Probability Distribution.

→ $P(W=\text{rain}) = 0.1$ and $\forall x \; P(X=x) \geq 0$, $\sum_x P(X=x) = 1$

Joint Distribution:

→ $P(X_1=x_1, X_2=x_2, \ldots, X_n=x_n)$, $P(x_1, x_2, \ldots x_n)$

↳ $P(x_1, x_2, \ldots x_n) \geq 0$, $\boxed{\sum_{x_1, x_2 \ldots x_n} P(x_1, x_2 \ldots x_n) = 1}$

Probabilistic Models

- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called *outcomes*
  - Joint distributions: say whether assignments (outcomes) are likely
  - *Normalized:* sum to 1.0
  - Ideally: only certain variables directly interact

- Constraint satisfaction problems:
  - Variables with domains
  - Constraints: state whether assignments are possible
  - Ideally: only certain variables directly interact

Events

- 简单事件（Simple Event）：只包含一个结果的事件。比如，掷骰子得到4就是一个简单事件。
- 复合事件（Compound Event）：包含两个或多个简单事件的组合。比如，掷骰子得到偶数就是一个复合事件，因为它包含了掷出2、4、6这三个简单事件的组合。

P(T,W)

$P(2.4.6)$

Marginal Distributions

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(t) = \sum_s P(t,s)$$

$$P(s) = \sum_t P(t,s)$$

$P(T)$

T	P
hot	0.5
cold	0.5

$P(W)$

W	P
sun	0.6
rain	0.4

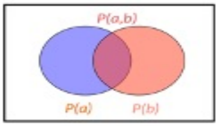
$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

$P(X_1 = x_1) =$

x_1 $\cancel{X_2}$

Conditional Probabilities

$$P(a|b) = \frac{P(a,b)}{P(b)}$$



$$P(a|b) = \frac{P(a,b)}{P(b)}$$

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

$$P(W=s|T=c) = \frac{P(W=s,T=c)}{P(T=c)} = \frac{0.2}{0.5} = 0.4$$

$$= P(W=s,T=c) + P(W=r,T=c)$$
$$= 0.2 + 0.3 \quad = 0.5$$

Normalization Trick

当分母复杂时，避免直接计算概率的分母，从而提高计算效率

$$P(x_1|x_2) = \frac{P(x_1,x_2)}{P(x_2)} = \frac{P(x_1,x_2)}{\sum P(x_1,x_2)}$$

Normalization

$$P(W=s|T=c) = \frac{0.2}{0.2+0.3} = 0.4$$

$P(T,W)$

T	W	P
hot	sun	0.4
hot	rain	0.1
cold	sun	0.2
cold	rain	0.3

SELECT the joint probabilities matching the evidence →

$P(c,W)$

T	W	P
cold	sun	0.2
cold	rain	0.3

NORMALIZE the selection (make it sum to one) →

$P(W|T=c)$

W	P
sun	0.4
rain	0.6

$$P(x_1|x_2) = \frac{P(x_1,x_2)}{P(x_2)} = \frac{P(x_1,x_2)}{\sum_{x_1} P(x_1,x_2)}$$

Probabilistic Inference

通过这些模型，我们可以计算出在给定某些已知信息（证据）的条件下，一个或多个未知事件发生的概率

Inference by Enumeration

拿已知e求Q通过过滤h

- General case:
  - Evidence variables: $E_1 \dots E_k = e_1 \dots e_k$
  - Query* variable: Q
  - Hidden variables: $H_1 \dots H_r$

  $X_1, X_2, \dots X_n$ All variables

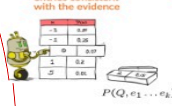
- We want: $P(Q|e_1 \dots e_k)$

  * Works fine with multiple query variables, too

$\to = 1$

$Q + e_{1 \dots k} + H = P(Q, e_1 \dots e_k)$

- Step 1: Select the entries consistent with the evidence



- Step 2: Sum out H to get joint of Query and evidence

$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} \underbrace{P(Q, h_1 \dots h_r, e_1 \dots e_k)}_{X_1, X_2, \dots X_n}$$

- Step 3: Normalize

$$\times \frac{1}{Z}$$

$$Z = \sum_q P(Q, e_1 \cdots e_k)$$

$$P(Q|e_1 \cdots e_k) = \frac{1}{Z} P(Q, e_1 \cdots e_k)$$

$Z = \sum_q P(Q=q, e_1 \dots e_k)$

$q - \{$ 几种可能性

Product rule:

$$P(y)P(x|y) = P(x,y)$$

$$P(x|y) = \frac{P(x,y)}{P(y)}$$

$$P(y)P(x|y) = P(x,y)$$

- Example:

