PROBABILITIES & RANDOM VARIABLES

"*Some of the Problems about Chance having a great appearance of Simplicity, the Mind is easily drawn into a belief, that their Solution may be attained by the meer Strength of natural good Sense; which generally proving otherwise and the Mistakes occasioned thereby being not infrequent, 'tis presumed that a Book of this Kind, which teaches to distinguish Truth from what seems so nearly to resemble it, will be looked upon as a help to good Reasoning.*"

Abraham de Moivre (1667-1754)

OUTLINE

- Probabilities: Introduction
- Odds and ratio
- Properties of Probabilities
- Conditional probabilities
- Random variables

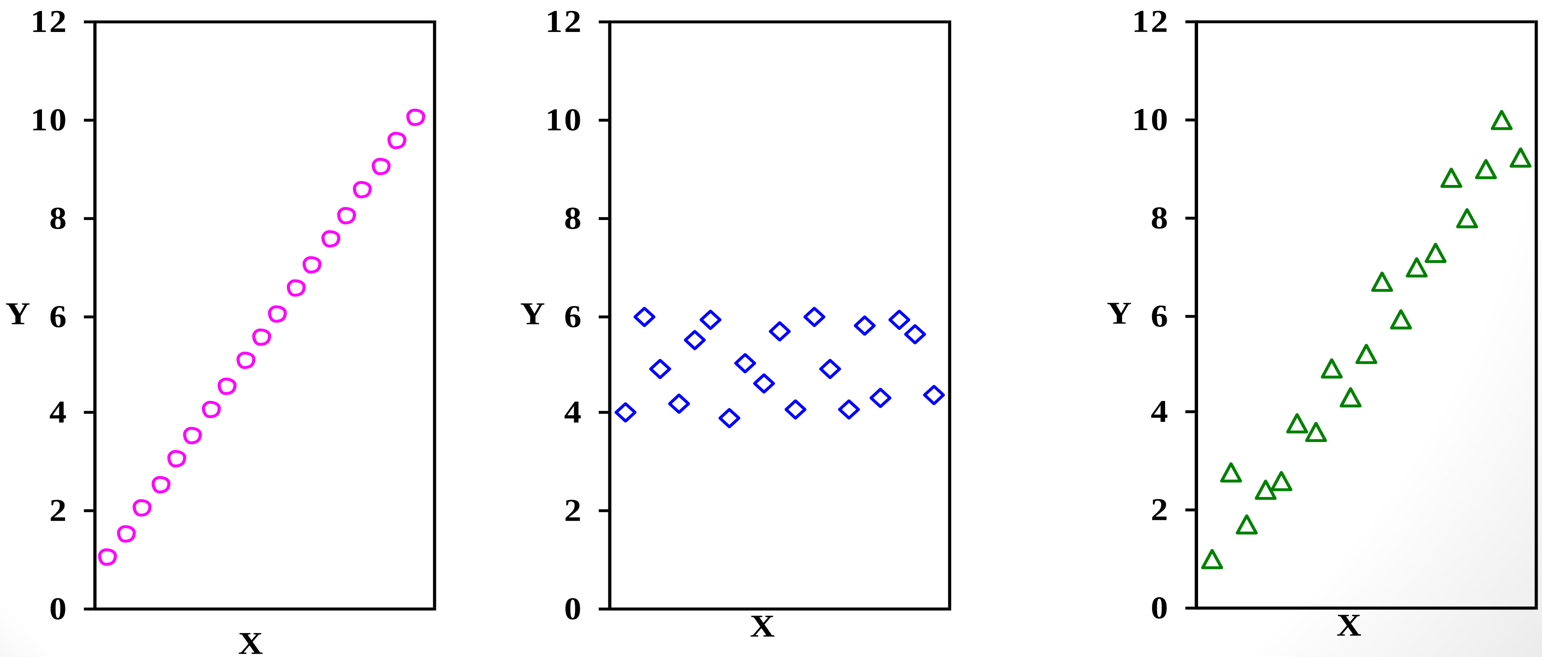
PROBABILITY

- Basic concepts
- Conditional probabilities
- Probability Properties
- Probability Rules

- Probability is a way of expressing knowledge or belief that an event will occur or has occurred

Probability

- Data could be generated by a:
  - Purely systemic process
  - Purely random process
  - Combination of systemic and random processes



