[AIMLBD] MACHINE LEARNING, BIG DATA, ARTIFICIAL INTELLIGENCE per medicina e chirurgia high tech

L01a: Probability

Dott. Giorgio De Magistris demagistris@diag.uniroma1.it

Corso di Laurea in Medicina e Chirurgia High Tech

Sapienza Università di Roma

Facoltà di Ingegneria dell'Informazione, Informatica e Statistica

Dipartimento di Ingegneria Informatica, Automatica e Gestionale

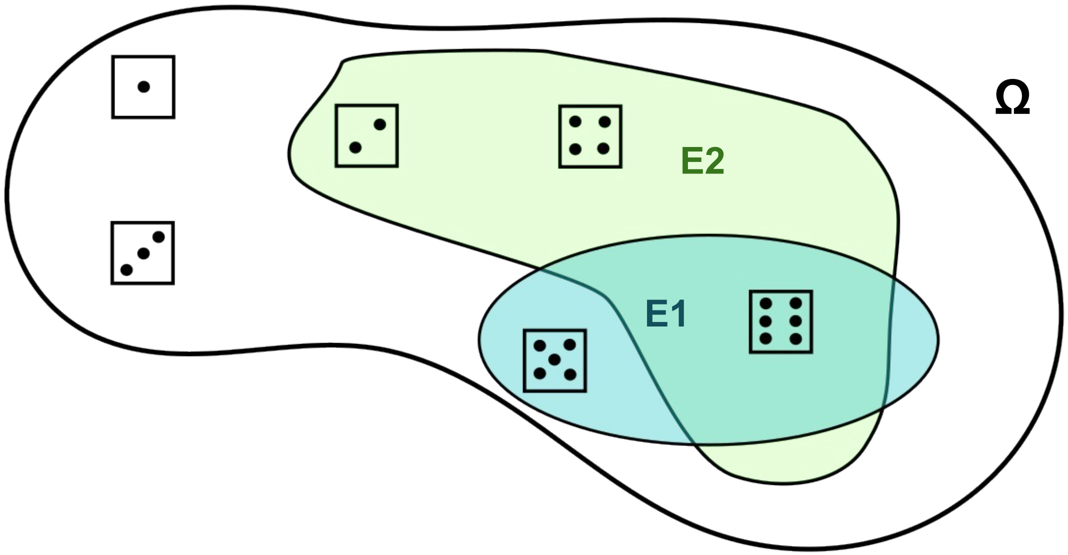
What is probability?

- Probability theory tells us how to measure uncertainty

- For example it allows to answer questions like:
  - What is the probability that rolling a die will produce 6?
  - What is the probability that, rolling two dice, the second produces 6, knowing that the first gave 6?
  - What is the probability that, rolling two dice, the sum of the outcomes is greater than 6?
  - What is the probability that, rolling two dice, the sum of the outcomes is greater than 6, knowing that the first die produced 3?

Sample Space

- An event is a set of outcomes of an experiment
- The sample space Ω is the set of all possible outcomes of the experiment
- The probability is the uncertainty of the outcomes
- It is a unitary mass that is spread out over the sample space

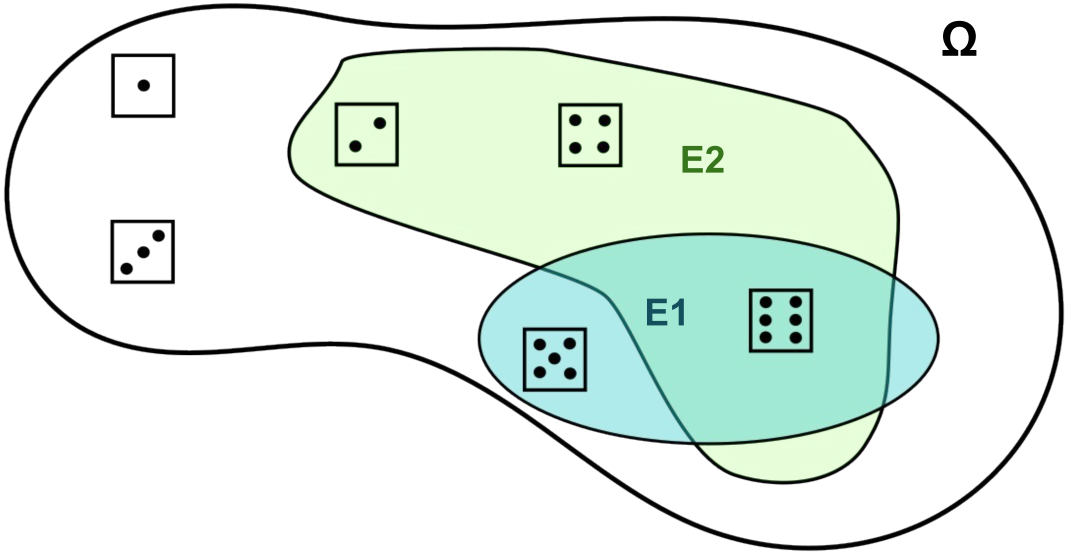


Ω = Sample space
= All the possible
outcomes of the
experiment: "rolling
of a die"

- Pr(E) = # of outcomes in Ei / # of outcomes in Ω
- 0 <= Pr(E) <= 1
- Pr(Ω) = 1

Events

- Consider the previous example
- E1 = “from the die comes a number greater than 4”
- E2 = “from the die comes an even number”



