Chapter 1. Probability Theory

1.1 Set theory

Defn: $S = \{$all possible outcomes of an experiment$\}$ is called a sample space.

(e.g.) (coin tossing) $S = \{H, T\}$

- S can be either countable or uncountable.
  (↳ $\exists$ 1-1 correspondence with a subset of the integers)

- The distinction betⁿ countable and uncountable sample space is important in that it dictates the way in which prob. can be assigned.

Defn: $A \subset S$ is an event.

$\Leftrightarrow \forall x \in A,\ x \in B$.

cf. $A = B \Leftrightarrow A \subset B$ and $B \subset A$.

Set operations: $A \cup B$, $A \cap B$, A^c, ϕ

Thm: $A, B, C \subset S$

a. Commutativity: $A \overset{\cap}{\cup} B = B \overset{\cap}{\cap} A$

b. Associativity: $A \overset{\cap}{\cup} (B \overset{\cap}{\cup} C) = (A \overset{\cap}{\cup} B) \overset{\cap}{\cup} C$

c. Distributive Laws: $A \overset{\cup}{\cap} (B \overset{\cap}{\cup} C) = (A \overset{\cup}{\cap} B) \overset{\cap}{\cup} (A \overset{\cup}{\cap} C)$

d. DeMorgan's laws: $(A \overset{\cap}{\cup} B)^c = A^c \overset{\cup}{\cap} B^c$

- $A_1, A_2, \cdots \subset S$

$\bigcup_{i=1}^{\infty} A_i \triangleq \{x \in S \mid x \in A_i \text{ for some } i\}$

$\bigcap_{i=1}^{\infty} A_i \triangleq \{x \in S \mid x \in A_i \text{ for all } i\}$

(e.g.) $S = (0, 1]$, $A_i = [1/i, 1]$

$\bigcup_{i=1}^{\infty} A_i = (0, 1] = S$, $\bigcap_{i=1}^{\infty} A_i = \{1\}$.

- $A_a \subset S$

$\bigcup_{a \in \Gamma} A_a \triangleq \{x \in S \mid x \in A_a \text{ for some } a\}$

$\bigcap_{a \in \Gamma} A_a \triangleq \{x \in S \mid x \in A_a \text{ for all } a\}$

(e.g.) $\Gamma = \{\text{all (+)ve real numbers}\}$, $A_a = (0, a]$

$\bigcup_{a \in \Gamma} A_a = (0, \infty)$

Defn: A and B are disjoint if $A \cap B = \phi$.

$A_1, A_2, \cdots$ are pairwise disjoint (or mutually exclusive) if $A_i \cap A_j = \phi \quad \forall i \neq j$.

Defn: $A_1, A_2, \cdots$ forms a partition of S if A_i's are pairwise disjoint and $\bigcup_{i=1}^{\infty} A_i = S$.

(e.g.) $S = [0, \infty)$, $A_i = [i, i+1)$, $i = 0, 1, 2, \cdots$

1.2 Basics of prob. theory

(1) Axiomatic foundations

(For) $\forall A \subset S$, $\quad A \xmapsto{P} P(A) \in [0,1]$, a set function

Defn: $\mathcal{B}$ = a collection of events is called a σ-algebra if

a. $\phi \in \mathcal{B}$

b. $A \in \mathcal{B}$ implies $A^c \in \mathcal{B}$

c. $A_1, A_2, \cdots \in \mathcal{B}$ implies $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

(e.g.) $\mathcal{B} = \{\phi, S\}$ is a σ-algebra.

(e.g.) $\mathcal{B} = \{$all subsets of $S\}$ is a σ-algebra.

(e.g.) $S = (-\infty, \infty)$, $\mathcal{B} = \{$all sets of the form $[a,b]$, $(a,b]$, $[a,b)$ and $(a,b)\}$ is a σ-algebra.

