

PROBABILITY THEORY

MATH 154

Unit 4: Probability Measures

4.1. (Ω, \mathcal{A}, P) is called a **probability space**, if \mathcal{A} is a σ algebra on Ω and $P : \mathcal{A} \rightarrow [0, 1]$ is (i) **non-negative** $P[A] \geq 0$, (ii) **normalized** $P[\Omega] = 1$, and (iii) **σ -additive**: $A_n \in \mathcal{A}$ disjoint $\Rightarrow P[\bigcup_n A_n] = \sum_n P[A_n]$.

4.2. A function λ from a π -system \mathcal{I} to $[0, 1]$ is called **monotone** if $\lambda(A \cap B) \leq \lambda(A)$. If $\lambda(\emptyset) = 0$ and $\lambda(\Omega) = 1$, we also call it a **probability measure on the π -system**. If \mathcal{I} is a π -system, let $\sigma(\mathcal{I})$ denote the smallest σ algebra containing \mathcal{I} .

4.3. $A, B \in \mathcal{A}$ are called **independent**, if $P[A \cap B] = P[A] \cdot P[B]$. With the **conditional probability** $P[A|B] = \frac{P[A \cap B]}{P[B]}$ **independence** means $P[A|B] = P[A]$.

4.4. Example: A fair dice produces a random number k in $\{1, 2, 3, 4, 5, 6\}$. Now toss a fair coin k times. You don't see the coins but are told that all coins are head. What is the probability that the dice showed 5? It can be solved using Bayes rule.

4.5. You look the following $\Pi\Lambda\Sigma$ result in the homework.

Theorem 1 (Sorority). *The smallest λ -system \mathcal{A} containing a π -system \mathcal{I} is $\sigma(\mathcal{I})$.*

4.6. Let $\mathcal{P} = 2^\Omega$ denote the set of all subsets of Ω . A map $\mu : \mathcal{P} \rightarrow [0, 1]$ is called an **outer measure** if $\mu(\emptyset) = 0$, $A, B \in \mathcal{A}$ with $A \subset B \Rightarrow \mu(A) \leq \mu(B)$ and **σ -sub-additivity** holds $A_n \in \mathcal{P} \Rightarrow \mu(\bigcup_n A_n) \leq \sum_n \mu(A_n)$ for all sets. Given an outer measure μ , a set A in \mathcal{P} is a **μ -set**, if $\mu(A \cap G) + \mu(A^c \cap G) = \mu(G)$ for all $G \in \mathcal{P}$. Given an outer measure μ , let \mathcal{A}_μ be the set of all μ -sets.

Theorem 2. *An outer measure μ defines a σ -algebra $\mathcal{A}_\mu \subset \mathcal{P}$ on which μ is σ -additive.*

Proof. (i) \mathcal{A}_μ is an algebra. First of all, $\Omega \in \mathcal{A}_\mu$. If $B \in \mathcal{A}_\mu$, then $B^c \in \mathcal{A}_\mu$. Given $B, C \in \mathcal{A}_\mu$. Then $A = B \cap C \in \mathcal{A}_\mu$. Since $C \in \mathcal{A}_\mu$, we get $\mu(C \cap A^c \cap G) + \mu(C^c \cap A^c \cap G) = \mu(A^c \cap G)$. This can be rewritten with $C \cap A^c = C \cap B^c$ and $C^c \cap A^c = C^c$ as $\mu(A^c \cap G) = \mu(C \cap B^c \cap G) + \mu(C^c \cap G)$. Because B is a μ -set, we get using $B \cap C = A$. $\mu(A \cap G) + \mu(B^c \cap C \cap G) = \mu(C \cap G)$. Since C is a μ -set, we have $\mu(C \cap G) + \mu(C^c \cap G) = \mu(G)$. Adding up these three equations shows that $B \cap C$ is a μ -set. If B and C are disjoint in \mathcal{A}_μ we deduce from the fact that B is a μ -set, $\mu(B \cap (B \cup C) \cap G) + \mu(B^c \cap (B \cup C) \cap G) = \mu((B \cup C) \cap G)$. This can be rewritten as $\mu(B \cap G) + \mu(C \cap G) = \mu((B \cup C) \cap G)$. By induction, $\sum_{k=1}^n \mu(A_k \cap G) = \mu((\bigcup_{k=1}^n A_k) \cap G)$ holds for all $\{A_k\}_{k=1}^n$ and all $G \in \mathcal{A}$.

(ii) Given a disjoint sequence $A_n \in \mathcal{A}_\mu$. We have to show that $A = \bigcup_n A_n \in \mathcal{A}_\mu$ and

$\mu(A) = \sum_n \mu(A_n)$. We know that $B_n = \bigcup_{k=1}^n A_k$ is in \mathcal{A}_μ . Because $\mu(G) = \mu(B_n \cap G) + \mu(B_n^c \cap G) \geq \mu(B_n \cap G) + \mu(A^c \cap G) = \sum_{k=1}^n \mu(A_k \cap G) + \mu(A^c \cap G) \geq \mu(A \cap G) + \mu(A^c \cap G)$ and $\mu(G) \leq \mu(A \cap G) + \mu(A^c \cap G)$, we have $\mu(G) = \mu(A \cap G) + \mu(A^c \cap G)$ showing $A \in \mathcal{A}_\mu$. Finally we show that μ is σ -additive on \mathcal{A}_μ : for any $n \geq 1$ we have $\sum_{k=1}^n \mu(A_k) \leq \mu(\bigcup_{k=1}^n A_k) \leq \sum_{k=1}^\infty \mu(A_k)$. Take the limit $n \rightarrow \infty$. \square

Theorem 3 (Carathéodory Extension Theorem). *A probability measure λ on a π -system \mathcal{I} extends uniquely to a probability measure μ on $\sigma(\mathcal{I})$.*

Proof. The function $\mu(A) = \inf\{\sum_{n \in \mathbb{N}} \lambda(A_n) \mid A_n \in \mathcal{I} \text{ with } A \subset \bigcup_n A_n\}$ defines an outer measure on \mathcal{P} . We will show that it defines a probability measure on \mathcal{A}_μ . This then extends to the smallest σ -algebra \mathcal{A} containing \mathcal{I} .

