# Probability

## Discrete Random Variables

- A is a Boolean-valued random variable if A denotes an event, and there is some degree of uncertainty as to whether A occurs. This uncertinity is stated in terms of probability

Examples:

A = The next toss of coin is Head

A = The next toss of a coin is Tail

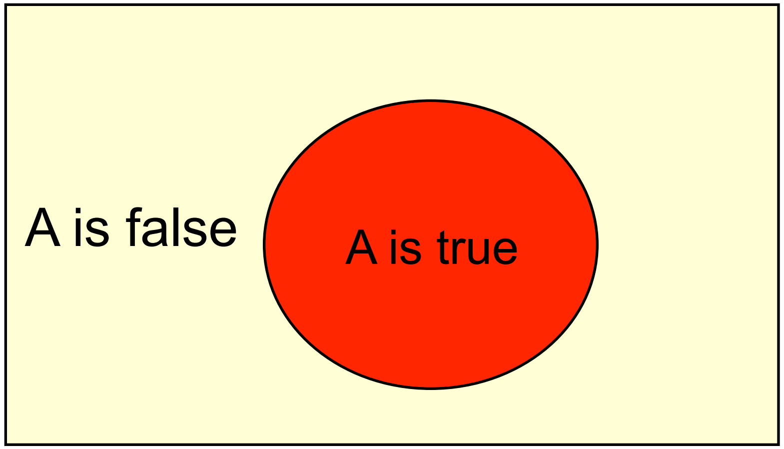
A = The flights will resume next day

# Probability

## Discrete Random Variables

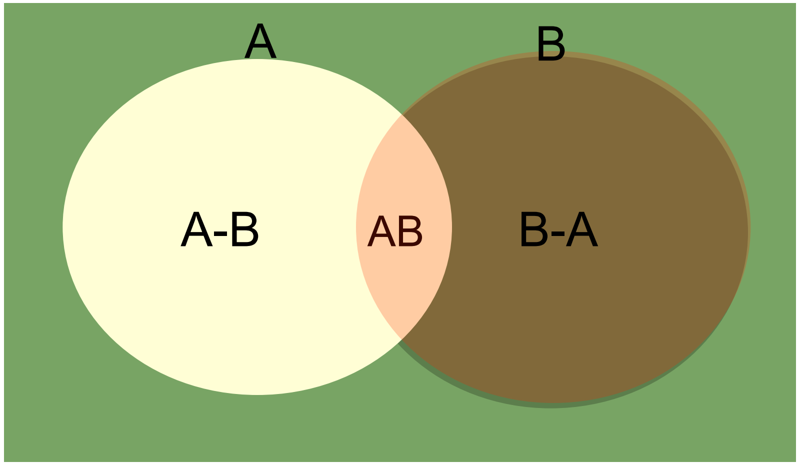
• P(A) = “the fraction of worlds in which A is true” or the fraction of times the event is true in independent trails

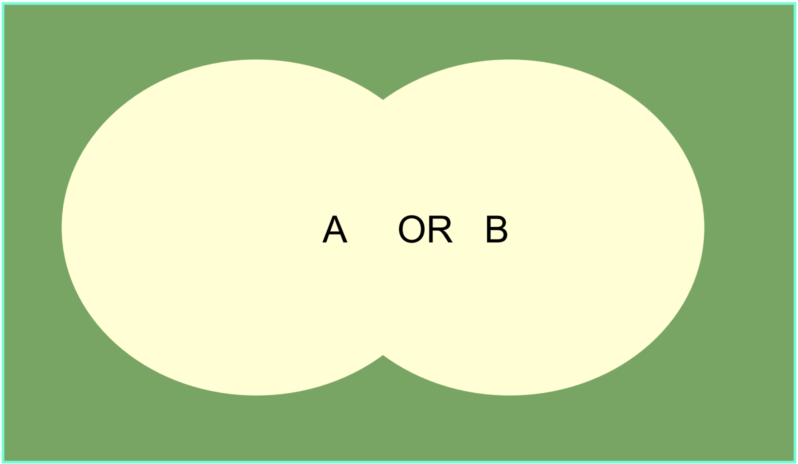
P(A) = Proportion of area of reddish oval.



