4. Probability and Bayes

4.1 Probability

Uncertainty: not secure about the outcomes

Omega sample space (set of possibilities)

omega in Omega sample point

Probability space: Function P : Omega → R such that:

- 0≤P(omega) ≤ 1
- Sum P(omega) = 1

Event: any subset of Omega

Probability of an event A is a function assigning A to [0,1]

P(A) = Sum P(omega)

A **random variable** (outcome of a random phenomenon) is a function from the sample space to some range X : Omega → R or B etc.

P induces a **probability distribution** for a random variable X:

P(X = xi) = Sum P(w)

A **proposition** is the event (subset of) where an assignment to a random variable holds.

event a = A = true = { omega in Omega such that A(omega) = true}

4.2 Syntax and Semantics

Prior or unconditional probabilities → normal probability (P(odd = true) = 0.5) without knowing anything

A **probability distribution** is a function assigning a probability value to all possible assignments of a random variable. (for Real is continuous)

e.g.: P(Weather) =< 0.72, 0.1, 0.08, 0.1 >

Joint probability distribution for a set of random variables gives the probability of every atomic joint event on those random variables.

Joint probability distribution of the random variables *Weather* and *Cavity*:
$P(Weather, Cavity) =$ a 4×2 matrix of values:

$Weather =$	$sunny$	$rain$	$cloudy$	$snow$
$Cavity = true$	0.144	0.02	0.016	0.02
$Cavity = false$	0.576	0.08	0.064	0.08

Conditional/Posterior Probability: I know the outcome of a random variable, how does this affect probability of other random variables?

P(a|b) = P(a^b)/P(b)   if P(b) ≠ 0

Product rule

P(a^b) = P(a|b)P(b) = P(b|a)P(a)

Total probabilities

P(a) = P(a|b)P(b) + P(a|¬b)P(¬b)

In general,

P(X) = Sum P(X|Y = yi)P(Y = yi)

Chain rule

P(X1,X2) = P(X1)P(X2|X1)

P(X1, . . . ,Xn) = Prod. P(Xi |X1, . . . ,Xi−1)

4.3 Inference by enumeration

	$toothache$		$\neg\ toothache$	
	$catch$	$\neg\ catch$	$catch$	$\neg\ catch$
$cavity$	.108	.012	.072	.008
$\neg\ cavity$	.016	.064	.144	.576

For any proposition ϕ, sum the atomic events where it is true:
$P(\phi) = \sum_{\omega:\omega\models\phi} P(\omega)$

e.g.: $P(\neg cavity|toothache) = P(\neg cavity \wedge toothache)/P(toothache)$

4.4 Independence

A and B are independent iff

$$P(A|B) = P(A) \text{ or } P(B|A) = P(B) \text{ or } P(A, B) = P(A)P(B)$$

$P(Toothache, Catch, Cavity, Weather) = P(Toothache, Catch, Cavity)P(Weather)$

- $P(Toothache, Cavity, Catch)$ has 23 − 1 = 7 independent entries
- If I have a cavity, the probability that the probe catches in it does not depend on whether I have a toothache:
(1) $P(catch|toothache, cavity) = P(catch|cavity)$
- The same independence holds if I haven't got a cavity:
(2) $P(catch|toothache, \neg cavity) = P(catch|\neg cavity)$
- Catch is conditionally independent of Toothache given Cavity:
$P(Catch|Toothache, Cavity) = P(Catch|Cavity)$
- Equivalent statements:
$P(Toothache|Catch, Cavity) = P(Toothache|Cavity)$
$P(Toothache, Catch|Cavity) = P(Toothache|Cavity)P(Catch|Cavity)$

General formulation:
X conditionally independent from Y given Z iff:

- $P(X|Y, Z) = P(X|Z)$

  $P(X, Y|Z) = P(X|Y, Z)P(Y|Z) = P(X|Z)P(Y|Z)$

  In general,
  $P(Y1, \ldots, Yn|Z) = P(Y1|Y2, \ldots Yn, Z)P(Y2|Y3 \ldots Yn, Z) \cdots P(Yn|Z)$

  Yi conditionally independent from Yj given Z

  $P(Y1, \ldots, Yn|Z) = P(Y1|Z)P(Y2|Z) \cdots P(Yn|Z)$

Chain rule + Conditional independence
$P(Toothache, Catch, Cavity) = P(Toothache|Catch, Cavity)P(Catch, Cavity)$
$= P(Toothache|Catch, Cavity)P(Catch|Cavity)P(Cavity)$
$= P(Toothache|Cavity)P(Catch|Cavity)P(Cavity)$ = 2 + 2 + 1 = 5 independent numbers (instead of 2^3 − 1)

4.5 Bayes' Rule

Product rule: P(a ^ b) = P(a|b)P(b) = P(b|a)P(a)

⇒ **Bayes' rule** P(a|b) =P(b|a)P(a)/P(b)

Or P(Y |X) = P(X|Y)P(Y)/P(X) = alfa*P(X|Y)P(Y)

P(Cause|Effect) = P(Effect|Cause)P(Cause) / P(Effect)

With conditional independence...

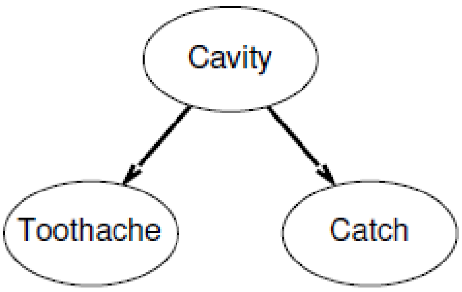
General/chained situation

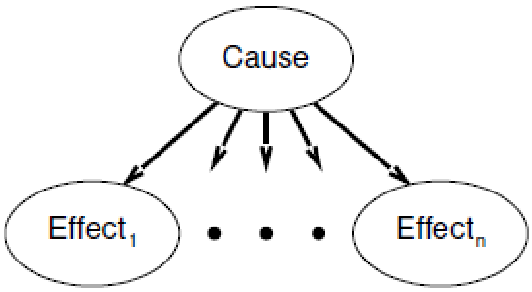
Yi , . . .Yn conditionally independent each other given Z

P(Z|Y1, . . . ,Yn) =P(Y1|Z) · · · P(Yn|Z) P(Z)

P(Cause|Effect1, . . . , Effectn) = alfa* P(Cause) PROD P(Effecti |Cause)

4.5.1 Bayesian networks





- a directed, acyclic graph (link “directly influences”)
- a conditional distribution for each node given its parents: P(Xi |Parents(Xi))

In the simplest case, conditional distribution represented as a **conditional probability table (CPT)** giving the distribution over Xi for each combination of parent values.

All **joint probabilities** computed with the chain rule:

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