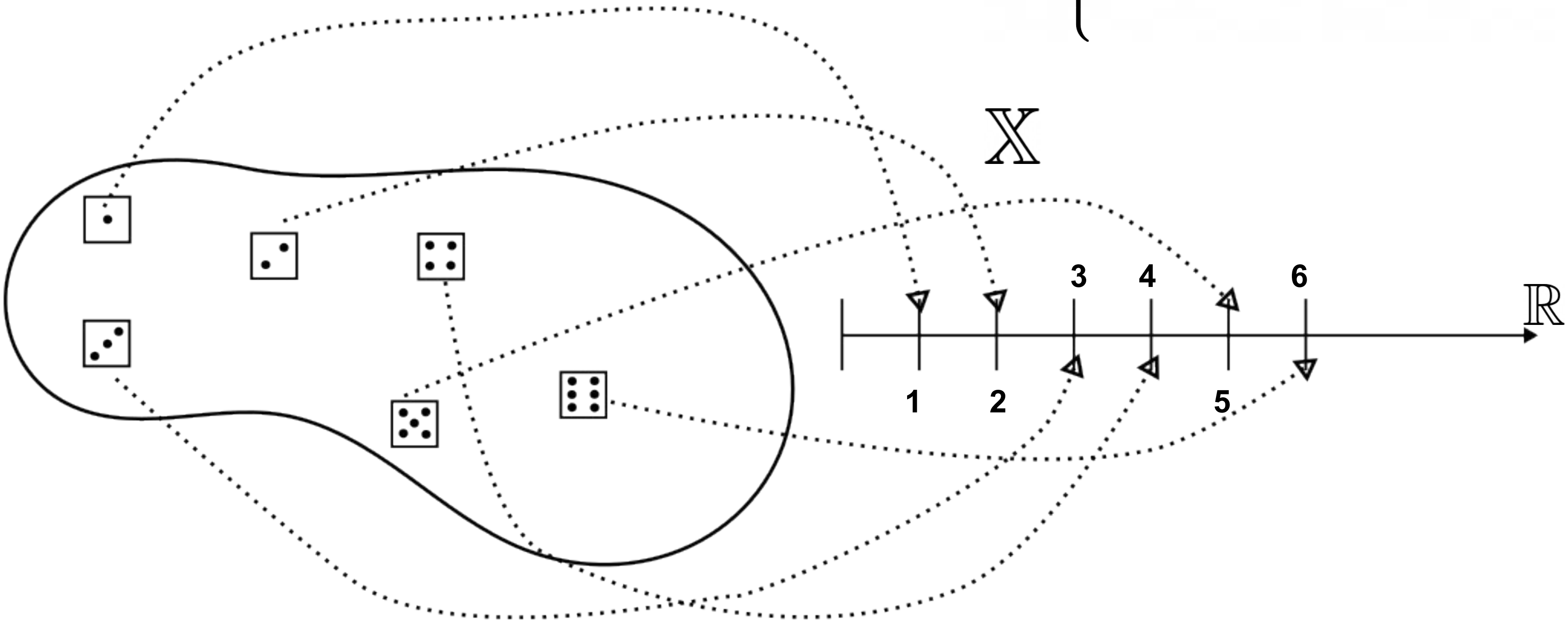
- Pr(E1) = # outcomes in E1 / # outcomes in Ω = 2/6
- Pr(E2) = # outcomes in E2 / # outcomes in Ω = 3/6

Random Variables

- A random variable (RV) X is a function from the sample space of an experiment E to the real numbers

$$\mathbb{X} : \Omega \to \mathbb{R}$$

$$\mathbb{X} = \begin{cases} 1 & if \quad E = Die\ produced\ 1 \\ \dots & \\ 6 & if \quad E = Die\ produced\ 6 \end{cases}$$



Probability Distribution

- A probability distribution tells how uncertainty is distributed among the outcomes of the RV

$$Pr(\{\text{“die produced 6”}\}) = Pr(\mathbb{X} = 6)$$

- Since the domain of the RV is the set of possible outcomes of the experiment (sample space) then:

$$\sum_x Pr(\mathbb{X} = x) = 1$$

- I can represent a discrete probability distribution (a probability distribution over a discrete RV) in tabular form:

X = {outcome of the roll of a die}					
1	*2*	*3*	*4*	*5*	*6*
1/6	1/6	1/6	1/6	1/6	1/6

Joint Probability Distribution

- Given two experiments, represented through the random variables X and Y
- The Joint probability distribution tells for each outcome of the RV X and for each outcome of the RV Y, what is the probability that the two outcomes happen together
- Again I can represent a discrete joint probability distribution in tabular form:

		Condition of the sky	
Presence of Rain		*Cloudy*	*Sunny*
	Raining	30%	1%
	Not Raining	9%	60%

Sum Rule (Marginalization)

- From the Joint distribution I can derive everything about the experiments, like the individual probability distribution (applying the sum rule) or the conditional probabilities (see later)

- Given two random variables X,Y and the joint probability distribution Pr(X=x,Y=y), we can obtain the individual probability distribution Pr(X=x) summing all the outcomes of the variable Y

$$Pr(\mathbb{X} = x) = \sum_{y} Pr(\mathbb{X} = x, \mathbb{Y} = y)$$

$$\mathbb{X} = \begin{cases} 1 \ \ if \ \ \text{cloudy} \\ 2 \ \ if \ \ \text{sunny} \end{cases}$$

$$\mathbb{Y} = \begin{cases} 1 \ \ if \ \ \text{raining} \\ 2 \ \ if \ \ \text{not raining} \end{cases}$$

X				**Pr(Y = y)**
Y		*X = 1*	*X = 2*	
	Y = 1	30%	1%	31%
	Y = 2	9%	60%	69 %
Pr(X = x)		39%	61%	

Conditional Probability

- Given two random variables X,Y the conditional probability Pr(X=x | Y=y) (probability of X=x given Y=y) expresses the probability of X=x given that the event Y=y happened

- It is obtained by $Pr(\mathbb{X}=x|\mathbb{Y}=y)=\dfrac{Pr(\mathbb{X}=x,\mathbb{Y}=y)}{Pr(\mathbb{Y}=y)}$

