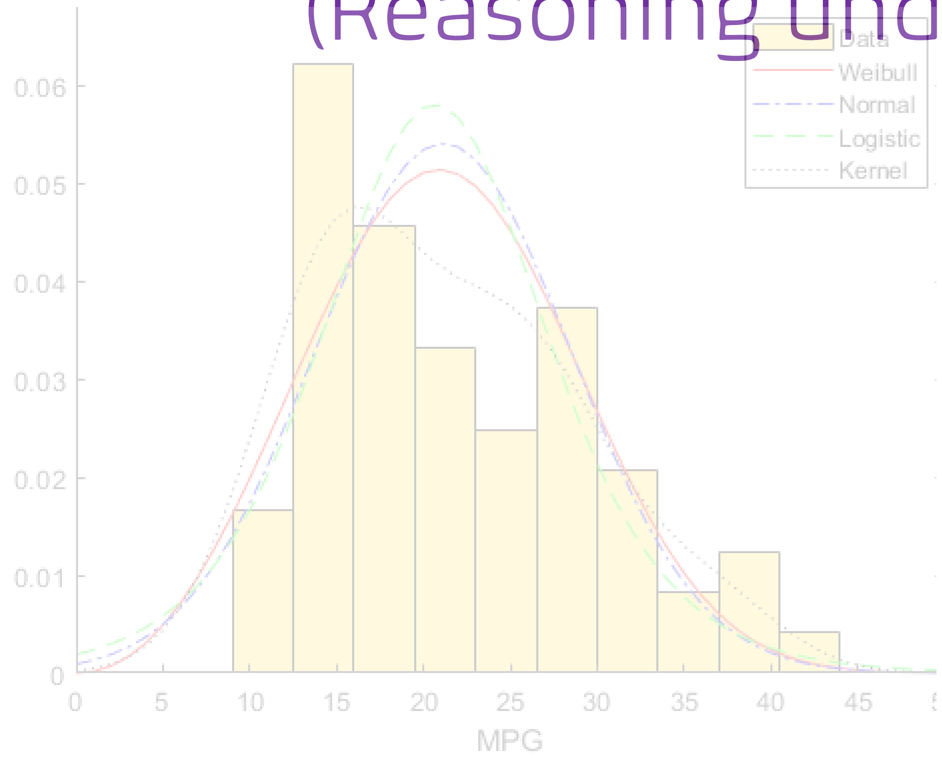
UNIVERSITÉ BISHOP'S UNIVERSITY

Probabilistic Reasoning

(Reasoning under uncertainty with Bayesian Nets)

CS417

R. Hedjam



Introduction

- **Problem**: suppose we are trying to determine if a patient has pneumonia. The following symptoms are observed:
    - Cough
    - Fever
    - Difficulty of breathing
- **Objective**: determine how likely the patient has pneumonia given the observed symptoms.
- Observing only the symptoms does not confirm 100% that the patient has pneumonia.

