**PROBABILITY THEORY**

MATH 154

# Unit 4: Probability Measures

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

**4.2.** A function $\lambda$ from a $\pi$-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 $\pi$-system**. If $\mathcal{I}$ is a $\pi$-system, let $\sigma(\mathcal{I})$ denote the smallest $\sigma$ algebra containing $\mathcal{I}$.

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

**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 prove the following $\Pi\Lambda\Sigma$ result in the homework. [1]

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

**4.6.** Let $\mathcal{P} = 2^\Omega$ denote the set of all subsets of $\Omega$. A map $\mu : \mathcal{P} \to [0,1]$ 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 **$\sigma$-subadditivity** 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 $\mu$, a set $A$ in $\mathcal{P}$ is a **$\mu$-set**, if $\mu(A \cap G) + \mu(A^c \cap G) = \mu(G)$ for all $G \in \mathcal{P}$. Given an outer measure $\mu$, let $\boxed{\mathcal{A}_\mu}$ be the set of all $\mu$-sets.

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

*Proof.* (i) $\mathcal{A}_\mu$ 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 $\mu$-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 $\mu$-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 $\mu$-set. If $B$ and $C$ are disjoint in $\mathcal{A}_\mu$ we deduce from the fact that $B$ is a $\mu$-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, $\sum_{k=1}^n \mu(A_k \cap G) = \mu((\bigcup_{k=1}^n A_k) \cap G)$

[1] Never mind that the sorority Pi Lambda Sigma at BU merged with Theta Phi Alpha in 1952.

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 $\mathcal{A}_\mu$. 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 $\mu$ is $\sigma$-additive on $\mathcal{A}_\mu$: 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 $n \to \infty$. □

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

*Proof.* The function $\mu(A) = \inf\{\sum_{n \in \mathbb{N}} \lambda(A_n) \mid A_n \in \mathcal{I}$ with $A \subset \bigcup_n A_n \}$ defines an outer measure on $\mathcal{P}$. We will show that it defines a probability measure on $\mathcal{A}_\mu$. This then extends to the smallest $\sigma$-algebra $\mathcal{A}$ containing $\mathcal{I}$.
(i) $\boxed{\mu(\emptyset) = 0 \text{ and } \mu(A) \leq \mu(B) \text{ for } A \subset B}$ follow from the outer measure properties of $\lambda$. To see $\boxed{\sigma\text{-sub-additivity of } \mu}$, 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 $\epsilon$ was arbitrary, the $\sigma$-subadditivity of $\mu$ is proven.
(ii) $\boxed{\lambda = \mu \text{ on } \mathcal{I}}$ 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 $\sigma$-additivity of $\mu$ 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) $\boxed{\mathcal{I} \subset \mathcal{A}_\mu}$. 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 $\epsilon$ is arbitrary, we get $\mu(G) \geq \mu(A \cap G) + \lambda(A^c \cap G)$. On the other hand, since $\mu$ is sub-additive, we have also $\mu(G) \leq \mu(A \cap G) + \mu(A^c \cap G)$ and $A$ is a $\mu$-set.
(iv) By (i), $\mu$ 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 $\mu$ on $\mathcal{A}$ is defined by restricting $\mu$ from $\mathcal{A}_\mu$ to $\mathcal{A} = \sigma(\mathcal{I})$.
(v) $\boxed{\text{Uniqueness.}}$ If two probability measures $\mu, \nu$ on $\sigma(\mathcal{I})$ satisfy $\mu(A) = \nu(A)$ for $A \in \mathcal{I}$, then $\boxed{\mu = \nu}$: the set $\mathcal{D} = \{A \in \mathcal{A} \mid \mu(A) = \nu(A) \}$ is a $\lambda$ 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 $\sigma$ 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 $\lambda$ system containing the $\pi$-system $\mathcal{I}$, we know that (ask sorority) that $\sigma(\mathcal{I}) =\subset \mathcal{D}$ which means that $\mu = \nu$ on $\sigma(\mathcal{I})$. □

**4.7.** Examples:
**1)** $\mu([a,b)) = b - a$ on the $\pi$-system $\mathcal{I} = \{[a,b) \subset [0,1]\}$ extends to $\sigma(I)$.
**2)** If $(\Omega_1, \mathcal{A}_1, \mathrm{P}_1), (\Omega_2, \mathcal{A}_2, \mathrm{P}_2)$ are probability spaces then $\lambda(A \times B) = \mathrm{P}_1[A]\mathrm{P}_2[B]$ extends from the $\pi$ 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.