$P(W)$

R	P
sun	0.8
rain	0.2

$P(D|W)$

D	W	P
wet	sun	0.1
dry	sun	0.9
wet	rain	0.7
dry	rain	0.3

$P(D,W)$

D	W	P
wet	sun	0.08
dry	sun	0.72
wet	rain	0.14
dry	rain	0.06

The chain rule:

- More generally, can always write any joint distribution as an incremental product of conditional distributions

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$

$$P(x_1, x_2, \ldots x_n) = \prod_i P(x_i|x_1 \ldots x_{i-1})$$

$= P(x_1) \quad \frac{P(x_1, x_2)}{P(x_1)} \quad \frac{P(x_1, x_2, x_3)}{P(x_1, x_2)}$

Bayes' rule:

- Two ways to factor a joint distribution over two variables:

$$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$

- Dividing, we get:

$$\boxed{P(x|y) = \frac{P(y|x)}{P(y)}P(x)}$$

- Why is this at all helpful?
  - Lets us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple
  - Foundation of many systems we'll see later (e.g. ASR, MT)

Inference with Bayes' Rule

- Example: Diagnostic probability from causal probability:

posterior conditional prior

$$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}|\text{cause})P(\text{cause})}{P(\text{effect})}$$

- Example:
  - M: meningitis, S: stiff neck

$$\left.\begin{array}{l} P(+m) = 0.0001 \\ P(+s|+m) = 0.8 \\ P(+s|-m) = 0.01 \end{array}\right\} \text{Example givens}$$

$$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m) + P(+s|-m)P(-m)} = \frac{0.8 \times 0.0001}{0.8 \times 0.0001 + 0.01 \times 0.999}$$

0.

  - Note: posterior probability of meningitis still very small
  - Note: you should still get stiff necks checked out! Why?

Independence:

1. **联合概率等于边缘概率的乘积**：
   $P(X = x, Y = y) = P(X = x)P(Y = y)$
   这意味着变量 X 的发生不影响变量 Y 的发生概率，反之亦然。
2. **条件概率等于边缘概率**：
   $P(X = x|Y = y) = P(X = x)$
   这意味着给定 Y 的任何值，X 的概率分布不会改变，即 X 的发生不依赖于 Y 的特定值。

在概率论的符号中，$X \perp\!\!\!\perp Y$ 表示 X 和 Y 是独立的。

Conditional Independence:

- *Conditional independence* is our most basic and robust form of knowledge about uncertain environments.

- X is conditionally independent of Y given Z: $X \perp\!\!\!\perp Y | Z$

if and only if:

$$\forall x, y, z : P(x, y|z) = P(x|z)P(y|z)$$

or, equivalently, if and only if

$$\forall x, y, z : P(x|z, y) = P(x|z)$$

火与火警 dependent

火与烟 inde

火警与烟 inde

所以，火与火警 C Inde

Summary:

To Summarize ...

- Basic laws: $0 \leq P(\omega) \leq 1$, $\sum_{\omega \in \Omega} P(\omega) = 1$, $P(A) = \sum_{\omega \in A} P(\omega)$
- Random variable $X(\omega)$ has a value in each ω
  - Distribution $P(X)$ gives probability for each possible value x
  - Joint distribution $P(X,Y)$ gives total probability for each combination x,y
- Summing out/marginalization: $P(X=x) = \sum_y P(X=x,Y=y)$
- Conditional probability: $P(X|Y) = P(X,Y)/P(Y)$
- Chain rule: $P(X_1,..,X_n) = \prod_i P(X_i | X_1,..,X_{i-1})$
- Bayes Rule: $P(X|Y) = P(Y|X)P(X)/P(Y)$
- Independence: $P(X,Y) = P(X) P(Y)$ or $P(X|Y) = P(X)$ or $P(Y|X) = P(Y)$
- Conditional Independence: $P(X|Y,Z) = P(X|Z)$ or $P(X,Y|Z) = P(X|Z) P(Y|Z)$

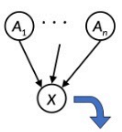
1. **节点（Nodes）**：
   - 表示变量，并带有各自的定义域（可能的取值范围）。
   - 节点可以是已赋值的（观察到的数据）或未赋值的（未观察到的数据）。
2. **弧（Arcs）**：
   - 表示变量之间的相互作用。
   - 与约束满足问题（CSP）中的约束类似，它们表明变量之间存在"直接影响"。
   - 在形式上，弧编码了变量之间的条件独立性。稍后将进一步解释这一点。
3. **箭头**：
   - 截图下方的注释说明，目前可以将箭头理解为表示直接因果关系。然而，通常情况下箭头并不总是代表直接的因果关系，而是表示一个变量的取值会影响另一个变量的概率分布。换句话说，箭头在这里用于指示条件概率的方向。

