# Probability and NumPy

AI+X
INSTITUTE FOR ARTIFICIAL INTELLIGENCE

# Probability

- Random Variables
- Joint and Marginal Distributions
- Conditional Distribution
- Product Rule, Chain Rule, Bayes' Rule
- Inference

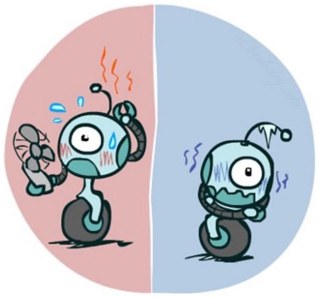
# Random Variables

- A random variable is some aspect of the world about which we (may) have uncertainty
  - R = Is it raining?
  - T = Is it hot or cold?
  - D = How long will it take to drive to work?
  - L = Where is the enemy?
- We denote random variables with capital letters
- Random variables have domains
  - R in {true, false}    (often write as {+r, –r})
  - T in {hot, cold}
  - D in [0, ∞)
  - L in possible locations, maybe {(0,0), (0,1), …}

# Probability Distributions

- Associate a probability with each value

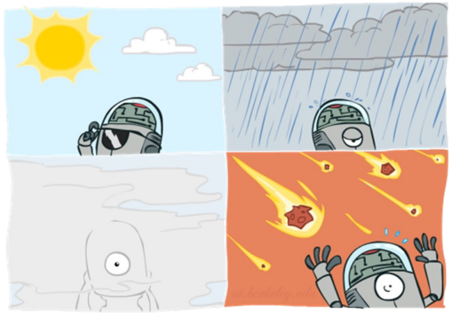
Temperature



$P(T)$

| T | P |
|---|---|
| hot | 0.5 |
| cold | 0.5 |

Weather



$P(W)$

| W | P |
|---|---|
| sun | 0.6 |
| rain | 0.1 |
| fog | 0.3 |
| meteor | 0.0 |

# Probability Distributions

- Unobserved random variables have distributions

$P(T)$

| T | P |
|---|---|
| hot | 0.5 |
| cold | 0.5 |

$P(W)$

| W | P |
|---|---|
| sun | 0.6 |
| rain | 0.1 |
| fog | 0.3 |
| meteor | 0.0 |

- A distribution is a TABLE of probabilities of values
- A probability (lower case value) is a single number
  - $P(W = rain) = 0.1$

# Probability Distributions

- Must have:

  $$\forall x \;\; P(X = x) \geq 0$$

  $$\sum_{x} P\,(\mathrm{X} = \mathrm{x}) = 1$$

- Shorthand notation (OK if all domain entries are unique):

  $P(hot) = P(T = hot),$

  $P(cold) = P(T = cold),$

  $P(rain) = P(W = rain),$

  $\ldots$

# Joint Distributions

- A joint distribution over a set of random variables $X_1, X_2, \ldots, X_n$ specifies a real number for each assignment (or outcome):

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

  $$P(x_1, x_2, \ldots x_n)$$

- Must obey:

  $$P(x_1, x_2, \ldots x_n) \geq 0$$

  $$\sum_{(x_1, x_2, \ldots x_n)} P(x_1, x_2, \ldots x_n) = 1$$

# Joint Distributions

$$P(T, W)$$

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

- Size of distribution if $n$ variables with domain sizes $d$?
  - For all but the smallest distributions, impractical to write out!

# Probabilistic Model

- A probabilistic model is a joint distribution over a set of random variables
- Probabilistic models:
  - (Random) variables with domains
  - Assignments are called outcomes
  - Joint distributions: say whether assignments (outcomes) are likely
  - Normalized: sum to 1.0

$$P(T,W)$$

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

# Events

- An event is a set E of outcomes

$$P(E) = \sum_{(x_1 \ldots x_n) \in E} P(x_1 \ldots x_n)$$

- From a join distribution, we can calculate the probability of any event
  - Probability that it's hot AND sunny?
  - Probability that it's hot?
  - Probability that it's hot OR sunny?
- Typically, the events we care about are partial assignments, like P(T=hot)