Def^n: Given $(S, \mathcal{B})$, a prob. function is a funtion P s.t. such that

a. $P(A) \geq 0$, $\forall A \in \mathcal{B}$

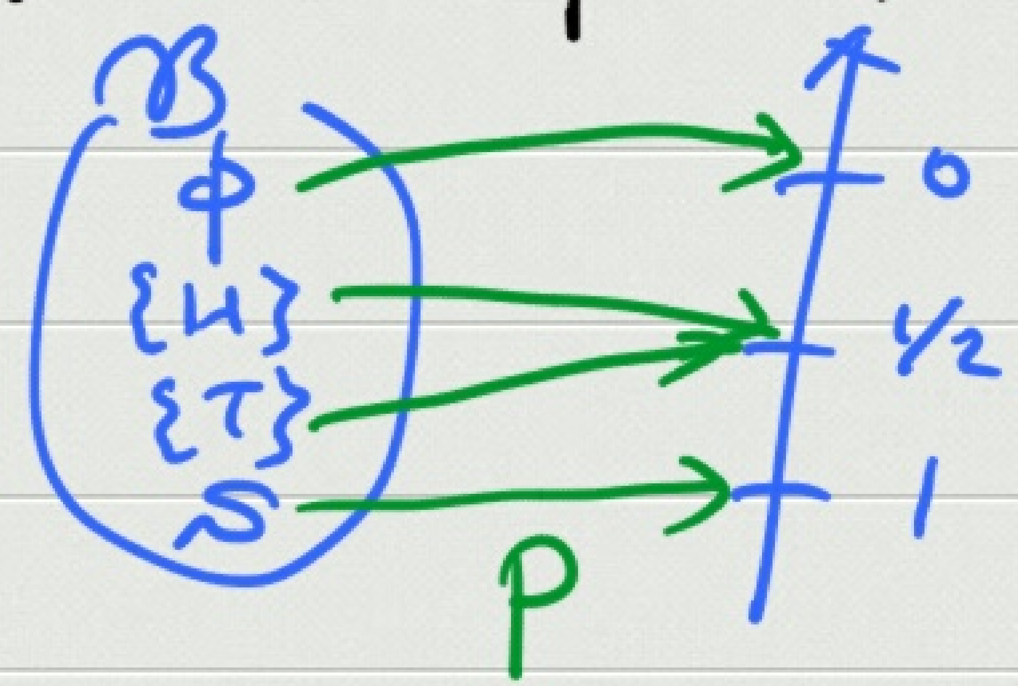
b. $P(S) = 1$

c. $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint $\Longrightarrow P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$.

(e.g.) $S = \{H, T\}$, $\mathcal{B} = \{\phi, \{H\}, \{T\}, S\}$

To find the prob. function P s.t. $P(\{H\}) = P(\{T\})$.

sol.



satisfies "a, b, c" in the defn of P.

Thm : $S = \{ s_1, \cdots, s_n \}$, a finite set [countable]

$\mathcal{B}$ = any σ-algebra of subsets of S

$p_1, p_2, \cdots, p_n \geq 0$ s.t. $\sum_{i=1}^{n} p_i = 1$.

Then, for any $A \in \mathcal{B}$, $\sum_{i : s_i \in A} p_i$ defines $P(A)$.

pf: a. For any $A \in \mathcal{B}$, $P(A) = \sum_{i : s_i \in A} p_i \geq 0$.

b. $P(S) = \sum_{i : s_i \in S} p_i = \sum_{i=1}^{n} p_i = 1$.

c. $A_1, A_2, \cdots, A_k$: pairwise disjoint events, $\in \mathcal{B}$

$$P\left(\bigcup_{i=1}^{k} A_i \right) = \sum_{j : s_j \in \bigcup_{i=1}^{k} A_i} p_j = \sum_{i=1}^{k} \sum_{j : s_j \in A_i} p_j = \sum_{i=1}^{k} P(A_i).$$ //

Thm: $(S, \mathcal{B}, P)$. Then

a. $P(\phi) = 0$.

b. $P(A) \leq 1$.

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

pf: a. $S = S \cup \phi$, $P(S) = P(S) + P(\phi)$ $\therefore P(\phi) = 0$

c. $S = A \cup A^c$, $1 = P(S) = P(A) + P(A^c)$ $\therefore P(A^c) = 1 - P(A)$.

b. By c, $0 \leq P(A^c) = 1 - P(A)$ $\therefore P(A) \leq 1$. //

$\underline{\text{Th}^m}$: $(S, \mathcal{B}, P)$, $A, B \in \mathcal{B}$.

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

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

c. $A \subset B$ implies $P(A) \le P(B)$

pf) (HW)

"generalization" $\rightarrow P\left(\bigcap_{i=1}^{n} A_i\right) \ge \sum_{i=1}^{n} P(A_i) - (n-1)$

- Bonferroni's ineq. : $P(A \cap B) \ge P(A) + P(B) - 1$

(e.g.) $P(A) = P(B) = 0.95$

then $P(A \cap B) \ge 0.90$ //

$\underline{\text{Th}^{\text{m}}}$: $(S, \mathcal{B}, P)$

a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, \cdots$.

b. $P(\bigcup_{i=1}^{\infty} A_i) \le \sum_{i=1}^{\infty} P(A_i)$ for any $A_1, A_2, \cdots \in \mathcal{B}$. (Boole's ineq.)