Introduction

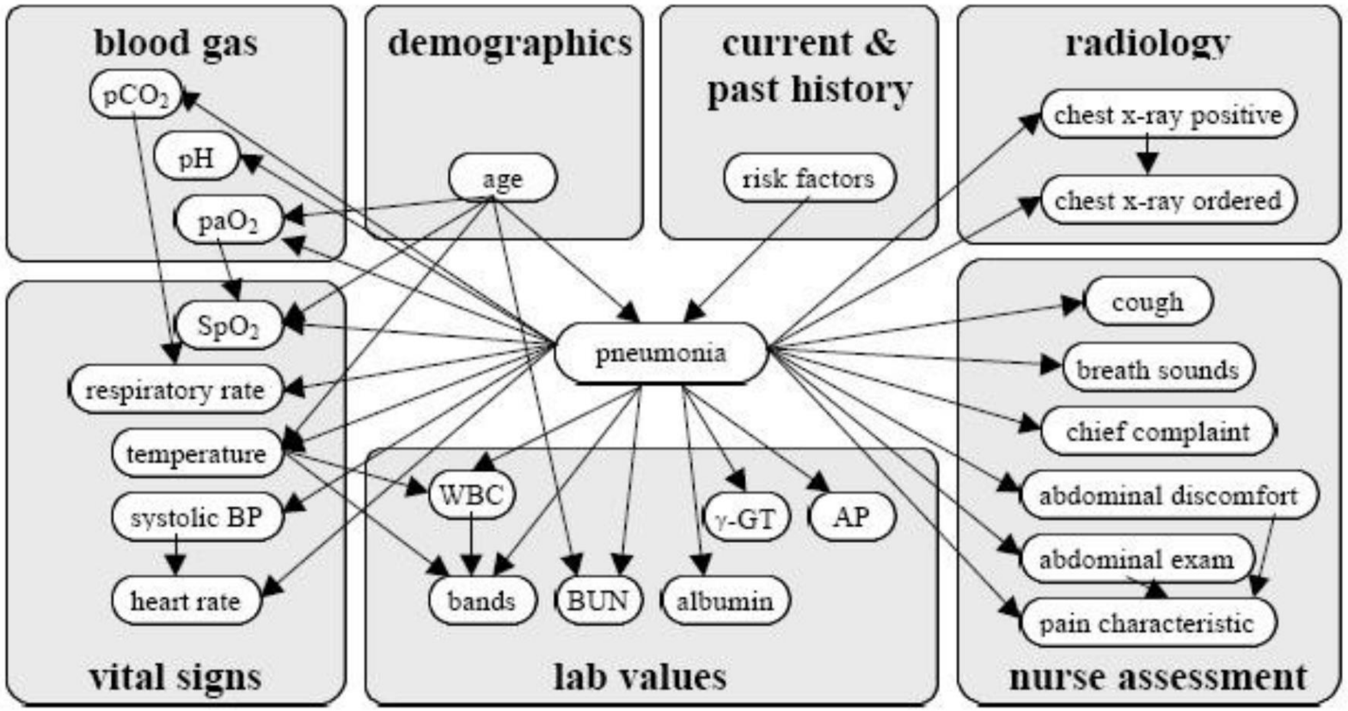
- Use a chest X-ray radiography and the result is positive.
- Our belief that the patient has pneumonia is now much higher.
- Thus, the observation affects our belief that the patient has pneumonia.
- This is called reasoning with uncertainty.
- **Bayes theory** and **Bayesian Networks** are great means to reason with uncertainty.

Introduction

- Bayesian networks are the most significant contribution in AI in the last 10 years.
- Domains of application:
  - Spam filtering / Text mining
  - Speech recognition
  - Robotics
  - Diagnostic systems
  - Syndromic surveillance
- …

***Judea Pearl** – ACM Turing Award winner, 2011.*
Championing the probabilistic approach to artificial intelligence and the development of Bayesian networks

Bayesian Networks - Example



- From: *Aronsky, D. and Haug, P.J., Diagnosing community-acquired pneumonia with a Bayesian network, In: Proceedings of the Fall Symposium of the American Medical Informatics Association, (1998) 632-636.*

Probability distributions

- For random variables we will discuss:
  - Joint probability distributions
  - Marginal probability distributions
  - Conditional probability distributions
  - Independence of random variables

Motivating Example

Experiment: deploy a smoke detector and see if it works. There could be four outcomes (events):

- Ω: events
- Ω = {(fire; smoke); (no fire; smoke); (fire; no smoke); (no fire; no smoke)}
- **Example** of one event: {(fire, smoke)}
- Note that these outcomes are mutually exclusive.
- Random variables, Two: fire_nofire (X) and smoke_nosmoke (Y)
- X has two values: {fire, nofire}
- Y has two values: {smoke, nosmoke}
- And we may choose:
- P({(fire; smoke); (no fire; smoke)}) = 0.005
- P({(fire; smoke); (fire; no smoke)}) = 0.003

Joint Probability distribution

- In general, if X and Y are two r.v., the probability distribution that defines their simultaneous behavior is called *a Joint probability distribution (JPD)*.
- **Example:**
- Let X denotes the r.v. *fire_nofire (X)*
  - The possible value of *X* are *fire* and *nofire*
- Let Y denotes the r.v. *smoke_nosmoke (Y)*
  - The possible values of *Y* are *smoke* and *nosmoke*.
- The table represents the Joint probability distribution of X and Y. The table cells are the *joint probabilities*.