Hypothesis

- We would like to know which of the three explanations is most likely correct
- The purely systematic process is easy to confirm or reject if we look at the graphical distribution of data (this kind of process is rarely seen in medical sciences)

- Thus, we have two questions:
  - Is the process purely random?
  - Is the process a combination of a systemic and random processes? (complex model)

Random Process

Definitions:

- **Test**: application of an experiment over an element of population or sample
- **Event**: the result of a test
- **Random event**: the event which appear as result of a single test

BINOMIAL RANDOM EVENT

- In throwing a coin we have two possible outcomes (heads or tails) associated with a specified probability (0.5)
- A probability of an event A is represented by a real number in the range from 0 to 1 and written as P(A), p(A) or Pr(A)
- We can not say much about the absolute frequency of one of two possible events for a few throws but we can say about the relative frequency obtained from the coin flipped several times.

RANDOM EVENTS

- When we say something can be described by a random generating process, we do NOT necessarily mean that it is caused by a mystical thing called "chance"

- There may be many independent systematic factors that combine together to create the observed random probability distribution. (e.g. coin tosses)

- When we say "random" we just mean that we ca not do any better than some basic (but characteristic) probability statements about how the outcomes will vary

PROBABILITY

- Probabilities are numbers which describe the likelihoods of random events.

$$\mathbf{Pr(A)} \in \mathbf{[0, 1]}$$

- Let A be an event:
  - Pr(A) = the probability of event A
  - If A is certain, then Pr(A) = 1
  - If A is impossible, then Pr(A) = 0

Subjective vs Objective Probability

Subjective probability:

- Established subjective (empiric) base on previous experience or on studying large populations
- Implies elementary that are not equipossible (equally likely)

Objective probability:

- Equiprobable outcomes
- Geometric probability

Formula of calculus:

- If an A event could be obtained in S tests out of n equiprobable tests, then the Pr(A) is given by the number of possible cases
- Pr(A) = (no of favorable cases)/(no of possible cases)

Chances and Odds

- Chances are probabilities expressed as percents.
  - Range from 0% to 100%.
  - Ex: a probability of 0.65 is the same as a 65% chance.
- The odds for an event is the probability that the event happens, divided by the probability that the event doesn't happen.
  - Can take any positive value
  - Let A be the event. Odds(A) = Pr(A)/[1-Pr(A)]
    - Where 1-Pr(A) = Pr(nonA)
  - Example: Pr(A) = 0.75; a probability of 0.75 is the same as 3-to-1 odds (0.75/(1-0.75)=0.75/0.25=3/1)

Events Space

- Is a list of all possible outcomes of a random process
  - When we roll a die, the events space is {1, 2, 3, 4, 5, 6}
  - When I toss a coin, the events space is {head, tail}.
- An event is a member of events space
  - "head" is a possible event when I toss a coin
  - "a number less than 4" is a possible event when I roll a die
- **Events are associated with probabilities!**

PROBABILITY PROPERTIES

- Take values between 0 and 1:

$$\mathbf{0 \leq \ Pr(A) \leq \ 1}$$

- Pr(evets space) = 1
- The probability that something happens is one minus the probability that it does not:

$$\mathbf{Pr(A) = 1 - Pr(nonA)}$$

Probability be Example

- If equally likely outcomes:
- Pr(A) = (outcomes favorable to event A)/(outcomes total)
- What is the probability of getting exactly 2 heads in three coin tosses? (Let A be the event of getting 2 heads in 3 coin tosses.)

HHH	HTT
HHT	TTT
HTH	TTH
THH	THT

- Outcomes with exactly 2 heads = 3
- Total possible outcomes = 8
- Pr(A) = 3/8

PROBABILITY

- Compatible events: events that can occur simultaneously:
  - A = {SBP < 140 mmHg}
  - B = {DBP < 90 mmHg}
  - SBP = systolic blood pressure; DBS = diastolic blood pressure
- Incompatible events: events that can not occur simultaneously:
  - A = {SBP < 140 mmHg}
  - B = {140 ≤ SBP < 200 mmHg}

Probability

- Event A imply event B IF the event B is produce any time when even A is produce:
  - Symbol $A \subset B$
    - $A = \{TBC\}$
    - B = {positive tuberculin test}
    - TBC = tuberculosis