(pf.) a. $A = A \cap S = A \cap (C_1 \cup C_2 \cup \cdots) = \bigcup_{i=1}^{\infty} (A \cap C_i)$

b. $A_1^* = A_1$, $A_2^* = A_2 \setminus A_1$, $A_3^* = A_3 \setminus (A_1 \cup A_2)$, $\cdots$

$A_i^* = A_i \setminus \bigcup_{j=1}^{i-1} A_j$ $\cdots$

Then $\bigcup_i A_i^* = \bigcup_i A_i$, so that $P(\bigcup_i A_i) = \sum_i P(A_i^*)$.

By construction, $P(A_i^*) \le P(A_i)$. //

1.3 Conditional prob. and independence.

Defn: $(S, \mathcal{B}, P)$

$A, B \in \mathcal{B}$ and $P(B) > 0$.

then, $P(A|B) \triangleq P(A \cap B) / P(B)$.

(the conditional prob. of A given B)

" Read Example 1.3.4 in p. 21. "

- $P(A \cap B) = P(A|B)\,P(B) = P(B|A)\,P(A)$.

Thm: (Bayes' rule) $(S, \mathcal{B}, P)$.

$A_1, A_2, \cdots$: a partition of S

$B \in \mathcal{B}$.

Then, for each $i = 1, 2, \cdots$

$$P(A_i \mid B) = \frac{P(B \mid A_i)\, P(A_i)}{\sum_{j=1}^{\infty} P(B \mid A_j)\, P(A_j)}.$$

disjoint union

(pf.) $B = \bigcup_{j=1}^{\infty} (B \cap A_j)$, $P(B) = \sum_{j=1}^{\infty} P(B \cap A_j) = \sum_{j=1}^{\infty} P(B \mid A_j)\, P(A_j)$

$P(A_i \cap B) = P(B \mid A_i)\, P(A_i)$ //

(e.g.) (Example 1.3.6)

$P(\text{dot sent}) = 3/7$, $P(\text{dash sent}) = 4/7$.

$P(\text{dash received} \mid \text{dot sent}) = P(\text{dot received} \mid \text{dash sent}) = 1/8$

$P(\text{dash received} \mid \text{dash sent}) = P(\text{dot received} \mid \text{dot sent}) = 7/8$

$P(\text{dot sent} \mid \text{dot received}) = \,?$

<u>sol</u>. $P(\text{dot received}) = \underbrace{P(\text{dot received} \mid \text{dot sent})\, P(\text{dot sent})}_{= P(\text{dot sent} \,\cap\, \text{dot received})}$

$+ P(\text{dot received} \mid \text{dash sent})\, P(\text{dash sent})$

$= \frac{7}{8} \times \frac{3}{7} + \frac{1}{8} \times \frac{4}{7} = \frac{25}{56}$

$P(\text{dot sent} \cap \text{dot received}) = \frac{21}{56}$

$\therefore$ The answer $= 21/25$. //

Defn : $(S, \mathcal{B}, P)$, $A, B \in \mathcal{B}$ are independent if $P(A \cap B) = P(A)P(B)$.

Thm : A, B : indep. then

a. A and B^c are indep.

b. A^c and B^c are indep.

(Pf) a. $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)\left(1 - P(B)\right)$.

b. $P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A)P(B)$

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

Defn: $A_1, A_2, \cdots$ are mutually indep., if for any subcollection $A_{i_1}, \cdots, A_{i_k}$ we have

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

(e.g.) (tossing two dice) $S = \{(1,1), \cdots, (6,6)\}$

$$\left.\begin{array}{l} A = \{\text{doubles appear}\} \\ B = \{\text{the sum is bet}^n\ 7 \text{ and } 10\} \\ C = \{\text{the sum is } 2 \text{ or } 7 \text{ or } 8\} \end{array}\right\} \text{are mutually indep.?}$$

$P(A) = 1/6$, $P(B) = 1/2$, $P(C) = 1/3$.