		X	
		fire	nofire
Y	smoke	0.002	0.003
	nosmoke	0.001	0.994

Joint Probability distribution

- The possible value of *X* are *fire* and *nofire*
- The possible values of *Y* are *smoke* and *nosmoke*.
- Therefore, there are *4* possible pairs *(X,Y)*.
- The sum of all the probabilities is 1.0.
- The event with the highest probability is *(nofire*, *nosmoke)*.
- The event with the lowest probability is *(fire, nosmoke)*
- **P(X=x, Y=y) >= 0**
- $\sum_x \sum_y P(X=x, Y=y) = 1$

		X	
		fire	nofire
Y	smoke	0.002	0.003
	nosmko e	0.001	0.994

Marginal probability distribution

- If we are given a joint probability distribution for X and Y, we can obtain the individual probability distribution for X or for Y, and these are called the *Marginal Probability Distribution (MPD)*.
- **Example:**
  - **$P(X=x) = \sum_y P(X=x, Y=y)$**
  - **$P(Y=y) = \sum_x P(X=x, Y=y)$**

		X		
		fire	nofire	P(Y=y)=
Y	smoke	0.002	0.003	0.005
	nosmkoe	0.001	0.994	0.995
	P(X=x)=	0.003	0.997	**1.00**

- The marginal distributions for each variable, formed by summing the joint probability over the other variable.
- Called marginal because they are written in the marginal.
- In table above, **JDP** is shown in green, while the two MPD are shown in purple

Marginal probability distribution

- **Q:** find the probability that the detector fires on.
- **X:** P(X = fire) = P(X=fire, Y=smoke)+P(X=fire, Y=nosmoke)

  = 0.002 + 0.001

  = 0.003
- **Q:** What is the MPD of **X**?
- **X:** The MPD for X appears in the column totals

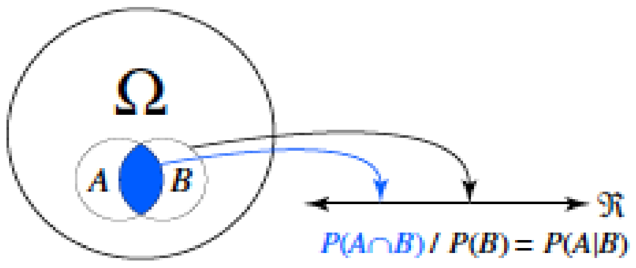
x	fire	nofire
P(X=x)	**0.003**	**0.997**

- **Q:** What is the MPD of **Y**?
- **X:** The MPD of Y appears in the row totals

y	smoke	nosmoke
P(Y=x)	**0.005**	**0.995**

		X		
		fire	nofire	P(Y=x)
Y	smoke	0.00 2	0.003	0.005
	nosmkoe	0.00 1	0.994	0.995
	P(X=x)	0.00 3	0.997	**1.00**

Conditional probability (CP)



- Random events: A and B.
- Ω: Sample probability space .
  - *A set of possible outcomes of an experiment*
- Conditional probability allows us to reason with partial information.
- When P(B)>0, the conditional probability of A given B is defined as

$$P(A \mid B) \triangleq \frac{P(A \cap B)}{P(B)}$$

- This is the probability that A occurs, given we have observed B, i.e., that we know the experiment's actual outcome will be in B. It is the fraction of probability mass in **B** that also belongs to A.
- P(B) is called the *priori* (or prior) probability of B and P(A|B) is called the a *posteriori* probability of A given B.

Conditional probability (CP)

- Given 2 r.v. X and Y with joint probability $P(X=x, Y=y)$, the conditional probability of $Y=y$ given $X=x$ is:

  $$P(Y=y \mid X = x) = P(X=x, Y=y)/P(X=x) \quad \text{for } P(X=x) > 0$$

- The conditional probability can be stated as the joint probability over the marginal probability.
- **Note**: we use the subscript $X \mid y$ for clarity to denote that this is a conditional distribution.