CONDITIONAL PROBABILITY

- Let A and B be two events:
  - The conditional probability of B, given A, is written as Pr(B|A). It is the probability of event B, given that A has occurred.
- Example: (Tuberculin Test+|TBC) is the probability of obtaining a positive tuberculin test to a patient with tuberculosis
- **Pr(B|A)** is not the same things as **Pr(A|B)**

Conditional Probability

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Let:
  - A = {TBC+}
  - B = {Tuberculin Test+}

- Pr(A) = (15+25)/(15+12+25+18) = 0.57 (prevalence)
- Pr(nonA) = (12+18)/(15+12+25+18) = 0.43
- Pr(B|A) = probability of a positive tuberculin test to a patient with TBC = 15/(15+25) = 0.38 = **Sensibility (Se)**

Conditional Probability

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Let:
  - A = {TBC+}
  - B = {Tuberculin Test+}

- Pr(nonB|nonA) = probability of obtaining a negative test to a patient without TBC = 18/(18+12) = 0.60 = **Specificity (Sp)**
- Pr(A|B) = probability that a person with TBC to have a positive tuberculin test = 15/(15+12) = 0.56 = Predictive Positive Value **(PPV)**

CONDITIONAL PROBABILITY

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Let:
  - A = {TBC+}
  - B = {Tuberculin Test+}

- Pr(nonA|nonB) = probability that a person without TBC to have a negative tuberculin test = 18/(18+25) = 0.42 = **Negative Predictive Value (NPV)**

CONDITIONAL PROBABILITY

	TBC+	TBC-
Test+	15	12
Test-	25	18

- Let:
  - A = {TBC+}
  - B = {Tuberculin Test+}

- Positive False Ratio: PFR = Pr(B|nonA)
- Negative False Ratio: RFN = Pr(nonA|B)

INDEPENDENT EVENTS: CONDITIONAL PROBABILITIES

- Events A and B are independent if the probability of event B is the same whether or not A has occurred.
- Two events A and B are Independent IF

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

- If (and only if) A and B are independent, then:
  - $\Pr(B|A) = \Pr(B|\text{non}A) = \Pr(B)$
  - $\Pr(A|B) = \Pr(A|\text{non}B) = \Pr(A)$
- It expressed the independence of the two events: the probability of event B (respectively A) did not depend by the realization of event A (respectively B)
  - Example: if a coin is toss twice the probability to obtain “head” to the second toss is always 0.5 and is not depending if at the first toss we obtained “head” or “tail”.

Joint Probability

- **Reunion (OR):**
  - Symbol: $A \cup B$
    - At least one event (A OR B) occurs
- **Intersection (AND)**:
  - The probability that A and B both occur
  - Use the multiplication rule
  - Symbol: $A \cap B$
    - the events A and B occur simultaneously
- **Negation:**
  - Symbol: nonA

Probability Rules

- **<u>Addition Rule</u>**: probability of A **or** B:

$$\mathbf{Pr(A \text{ or } B) = Pr(A) + Pr(B)}$$

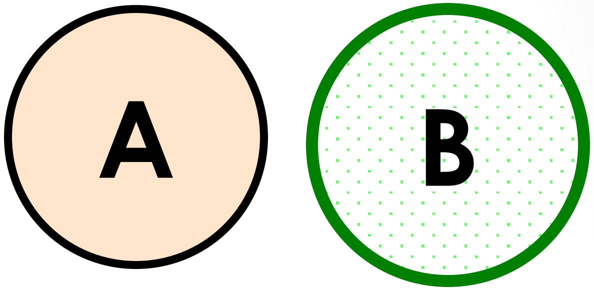
when A and B are mutually exclusive

- **<u>Multiplication Rule</u>**: probability of A **and** B:

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

when A and B are independent

Probability Rules: Addition Rule



- Let A and B be two events:

$$\mathbf{Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)}$$

$$\mathbf{Pr(A\ or\ B) = Pr(A) + Pr(B) - Pr(A\ and\ B)}$$

- A and B mutually exclusive:
  - **$Pr(A \cap B) = 0$**
  - **$Pr(A\ and\ B) = 0$**

Probability Rules: Addition Rule

- A = {SBP of mother > 140 mmHg}
  - $Pr(A) = 0.25$
- B = {SBP of father > 140 mmHg}
  - $Pr(B) = 0.15$
- What is the probability that mother or father to have hypertension?

