Supplemental Notes

EE503 Week 02 Dr. Franzke

HW # 2

1. HW # 2 handout.

   Argument assignment.

2. Leon-Garcia

3. Gubner

Topics

- Rules of Inference (Mathematical argumentation)
- Borel Sigma-Algebra CIA.
- Conditional Probability
- Theorem of total probability and Bayes Theorem

Rules of Inferance (the logical schema)

Modus Ponens

Premise 1: $P \to Q$

Premise 2: P

Conclusion: Q

Theorem: $[(P \to Q) \& P] \to Q$

Prf: by truth table

Modus Tollens

Premise 1: $P \to Q$

Premise 2: $\sim Q$

Conclusion: $\sim P$

Theorem: $[(P \to Q) \& \sim Q] \to \sim P$

Conditional: $P \to Q$

Contra positive: $\sim Q \to \sim P$

Converse: $Q \to P$

Theorem: "Material implication"

$$P \to Q = \sim P \vee Q$$

Defn: An argument is VALID iff the premises logically imply the conclusion (can't make if-part TRUE and then-part FALSE)

Defn: An argument is SOUND iff it is valid and all premises are true

CIA CUT CAT

- mnemonic for generating sigma-algebras.

Thrm: $\mathcal{A}_\alpha \subset 2^X$ are S.A $\longrightarrow$ $\bigcap_\alpha \mathcal{A}_\alpha$ is a S.A.

$\therefore$ S.A.'s are closed under intersection

CIA: Pick any $\mathcal{A} \subset 2^X$. Then $\mathcal{A}$ generates the S.A $\sigma(\mathcal{A})$:

$$\sigma(\mathcal{A}) = \bigcap_\alpha \mathcal{A}_\alpha \quad , \quad \text{where}$$

C : Containment : $\mathcal{A} \subset \mathcal{A}_\alpha \quad \forall \alpha$

I : Intersection : $\bigcap_\alpha \mathcal{A}_\alpha$

A : Sigma-algebras : $\mathcal{A}_\alpha$ is a S.A. (CUT) $\forall \alpha$.

Defn: $\mathcal{B}(\mathbb{R})$ is the Borel S.A. on $\mathbb{R}$. iff

$$\mathcal{B}(\mathbb{R}) = \sigma\Big(\big\{(-\infty, a] : a \in \mathbb{R}\big\}\Big)$$

$\therefore$ $\mathcal{B}(\mathbb{R})$ is the "interval S.A."

Fact: $|\mathcal{B}(\mathbb{R})| = \underset{\substack{\uparrow \\ \text{"power of the continuum"}}}{c} = |\mathbb{R}| < 2^c = |2^{\mathbb{R}}|$.

$\therefore$ $\mathcal{B}(\mathbb{R})$ is not "too-big" for a prob. space

(unlike $2^{\mathbb{R}}$)

Defn: $f: X \to Y$ is 1-to-1 ("Injective") iff

$$\forall x \, \forall z : f(x) = f(z) \to x = z.$$

Defn: $f: X \to Y$ is ONTO ("Surjective") iff

$$\forall y \in Y \, \exists x \in X : y = f(x) \qquad (f(X) = Y)$$

Defn: $f: X \to Y$ is a BIJECTION ("1-to-1 correspondence") iff f is 1-to-1 and onto.

$|A|$ = cardinality ("size") of set A.

$\mathbb{N} = \{1, 2, 3, \dots\}$ natural numbers

$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \dots\}$ integers

$\mathbb{Z}^+ = \mathbb{N}$

$\mathbb{R} = (-\infty, \infty)$ reals

Defn: A is finite iff $A \xleftrightarrow[\text{onto}]{\text{1-to-1}} S$ for some $S \subset \mathbb{N}$ such that $A = \{a_1, \dots, a_n\}$ for some $n \in \mathbb{N}$ (or $A = \emptyset$)

Else A is infinite.