Example of conditional probability

- If P is defined by

	fire	*no fire*
smoke	0.002	0.003
no smoke	0.001	0.994

- Then

$$
\begin{aligned}
&P(\{(\mathit{fire}, \mathit{smoke})\} \mid \{(\mathit{fire}, \mathit{smoke}), (\mathit{no\ fire}, \mathit{smoke})\}) \\
&= \frac{P(\{(\mathit{fire}, \mathit{smoke})\} \cap \{(\mathit{fire}, \mathit{smoke}), (\mathit{no\ fire}, \mathit{smoke})\})}{P(\{(\mathit{fire}, \mathit{smoke}), (\mathit{no\ fire}, \mathit{smoke})\})} \\
&= \frac{P(\{(\mathit{fire}, \mathit{smoke})\})}{P(\{(\mathit{fire}, \mathit{smoke}), (\mathit{no\ fire}, \mathit{smoke})\})} \\
&= \frac{0.002}{0.005} = 0.4
\end{aligned}
$$

Example of conditional probability

- **Q**: Find the probability that there is a smoke given that there is no fire.
- **X:** P(Y=smoke | X=nofire) = P(Y=smoke, X=nofire)/P(X=nofire)

  = 0.003/0.997 = 0.003
- **Q**: Find the probability that there is no fire given that there is a smoke.
- **X:** P(X=nofire|Y=smoke) = P(X=nofire, Y=smoke)/P(Y=smoke)

  = 0.003/0.005 = 0.6
- **Q**: Find the probability that there is fire given that there is a smoke.
- **X:** P(X=fire|Y=smoke) = P(X=fire, Y=smoke)/P(Y=smoke)

  = 0.002/0.005 = 0.4

		X		
		fire	nofire	$P(Y=y)=$
Y	smoke	**0.002**	0.003	0.005
	nosmkoe	**0.001**	0.994	0.995
	$P(X=x)=$	0.003	0.997	**1.00**

Example of conditional probability

- **Q**: Find the conditional distribution of X given Y = smoke.
- **X:** P(X=fire|Y=smoke) = 0.002/0.005 = 0.4
- P(X=nofire|Y=smoke) = 0.003/0.005 = 0.6
- P(X|Y=smoke) = {0.4, 0.6}
- The sum of these probabilities is 1, and it is a legitimate probability distribution

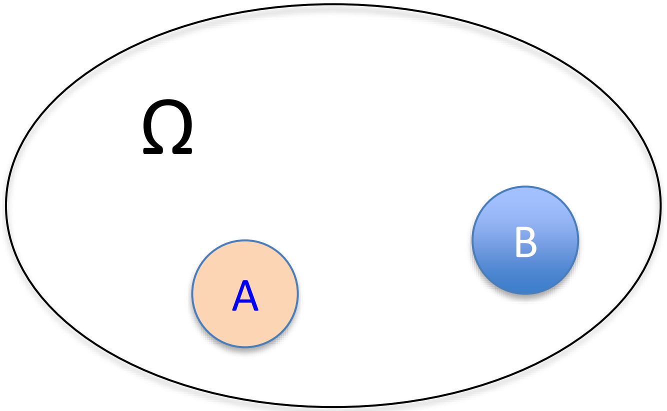
		X		
		fire	nofire	P(Y=y)
Y	smoke	**0.002**	0.003	0.005
	nosmkoe	**0.001**	0.994	0.995
	P(X=x)	0.003	0.997	**1.00**

Properties of conditional probability

- X and Y are discrete r.v.
- $P(Y=y|X=x) = P(X=x, Y=y)/P(X=x)$ for $P(X=x) > 0$
  1. $P(Y=y|X=x) >= 0$
  2. $\sum_y P(Y=y|X=x) = 1$

Independence

- **Ω**: Sample probability space .
  - *A set of possible outcomes of an experiment.*
- Random events: **A** and **B**.
- In some cases an event **A** doesn't tell us anything about the event **B**.
- $P(A|B) = P(A \cap B)/P(B)=P(A)P(B)/P(B) = P(A)$ or
- $P(B|A)= P(A \cap B)/P(A) = P(B)$ or
- Because, $P(A \cap B)=P(A) \cdot P(B)$