$$\mathbf{Pr(A \cup B) = 0.25 + 0.15 - 0 = 0.40}$$

$$\mathbf{Pr(A\ or\ B) = 0.25 + 0.15 - 0 = 0.40}$$

Probability Rules: Addition Rule

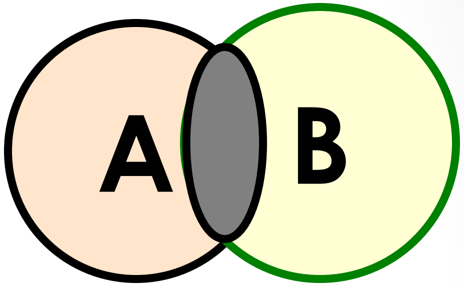
- In a cafe are at a moment 20 people, 10 like tea, 10 like coffee and other 2 like tea and coffee.
- What is probability to random extract from this population one person who like tea or coffee?

$$Pr(tea \cup coffee) = Pr(tea) + Pr(coffee) - Pr(tea \cap coffee)$$

$$Pr(\text{tea or coffee}) = Pr(\text{tea}) + Pr(\text{coffee}) - Pr(\text{tea and coffee})$$

$$Pr(\text{tea or coffee}) = 0.50 + 0.50 - 0.10 = 0.90$$

PROBABILITY RULES: MULTIPLICATION RULE



- Let A and B be two events:

$$\mathbf{Pr(A \cap B) = Pr(A) \cdot Pr(B|A)}$$

$$\mathbf{Pr(A\ and\ B) = Pr(A) \cdot Pr(B|A)}$$

- Independent events: $\mathbf{Pr(B|A) = Pr(B)}$

Probability Rules: Multiplication Rule

- A = {SBP of mother > 140 mmHg}
  - $Pr(A) = 0.10$
- B = {SBP of father > 140 mmHg}
  - $Pr(B) = 0.20$
- $Pr(A \cap B) = 0.05$; $Pr(A \text{ and } B) = 0.05$
- The two events are dependent or independent?

$Pr(A \cap B) = Pr(A) \cdot Pr(B)$ – independent events

$0.05 \neq 0.10 * 0.20 \rightarrow$ the events are dependent

Binomial Random Processes

- Two possible outcomes
- Heads or tails
- Make basket or miss basket
- Fatality, no fatality

- With probability p (or 1-p)

BINOMIAL RANDOM PROCESSES

- Predicting a specific versus a general pattern
- Which lotto ticket would you buy?
  - Equally likely (or unlikely) to win:
    - 26 45 8 72 91
    - 26 26 26 26 26 – less likely to be bought
  - Each specific ticket is equally (un)likely to win
  - A ticket that "looks like" ticket A (with alternating values) is more likely than one that "looks like" ticket B (with identical values).

Binomial Random Processes

- Probabilities for specific patterns get smaller as you run more tests
- What is the probability of getting heads on the second test and the tails on all other trials?
  - P(T,H) = 0.25
  - P(T, H, T) = 0.125
  - P(T, H, T, T) = 0.0625

Binomial Random Processes

- What is the probability of getting at least one heads when you toss a coin multiple times?
  - Two tosses: Pr(HT or TH or HH) = 0.75
  - Three tosses: Pr(HTT or THT or TTH or THH or HHT or HHH) = 0.875
  - Four tosses: 0.9375

SUMMARY

- **Addition rules:**
  - $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
  - Pr(A or B) = Pr(A) + Pr(B) – Pr(A and B)
  - **Mutually exclusive events:**
    - $\Pr(A \cup B) = \Pr(A) + \Pr(B)$
    - Pr(A or B) = Pr(A) + Pr(B):
- **Multiplication Rule:**
  - $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B|A)$
  - Pr(A and B) = Pr(A)·Pr(B|A)
  - **Independent events:**
    - $\Pr(A \cap B) = \Pr(A) \cdot \Pr(B)$

Random Variables
&
Discrete Probability Distribution

Random variable … What about them?

- Definition
- Introduction: probability distribution
  - Metric data:
    - Discrete
    - Continuous

DEFINITION

- Let X be a quantitative variable measured or observed from an experiment
- The value of X is a random variable

- **Example**:
  - Number of red cells in blood
  - Number of bacteria from the students hands
  - Average of depression score obtained from the application of a test on a sample of patients with malign tumors

Random Variables

- Arithmetic mean
- Standard deviation
- Proportion
- Frequency
  - All are random variables

TYPES OF RANDOM VARIABLES

Discrete:

- Can take a finite number of values
  - The number of peoples with RH- from a sample
  - The number of children with flue from a collectivity
  - The number of anorexic students from university
  - Pulse

Continuous:

- Can take an infinite number of values into a defined range
- Vary continuously in defined range
  - Body temperature
  - Blood sugar concentration
  - Blood pressure

Types of Random Variables

- Generally, means are continuous random variables and frequencies are discrete random variables
- Examples:
  - The mean of lung capacity of a people who work in coal mine
  - The number of patients with chronic B hepatitis hospitalized in Cluj-Napoca between 01/11-05/11/2008.