$P(T, W)$

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

# Knowledge Check 1

$P(+x, +y)?$

$P(+x)?$

$P(-y, \text{OR} + x)?$

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

| X | Y | P |
|---|---|---|
| +x | +y | 0.2 |
| +x | -y | 0.3 |
| -x | +y | 0.4 |
| -x | -y | 0.1 |

# Marginal Distributions

- Marginal distributions are sub-tables which eliminate variables
- Marginalization (summing out): Combine collapsed rows by adding

$P(T,W)$

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

$$P(t) = \sum_{s} P(t,s)$$

$P(T)$

| T | P |
|---|---|
| hot | 0.5 |
| cold | 0.5 |

$$P(s) = \sum_{t} P(t,s)$$

$P(W)$

| W | P |
|---|---|
| sun | 0.6 |
| rain | 0.4 |

$$P(\mathrm{X}_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2)$$

# Knowledge Check 2

$P(X,Y)$

| X | Y | P |
|---|---|---|
| +x | +y | 0.2 |
| +x | -y | 0.3 |
| -x | +y | 0.4 |
| -x | -y | 0.1 |

$$P(t) = \sum_{s} P(x,y)$$

$$P(s) = \sum_{t} P(x,y)$$

$P(X)$

| T | P |
|---|---|
| hot | |
| cold | |

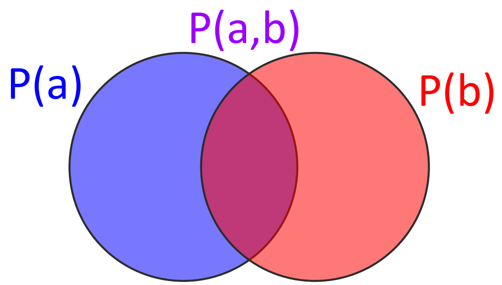
$P(Y)$

| W | P |
|---|---|
| sun | |
| rain | |

# Conditional Probabilities

- A simple relation between joint and conditional probabilities

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




- P(sun|cold)?

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

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

$P(T,W)$

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

# Knowledge Check 3

$P(+x|+y)?$

$P(-x| + y)?$

$P(-y| + x)?$

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

| X | Y | P |
|---|---|---|
| +x | +y | 0.2 |
| +x | -y | 0.3 |
| -x | +y | 0.4 |
| -x | -y | 0.1 |

# Conditional Distributions

- Conditional distributions are probability distributions over some variables given fixed values of others

$P(W|T)$

$P(W|T = hot)$

| W | P |
|---|---|
| sun | 0.8 |
| rain | 0.2 |

$P(W|T = cold)$

| W | P |
|---|---|
| sun | 0.4 |
| rain | 0.6 |

Conditional distributions

$P(T, W)$

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

Joint distributions

# Normalization Trick

$P(T, W)$

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

$$\begin{aligned} P(W=s|T=c) &= \frac{P(W=s,T=c)}{P(T=c)} \\ &= \frac{P(W=s,T=c)}{P(W=s,T=c)+P(W=r,T=c)} \\ &= \frac{0.2}{0.2+0.3} = 0.4 \end{aligned}$$

$$\begin{aligned} P(W=r|T=c) &= \frac{P(W=r,T=c)}{P(T=c)} \\ &= \frac{P(W=r,T=c)}{P(W=s,T=c)+P(W=r,T=c)} \\ &= \frac{0.3}{0.2+0.3} = 0.6 \end{aligned}$$

$P(W|T = cold)$

| W | P |
|---|---|
| sun | 0.4 |
| rain | 0.6 |

# Normalization Trick

$P(T,W)$

| 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)

$P(W|T = cold)$

| W | P |
|---|---|
| sun | 0.4 |
| rain | 0.6 |

# Knowledge Check 4

- P(X|Y = -y)?