Bayes' Net Semantics

- A set of nodes, one per variable X

- A directed, acyclic graph

- A conditional distribution for each node
  - A collection of distributions over X, one for each combination of parents' values $P(X|a_1 \ldots a_n)$



$$P(X|A_1 \ldots A_n)$$

  - CPT: conditional probability table

A Bayes' net = Topology (graph) + Local Conditional Probabilities

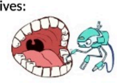
  - Description of a noisy "causal" process

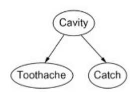
- Bayes' nets *implicitly* encode joint distributions
  - As a product of local conditional distributions
  - To see what probability a BN gives to a full assignment, multiply all the relevant conditionals together:

$$P(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} P(x_i|parents(X_i))$$

  - Chain rule gives:

  - Example:



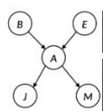


$$P(+cavity, +catch, -toothache)$$

Example: Alarm Network

B	P(B)
+b	0.001
-b	0.999

E	P(E)
+e	0.002
-e	0.998



A	J	P(J|A)
+a	+j	0.9
+a	-j	0.1
-a	+j	0.05
-a	-j	0.95

A	M	P(M|A)
+a	+m	0.7
+a	-m	0.3
-a	+m	0.01
-a	-m	0.99

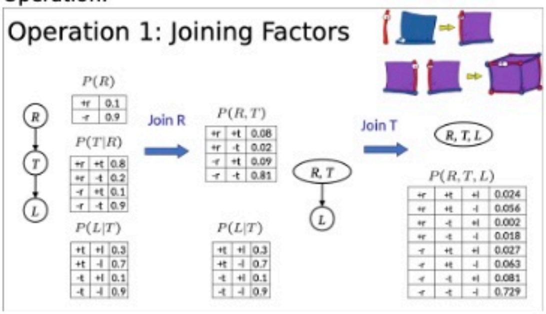
B	E	A	P(A|B,E)
+b	+e	+a	0.95
+b	+e	-a	0.05
+b	-e	+a	0.94
+b	-e	-a	0.06
-b	+e	+a	0.29
-b	+e	-a	0.71
-b	-e	+a	0.001
-b	-e	-a	0.999

$$P(+b, -e, +a, -j, +m) =$$
$$P(+b)P(-e)P(+a|+b, -e)P(-j|+a)P(+m|+a) =$$
$$0.001 \times 0.998 \times 0.94 \times 0.1 \times 0.7$$

Inference:

1. **推理（Inference）**：
   - 推理是指从联合概率分布中计算出一些有用的量。在概率论和统计学中，推理通常涉及到根据已知的信息来计算未知量的概率。
2. **例子**：
   - **后验概率（Posterior Probability）**：
     表示为 `P(Q|E1=e1, ..., Ek=ek)`，这是在给定一组证据 $E_1 = e_1, ..., E_k = e_k$ 的情况下，查询变量 Q 的条件概率。这是贝叶斯推理的核心，即在已知一些证据后，更新我们对某个假设的信念。
   - **最可能解释（Most Likely Explanation）**：
     这用到了 `argmax` 函数，表达式 `argmax_q P(Q=q|E1=e1 ...)` 意味着我们在寻找使条件概率 `P(Q=q|E1=e1 ...)` 最大的查询变量 Q 的特定值 q。换句话说，我们想要找到在给定证据的条件下最可能的 Q 的取值。

Operation:



$P(R)$

+r	0.1
-r	0.9

$P(T|R)$

+r	+t	0.8
+r	-t	0.2
-r	+t	0.1
-r	-t	0.9

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

$P(R,T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

$P(L|T)$

+t	+l	0.3
+t	-l	0.7
-t	+l	0.1
-t	-l	0.9

$P(R,T,L)$

+r	+t	+l	0.024
+r	+t	-l	0.056
+r	-t	+l	0.002
+r	-t	-l	0.018
-r	+t	+l	0.027
-r	+t	-l	0.063
-r	-t	+l	0.081
-r	-t	-l	0.729

Operation 2: Eliminate

- Second basic operation: marginalization
- Take a table and sum out a variable
  - Shrinks a table to a smaller one
  - A projection operation
- Example:

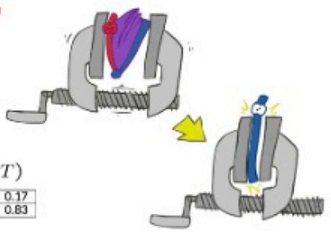
$P(R,T)$

+r	+t	0.08
+r	-t	0.02
-r	+t	0.09
-r	-t	0.81

sum R →

$P(T)$

+t	0.17
-t	0.83



Causal Chains

这个因果链被表示为：
$P(x, y, z) = P(x)P(y|x)P(z|y)$

这个公式说明：

1. 低气压（X）是一个独立的事件，它有自己的概率 $P(x)$。
2. 降雨（Y）的发生依赖于低气压（X），其概率为 $P(y|x)$。
3. 交通（Z）的状态依赖于是否下雨（Y），其概率为 $P(z|y)$。