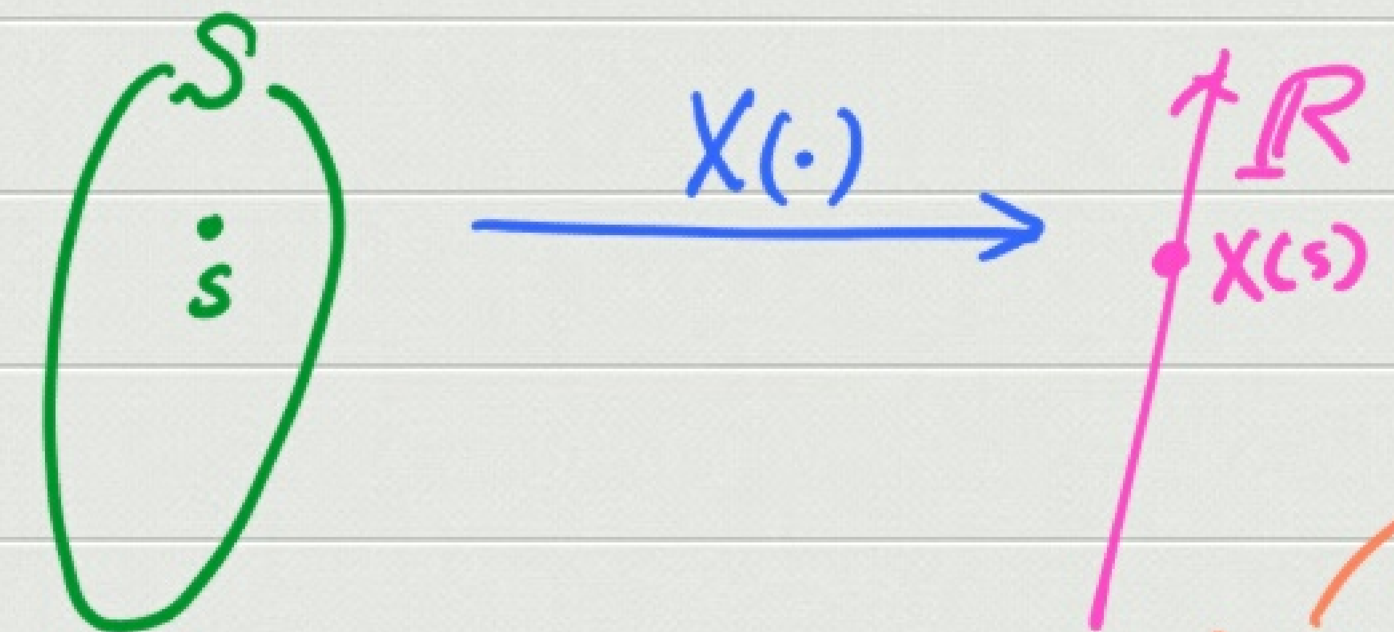
$P(A \cap B \cap C) = P(\text{the sum is } 8 \text{ and doubles}) = P(\{4,4\}) = 1/36$

$= \frac{1}{6} \times \frac{1}{2} \times \frac{1}{3} = P(A)P(B)P(C)$.

However, $P(B \cap C) = P(\text{the sum is } 7 \text{ or } 8) = \frac{11}{36}$ & $P(B)P(C) = \frac{1}{6}$ //

1.4 Random variables

Defn: A random variable is a function from S to $\mathbb{R}$.



$(S, \mathcal{B}, P)$

range of $X(\cdot) \subset \mathbb{R}$: a new sample space

"induced" prob. function

$(X(S), \mathcal{B}(X(S)), P_X)$

$A \in \mathcal{B}(X(S))$, $P_X(A) \triangleq P(\underbrace{\{s \mid X(s) \in A\}}_{\in \mathcal{B}(?)})$

$\rightarrow$ X should be a measurable func. w.r.t. $\mathcal{B}$.

1.5 Distribution functions

Defn : The cumulative dist. func. (cdf) of a r.v. X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X\big(\underbrace{X \in (-\infty, x]}_{X \le x}\big), \quad x \in \mathbb{R}.$$

Thm : $F(x)$ is a cdf iff "if and only if"

a. $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

b. $F(x)$ is a nondecreasing func. of x.

c. $F(x)$ is right-conti.

Defn: A r.v. X is continuous [discrete] if $F_X(x)$ is a continuous [step] func. of x.

Defn: X and Y are identically distributed if for every $A \in \mathcal{B}(\mathbb{R})$,
$$P(\underbrace{X \in A}_{= \{s \mid X(s) \in A\}}) = P(\underbrace{Y \in A}_{= \{s \mid Y(s) \in A\}})!$$

(e.g.) To toss a fair coin.
$X =$ # of heads, $Y =$ # of tails.

Thm: TFAE: "the followings are equivalent."

a. X and Y are identically distributed.

b. $F_X(x) \overset{\forall x}{\equiv} F_Y(x)$

1.6 Density and mass functions

Defn The prob. mass fun. (pmf) of a dicrete r.v. X is given by

$$f_X(x) = P(X = x) \quad \text{for all } x.$$

Defn The prob. density fun. (pdf) of a continuous r.v. X is the function $f_X(\cdot)$ sat'g satisfying

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt \quad \text{for all } x.$$

Thm A fun. $f_X(\cdot)$ is a pdf [pmf] of a r.v. X iff

a. $f_X(x) \geq 0 \quad \forall x$

b. $\sum_x f_X(x) = 1$ or $\int_{-\infty}^{+\infty} f_X(x)\, dx = 1$.