(i) $\mu(\emptyset) = 0$ and $\mu(A) \leq \mu(B)$ for $A \subset B$ follow from the outer measure properties of λ . To see σ -sub-additivity of μ , take a sequence $A_n \in \mathcal{P}$ and fix $\epsilon > 0$. For all $n \in \mathbb{N}$, one can find a sequence $\{B_{n,k}\}_{k \in \mathbb{N}}$ in \mathcal{I} such that $A_n \subset \bigcup_{k \in \mathbb{N}} B_{n,k}$ and $\sum_{k \in \mathbb{N}} \lambda(B_{n,k}) \leq \mu(A_n) + \epsilon 2^{-n}$. Define $A = \bigcup_{n \in \mathbb{N}} A_n \subset \bigcup_{n,k \in \mathbb{N}} B_{n,k}$, so that $\mu(A) \leq \sum_{n,k} \lambda(B_{n,k}) \leq \sum_n \mu(A_n) + \epsilon$. Since ϵ was arbitrary, the σ -subadditivity of μ is proven.

(ii) $\lambda = \mu$ on \mathcal{I} . Given $A \in \mathcal{I}$. Clearly $\lambda(A) \leq \mu(A)$. Suppose that $A \subset \bigcup_n A_n$, with $A_n \in \mathcal{R}$. Define a sequence $\{B_n\}_{n \in \mathbb{N}}$ of disjoint sets in \mathcal{R} inductively. That is $B_1 = A_1$, $B_n = A_n \cap (\bigcup_{k < n} A_k)^c$ such that $B_n \subset A_n$ and $\bigcup_n B_n = \bigcup_n A_n \supset A$. From the σ -additivity of μ on \mathcal{I} follows $\mu(A) \leq \mu(\bigcup_n A_n) = \mu(\bigcup_n B_n) = \sum_n \mu(B_n)$. Because the choice of A_n is arbitrary, this gives also $\mu(A) \leq \lambda(A)$.

(iii) $\mathcal{I} \subset \mathcal{A}_\mu$. Given $A \in \mathcal{I}$ and $G \in \mathcal{P}$. There exists a sequence $\{B_n\}_{n \in \mathbb{N}}$ in \mathcal{I} such that $G \subset \bigcup_n B_n$ and $\sum_n \mu(B_n) \leq \lambda(G) + \epsilon$. By definition $\sum_n \mu(B_n) = \sum_n \mu(A \cap B_n) + \sum_n \mu(A^c \cap B_n) \geq \mu(A \cap G) + \mu(A^c \cap G)$ because $A \cap G \subset \bigcup_n A \cap B_n$ and $A^c \cap G \subset \bigcup_n A^c \cap B_n$. Since ϵ is arbitrary, we get $\mu(G) \geq \mu(A \cap G) + \mu(A^c \cap G)$. On the other hand, since μ is sub-additive, we have also $\mu(G) \leq \mu(A \cap G) + \mu(A^c \cap G)$ and A is a μ -set.

(iv) By (i), μ is an outer measure on \mathcal{P} . Since by step (iii), $\mathcal{I} \subset \mathcal{A}_\mu$, we know that $\mathcal{A} = \sigma(\mathcal{I}) \subset \mathcal{A}_\mu$, so that μ on \mathcal{A} is defined by restricting μ from \mathcal{A}_μ to $\mathcal{A} = \sigma(\mathcal{I})$.

(v) Uniqueness. If two probability measures μ, ν on $\sigma(\mathcal{I})$ satisfy $\mu(A) = \nu(A)$ for $A \in \mathcal{I}$, then $\mu = \nu$: the set $\mathcal{D} = \{A \in \mathcal{A} \mid \mu(A) = \nu(A)\}$ is a λ system: $\Omega \in \mathcal{D}$. Given $A, B \in \mathcal{D}$, $A \subset B$. Then $\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A)$ so that $B \setminus A \in \mathcal{D}$. Given $A_n \in \mathcal{D}$ with $A_n \nearrow A$, then the σ additivity gives $\mu(A) = \limsup_n \mu(A_n) = \limsup_n \nu(A_n) = \nu(A)$, so that $A \in \mathcal{D}$. Since \mathcal{D} is a λ system containing the π -system \mathcal{I} , we know that (ask sorority) that $\sigma(\mathcal{I}) \subset \mathcal{D}$ which means that $\mu = \nu$ on $\sigma(\mathcal{I})$. \square

4.7. Examples:

1) $\mu([a, b]) = b - a$ on the π -system $\mathcal{I} = \{[a, b] \subset [0, 1]\}$ extends to $\sigma(\mathcal{I})$.

2) If $(\Omega_1, \mathcal{A}_1, P_1), (\Omega_2, \mathcal{A}_2, P_2)$ are probability spaces then $\lambda(A \times B) = P_1[A]P_2[B]$ extends from the π system \mathcal{I} of “rectangles” $A \times B$ to the product. Product spaces are a source for independence as $A \times \Omega_2$ and $\Omega_1 \times B$ are always independent.