# The Axioms of Probability

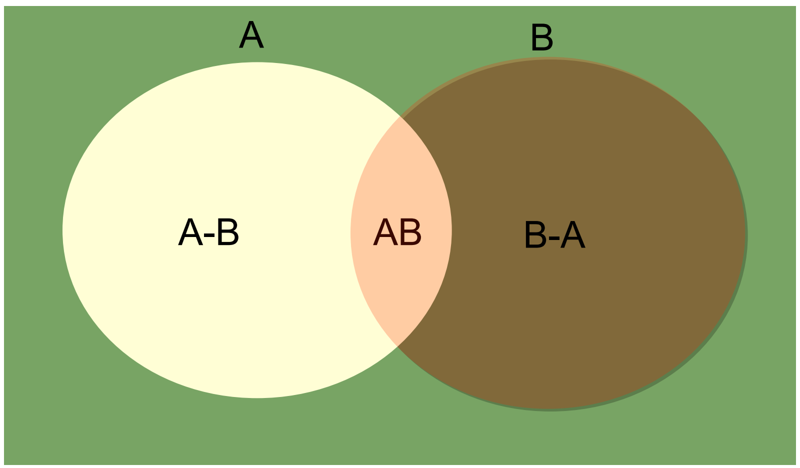
0 <= P(A) <= 1, Head and Tail are mutually exclusive.
P(True) = 1 e.g., p(A = Head or A = Tail) = 1 (we also write P(Head or Tail))
P(False) = 0 e.g. P(A = Head and A = Tail) = 0 (we also write P(Head and Tail))
If A and B are not mutually exclusive
P(A or B) = P(A) + P(B) - P(A and B)



# The Axioms of Probability

P(A) = P(A and B) + P(A and Not B)

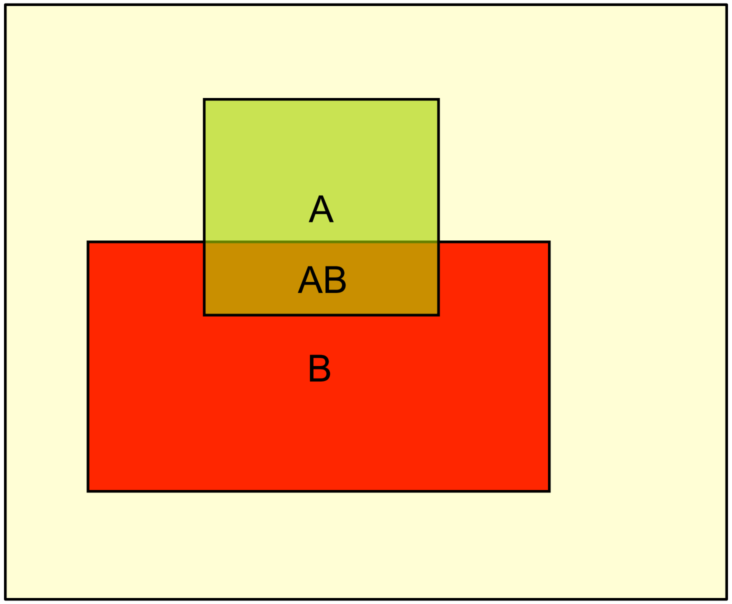


# Conditional Probability

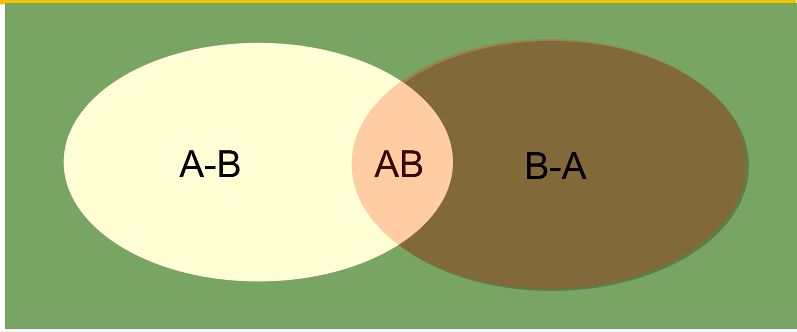
P(A|B) = Fraction of time A is true knowing B is true
= P(A and B)/P(B)

P(A|B) = P(A and B)/P(B)