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

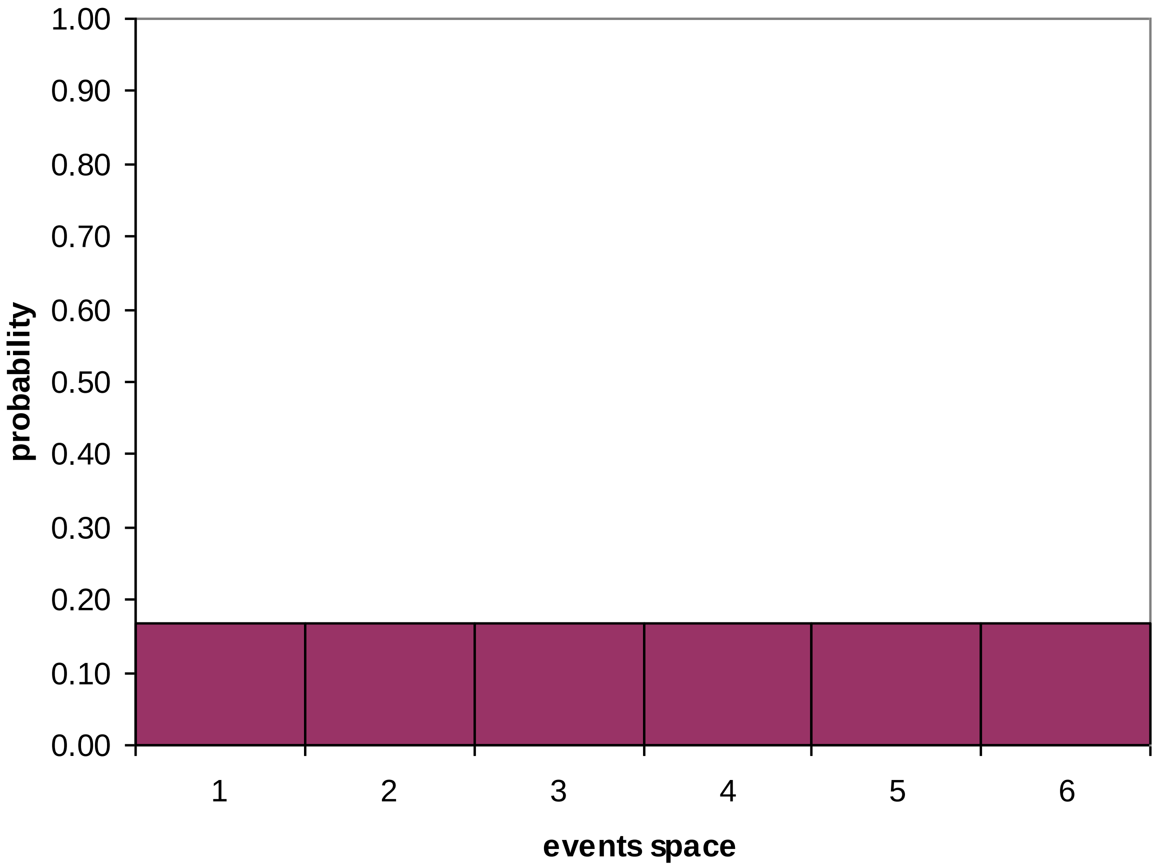
- The probability of X distribution: list of values from the events space and associated probabilities