Fact: A infinite $\longleftrightarrow$ $A \xleftrightarrow[\text{onto}]{\text{1-to-1}} B$ for some $B \subset A$ and $B \neq A$.

Cantor's theorem: $|X| < |2^X|$

Defn: A is denumerable iff $A \xleftrightarrow[\text{onto}]{\text{1-to-1}} \mathbb{N}$ (e.g. $\mathbb{Z}$)

Defn: A is countable iff (1) A is finite, or

(2) A is denumerable

Facts: $|\mathbb{Z}| = |\mathbb{N}| = \aleph_0$ (aleph-nought)

$|\mathbb{R}| = |2^{\mathbb{Z}}| = c = \aleph_1$

power of the continuum

$|A| = \aleph_k \implies |2^A| = \aleph_{k+1}$ $(k = 0, 1, \dots)$

Cantor's Continuum Hypothesis There is no ω such that

$$\aleph_k < \omega < \aleph_{k+1}$$

Fact: There is no S.A. $\mathbb{Q}: |\mathbb{Q}| = |\mathbb{Z}|$.

Ex: Cannot pick an integer "at random"

Defn: Suppose $(\Omega, \mathcal{A}, P)$ is a probability space.
Events A and B are (statistically) independent w.r.t. P iff $P(A \cap B) = P(A) \cdot P(B)$

"joint factors into marginals"

Defn: Conditional Probability $P(B|A) = \dfrac{P(A \cap B)}{P(A)}$ if $P(A) > 0$.

$\therefore$ A & B independent $\longleftrightarrow$ $P(B|A) = P(B)$
$\longleftrightarrow$ $P(A|B) = P(A)$

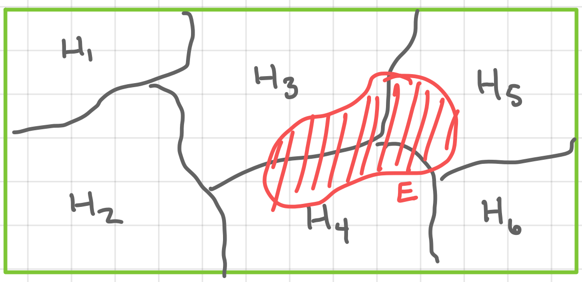
Defn: Suppose $(\Omega, \mathcal{A})$ is a measureable space. Then $\{H_k\} \subset \mathcal{A}$ partition Ω ("is a partition") iff

(1) $\bigcup_k H_k = \Omega$

(2) pairwise disjont $H_i \cap H_j = \phi$ if $i \neq j$

Theorem: (of Total Probability)

$$P(E) = \sum_k P(H_k) \cdot P(E | H_k) \quad \text{if } \{H_k\} \text{ partition } \Omega$$



x

Hypotheses: H_k

Evidence: E

Prf: $P(E) = P[E \cap \Omega]$
$= P[E \cap (\bigcup_k H_k)]$ since $\{H_k\}$ partiton Ω

$$= P\left(\bigcup_k (E \cap H_k)\right)$$

$$= \sum_k P(E \cap H_k) \quad \text{since } \{H_k\} \text{ partition } \Omega$$

$$= \sum_k P(H_k) \cdot P(E \mid H_k) \quad \text{defn of cond. prob. } P(B \mid A) = \frac{P(A \cap B)}{P(A)}$$

QED

Later in this course:

- Total Expectation $\quad E_X[X] = E_Y[E[X \mid Y]]$
- Total Variance $\quad V_X[X] = E_Y[V[X \mid Y]] + V_Y[E[X \mid Y]]$.

extremely important for problem solving

Special case: $\Omega = A \cup A^c$

$$\therefore P(B) \overset{\text{total prob.}}{=} P(A) \cdot P(B \mid A) + P(A^c) \cdot P(B \mid A^c)$$

Note: for any B.