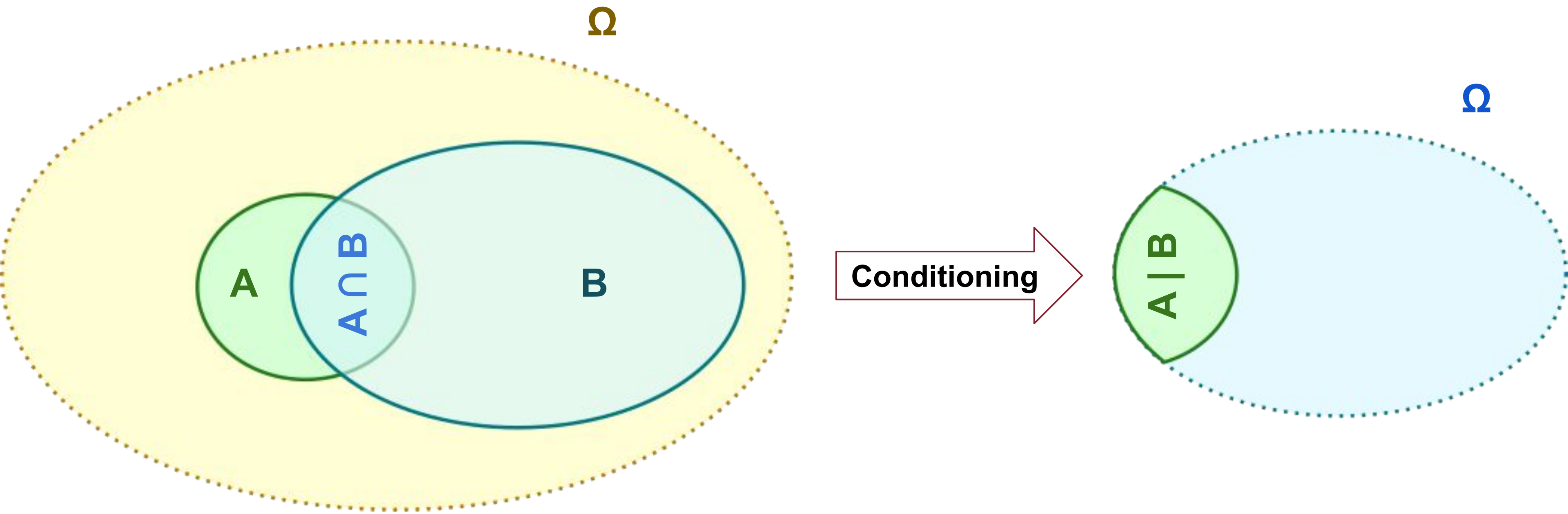
- Using the same example as before:

$$Pr(rain|cloudy)=Pr(\mathbb{Y}=y|\mathbb{X}=1)=\frac{Pr(\mathbb{Y}=y,\mathbb{X}=1)}{Pr(\mathbb{X}=1)}$$

Pr(Y=y | X=1)	
Y=1	*Y=2*
0.3/0.39 = 0.77	0.09/0.39 = 0.23

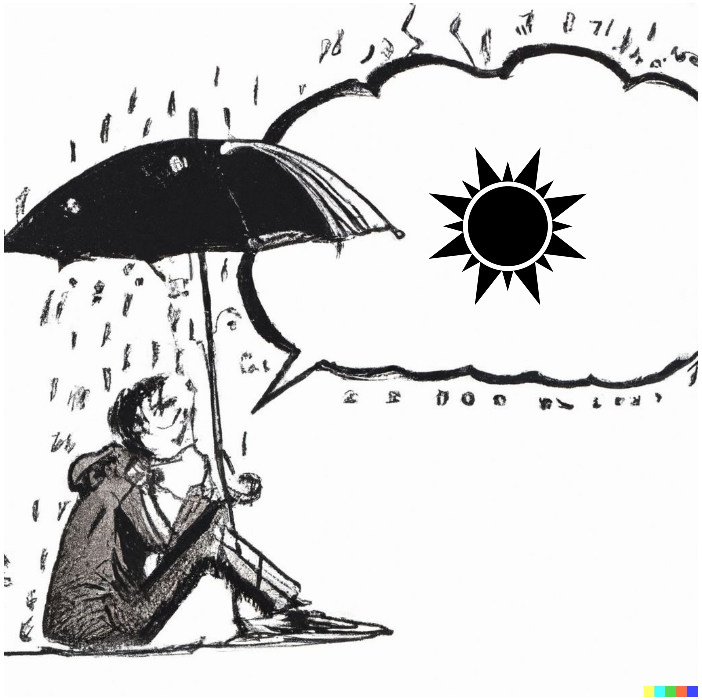
Conditional Probability Visualized

- Conditioning an event A with an event B consists in restricting the sample space to B and normalize in order to make the probability sum to one



