Monte Carlo Artificial Intelligence
Probability I

1

Random Variables

- The basic element of probability is the random variable
- Think of the random variable as an event with some degree of uncertainty as to whether that event occurs
- Random variables have a domain of values it can take on

2

Random Variables

Example:

- *ProfLate* is a random variable for whether your prof will be late to class or not
- The domain of *ProfLate* is {*true, false}*
  - *ProfLate* = *true*: proposition that prof will be late to class
  - *ProfLate* = *false*: proposition that prof will not be late to class

3

Random Variables

Example:

- *ProfLate* is a random variable for whether your prof will be late to class or not
- The domain of *ProfLate* is <*true*, *false*>
  - *ProfLate* = *true*: proposition that prof will be late to class
  - *ProfLate* = will not be l

You can assign some degree of belief to this proposition eg. *P(ProfLate = true)* = 0.9

4

Random Variables

Example:

- *ProfLate* is a random variable for whether your prof will be late to class or not
- The domain of *ProfLate* is <*true*, *false*>
  - *ProfLate* = *true*: proposition that prof will be late to class
  - *ProfLate* = *false*: proposition that prof will not be late to class

And to this one eg.
P(ProfLate = *false)* = 0.1

5

Random Variables

- We will refer to random variables with capitalized names eg. *X*, *Y*, *ProfLate*
- We will refer to names of values with lower case names eg. *x*, *y*, *proflate*
- This means you may see a statement like *ProfLate* = *proflate*
  - This means the random variable *ProfLate* takes the value *proflate* (which can be *true* or *false*)
- Shorthand notation:

  ProfLate = *true* is the same as *proflate* and
  ProfLate = *false* is the same as ¬*proflate*

6

Random Variables

Boolean random variables

- Take the values *true* or *false*
- Eg. Let *A* be a Boolean random variable
  - $P(A = false) = 0.9$
  - $P(A = true) = 0.1$

Random Variables

Discrete Random Variables

- Allowed to taken on a finite number of values eg.
  - $P(DrinkSize=Small) = 0.1$
  - $P(DrinkSize=Medium) = 0.2$
  - $P(DrinkSize=Large) = 0.7$

Values must be 1) Mutually exhaustive and 2) Exclusive

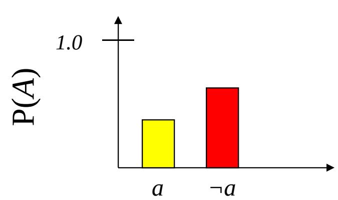
Random Variables

Continuous Random Variables

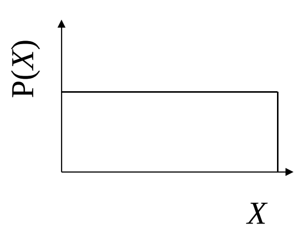
- Can take values from the real numbers
- eg. They can take values from [0, 1]
- Note: We will primarily be dealing with discrete random variables

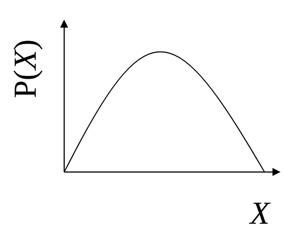
Probability Density Functions

Discrete random variables have probability distributions:



Continuous random variables have probability density functions eg:





Probability

- We will write *P(A=true)* as "the fraction of possible worlds in which *A* is true"
- We will sometimes talk about the probabilities distribution of a random variable
- Instead of writing
  - *P(A=false)* = 0.25
  - *P(A=true)* = 0.75
- We will write ***P***(*A*) = (0.25, 0.75)

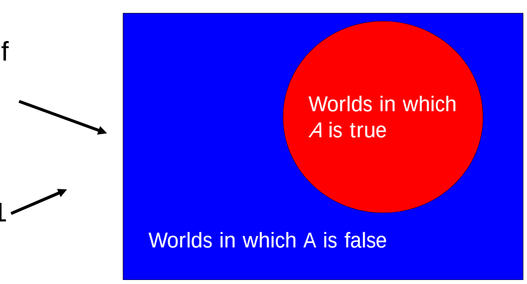
Note the boldface!

11

Probability

Event space of all possible worlds

Its area is 1



P(*a*) = Area of reddish oval

12

Probability

Axioms of Probability

- $0 \leq P(a) \leq 1$
- $P(true) = 1$
- $P(false) = 0$
- $P(a \text{ OR } b) = P(a) + P(b) - P(a \text{ AND } b)$

This OR is equivalent to set union $\cup$.