Theorem: (Baye's Theorem)

$$P(H_j \mid E) = \frac{P(E \mid H_j) \cdot P(H_j)}{\sum_k P(E \mid H_k) \cdot P(H_k)} \quad \text{if } \{H_k\} \text{ partitions } \Omega.$$

Prf: $P(H_j \mid E) = \dfrac{P(E \cap H_j)}{P(E)}$ defn of cond. prob

$$= \frac{P(H_j) \cdot P(E \mid H_j)}{P(E)} \quad \text{" " "}$$

$$= \frac{P(H_j) \cdot P(E|H_j)}{\sum_k P(H_k) \cdot P(E|H_k)}$$ total prob.

QED

$P(H_j)$: Prior

$P(H_j|E)$: Posterior

$P(E|H_j)$: Likelihood

$\therefore P(H|E) = \frac{a}{a+b}$

$P(H^c|E) = \frac{b}{a+b}$

Mnemonic:

"Party Unconditionally To Conquer Bayes"

P : Partition $\{H_k\}$ or $A \cup A^c$

U : Unconditional probability $P(B)$

T : Total Probability

C : Conditional Probability $P(B|A)$

B : Bayes Theorem

Issue Spotting Sequence

- Is there a partition ?
  - Unconditional probability $P(B)$?
    $\therefore$ Use total probability
  - Conditional probability $P(B|A)$?
    $\therefore$ Use Bayes Theorem

Theorem: ("Booles inequality") $P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k)$

Theorem: ("Multiplication theorem")

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)\cdot P(A_2 \mid A_1)\cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

Pf: (by induction on $n = 2, 3, 4, \ldots$)

Basis Step: $n = 2$.

$$P(A_1 \cap A_2) = P(A_1)\cdot P(A_2 \mid A_1)$$ by defn cond. prob.

QED - basis

Induction Step: - Assume holds for n.

- Derive that it holds for $n+1$

Induction Hypothesis (IH).

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)\cdot P(A_2 \mid A_1)\cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1})$$

$$\Rightarrow P\left(\bigcap_{k=1}^{n+1} A_k\right) = P\left(\bigcap_{k=1}^{n} A_k \cap A_{n+1}\right)$$

$$\overset{\text{assoc.}}{=} P\left(\left(\bigcap_{k=1}^{n} A_k\right) \cap A_{n+1}\right)$$

$$= P\left(\bigcap_{k=1}^{n} A_k\right)\cdot P\left(A_{n+1} \,\middle|\, \bigcap_{k=1}^{n} A_k\right)$$

$$= P(A_1)\cdot P(A_2 \mid A_1)\cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}) \cdot P\left(A_{n+1} \,\middle|\, \bigcap_{k=1}^{n} A_k\right)$$

QED.

Odds form of Baye's Theorem:

$$\frac{P(H|E)}{P(H^c|E)} = \underbrace{\frac{P(E|H)}{P(E|H^c)}}^{\text{likelihood ratio}} \cdot \frac{P(H)}{P(H^c)}$$

or $$O(H|E) = \underbrace{L(H|H^c)}_{\text{likelihood ratio}} \cdot O(H)$$

where $O(A) = \text{Odds}(A) = \dfrac{P(A)}{P(A^c)}$ $\quad \therefore P(A) = \dfrac{O(A)}{1+O(A)}$

Two-class Classification: H vs. H^c

Then can write the Baye's Theorem posterior $P(H|E)$ in terms of the odds form Baye's theorem term:

$$O(H|E) = \frac{P(H|E)}{P(H^c|E)} = \frac{P(E|H)}{P(E|H^c)} \cdot \frac{P(H)}{P(H^c)}$$

Theorem: The posterior $P(H|E)$ has the logistic form.