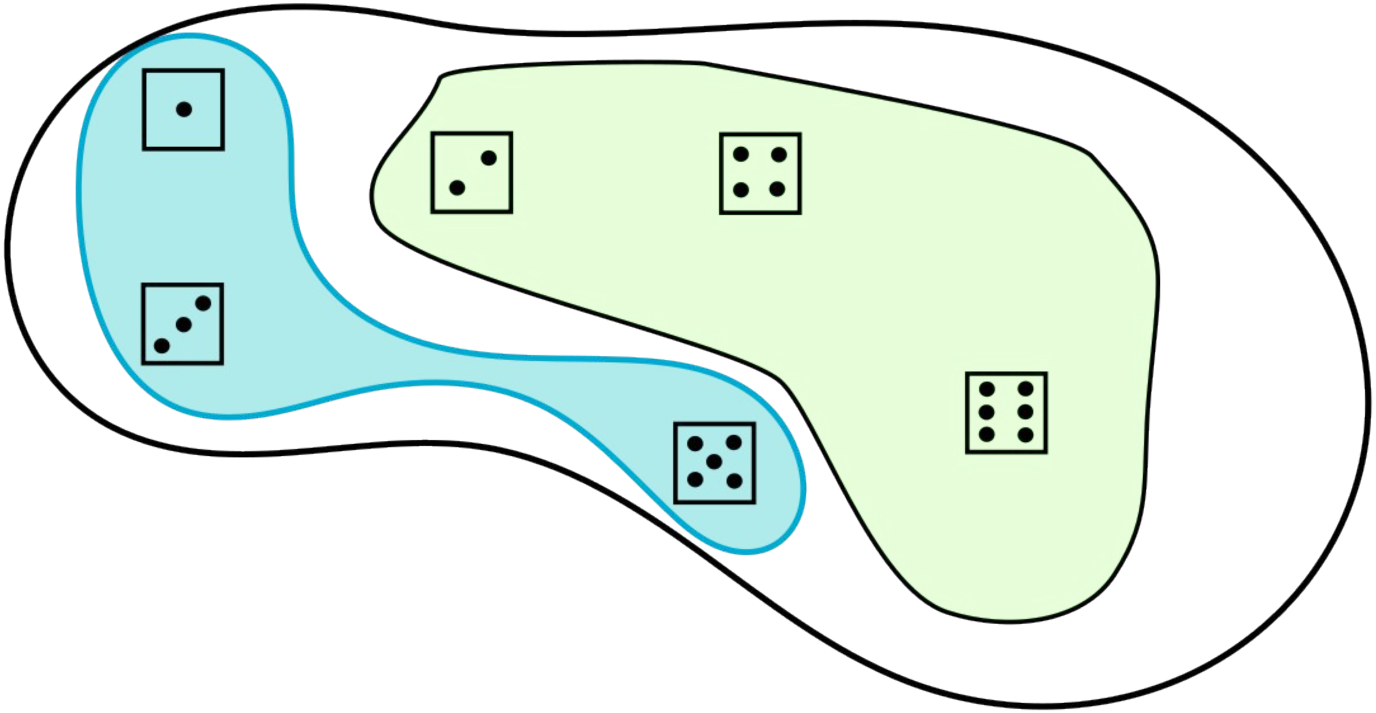
Independence

- Two RV X and Y are independent if and only if the outcome of Y does not give me information about the outcome of X and vice versa:
  - $\Pr(X=x \mid Y=y) = \Pr(X=x)$
- By applying the definition of conditional probability I get
  - $\Pr(X=x \mid Y=y) = \Pr(X=x, Y=y) / \Pr(Y=y)$
- Then if X and Y are independent I get
  - $\Pr(X=x, Y=y) = \Pr(X=x) \cdot \Pr(Y=y)$
- From the previous example
  - $\Pr(\text{rain} \mid \text{cloud}) = 0.77$
- That is different from
  - $\Pr(\text{rain}) = 0.31$
- Hence the two RV are not independent



Disjoint Events

- Two events A and B are disjoint if they have no outcome in common



Disjoint Events and Conditional Independence

- If two events A and B are disjoint, what can I say about the conditional probability
  - $Pr(A \mid B)$ or $Pr(B \mid A)$?

Disjoint Events and Conditional Independence

- If two events A and B are disjoint, what can I say about the conditional probability
  - $Pr(A \mid B)$ or $Pr(B \mid A)$?

- If two events are disjoint they are NOT Independent, indeed if A happens I'm sure that B will not happen and vice versa
  - $Pr(A \mid B) = Pr(B \mid A) = 0$

Example 1

- What is the probability that, rolling two dice, they both get 6?

Example 1

- What is the probability that, rolling two dice, they both get 6?

- Pr(X=6,Y=6) = 1/36

		X = {outcome of the first die}					
Y = {outcome of the second die}		*1*	*2*	*3*	*4*	*5*	*6*
	1	1/36	1/36	1/36	1/36	1/36	1/36
	2	1/36	1/36	1/36	1/36	1/36	1/36
	3	1/36	1/36	1/36	1/36	1/36	1/36
	4	1/36	1/36	1/36	1/36	1/36	1/36
	5	1/36	1/36	1/36	1/36	1/36	1/36
	6	1/36	1/36	1/36	1/36	1/36	1/36

Example 2

- What is the probability that, rolling two dice, the second produces 6, given that the first gave 6?

Example 2

- What is the probability that, rolling two dice, the second produces 6, given that the first gave 6?

$$Pr(\mathbb{X}=6|\mathbb{Y}=6)=\frac{Pr(\mathbb{X}=6,\mathbb{Y}=6)}{Pr(\mathbb{Y}=6)}=\frac{1/36}{1/6}=1/6=Pr(\mathbb{X}=6)$$

		X = {outcome of the first die}						Pr(Y)
Y = {outcome of the second die}		*1*	*2*	*3*	*4*	*5*	*6*	
	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Pr(X)		1/6	1/6	1/6	1/6	1/6	1/6	

Example 3

- What is the probability that rolling two dice, the sum is greater or equal to 6, given that the first die produced 3?

Example 3

- What is the probability that rolling two dice, the sum is greater or equal to 6, given that the first die produced 3?
- We call the RV Z = X + Y

$$Pr(\mathbb{Z} \geq 6 | \mathbb{X} = 3) = \frac{Pr(\mathbb{Z} \geq 6, \mathbb{X} = 3)}{Pr(\mathbb{X} = 3)} = \frac{\sum_{x+y \geq 6, x=3} Pr(\mathbb{X} = x, \mathbb{Y} = y)}{Pr(\mathbb{X} = 3)} = \frac{4/36}{1/6} = \frac{2}{3}$$

		X = {outcome of the first die}						**Pr(Y)**
Y = {outcome of the second die}		*1*	*2*	*3*	*4*	*5*	*6*	
	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Pr(X)		1/6	1/6	1/6	1/6	1/6	1/6	

Example 3

- What is the probability that rolling two dice, the sum is greater or equal to 6, given that the first die produced 3?
- We call the RV Z = X + Y

$$Pr(\mathbb{Z} \geq 6 | \mathbb{X} = 3) = \frac{Pr(\mathbb{Z} \geq 6, \mathbb{X} = 3)}{Pr(\mathbb{X} = 3)} = \frac{\sum_{x+y \geq 6, x=3} Pr(\mathbb{X} = x, \mathbb{Y} = y)}{Pr(\mathbb{X} = 3)} = \frac{4/36}{1/6} = \frac{2}{3}$$

		X = {outcome of the first die}						Pr(Y)
Y = {outcome of the second die}		*1*	*2*	*3*	*4*	*5*	*6*	
	1	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	2	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	3	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	4	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	5	1/36	1/36	1/36	1/36	1/36	1/36	1/6
	6	1/36	1/36	1/36	1/36	1/36	1/36	1/6
Pr(X)		1/6	1/6	1/6	1/6	1/6	1/6	

The chain rule

- The chain rule allows to factor a complex probability distribution into the product of conditional distributions

$$Pr(X_1, ..., X_n|D) = Pr(X_1|X_2, ..., X_n, D)Pr(X_2|X_3, ..., X_n, D)...Pr(X_n|D)$$

- Or using a more compact notation:

$$Pr(X_1, ..., X_n|D) = \prod_{i=1}^{n} Pr(X_i|X_{i+1}...X_n, D)$$

Example 4

- Now assume I have a box with three kinds od die
  - Type 1 biased towards 6 (6 is twice as probable then the others)
  - Type 2 biased towards 1 (1 is twice as probable then the others)
  - Type 3 unbiased