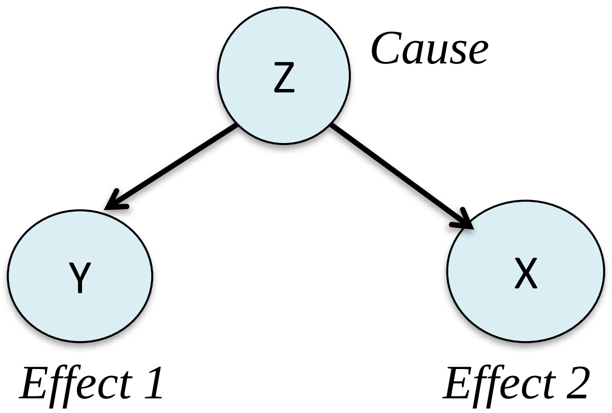


Independence

- In some cases a r.v. X doesn't tell us anything about the r.v. Y.
- Random variable independence means that knowledge of the value of X does not change any of the probabilities associated with the values of Y.
- Two r.v. are dependent if the values of one are influenced by the values of the other one.
- If X and Y are independent, then
  - $P(Y=y \mid X=x) = P(Y=y)$ for any x
  - $P(X=x \mid Y=y) = P(X=x)$ for any y
  - Because $P(X=x, Y=y) = P(X=x).P(Y=y)$ for any x and y

Independence

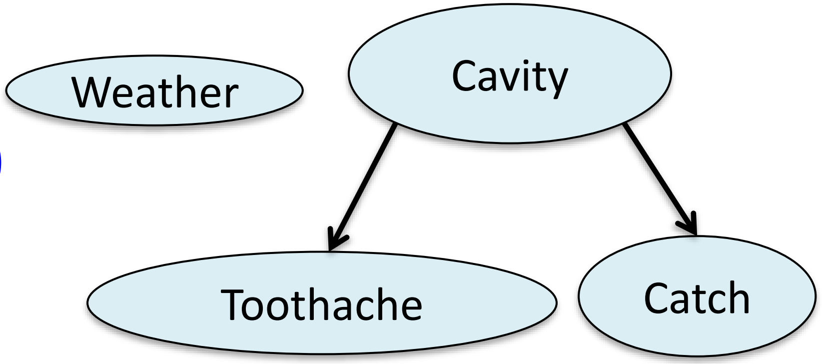
- Variable X and Y are conditionally independent given Z if the following holds:
  - $P(X,Y \mid Z) = P(X|Z)P(Y|Z)$
  - $P(X|Y,Z) = P(X|Z)$
  - $P(Y|X,Z) = P(Y|Z)$



Knowing Z tells us everything about Y. We don't gain anything by knowing X (either because X doesn't influence Y or because knowing Z provides all the information knowing X would give)

Independence

- Variable Catch and Toothache are conditionally independent given Cavity if the following holds:
- Knowing that there is Cavity tells us everything about Toothache. We don’t gain anything by knowing Catch (either because Catch doesn't influence Toothache or because knowing Cavity provides all the information knowing Catch would give)
  - P(Catch,Toothache|Cavity) = P(Catch|Toothache)P(Toothache|Cavity)
  - P(Catch|Toothache,Cavity) = P(Catch|Cavity)
  - P(Toothache|Catch,Cavity) = P(Toothache|Cavity)



Independence

- **Q.** I toss a coin twice (*experiment 1*) and define X to be the number of heads I observe. Then, I toss the coin two more times (*experiment 2*) and define Y to be the number of heads that I observe this time. Find $P((X<2) \text{ and } (Y>1))$.

Independence

- **Q.** I toss a coin twice (*experiment 1*) and define X to be the number of heads I observe. Then, I toss the coin two more times (*experiment 2*) and define Y to be the number of heads that I observe this time. Find P((X<2) and (Y>1)).
- Event spaces are:
  - E1={HH, HT, TH, TT}
  - E2={HH, HT, TH, TT}
  - Since X and Y are the result of independent experiments, the two r.v. X and Y are independent. Thus,
  - P((X<2) and (Y>1)) = P(X<2)P(Y>1)

    = (PX(0) + PX(1))PY(2)

    = (1/4 + 2/4)(1/4)