- 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

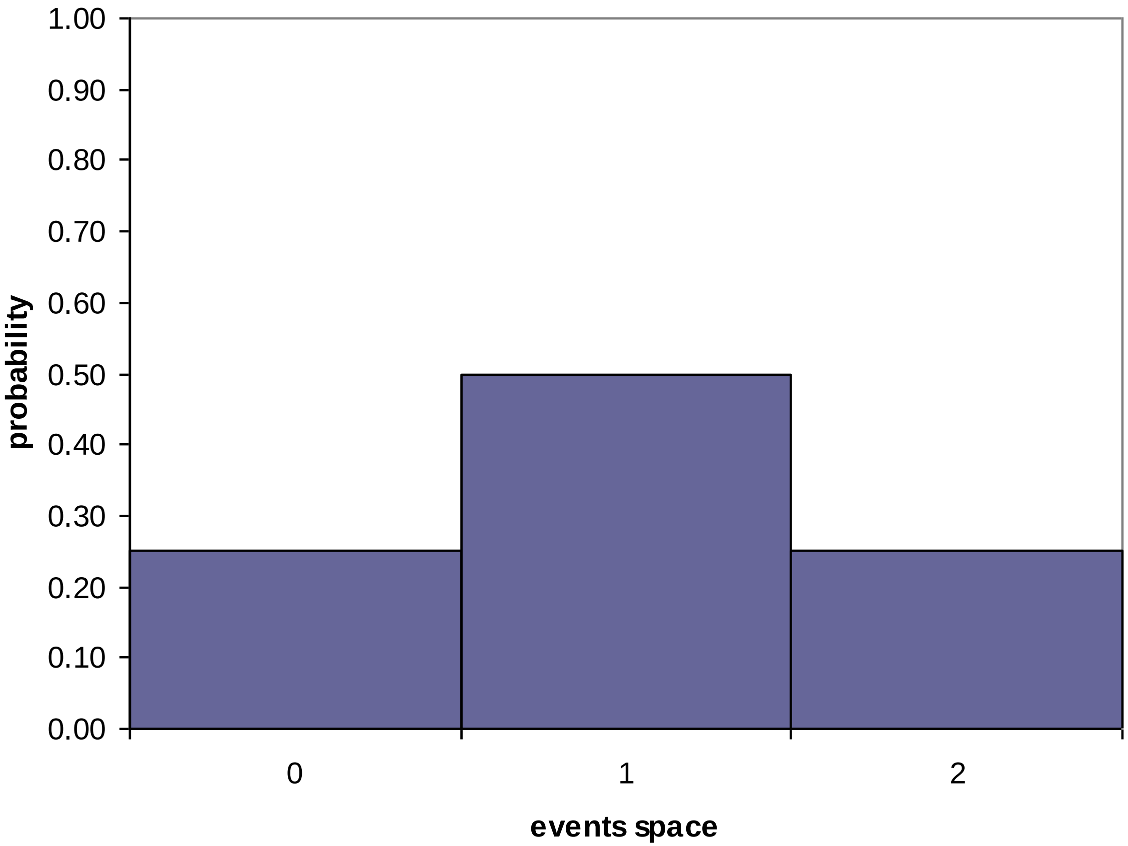
- The probability of X lists the values in the events space and their associated probabilities



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: V(X), σ(X)

X_i	$Pr(X_i)$	$X_i*Pr(X_i)$	X_i-M(X)	$(X_i\text{-}M(X))^2$	$(X_i\text{-}M(X))^2*Pr(X_i)$
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**

KNOWN DISCRETE DISTRIBUTIONS

- **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

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

- 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

- The mean or expected value of a binomial distribution is:

$$M(X) = n \cdot p$$

- Variance:

$$V(X) = n \cdot p \cdot q$$

- Standard deviation:

$$\sigma(X) = \sqrt{(n \cdot p \cdot q)}$$

Binomial Distribution

- 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)=10·0.47^2·0.53^3**
- **Pr(X=2) = 0.33**

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

$$C_5^2=\frac{5!}{2!\cdot(5-2)!}=\frac{5\cdot4\cdot3\cdot2\cdot1}{2\cdot1\cdot(3\cdot2\cdot1)}=\frac{120}{12}=10$$

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:

$$X : \begin{pmatrix} k \\ e^{-\theta} \cdot \frac{\theta^k}{k!} \end{pmatrix} \qquad \Pr(X = k) = \frac{e^{-\theta} \cdot \theta^k}{k!}$$

POISSON DISTRIBUTION

- Mean of expected values:

$$M(X) = \theta$$

- Variance:

$$V(X) = \theta$$

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?
- $n=400$
- $p=7/1000=0.007$

- $\theta=n\cdot p=400\cdot 0.007=2.8$
- $e=2.718281828=2.72$

$Pr(X=5) =$

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

$=10.45/120$

$=0.09$

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.