Example 4

- Now assume I have a box with three kinds od die
    - Type 1 biased towards 6 (6 is twice as probable then the others)
    - Type 2 biased towards 1 (1 is twice as probable then the others)
    - Type 3 unbiased
- I can use the random variable H to model the outcomes of the event “I extract one die from the box”

$$\mathbb{H} = \begin{cases} 1 & \textit{if I picked a die of type 1} \\ 2 & \textit{if I picked a die of type 2} \\ 3 & \textit{if I picked a die of type 3} \end{cases}$$

Example 4

- Assume that I also know how many dice for each type there are in the box
  - 10 of type 1
  - 20 of type 2
  - 20 of type 3

Example 4

- Assume that I also know how many dice for each type there are in the box
    - 10 of type 1
    - 20 of type 2
    - 20 of type 3
- Knowing the frequencies I can compute the probability distribution of the random variable H

Example 4

- Assume that I also know how many dice for each type there are in the box
  - 10 of type 1 → $\Pr(H=1) = 1/5$
  - 20 of type 2 → $\Pr(H=2) = 2/5$
  - 20 of type 3 → $\Pr(H=3) = 2/5$
- Knowing the frequencies I can compute the probability distribution of the random variable H

Example 4

- Now I randomly pick and roll a die, what is the probability that the die was unbiased, given that the outcome is 6?

Example 4

- Now I randomly pick and roll a die, what is the probability that the die was unbiased, given that the outcome is 6?

- I can model the event “outcome of the toss” with the random variable X

- I can also write the conditional probabilities
  - $\Pr(X=6 \mid H=1) = 2/7$
  - $\Pr(X=1 \mid H=2) = 2/7$

- Then I’m asking:
  - $\Pr(H=3 \mid X=6)$

Example 4

- Now I randomly pick and roll a die, what is the probability that the die was unbiased, given that the outcome is 6?

- I can model the event “outcome of the toss” with the random variable X

- Then I’m asking:
    - Pr(H=3 | X=6)

- By applying the definition of conditional probability I get

$$Pr(\mathbb{H} = 3|\mathbb{X} = 6) = \frac{Pr(\mathbb{H} = 3, \mathbb{X} = 6)}{Pr(\mathbb{X} = 6)}$$

Example 4

- Then applying the same definition in the opposite direction

$$Pr(\mathbb{H}=3|\mathbb{X}=6)=\frac{Pr(\mathbb{H}=3,\mathbb{X}=6)}{Pr(\mathbb{X}=6)}=\frac{Pr(\mathbb{X}=6|\mathbb{H}=3)Pr(\mathbb{H}=3)}{Pr(\mathbb{X}=6)}$$

- Then I can use marginalization and the product rule to write

$$Pr(\mathbb{X}=6)=\sum_h Pr(\mathbb{X}=6,\mathbb{H}=h)=\sum_h Pr(\mathbb{X}=6|\mathbb{H}=h)Pr(\mathbb{H}=h)$$

- And replacing this expansion into the first equation I get

$$Pr(\mathbb{H}=3|\mathbb{X}=6)=\frac{Pr(\mathbb{X}=6|\mathbb{H}=3)Pr(\mathbb{H}=3)}{\sum_h Pr(\mathbb{X}=6|\mathbb{H}=h)Pr(\mathbb{H}=h)}=\frac{1/6\cdot 2/5}{1/6\cdot 2/5+1/7\cdot 2/5+2/7\cdot 1/5}=0.37$$

Bayes Theorem

- In the previous exercise we derived the Bayes theorem

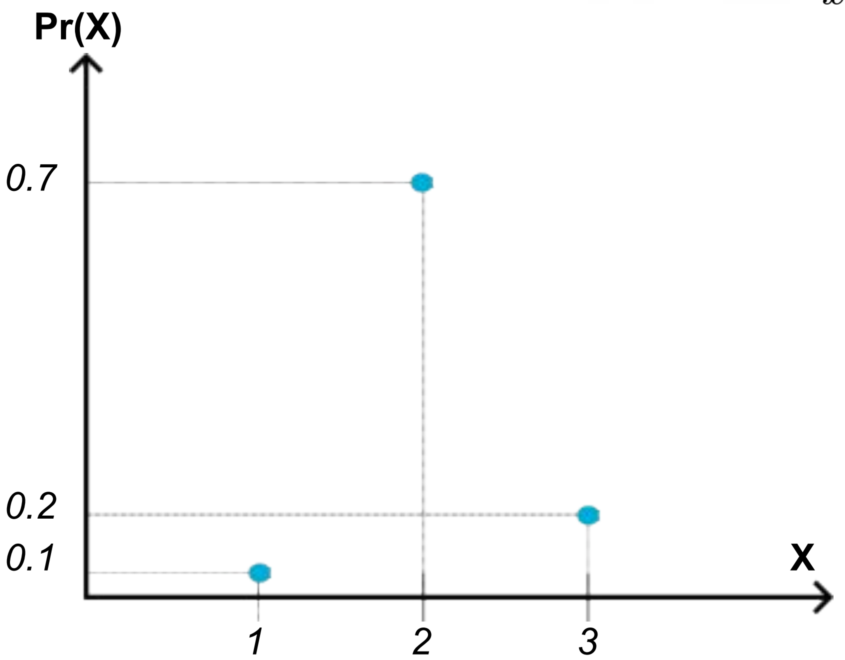
- Bayes' theorem is used to convert a prior probability into a posterior probability by incorporating the evidence provided by the observed data