$$P(H|E) = \frac{1}{1+e^{-\phi(E)}} \quad \text{where } \phi(E) = \ln O(H|E)$$

Prf:

$$P(H|E) = \frac{P(H)\cdot P(E|H)}{P(H)\cdot P(E|H) + P(H^c)\cdot P(E|H^c)}$$

$$= \frac{1}{1 + \dfrac{P(H^c)\cdot P(E|H^c)}{P(H)\cdot P(E|H)}}$$

$$= \frac{1}{1 + \dfrac{1}{O(H|E)}}$$

$$= \frac{1}{1 + e^{\ln(1/o(H|E))}}$$

$$= \frac{1}{1 + e^{-\ln o(H|E)}}$$

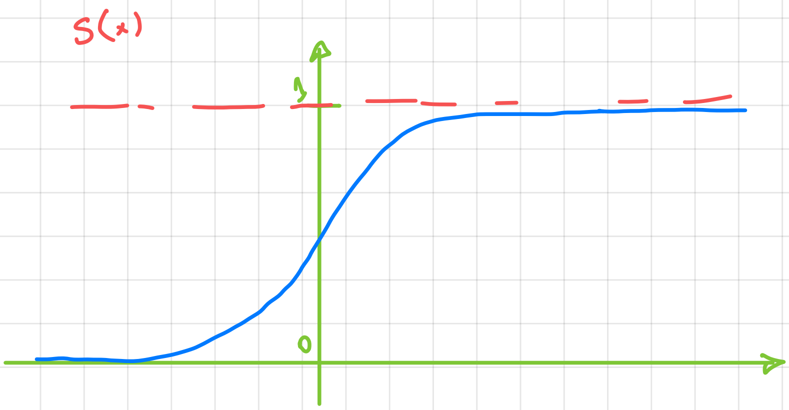
$$= \frac{1}{1 + e^{-\phi(E)}}$$

QED.

Ex: Logistic "activation" function

$$S(x) = \frac{1}{1 + e^{-cx}} \quad \text{for } c > 0.$$

$$\therefore \quad S' = c \cdot S\,(1 - S) > 0$$



$\therefore$ S is a "soft threshold"

Thrm: Probabilistic Modus Ponens

Premise 1: $P(B|A) \geq c$

Premise 2: $P(A) \geq a$

Conclusion: $P(B) \geq a \cdot c$

Pf:

$$
\begin{aligned}
P(B) &\geq P(A \cap B) && \text{monotonocity} \\
&= P(A) \cdot \frac{P(A \cap B)}{P(A)} && P(A) > 0 \\
&= P(A) \cdot P(B|A) && \text{defn cond. prob.} \\
&\geq a \cdot c && \text{premise 1 and 2.}
\end{aligned}
$$

QED

Check: $a = c = 1$ $\therefore P(B) = 1$

Thrm: Probabilistic Modus Tollens

Premise 1: $P(B|A) \geq c > 0$

Premise 2: $P(B) \leq b.$

Conclusion: $P(A) \leq \min\left(1, \frac{b}{c}\right)$

Pf:

$$
\begin{aligned}
P(A) &\leq P(A)\left(\frac{P(B)}{P(A \cap B)}\right) && \text{since } P(A \cap B) \leq P(B) \\
&= \frac{P(B)}{P(B|A)} \\
&\leq \frac{b}{P(B|A)} && \text{premise 1.} \\
&\leq \frac{b}{c} && \text{premise 2}
\end{aligned}
$$

Always $P(A) \leq 1$ $\therefore P(A) \leq \min\left(1, \frac{b}{c}\right)$

QED.

Independent Sigma-Algebras (optional)

Defn: Sigma-algebras $\mathcal{A}$ and $\mathcal{B}$ on $(\Omega, \mathcal{Q}, P)$ are independent iff $P(A \cap B) = P(A) \cdot P(B)$ for all $A \in \mathcal{A}$ and all $B \in \mathcal{B}$.