This AND is equivalent to set intersection ($\cap$). I'll often write it as *P(a, b)*

13

Conditional Probability

- We can consider *P(A)* as the unconditional or prior probability
  - eg. *P(ProfLate = true)* = 1.0
- It is the probability of event *A* in the absence of any other information
- If we get new information that affects *A*, we can reason with the conditional probability of *A* given the new information.

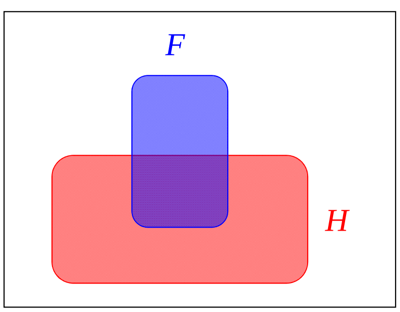
14

Conditional Probability

- P(*A* | *B*) = Fraction of worlds in which *B* is true that also have *A* true
- Read this as: "Probability of *A* conditioned on *B*"
- Prior probability P(*A*) is a special case of the conditional probability P(*A* |) conditioned on no evidence

15

Conditional Probability



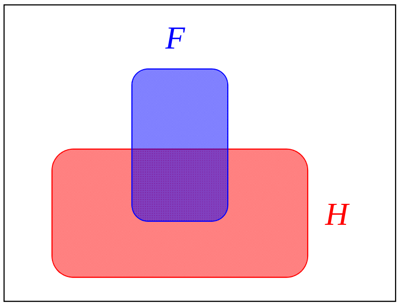
H = "Have a headache"
F = "Coming down with Flu"

P(*H*) = 1/10
P(*F*) = 1/40
P(*H* | *F*) = 1/2

"Headaches are rare and flu is rarer, but if you're coming down with 'flu there's a 50-50 chance you'll have a headache."

16

Conditional Probability



H = "Have a headache"
F = "Coming down with Flu"

P(*H*) = 1/10
P(*F*) = 1/40
P(*H* | *F*) = 1/2

P(*H*|*F*) = Fraction of flu-inflicted worlds in which you have a headache

$$= \frac{\text{\# worlds with flu and headache}}{\text{\# worlds with flu}}$$

$$= \frac{\text{Area of "H and F" region}}{\text{Area of "F" region}}$$

$$= \frac{\mathrm{P(H,F)}}{\mathrm{P(F)}}$$

17

Conditional Probability

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

Corollary: The Chain Rule (aka The Product Rule)

$$P(A,B) = P(A|B)P(B)$$

18

Joint Probability Distribution

- P(*A*, *B*) is called the joint probability distribution of *A* and *B*
- It captures the probabilities of all combinations of the values of a set of random variables

19

Joint Probability Distribution

- For example, if *A* and *B* are Boolean random variables, then P(*A*,*B*) could be specified as:

P(*A*=*false*, *B*=*false*)	0.25
P(*A*=*false*, *B*=*true*)	0.25
P(*A*=*true*, *B*=*false*)	0.25
P(*A*=*true*, *B*=*true*)	0.25

20

Joint Probability Distribution

- Now suppose we have the random variables:
  - *Drink* = {*Coke*, *Sprite*}
  - *Size* = {*Small*, *Medium*, *Large*}
- The joint probability distribution for P(*Drink*,*Size*) could look like:

P(*Drink*=*Coke*, *Size*=*Small*)	0.1
P(*Drink*=*Coke*, *Size*=*Medium*)	0.1
P(*Drink*=*Coke*, *Size*=*Large*)	0.3
P(*Drink*=*Sprite*, *Size*=*Small*)	0.1
P(*Drink*=*Sprite*, *Size*=*Medium*)	0.2
P(*Drink*=*Sprite*, *Size*=*Large*)	0.2

Joint Probability Distribution

- Suppose you have the complete set of random variables used to describe the world
- A joint probability distribution that covers this complete set is called the full joint probability distribution
- Is a complete specification of one's uncertainty about the world in question
- Very powerful: Can be used to answer any probabilistic query

Joint Probability Distribution

Toothache	Cavity	Catch	P(Toothache, Cavity, Catch)
false	false	false	0.576
false	false	true	0.144
false	true	false	0.008
false	true	true	0.072
true	false	false	0.064
true	false	true	0.016
true	true	false	0.012
true	true	true	0.108

"Catch" means the dentist's probe catches in my teeth

This cell means P(Toothache = true, Cavity = true, Catch = true) = 0.108

23

Joint Probability Distribution

Toothache	Cavity	Catch	P(Toothache, Cavity, Catch)
false	false	false	0.576
false	false	true	0.144
false	true	false	0.008
false	true	true	0.072
true	false	false	0.064
true	false	true	0.016
true	true	false	0.012
true	true	true	0.108

The probabilities in the last column sum to 1

24

Joint Probability Distribution

From the full joint probability distribution, we can calculate any probability involving the three random variables in this world eg.