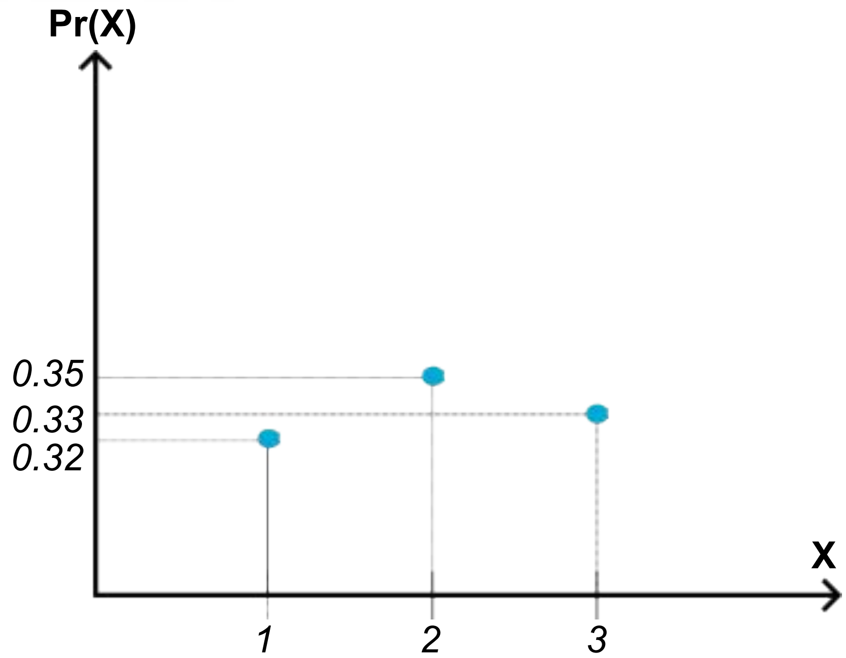
$$\underbrace{Pr(H|E)}_{\text{Posterior Distribution}} = \frac{\overbrace{Pr(E|H)}^{\text{Likelihood}}\overbrace{Pr(H)}^{\text{Prior}}}{Pr(E)}$$

Expected Value

- The expected value is the average outcome of a probabilistic event over multiple trials, weighted by their respective probabilities:

$$\mathbf{E}[X] = \sum_x x Pr(X = x)$$

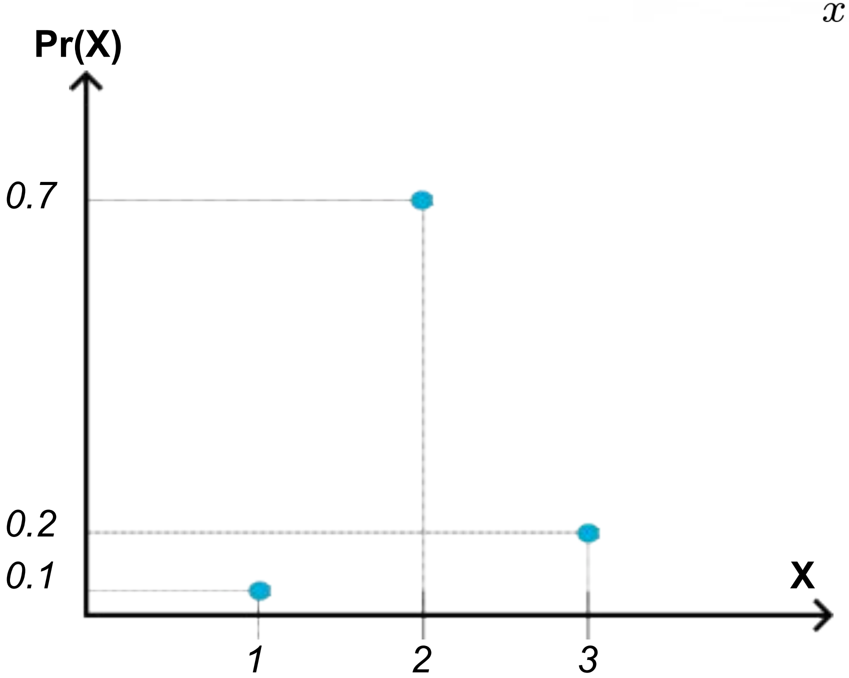


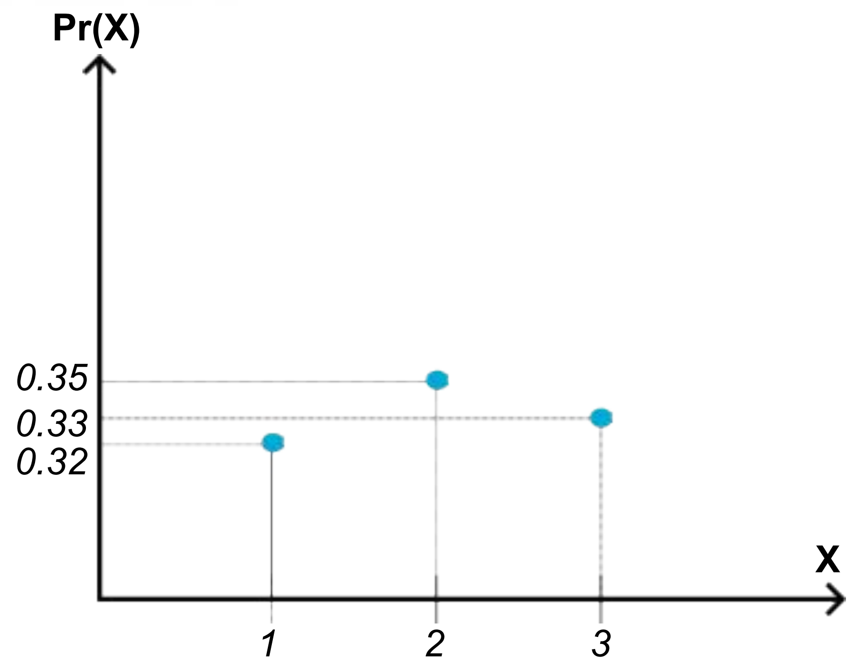


Expected Value

- The expected value is the average outcome of a probabilistic event over multiple trials, weighted by their respective probabilities:

