36 705 WORK - WEEK 1, L1, L2

20/04/2020

- Review CH 1-3 ; with a view to completing problems
- Wasserman

CH 1 - Probability - key points/definitions

1.2 - Sample spaces, events

. experiment
. outcomes
. events

. Sample space Ω - set of possible outcomes of experiment
. Points $w \in \Omega$ -
(sample outcomes/ realis. /elements)
. Subsets of Ω - events
. Define /set up your problem using these as your building blocks

- Ex 1.1 ✓
- Ex 1.2 ✓
- Example 1.3
- Toss coin forever (experiment)
- Sample space $\Omega = \{(w = (w_1, w_2, w_3, \dots) : w_i \in \{H, T\}\}$
- The event E that heads on 3rd toss:-
- $E = \{(w_1, w_2, w_3, \dots) : w_1 = T, w_2 = T, w_3 = H, w_i \in \{H,T\} \text{ for } i > 3\}$
- Complements, unions, intervals and other properties

. For an event A
. The complement is informally "not A". (set of outcomes (w) in sample space (Ω), but not in subset (event) A)
. Formally, the complement $A^c = \{w \in \Omega : w \notin A\}$
. The complement of the sample space Ω is the empty set $\emptyset$.

. For events A, B
. The union of A and B is informally "A or B" (or both")
. The union of events A and B is defined as:-

$$A \cup B = \{w \in \Omega : w \in A \text{ or } w \in B \text{ or } w \in \text{both}\}$$

. For a sequence of sets $A_1, A_2, \dots$ (union-sequence)

$$\bigcup_{i=1}^{\infty} A_i = \{w \in \Omega : w \in A_i \text{ for at least one } i\}$$

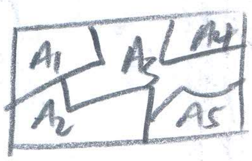
- For events A, B
- The <u>intersection</u> is informally 'A and B'
- Formally, the <u>intesection</u> is defined as

$$A \cap B = \{\omega \in \Omega : \omega \in A \text{ and } \omega \in B\}$$

- Notationally, $A \cap B = AB = (A, B)$
- For a <u>sequence of events</u> $A_1, A_2, \dots$ (<u>intesection-sequence</u>)

$$\bigcap_{i=1}^{\infty} A_i = \{\omega \in \Omega : \omega \in A_i \ \forall i\}$$

- The <u>set difference</u>: $A - B = \{\omega \in \Omega : \omega \in A, \omega \notin B\}$
- If every element of A is also contained in B; $A \subset B$ (subset) or $B \supset A$ (supeset)
- If A is a finite set; $|A|$ is its cardinality (no. of <u>elements</u>)
- $A_1, A_2, \dots$ are <u>disjoint</u>/<u>mutually exclusive</u> if $A_i \cap A_j = \emptyset$ for $i \neq j$
  (e.g. $A_1 = [0,1)$, $A_2 = [1,2)$, ... $A_n = [n-1, n)$)
- A <u>partition</u> of sample space Ω is a <u>sequence of disjoint sets</u> $A_1, A_2, \dots$ such that $\bigcup_{i=1}^{\infty} A_i = \Omega$



- Given an event A, an <u>indicator function</u> of A:

$$I_A(\omega) = I(\omega \in A) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

- A sequence of sets $A_1, A_2, \dots$ is <u>monotone increasing</u> if:-
  i) $A_1 \subset A_2 \subset \dots$
  ii) we define $\lim_{n \to \infty} A_n = \bigcup_{i=1}^{\infty} A_i$

'$A_n \to A$'

- A sequence of sets $A_1, A_2, \dots$ is <u>monotone decreasing</u> if:-
  i) $A_1 \supset A_2 \supset A_3, \dots$
  ii) we define $\lim_{n \to \infty} A_n = \bigcap_{i=1}^{\infty} A_i$

1.3 - Probability

- Assign a real no. $P(A)$ to every event A, called the probability of A
- We call P a probability distribution or probability measure
- To qualify as a probability, P must satisfy 3 axioms:-

1.5 Definition

- A function P that assigns a real number $P(A)$ to each event A is a probability distr/measure if it satisfies:-

1) $P(A) \geq 0$ for every event (A) (pls name)

2) $P(\Omega) = 1$

3) If $A_1, A_2, \ldots$ are disjoint then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- 2 interpretations of $P(A)$ - frequentist - asymptotic relative frequency a ct
  Bayesian - observer strength of belief that, A is true.
- Both rely on above axioms; only make a difference in inference
- Properties of P:- (from axioms)

$P(\emptyset) = 0$

$A \subset B \Rightarrow P(A) \leq P(B)$

$0 \leq P(A) \leq 1$

$P(A^c) = 1 - P(A)$

$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$

1.6 Lemma

- For any events A and B

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- Proof:- Rewrite $A \cup B = AB^c \cup AB \cup A^cB$ (i.e. mechanistically decompose $A \cup B$)
  - Note that each is disjoint
  - Apply a trick

1.8 Theorem (continuity of Probabilities)

- If $A_n \to A$ then $P(A_n) \to P(A)$ as $n \to \infty$

Proof - see scribblings/book
- Assume A_n monotone increasing
- Define B_i appropriately
- make an argument via disjointness of B_i ; and $\hat{\cup}$, $\overset{\infty}{\cup}$
- Apply Axiom 3 (countable add.) and limits.