P(*Toothache* = *true* OR *Cavity* = true) =

P(*Toothache=true*, *Cavity=false*, *Catch=false*) +
P(*Toothache=true*, *Cavity=false*, *Catch=true*) +
P(*Toothache=false*, *Cavity=true*, *Catch=false*) +
P(*Toothache=false*, *Cavity=true*, *Catch=true*) +
P(*Toothache=true*, *Cavity=true*, *Catch=false*) +
P(*Toothache=true*, *Cavity = true*, *Catch=true*) +

= 0.064 + 0.016 + 0.008 + 0.072 + 0.012 + 0.108 = 0.28

Marginalization

We can even calculate marginal probabilities (the probability distribution over a subset of the variables) eg:

P(*Toothache=true*, *Cavity=true*) =

P(*Toothache=true*, *Cavity=true*, *Catch=true*) +

P(*Toothache=true*, *Cavity=true*, *Catch=false*)

= 0.108 + 0.012 = 0.12

Marginalization

Or even:

P(*Cavity=true*) =
P(*Toothache=true, Cavity=true, Catch=true*) +
P(*Toothache=true, Cavity=true, Catch=false*)
P(*Toothache=false, Cavity=true, Catch=true*) +
P(*Toothache=false, Cavity=true, Catch=false*)

= 0.108 + 0.012 + 0.072 + 0.008 = 0.2

27

Marginalization

The general marginalization rule for any **sets** of variables ***Y*** and ***Z***:

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\mathbf{z}} \boldsymbol{P}(\boldsymbol{Y}, \mathbf{z})$$

z is over all possible combinations of values of ***Z*** (remember ***Z*** is a set)

or

$$\boldsymbol{P}(\boldsymbol{Y}) = \sum_{\mathbf{z}} \boldsymbol{P}(\boldsymbol{Y} \mid \mathbf{z}) P(\mathbf{z})$$

28

Normalization

$$P(Cavity = true \mid Toothache = true)$$
$$= \frac{P(Cavity = true, Toothache = true)}{P(Toothache = true)}$$
$$= \frac{0.108 + 0.012}{0.108 + 0.012 + 0.016 + 0.064} = 0.6$$

$$P(Cavity = false \mid Toothache = true)$$
$$= \frac{P(Cavity = false, Toothache = true)}{P(Toothache = true)}$$
$$= \frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

Note that 1/P(*Toothache*=*true*) remains constant in the two equations.

Normalization

- In fact, 1/P(*Toothache*=*true*) can be viewed as a normalization constant for ***P***(*Cavity* =true| *Toothache*=*true*), ensuring it adds up to 1
- We will refer to normalization constants with the symbol α

$$\mathbf{P}(Cavity = true \mid Toothache = true)$$
$$= \alpha \mathbf{P}(Cavity = true, Toothache = true)$$

30

Inference

- Suppose you get a query such as
  P(*Cavity = true* | *Toothache = true*)

Toothache is called the evidence variable because we observe it. More generally, it's a set of variables.

Cavity is called the query variable (we'll assume it's a single variable for now)

There are also unobserved (aka hidden) variables like *Catch*

31

Inference

- We will write the query as $\boldsymbol{P}(X \mid \boldsymbol{e})$

This is a probability distribution hence the boldface

X = Query variable (a single variable for now)

$\boldsymbol{E}$ = Set of evidence variables

$\boldsymbol{e}$ = the set of observed values for the evidence variables

$\boldsymbol{Y}$ = Unobserved variables

32

Inference

We will write the query as $\mathbf{P}(X \mid \mathbf{e})$

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Summation is over all possible combinations of values of the unobserved variables ***Y***

X = Query variable (a single variable for now)

E = Set of evidence variables

e = the set of observed values for the evidence variables

Y = Unobserved variables

Inference

$$\mathbf{P}(X \mid \mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{e}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y})$$

Computing $\mathbf{P}(X \mid \mathbf{e})$ involves going through all possible entries of the full joint probability distribution and adding up probabilities with $X=x_i$, $\mathbf{E}=\mathbf{e}$, and $\mathbf{Y}=\mathbf{y}$

Suppose you have a domain with n Boolean variables. What is the space and time complexity of computing $P(X \mid \mathbf{e})$?