图片中的关键问题是："在给定Y的情况下，X和Z是否独立？"按照条件概率的定义和链式规则：

$$P(z|x, y) = \frac{P(x,y,z)}{P(x,y)}$$
$$= \frac{P(x)P(y|x)P(z|y)}{P(x)P(y|x)}$$
$$= P(z|y)$$

这里，$P(z|y)$ 不依赖于X的值，因此在给定Y（雨）的情况下，X（低气压）和Z（交通）是条件独立的。换句话说，一旦我们知道了是否下雨，知道低气压的信息就不会对我们关于交通状态的知识产生额外的影响。

Common Causes

在这个特定的例子中，有一个项目截止（Y：Project due），它导致了论坛繁忙（X：Forums busy）和实验室满员（Z：Lab full）。这个项目截止是共同原因，因为它同时影响论坛和实验室的状态。这可以用以下概率公式表示：
$P(x, y, z) = P(y)P(x|y)P(z|y)$

这意味着：

- $P(y)$ 是项目截止的概率。
- $P(x|y)$ 是在项目截止的条件下，论坛变得繁忙的概率。
- $P(z|y)$ 是在项目截止的条件下，实验室满员的概率。

图片还提出了一个问题："在给定Y的情况下，X和Z是否独立？"在"共同原因"结构中，两个子节点（在本例中为X和Z）在没有考虑共同原因时通常不是独立的。但如果考虑了共同原因Y（即知道了项目是否截止），子节点X和Z之间的关联可能会减弱或消失，使它们在给定Y的情况下成为条件独立的。

然而，图片中指出，X和Z是否在给定Y的情况下独立，并没有一个保证。为了证明非独立性，举了一个具体的例子，其中列出了几个概率条件，表明即使知道了Y的状态，X和Z仍然可能相互关联：

- $P(+x|+y) = 1$，如果项目截止，则论坛一定繁忙。
- $P(-x|-y) = 1$，如果项目没有截止，则论坛一定不繁忙。
- $P(+z|+y) = 1$，如果项目截止，则实验室一定满员。
- $P(-z|-y) = 1$，如果项目没有截止，则实验室一定不满员。

- Guaranteed x and z independent given y

$$P(z|x,y) = \frac{P(x,y,z)}{P(x,y)} = \frac{P(y)P(x|y)P(z|y)}{P(y)P(x|y)} = P(z|y)$$

$\therefore$ Yes

Common Effect

这里的例子具体涉及到三个变量：

- X：下雨（Raining）
- Y：球赛（Ballgame）
- Z：交通（Traffic）

下雨和球赛是交通堵塞的两个独立原因。在没有考虑交通这一共同效应的情况下，下雨和球赛彼此独立。它们分别增加了交通拥堵的可能性，但下雨本身并不影响是否有球赛，反之亦然。

图片中的问题是“X和Y是否独立？”，答案是“是的”，意味着球赛和下雨这两个事件互不影响。尽管它们都会影响交通，但在不考虑交通的情况下，它们之间没有相关性。

为了验证这一点，给出了一个概率计算：

$P(x, y) = \sum_z P(x, y, z)$

这里，$P(x, y)$ 表示同时考虑下雨和有球赛的概率，而 $P(x, y, z)$ 表示考虑下雨、有球赛以及交通情况的联合概率。由于下雨和球赛独立，你可以分别对它们进行求和（对所有可能的交通情况），结果显示下雨和有球赛的联合概率仅仅是它们各自边缘概率的乘积。

您上传的图像描述了贝叶斯网络中“共同效应”或“V-结构”的情况。在这个特定的例子中，两个原因（X：下雨和Y：球赛）共同影响了一个效应（Z：交通）。

- 当我们没有观察到效应Z时，原因X和Y是独立的。换句话说，在不考虑交通的情况下，下雨的情况不会告诉我们关于是否有球赛的任何信息，反之亦然。
- 然而，一旦我们观察到效应Z（例如，我们看到交通拥堵了），原因X和Y不再独立。如果我们知道交通拥堵了，下雨可能会影响我们关于是否有球赛的推断，因为这两个因素都可能是交通拥堵的原因。这种现象被称为“解释消除”或“解释迷失”。

图中的文字还提到了一个有趣的概念，即观察效应“激活”了原因之间的影响。这意味着当我们注意到共同效应（本例中为交通）时，原因变量之间的条件独立性就会消失，它们就变得相互关联了。这是因为观察到的效应提供了关于其潜在原因的信息，这些原因现在必须相互考虑，以解释观察到的效应。