$P(X,Y)$

| X | Y | P |
|---|---|---|
| +x | +y | 0.2 |
| +x | -y | 0.3 |
| -x | +y | 0.4 |
| -x | -y | 0.1 |

→

**SELECT** the joint probabilities matching the evidence

→

**NORMALIZE** the selection (make it sum to one)

# Normalization

- To bring or restore to a normal condition (all entries sum to 1)
- Procedure:
  - Step 1: Computer Z – sum over all entries
  - Step 2: Divide every entry by Z

| W | P |
|---|---|
| sun | 0.2 |
| rain | 0.3 |

NORMALIZE
Z = 0.5

| W | P |
|---|---|
| sun | 0.4 |
| rain | 0.6 |

| T | W | P |
|---|---|---|
| hot | sun | 20 |
| hot | rain | 5 |
| cold | sun | 10 |
| cold | rain | 15 |

NORMALIZE
Z = 50

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

# Probabilistic Inference

- Probabilistic inference: compute a desired probability from other known probabilities (e.g. conditional from joint)
- We generally compute conditional probabilities
  - P(on time | no reported accidents) = 0.90
  - These represent the agent's beliefs given the evidence
- Probabilities change with new evidence:
  - P(on time | no accidents, 5 a.m.) = 0.95
  - P(on time | no accidents, 5 a.m., raining) = 0.80
  - Observing new evidence causes beliefs to be updated

# Inference by Enumeration

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

  $\left.\begin{array}{l} E_1 \dots E_k = e_1 \dots e_k \\ Q \\ H_1 \dots H_r \end{array}\right\} X_{1,}X_{2,\dots}X_n$
- We want: $P(Q|e_{1\dots}e_k)$

- Step 1: Select the entries consistent with the evidence
- Step 2: Sum out H to get joint of Query and evidence
- Step 3: Normalize

$$P(Q, e_1 \dots e_k) = \sum_{h_1 \dots h_r} P(Q, h_1 \dots h_r, e_1 \dots e_k)$$

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

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

# Inference by Enumeration

- Obvious problems:
  - Worst-case time complexity $O(d^n)$
  - Space complexity $O(d^n)$ to store the joint distribution

# Knowledge Check 5

$P(W)$?

$P(W|\text{winter}, \text{hot})$?

$P(W|\text{winter})$?

| S | T | W | P |
|---|---|---|---|
| summer | hot | sun | 0.30 |
| summer | hot | rain | 0.05 |
| summer | cold | sun | 0.10 |
| summer | cold | rain | 0.05 |
| winter | hot | sun | 0.10 |
| winter | hot | rain | 0.05 |
| winter | cold | sun | 0.15 |
| winter | cold | rain | 0.20 |

# The Product Rule

- Sometimes have conditional distributions but want the joint

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

- Example:

$P(W)$

| W | 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

$$P(x_1, x_2, x_3) = P(x_1)P(x_2|x_1)P(x_3|x_1, x_2)$$
$$P(x_1, x_3, \dots x_n) = \prod_i P(x_i|x_1 \dots x_i - 1)$$

# Bayes' Rule

- Two ways to factor a join distribution over two variables:
- $$P(x, y) = P(x|y)P(y) = P(y|x)P(x)$$
- Dividing both sides by P(y) we get:
- $$P(x|y) = \frac{P(y|x)}{P(y)} P(x)$$
- Why is this at all helpful?
  - Let us build one conditional from its reverse
  - Often one conditional is tricky but the other one is simple

# Inference with Bayes' Rule

- Example: Diagnostic probability from casual probability:
- $$P(\text{cause}|\text{effect}) = \frac{P(\text{effect}\,|\,\text{cause})P(\text{cause})}{P(\text{effect})}$$
- M: meningitis, S: stiff neck
- $P(+m) = 0.0001$
- $P(+s|+m) = 0.8$
- $P(+s|-m) = 0.01$
- $$P(+m|+s) = \frac{P(+s|+m)P(+m)}{P(+s)} = \frac{P(+s|+m)P(+m)}{P(+s|+m)P(+m)+P(+s|-m)P(-m)} = \frac{0.8\times 0.0001}{0.8\times 0.0001+0.01\times 0.999}$$
- Posterior probability of meningitis still very small!!!

# Knowledge Check 6

- What is P(W|dry)?

$P(W)$

| W | 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 |

You have reached the end of the lecture.