P(A and B) = P(A|B)*P(B) -- product rule



# Probability




**Product rule** $$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$$

**Sum rule** $$p(\mathbf{x}) = \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y})$$

$$p(\mathbf{x}) = \sum_{\mathbf{y}} p(\mathbf{x}|\mathbf{y})p(\mathbf{y})$$

**From these we have Bayes theorem**

$$p(y/x) = \frac{p(x,y)}{p(x)} \qquad p(\mathbf{y}|\mathbf{x}) = \frac{p(\mathbf{x}|\mathbf{y})p(\mathbf{y})}{p(\mathbf{x})}$$

$$p(y|x) = \frac{p(x|y)p(y)}{\sum_{y} p(x|y)p(y)}$$

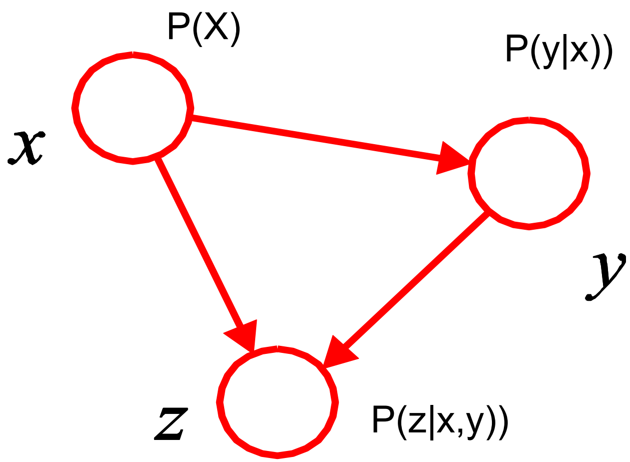
# Probabilistic Graphical Models (PGM)

**PGM provides new insights into existing models**

**Consider an arbitrary joint distribution**

**By successive application of the product rule** $p(x, y, z)$

$$\begin{aligned} p(x, y, z) &= p(x)p(y, z|x) \\ &= p(x)p(y|x)p(z|x, y) \end{aligned}$$

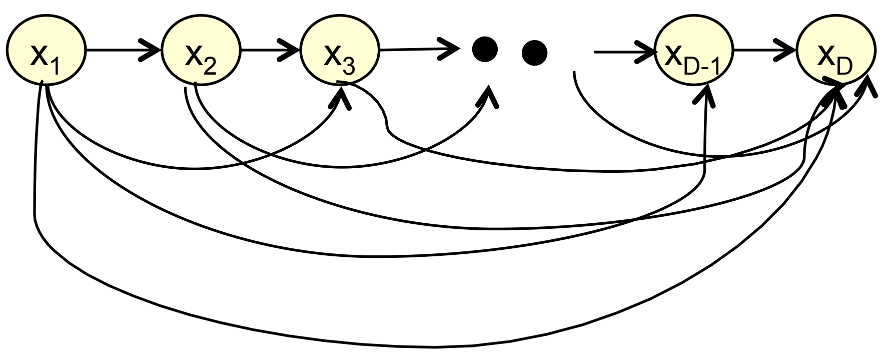


# Directed Acyclic Graphs

**Joint distribution**
**where** pa$_i$ **denotes the parents of i**

$$p(x_1,\ldots,x_D) = \prod_{i=1}^{D} p(x_i|\mathrm{pa}_i)$$

$$\begin{aligned} p(x_1,x_2,\ldots,x_D) &= p(x_1)p(x_2,x_3,\ldots x_D \mid x_1) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3,x_4,\ldots x_D \mid x_1,x_2) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1,x_2)p(x_4,\ldots x_D \mid x_1,x_2,x_3) \\ &= p(x_1)p(x_2 \mid x_1)p(x_3 \mid x_1,x_2)\ldots p(x_D \mid x_1,x_2,\ldots,x_{D-1}) \end{aligned}$$

