# **MATH356** - Probability

Based on lectures from Fall 2024 by Prof. Asoud Asgharian.
Notes by Louis Meunier

## Contents

# §1 Prerequisites

↪**Definition 1.1** (limsup, liminf of sets): Let $\{A_n\}_{n\geq 1}$ be a sequence of sets. We define

$$\overline{\lim}_{n\to\infty} = \limsup_{n\to\infty} A_n := \{x : x \in A_n \text{ for infinitely many } n\} = \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$$

and

$$\underline{\lim}_{n\to\infty} = \liminf_{n\to\infty} A_n := \{x : x \in A_n \text{ for all but finitely many } n\} = \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k.$$

If $\liminf A_n = \limsup A_n$, we say $A_n$ *converges* to this value and write $\lim_{n\to\infty} A_n = \liminf A_n = \limsup A_n$

↪**Proposition 1.1**: $\liminf A_n \subseteq \limsup A_n$

⊛ **Example 1.1**: Let $A_n = \{n\}$. Then $\liminf A_n = \limsup A_n = \varnothing = \lim A_n$. Let $A_n = \left\{(-1)^n\right\}$. Then $\liminf A_n = \varnothing, \limsup A_n = \{-1, 1\}$.

↪**Definition 1.2** (sigma-field): A non-empty class of subsets of a set $\Omega$ which is closed under countable unions and complement, and contains $\varnothing$ is called a *$\sigma$-field* or *$\sigma$-algebra*.

↪**Definition 1.3** (Borel sigma-algebra): The $\sigma$-algebra generated by the class of all bounded, semi-closed intervals is called the *Borel algebra* of subsets of $\mathbb{R}$, denoted $\mathfrak{B}, \mathfrak{B}(\mathbb{R})$.

↪**Theorem 1.1**: Every countable set is Borel.

Proof. $\{x\} = \bigcap_{n=1}^{\infty}\left(x - \frac{1}{n}, x\right]$ for any $x \in \mathbb{R}$, so $A := \{x_n : n \in \mathbb{N}\} = \bigcup_{n=1}^{\infty}\{x_n\} \in \mathfrak{B}$. ■

↪**Theorem 1.2**: $\mathfrak{B} = \sigma(\{\text{open sets in} \mathbb{R}\})$.

# §2 Probability

## §2.1 Sample Space

↪**Definition 2.1** (Random/statistical experiment): A *random/statistical experiment* (stat. exp.) is one in which

1. all outcomes are known in advance;
2. any performance of the experiment results in an outcome that is not known in advance;
3. the experiment can be repeated under identical conditions.

↪**Definition 2.2** (Sample space): The *sample space* of a stat. exp. is the pair $(\Omega, \mathcal{F})$ where $\Omega$ the set of all possible outcomes and $\mathcal{F}$ a $\sigma$-algebra of subsets of $\Omega$.

We call points $\omega \in \Omega$ *sample points*, $A \in \mathcal{F}$ *events*. If $\Omega$ countable, we call $(\Omega, \mathcal{F})$ a *discrete sample space*.

↪**Definition 2.3**: Let $(\Omega, \mathcal{F})$ be a sample space. A set function $P$ is called a *probability measure* or simply *probability* if

1. $P(A) \geq 0$ for all $A \in \mathcal{F}$
2. $P(\Omega) = 1$
3. For $\{A_n\} \subseteq \mathcal{F}$, disjoint, then $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$.

↪**Theorem 2.1**: $P$ monotone ($A \subseteq B \Rightarrow P(A) \leq P(B)$) and subtractive $P(B \setminus A) = P(B) - P(A)$.

↪**Theorem 2.2**: For all $A, B \in \mathcal{F}$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

↪**Corollary 2.1**: $P$ subadditive; for any $A, B \in \mathcal{F}$, $P(A \cup B) \leq P(A) + P(B)$.

↪**Corollary 2.2**: $P(A^c) = 1 - P(A)$.