$$\mathbf{E}[X] = \sum_{x} x Pr(X = x)$$





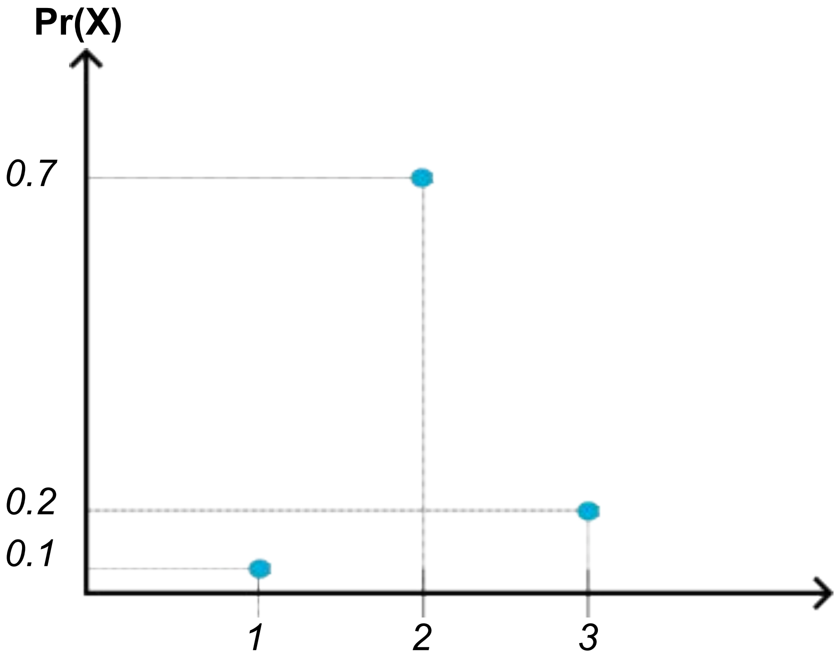
E[X] = 1x 0.10 + 2 x 0.7 + 3 x 0.2 = 2.1

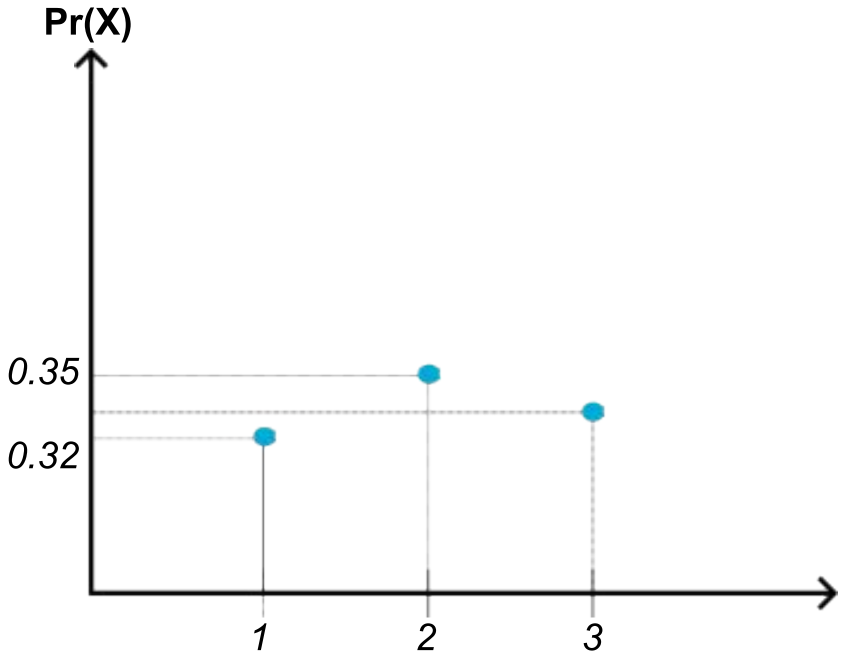
E[X] = 1x 0.32 + 2 x 0.35 + 3 x 0.33 = 2.01

Variance

- Variance is a measure of how much the values of a random variable deviate from its expected value. It quantifies the spread or dispersion of a random variable's values around its mean or expected value.

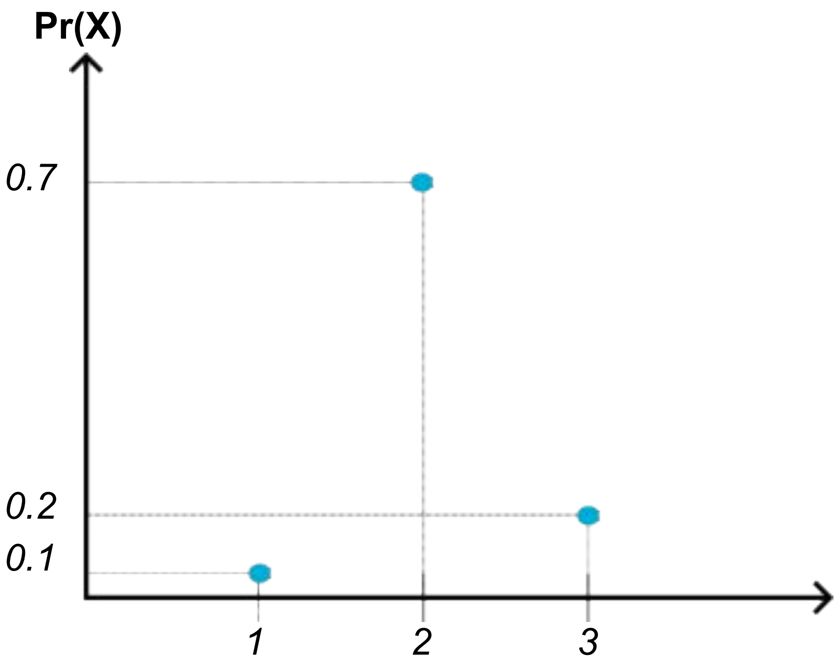
$$\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$