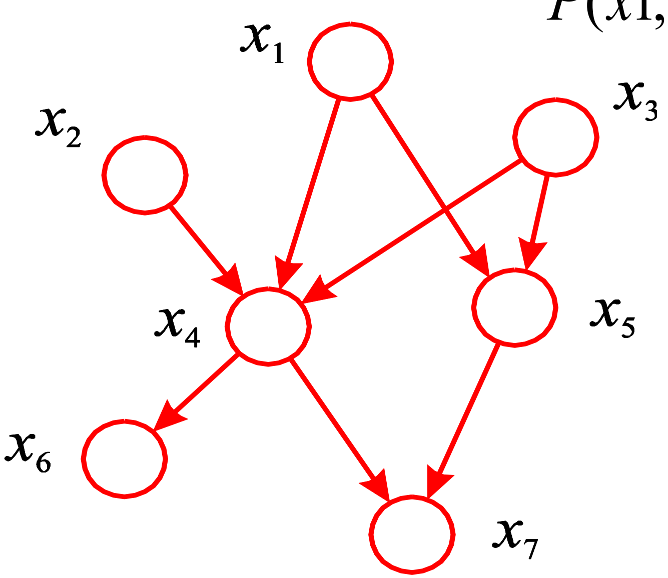


# Directed Acyclic Graphs

**Joint distribution**
**where** pa$_i$ **denotes the parents of i**

$$p(x_1, \ldots, x_D) = \prod_{i=1}^{D} p(x_i|\mathrm{pa}_i)$$

$$P(x1, x2, x3, x4, x5, x6, x7) = \left\{ \begin{array}{l} P(x1) * P(x2) * p(x3) * \\ P(x4 \mid x1, x2, x3) * \\ P(x5 \mid x1, x3) * \\ P(x6 / x4) * \\ p(x7 \mid x4, x5) \end{array} \right\}$$



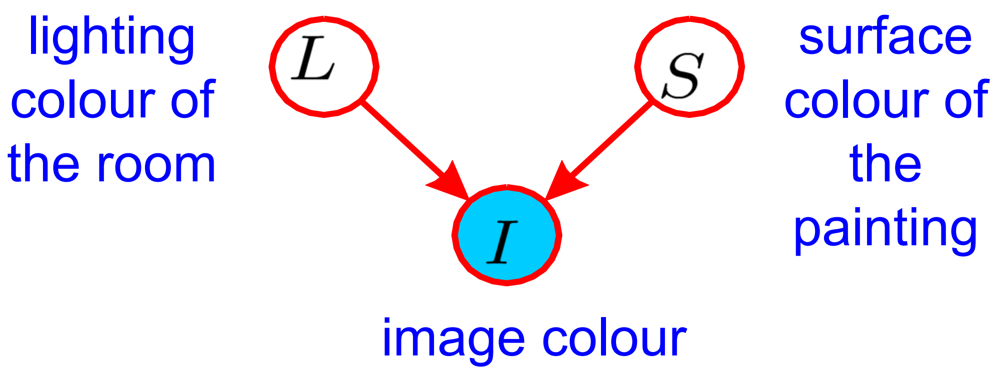
# "Explaining Away"

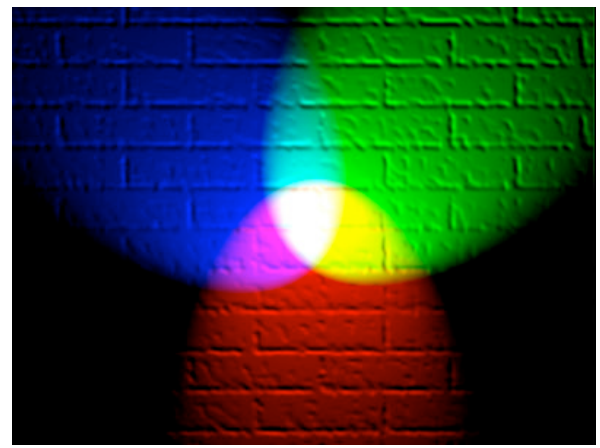
**Conditional independence for directed graphs is similar, but with one subtlety**

**Illustration: pixel colour in an image**

$$p(I,L,S) = p(L,S)p(I|L,S)$$

$$p(I,L,S) = p(L)p(S)p(I|L,S)$$



$$p(L,S) = p(L)p(S)$$

$$p(L,S|I) \neq p(L|I)p(S|I)$$

We use “~” to represent “not”, E.g., Not B is represented by ~B

$P(A) + P(\sim A) = 1$;  $P(A) = P(A,B) + P(A, \sim B)$

$P(x|y) + P(\sim x|y) = 1$;  But $P(x|y) + P(x|\sim y) \sim= 1$;

Example:

p( restaurant = bad) = p(B) = ½ = 0.5

p( menu = smudged | restaurant = bad) = p(S|B) = 3/4

p( menu = smudged | restaurant = ~ bad) = p(S| ~B) = 1/3

p( restaurant = bad | menu = smudged ) = P(B|S)?

$$p(B|S) = p(B,S)/p(S) = p(B, S)/[(p(S, B) + p(S,\sim B)]$$

$$= p(S|B)P(B)/[p(S|B)P(B) + p(S|\sim B)P(\sim B)]$$

$$= (3/4)*(1/2)/[(3/4)*(1/2) + (1/3)*(1/2)]$$

$$= 9/13 = 0.69$$