E.g. coin, $\mathbb{P}(H) = 0.7$

$\Rightarrow \mathbb{P}(T) = 0.3$

Conditional probability

Modeling knowledge of uncertainty

Michael Psenka

What we aim to model

$\mathbb{P}(x \mid z)$

what we're modelling

information we know

$\{52\}$ (1) $\mathbb{P}(\text{card} > 7?) \approx 0.46$

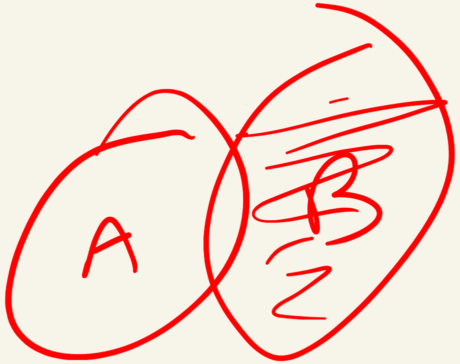
$\{30\}$ (2) $\mathbb{P}(\text{card} > 7)?$ (already drew 22 cards above 7)

$(1): \dfrac{|A|}{|\Omega|} \approx 0.46$

$(2): \dfrac{|A|}{|\{30\}|} = 0.80$

Conditional probability

Definition. *The conditional probability* $\mathbb{P}(A \mid B)$ *(read "probability of* A *given* B*") is defined as the following:*



$$\mathbb{P}(A \mid B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

$$\mathbb{P}(A \cap B) = \mathbb{P}(B)\,\mathbb{P}(A \mid B)$$

Bayes' Theorem

Theorem. *For a distribution $\mathbb{P}$, the following equation holds:*

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

<u>Proof</u> By def., $\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$. Multiply over, we get $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$. Since $\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A)$, we get $\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(B|A)\mathbb{P}(A)$. Divide $\mathbb{P}(B)$ to get desired equation. $\blacksquare$

Why Bayes'?

$$\mathbb{P}(I \mid \text{Pos}) = \frac{\mathbb{P}(\text{Pos} \mid I)\,\mathbb{P}(I)}{\mathbb{P}(\text{Pos})}$$

$$= \frac{\mathbb{P}(\text{Pos} \mid I)\,\mathbb{P}(I)}{\mathbb{P}(\text{Pos} \mid I)\,\mathbb{P}(I) + \mathbb{P}(\text{Pos} \mid \text{not } I)\,\mathbb{P}(\text{not } I)}$$

$$\text{Partition} = \{I, \text{not } I\}$$

$$\left(\mathbb{P}(\text{not } I) = 1 - \mathbb{P}(I)\right)$$

A_1, A_2, A_3

$\{A_i\}_{i=1}^{n}$:= "collection of n sets"

Total probability rule

$\bigcup_{i=1}^{n} A_i := A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n$

Theorem. *Let* $(\Omega, \mathbb{P})$ *be a random variable. If we have a collection of sets* $\{A_i\}_{i=1}^{n}$ *that "partition" the sample space* Ω*, i.e.:*

1. $A_i \cap A_j = \emptyset$ *for all* $i, j \in \{1, \ldots, n\}$,
2. $\cup_{i=1}^{n} A_i = \Omega$,

Then the following equation holds for any event $B \subset \Omega$*:*

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\mathbb{P}(A_i).$$

$a \cdot (b + c) = a \cdot b + a \cdot c$ "distributive property"

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Total probability rule (1)

Proof. We know $\bigcup_{i=1}^{n} A_i = \Omega$, so since $\Omega \cap B = B$,

$\Rightarrow$ $B = B \cap \left(\bigcup_{i=1}^{n} A_i\right)$. Distributive property, write

$B = \bigcup_{i=1}^{n} B \cap A_i$. Since A_i's (2) disjoint, $\bigcup_{i=1}^{n} B \cap A_i$ is

a disjoint union, thus use additivity (from lecture)

to conclude $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \cap A_i)$. Since

$\mathbb{P}(B \cap A_i) = \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)$, $\Rightarrow$ $\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(B \mid A_i)\,\mathbb{P}(A_i)$. ∎

Joint distributions

"Definition": a random variable ~~of~~ the form $(A \times B, \mathbb{P})$ for some sets A, B.

E.g. $A = \{Pos, Neg\}$, $B = \{I, \text{not } I\}$

$\mathbb{P}((Pos, I))$

	Pos	Neg
I	0.2	0.3
not I	0	0.5

Applying Bayes' joint distribution

	I	not I
Pos	A	B
Neg	C	D

1st attempt: solve for unknowns in:

$$P(Pos \mid I) = \frac{P(Pos \cap I)}{P(I)} = \frac{A}{A+C} = 0.9$$

$$P(I) = A+C = 0.05$$

$$P(Pos) = A+B = 0.2$$

Conclusion Can be done, but hard to solve for distributions purely algebraically: this is the power of formulas like Bayes' Theorem

Simpler, using equations, just plug in:

"5% of US population is affected": $\mathbb{P}(I) = 0.05$, $\mathbb{P}(\text{not } I) = 0.95$

"90% of affected people test positive": $\mathbb{P}(\text{Pos} \mid I) = 0.9$

"20% of healthy people test positive": $\mathbb{P}(\text{Pos} \mid \text{not } I) = 0.2$

"Probability of being affected": $\mathbb{P}(I \mid \text{Pos})$

$$= \frac{\mathbb{P}(\text{Pos} \mid I)\,\mathbb{P}(I)}{\mathbb{P}(\text{Pos} \mid I)\,\mathbb{P}(I) + \mathbb{P}(\text{Pos} \mid \text{not } I)\,\mathbb{P}(\text{not } I)} = \frac{0.9 \cdot 0.05}{0.9 \cdot 0.05 + 0.2 \cdot 0.95} \simeq 0.19$$