34

Independence

- How do you avoid the exponential space and time complexity of inference?
- Use independence (aka factoring)

35

Independence

Suppose the full joint distribution now consists of four variables:

Toothache = {*true*, *false*}

Catch = {*true*, *false*}

Cavity = {*true*, *false*}

Weather = {*sunny*, *rain*, *cloudy*, *snow*}

There are now 32 entries in the full joint distribution table

36

Independence

Does the weather influence one's dental problems?

Is P(*Weather=cloudy* | *Toothache* = *toothache*, *Catch* = *catch*, *Cavity* = *cavity*) = P(*Weather=cloudy*)?

In other words, is *Weather* independent of *Toothache, Catch* and *Cavity*?

37

Independence

We say that variables X and Y are independent if any of the following hold: (note that they are all equivalent)

$$\boldsymbol{P}(X \mid Y) = \boldsymbol{P}(X) \text{ or}$$

$$\boldsymbol{P}(Y \mid X) = \boldsymbol{P}(Y) \text{ or}$$

$$\boldsymbol{P}(X, Y) = \boldsymbol{P}(X)\boldsymbol{P}(Y)$$

38

Why is independence useful?

Assume that Weather is independent of toothache, catch, cavity ie.

P(*Weather=cloudy* | *Toothache = toothache*, *Catch = catch*, *Cavity = cavity*) = P(*Weather=cloudy*)

Now we can calculate:

P(*Weather=cloudy*, *Toothache = toothache*, *Catch = catch*, *Cavity = cavity)*

= P(*Weather=cloudy* | *Toothache = toothache*, *Catch = catch*, *Cavity = cavity*) * P(*toothache*, *catch*, *cavity*)

= P(*Weather=cloudy*) * P(*Toothache = toothache*, *Catch = catch*, *Cavity = cavity*)

39

Why is independence useful?

P(*Weather=cloudy*, *Toothache = toothache*, *Catch = catch*, *Cavity = cavity)*

= P(*Weather=cloudy*) * P(*Toothache = toothache*, *Catch = catch*, *Cavity = cavity*)

This table has 4 values

This table has 8 values

- You now need to store 12 values to calculate ***P***(*Weather*, *Toothache*, *Catch*, *Cavity*)
- If *Weather* was not independent of *Toothache*, *Catch*, and *Cavity* then you would have needed 32 values

40

Independence

Another example:

- Suppose you have n coin flips and you want to calculate the joint distribution $\mathbf{P}(C_1, \ldots, C_n)$
- If the coin flips are not independent, you need 2^n values in the table
- If the coin flips are independent, then

$$P(C_1, \ldots, C_n) = \prod_{i=1}^{n} P(C_i)$$

Each $P(C_i)$ table has 2 entries and there are n of them for a total of $2n$ values

41

Independence

- Independence is powerful!
- It required extra domain knowledge. A different kind of knowledge than numerical probabilities. It needed an understanding of causation.

42

Conditional Independence

Are *Toothache* and *Catch* independent?

No – if probe catches in the tooth, it likely has a cavity which causes the toothache.

But given the presence or absence of the cavity, they are independent (since they are directly caused by the cavity but don't have a direct effect on each other)

Conditional independence:

P(*Toothache* = *true*, *Catch* = *catch* | *Cavity*) =

P(*Toothache* = *true* | *Cavity*) * ***P***(*Catch* = *true* | *Cavity*)

43

Conditional Independence

General form:

$$\boldsymbol{P}(A, B \mid C) = \boldsymbol{P}(A \mid C)\boldsymbol{P}(B \mid C)$$

Or equivalently:

$$\boldsymbol{P}(A \mid B, C) = \boldsymbol{P}(A \mid C) \quad \text{and}$$

$$\boldsymbol{P}(B \mid A, C) = \boldsymbol{P}(B \mid C)$$

How to think about conditional independence:

In P(*A* | *B*, *C*) = P(*A* | *C*): if knowing *C* tells me everything about A, I don't gain anything by knowing *B*

44

Conditional Independence

7 independent values in table (have to sum to 1)

$\mathbf{P}(Toothache, Catch, Cavity)$

$= \mathbf{P}(Toothache, Catch \mid Cavity)\ \mathbf{P}(Cavity)$

$= \mathbf{P}(Toothache \mid Cavity)\ \mathbf{P}(Catch \mid Cavity)\ \mathbf{P}(Cavity)$

2 independent values in table

2 independent values in table

1 independent value in table

Conditional independence permits probabilistic systems to scale up!

45