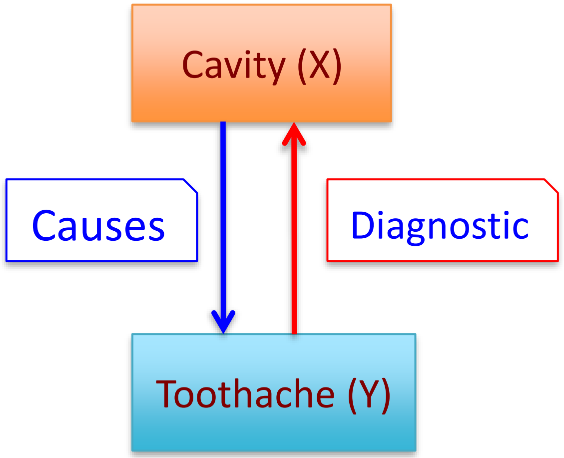
    = 3/16

Exper. 1	X
HH	2
TT	0
HT	1
TH	1

Exper. 2	Y
HH	2
TT	0
HT	1
TH	1

Bayes' rule

- The product rule gives us two ways to factor a joint probability:
  $P(X,Y) = P(X|Y)P(Y) = P(Y|X)P(X)$
- Therefore,
  $P(X \mid Y) = P(Y \mid X)P(X)/P(Y)$
- **Example:**
  - X: event that a patient has a disease.
  - Y: event that a patient displays a symptom, then:
  - $P(Y|X)$ describes a causal relationship.
  - $P(X|Y)$ describes a diagnostic one (hard to assess).
  - If $P(Y|X)$ and $P(Y)$ can be assessed easily, then we get $P(X|Y)$ for free.
- Bayes' rule translate causal knowledge into diagnostic knowledge.
  - Can get diagnostic probability P(Cavity|Toothache) from causal probability P(Toothache|Cavity)
  - Can update our beliefs based on evidence
  - Key tool probabilistic inference



Bayes' rule

Conditional prob. of X given Y

Conditional prob. of Y given X

Marginal of X

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Marginal of Y

- **P(X)**: marginal probability of X or Prior probability of X
- **P(Y)**: marginal probability of Y or prior probability of Y or evidence
- **P(Y|X):** conditional probability of Y given X or likelihood.
- **P(X|Y)**: conditional probability of X given Y. It is also called *posterior probability* of Y given X.
- Called posterior probability because it is computed after the prior probability is computed.

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- What is the marginal PY(Y=0)?
  1. 1/6
  2. 2/6
  3. 3/6
  4. 4/6
  5. else

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- What is the marginal PY(Y=0)?
  1. 1/6
  2. 2/6
  3. 3/6
  4. $\boxed{4/6}$
  5. else

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- What is the conditional $P(X=0|Y=0)$?
  1. 2/6
  2. 1/2
  3. 1/6
  4. 4/6
  5. else

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- What is the conditional P(X=0|Y=0)?
  1. 2/6
  2. $\boxed{1/2}$
  3. 1/6
  4. 4/6
  5. else

P(X=0|Y=0) = P(X=0,Y0)/P(Y=0)

= (2/6)/(4/6) =2 / 4 = 1/2

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- Are they independent?
  1. yes
  2. no
  3. I don’t know

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	2/6	1/6
X=1	2/6	1/6

- Are they independent?
  1. yes
  2. no
  3. I don't know

*Check **if P(X=0,Y=0) = P(X=0)P(Y=0)***
If yes, then X and Y are independent
Else, they are not.

P(X=0, Y=0) = 2/6
P(X=0)P(Y=0) = 3/6*4/6 = 12/36 = 2/6
P(X=0,Y=0) = P(X=0)P(Y=0)

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	1/2	0
X=1	0	1/2

- Are they independent?
  1. yes
  2. no
  3. I don’t know

Exercise

- Given two binary r.v. X and Y (Bernoulli variables).
- Given the joint probability of X and Y.

	Y=0	Y=1
X=0	1/2	0
X=1	0	1/2

- Are they independent?
  1. yes
  2. $\boxed{\text{no}}$
  3. I don't know

*Check **if P(X=0,Y=0) = P(X=0)P(Y=0)***
If yes, then X and Y are independent
Else, they are not.