1.4 Probability on Finite Sample Spaces

- For a finite sample space $\Omega = \{w_1, w_2, \ldots, w_n\}$ and if each outcome w_i is equally likely
- $P(A) = \dfrac{|A|}{|\Omega|}$ - uniform probability distribution (remember Laplace (19th from Mauldin)
- compute probabilities $\Rightarrow$ combinatorial methods for counting no of w_i in event A.

- given n objects, there are $n!$ ways of ordering them ; $0! = 1$
- given n objects, there are $\binom{n}{k}$ distinct ways of choosing k objects
- Properties of binomial coefficient:-

$$\binom{n}{0} = \binom{n}{n} = 1 \quad \binom{n}{k} = \binom{n}{n-k}$$

1.5 Independent Events

1.9 Definition

- Two events A and B are independent if:-

$$P(AB) = P(A)P(B)$$

And we write $A \perp B$
- A set of events $\{A_i : i \in I\}$ is independent if

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

- for every finite subset J of I. If A and B are not independent, we write
- $A \not\perp B$

- We can 1) Assume independence as part of our probabilistic model (e.g. coin has no 'memory')
2) Derive by verification of $P(AB) = P(A)P(B)$

- Suppose A and B are disjoint events, each with positive prob.
- Can they be independent? No
- $P(A)P(B) > 0$ but $P(AB) = P(\emptyset) = 0$
- Except in this case, no way to judge independence by looking at sets in venn diagram ?

(*)

1.6. Conditional Probability

- Assuming that $P(B) > 0$, we define the conditional probability of an event A, given B has occurred as follows:-

1.12 Definition

- If $P(B) > 0$, then the conditional probability of A given B is:-

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (1.4)$$

(*): Intuition
- $P(A|B)$ as fraction of times A occurs among those in which event B occurs ✓
- For any fixed B such that $P(B) > 0$, $P(\cdot|B)$ is a probability (i.e. satisfies axioms)
- In particular:- (Axioms applied to conditional probability)

i) $P(A|B) \geq 0 \ \forall$ events A

ii) $P(\Omega|B) = 1$

iii) If $A_1, A_2, \ldots,$ are disjoint then $P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} P(A_i|B)$

(*). Rules of probability apply to the left of the bar ✓

(!!): $P(A|B \cup C) \neq P(A|B) + P(A|C)$ } in general

$P(A|B) \neq P(B|A)$

- Be really careful with these e.g. - spots/measles
- (*) - Example 1.13 - excellent → check you've mastered; (*) definition (1.12) C.D. definition does NOT require independence of events!

1.14 Lemma

- If A and B are independent events then $P(A|B) = P(A)$
- For any pair of independent events $P(AB) = P(A|B)P(B) = P(B|A)P(A)$

- Intuitive interpretation of independence (other than multip.)
- Knowing/observing event B does not change the probability of event A
- Helpful for calculating probabilities

1.7 Bayes Theorem

1.16. Theorem (Law of Total Probability)

- Let $A_1, A_2, \dots, A_k$ be a partition of Ω
- For any event B,

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

1.16 Proof
- Define $C_j = BA_j$
- Note disjointness of C_j
- Specify B in terms of C_j, apply additivity, cp definition.

1.17 Theorem (Bayes Theorem)

- Let $A_1, A_2, \dots, A_k$ be a partition of Ω such that $P(A_i) > 0 \ \forall i$
- If $P(B) > 0$, then for each $i = 1, \dots, k$:-

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)} \quad (1.5)$$

1.18 Remark

- $P(A_i)$ is the prior probability of A
- $P(A_i|B)$ is the posterior probability of A

1.18 - Proof

- Apply c.p. formula twice; then law of total expectation

1.9 Appendix

- Not always possible to assign a probability to every event A if sample space is large, such as whole real line
- Instead, assign probabilities to limited class of a set called a σ-field

- Not feasible to assign probabilities to all subsets of a sample space Ω
- Restriction attention to a set of events called a σ-algebra / σ-field; a class $\mathcal{A}$ that satisfies:-
  i) $\emptyset \in \mathcal{A}$
  ii) If $A_1, A_2, \ldots \in \mathcal{A}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$
  iii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
- The sets in $\mathcal{A}$ are said to be measurable
- $(\Omega, \mathcal{A})$ is a measurable space
- If P is a probability measure defined on $\mathcal{A}$, then $(\Omega, \mathcal{A}, P)$ is called a probability space
- When Ω is real-line; $\mathcal{A}$ is smallest σ-field that contains all open subsets i.e. the Borel σ-field