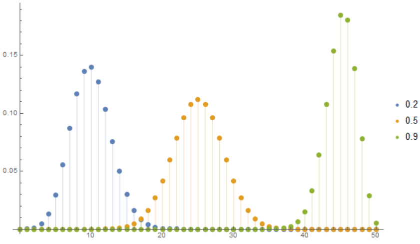
$$f_X(x) = P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, 2, \ldots, n$$

## Expected Value and Variance:

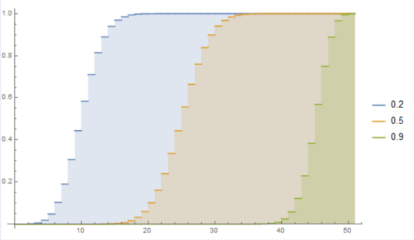
$$E[X] = np \qquad V[X] = np(1-p)$$

# Binomial Distribution Graphs – $n = 50, p = ??$

pdf:




cdf:




Calculation of binomial probabilities becomes tedious when $n$ is very large.

# Poisson Distribution

A r.v. $X$ follows a *Poisson distribution* with parameter $\lambda$ if $X$ counts the number of occurrences of some random event in an interval of time or space. The pmf of $X$ is then given by:

**Probability mass function:**

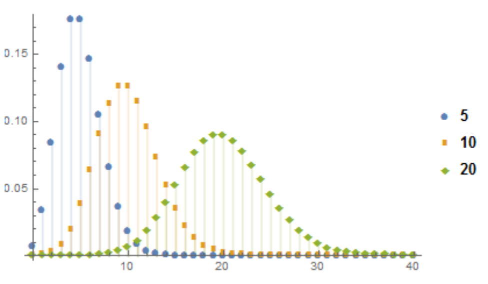
$$f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \text{ for } x = 0, 1, 2, \ldots,$$

Expected Value and Variance:

$$E[X] = \lambda \qquad V[X] = \lambda$$

# Poisson Distribution Graph – $\lambda = ??$

## pdf:

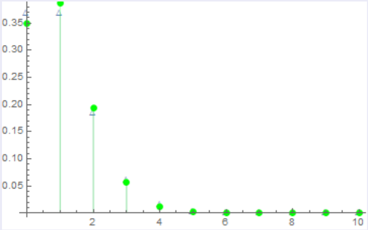



For small values of $\lambda$ Poisson distribution is asymmetric, but for larger values it becomes symmetric.

# Poisson vs Binomial

A binomial distribution with large $n$ and small $p$ (practically, if $n > 30$ and $np < 5$) can be approximated by a Poisson with $\lambda = np$.

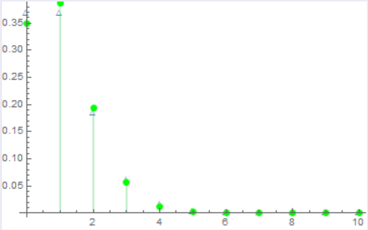
## Binom(10,0.1) vs Poisson(1)



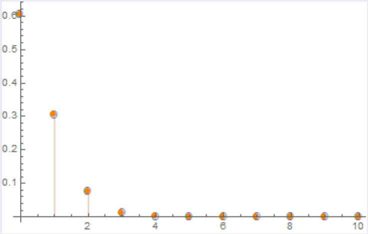
# Poisson vs Binomial

A binomial distribution with large $n$ and small $p$ (practically, if $n > 30$ and $np < 5$) can be approximated by a Poisson with $\lambda = np$.

**Binom(10,0.1) vs Poisson(1)**




**Binom(50,0.01) vs Poisson(0.5)**



# Poisson Process

Let $N(t)$ be the number of random events that take place in the interval $(0, t]$. Then the set of r.v. $\{N(t), t > 0\}$ is a *Poisson process* with mean rate $\nu$ iff:

1. $N(0) = 0$
2. The number of random events that occur in an interval is independent of the number of random events that occur in any other non overlapping interval
3. For any interval $(s, s+t)$ with $s \geq 0$ and $t \geq 0$, the number of random events within this interval follows a Poisson distribution with $\lambda = \nu t$

$$P\left[N(s+t) - N(s) = x\right] = \frac{e^{-\nu t}(\nu t)^x}{x!} \text{ for } x = 0, 1, 2, \ldots,$$

# Exponential Distribution

A r.v. $X$ follows an *exponential distribution* with parameter $\lambda$ if its pdf is:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x > 0 \\ 0 & \textit{elsewhere} \end{cases}$$

Expected Value and Variance:

$$E[X] = \frac{1}{\lambda} \qquad V[X] = \frac{1}{\lambda^2}$$

# Exponential Distribution – Memoryless Property

Any exponential r.v. $X$ carries the *memoryless property*:

$$P(X > s + t | X > t) = P(X > s)$$

Proof:

$$P(X > t + s | X > t) = \frac{P(X > t + s \cap X > t)}{P(X > t)} = \frac{P(X > t + s)}{P(X > t)}$$

$$= \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = P(X > s)$$

*Out of all continuous distributions only exponential carries the memoryless property.*

# Exponential Distribution and Poisson Process

The exponential distribution is connected to a Poisson Process since the times between sequential events are exponentially distributed.

Let $\{N(t), t > 0\}$ be a Poisson process with mean rate $\nu$, then the r.v. $T_i$, for $i = 1, 2, \ldots$, where $T_i$ is the time between the $(i-1)$-th and $i$-th events, are independent and exponentially distributed with parameter $\nu$.

We conclude that in a Poisson Process the # of events in any interval is distributed according to a Poisson distribution and the times between sequential events is distributed according to an exponential distribution.

# Normal Distribution

A r.v. $X$ follows a *normal distribution* with parameter $\mu$ and $\sigma$ if its pdf is:

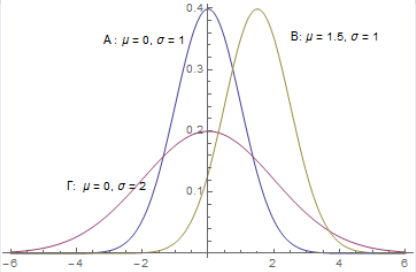
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} \text{ for } -\infty < x < \infty$$

Expected Value and Variance:

$$E[X] = \mu \qquad V[X] = \sigma^2$$

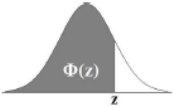
# Normal Distribution Graph – $\mu =??$, $\sigma =??$

pdf:




- symmetric around its mean $\mu$.
- if $\mu = 0$ and $\sigma^2 = 1$ it is called *standard normal distribution*
- cdf cannot be written in closed form
- use of tables for calculating probabilities

# Standard Normal Distribution – Table




$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 | .09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |
| 3.5 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 | 0.9998 |
| 3.6 | 0.9998 | 0.9998 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 | 0.9999 |

# Standardization of Normal Distribution

1. If a r.v. $X$ is normally distributed with parameter $\mu$ and $\sigma$, then $Z = (X - \mu)/\sigma$ has the standard normal distribution
2. Using the table for the standard normal distribution we can calculate any probability for all normal distributions

$$
\begin{aligned}
P(a \leq X \leq b) &= P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) \\
&= P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right) \\
&= \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)
\end{aligned}
$$