Thrm: Events A and B are independent on probability space $(X, \mathcal{Q}, P)$ iff $\sigma(\{A\})$ and $\sigma(\{B\})$ are independent sigma-algebras

Prf: $\sigma(\{A\}) \overset{CIA}{=} \{\emptyset, A, A^c, X\}$ $\qquad \sigma(\{B\}) = \{\emptyset, B, B^c, X\}$.

" $\Leftarrow$ "

Say $\sigma(\{A\})$ and $\sigma(\{B\})$ are independent.

$\therefore$ A and B are independent because $A \in \sigma(\{A\})$ and $B \in \sigma(\{B\})$.

QED.

" $\Rightarrow$ "

Say A and B are independent on $(\Omega, \mathcal{Q}, P)$

$\therefore P(A \cap B) = P(A) \cdot P(B)$

- $P(\Omega) = 1 = P(\Omega \cap \Omega)$ $\qquad \therefore P(\Omega \cap \Omega) = 1 = 1 \cdot 1 = P(\Omega) \cdot P(\Omega)$

  $\therefore \Omega \in \sigma(\{A\})$ and $\Omega \in \sigma(\{B\})$ are independent.
- $0 = P(\emptyset) = P(\emptyset \cap \emptyset)$ $\qquad \therefore P(\emptyset \cap \emptyset) = 0 = 0 \cdot 0 = P(\emptyset) \cdot P(\emptyset)$.
- $P(A \cap \Omega) = P(A) = P(A) \cdot 1 = P(A) \cdot P(\Omega)$

  $\therefore A \in \sigma(\{A\})$ and $\Omega \in \sigma(\{B\})$ are independent.

  and same for $\Omega \in \sigma(\{A\})$ and $B \in \sigma(\{B\})$

- $P(A^c \cap B^c) \overset{DeM}{=} P((A \cup B)^c) = 1 - P(A \cup B)$

$$\overset{Add}{=} 1 - P(A) - P(B) + P(A \cap B)$$

$$\overset{Ind}{=} 1 - P(A) - P(B) + P(A) \cdot P(B)$$

$$= (1 - P(A))(1 - P(B))$$

$$= P(A^c) \cdot P(B^c)$$

$\therefore$ $A^c \in \sigma(\{A\})$ and $B^c \in \sigma(\{B\})$

- $A = A \cap \Omega = A \cap (B \cup B^c) \overset{dist}{=} (A \cap B) \cup (A \cap B^c)$

and $(A \cap B) \cap (A \cap B^c) = A \cap (B \cap B^c) = A \cap \phi = \phi$.

$\therefore$ $P(A) \overset{CAT}{=} P(A \cap B) + P(A \cap B^c)$

$\therefore$ $P(A \cap B^c) = P(A) - P(A \cap B)$

$$\overset{ind}{=} P(A) - P(A) \cdot P(B)$$

$$= P(A)(1 - P(B))$$

$$= P(A) \cdot P(B^c).$$

$\therefore$ $A \in \sigma(\{A\})$ and $B^c \in \sigma(\{B\})$ are independent.

Same argument, $A^c \in \sigma(\{A\})$ and $B \in \sigma(\{B\})$ are independent.

$\therefore$ $P(C \cap D) = P(C) \cdot P(D) \quad \forall C \in \sigma(\{A\}), \forall D \in \sigma(\{B\})$

$\therefore$ $\sigma(\{A\})$ and $\sigma(\{B\})$ are independent.

QED.

Defn: A π-system is a non-empty set collection closed under finite intersections

Thrm: If probability measures P and Q agree on a π-system $\mathcal{I}$ (i.e $P(A) = Q(A)\ \forall A \in \mathcal{I}$) then P and Q agree on the generated sigma-algebre $\sigma(\mathcal{I})$.

Analogy: for functions continuous on $[0,T]$ you can uniquely represent f. from samples $f(x)$ on rationals in the interval (or any set "dense" in $[0,T]$)

So the above theorem acts like a sampling theorem for probability measures on $(X, \mathcal{Q})$