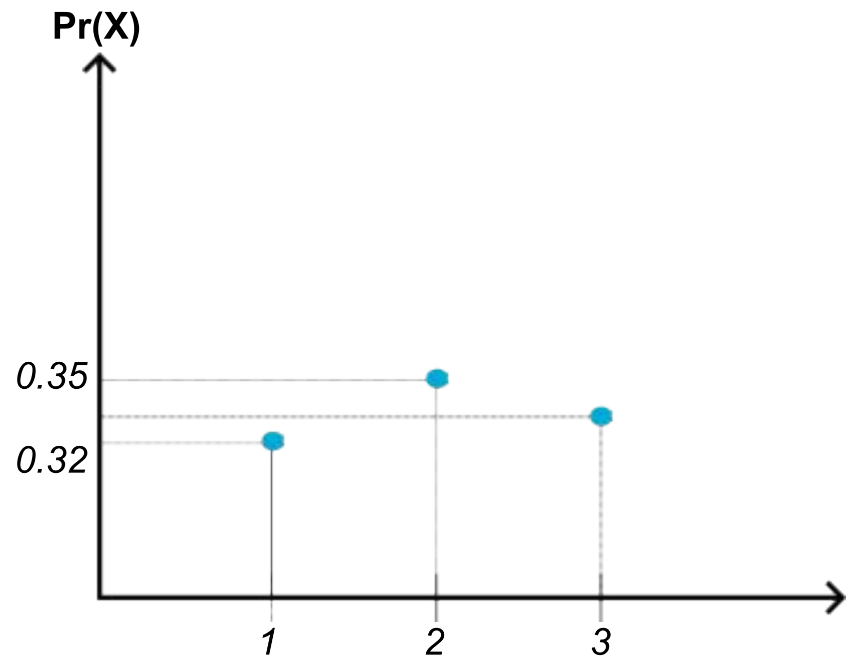




Variance

$$\mathbf{var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2]$$





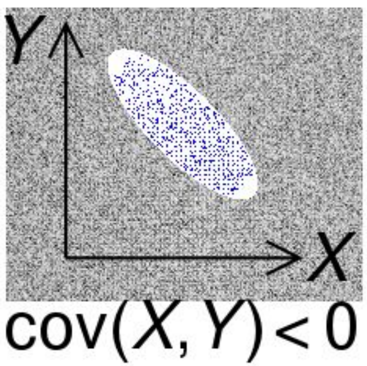
var[X] = $(1-2.1)^2 \times 0.1 + (2-2.1)^2 \times 0.7 + (3-2.1)^2 \times 0.2 = 0.37$

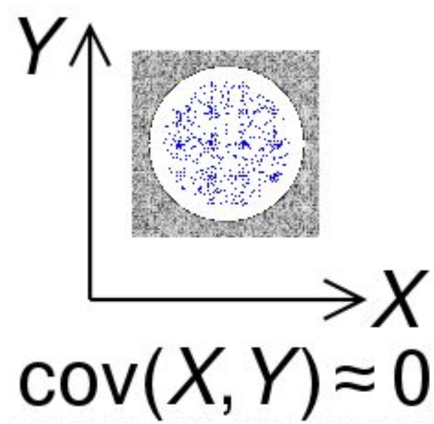
var[X] = $(1-2.01)^2 \times 0.32 + (2-2.01)^2 \times 0.35 + (3-2.01)^2 \times 0.33 = 0.66$

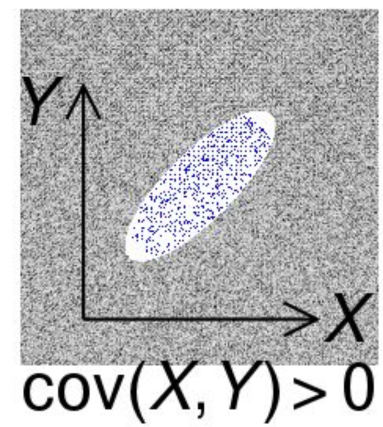
Covariance

- Covariance is a measure of the degree to which two random variables change together.
- It quantifies the **linear** association or relationship between two random variables
  - when they tend to increase or decrease in value together we have positive covariance
  - when they move in opposite directions we have negative covariance
  - when they show no significant pattern of change we have zero covariance

$$\mathbf{Cov}(X,Y) = \mathbf{Cov}(Y,X) = \mathbf{E}[(X-\mathbf{E}[X])(Y-\mathbf{E}[Y])]$$



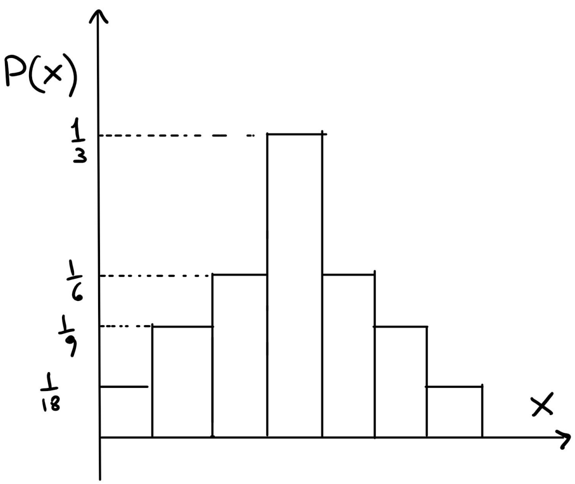




From discrete to continuous

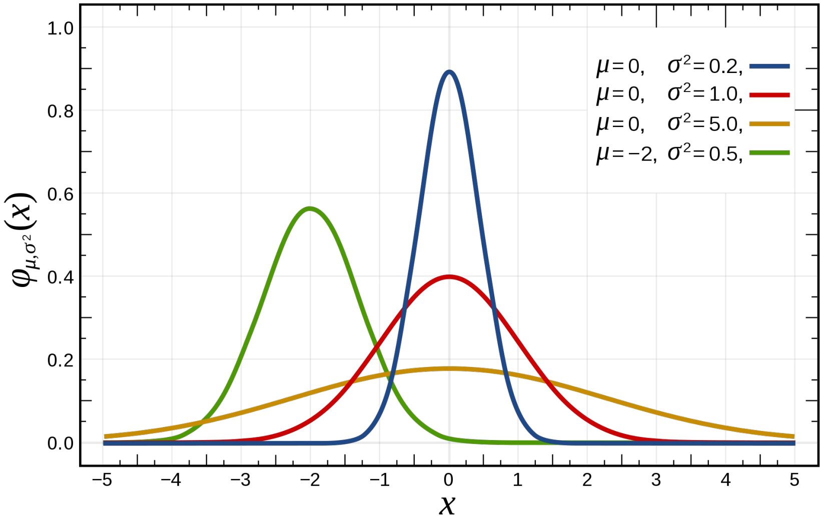
- If the RV is continuous the probability distribution is obtained integrating a density function
- The density function tells the density of probability in the domain of the RV

Discrete RV



$$\sum_{x \in \mathbb{X}} P(x) = 1$$

Continuous RV



$$\int_{\mathcal{X}} \phi_{\mu,\sigma^2}(x) dx = 1$$