↪**Theorem 2.3** (Principle of Inclusion/Exclusion): Let $A_1, \ldots, A_n \in \mathcal{F}$. Then

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k) - \sum_{k_1 < k_2} P\left(A_{k_1} \cap A_{k_2}\right) + \sum_{k_1 < k_2 < k_3} P\left(A_{k_1} \cap A_{k_2} \cap A_{k_3}\right) + \ldots + (-1)^n P\left(\bigcap_{k=1}^{n} A_k\right).$$

↪**Theorem 2.4** (Bonferroni's Inequality): For $A_1, \ldots, A_n$,

$$\sum_{i=1}^{n} P(A_i) - \sum_{i<j} P\left(A_i \cap A_j\right) \leq P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P(A_i).$$

↪**Theorem 2.5** (Boole's Inequality): $P(A \cap B) \geq 1 - P(A^c) - P(B^c)$.

↪**Corollary 2.3**: For $\{A_n\} \subseteq \mathcal{F}$,

$$P(\cap_{n=1}^{\infty} A_n) \geq 1 - \sum_{n=1}^{\infty} P(A_n^c)$$

↪**Theorem 2.6** (Implication Rule): If $A, B, C \in \mathcal{F}$ and $A$ and $B$ imply $C$ (i.e. $A \cap B \subseteq C$) then $P(C^c) \leq P(A^c) + P(B^c)$.

↪**Theorem 2.7** (Continuity): Let $\{A_n\} \subseteq \mathcal{F}$ non-decreasing i.e. $A_n \supseteq A_{n-1} \; \forall \, n$, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

Let $\{A_n\}$ non-increasing, then

$$\lim_{n\to\infty} P(A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right).$$

Finally, more generally, for $\{A_n\}$ such that $\lim_{n\to\infty} A_n = A$ exists, then

$$P(A) = \lim_{n\to\infty} P(A_n).$$

# §3 Combinatorics - Finite $\sigma$-fields

## §3.1 Counting

We consider now $\Omega = \{\omega_1, ..., \omega_n\}$ finite sample spaces, and consider $\mathcal{F} = 2^\Omega$.

↪**Definition 3.1** (Permutation): An ordered arrangement of $r$ distinct objects is called a permutation. The number of ways to order $n$ distinct objects taken $r$ at a time is

$$P_r^n = \frac{n!}{(n-r)!}.$$

↪**Definition 3.2** (Combination): The number of combinations of $n$ objects taken $r$ at a time is the number of subsets of size $r$ that can be formed from $n$ objects,

$$C_r^n = \binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}.$$

↪**Theorem 3.1**: The number of unordered arrangements of $r$ objects out of a total of $n$ objects when sampling with replacement is

$$\binom{n+r-1}{r}.$$

## §3.2 Conditional Probability

↪**Theorem 3.2**: Let $A, H \in \mathcal{F}$. We denote by $P(A \mid H)$ the probability of $A$ given $H$ has occured. We have, in particular,

$$P(A \mid H) = \frac{P(A \cap H)}{P(H)},$$

if $P(H) \neq 0$.

↪**Definition 3.3**: We say two events $A, B$ are independent if $P(A \mid B) = P(A)$, or equivalently $P(A \wedge B) = P(A)P(B)$.

↪**Proposition 3.1** (Multiplication Rule):

$$P\left(\bigcap_{j=1}^n A_j\right) = \prod_{i=1}^n P\left(A_i \mid \cap_{j=0}^{i-1} A_j\right),$$

taking $A_0 := \Omega$ by convention.

↪**Proposition 3.2** (Law of Total Probability): Let $\{H_n\} \subseteq \mathcal{F}$ be a partition of $\mathcal{F}$, namely $H_i \cap H_j = \varnothing$ for all $i \neq j$, and $\cup_{j=1}^{\infty} H_j = \Omega$. If $P(H_n) > 0 \,\forall\, n$, then

$$P(B) = \sum_{n=1}^{\infty} P(B \mid H_n) P(H_n) \,\forall\, B \in \mathcal{F}.$$

↪**Theorem 3.3** (Baye's): Let $\{H_n\}$ be a partition of $\Omega$ with all strictly nonzero measure and let $B \in \mathcal{F}$ with nonzero measure. Then

$$P(H_n \mid B) = \frac{P(H_n)P(B \mid H_n)}{\sum_{n=1}^{\infty} P(H_n)P(B \mid H_n)}.$$

↪**Definition 3.4** (Mutual Independence): A family of sets $\mathcal{A}$ is said to be *mutually independent* iff $\forall$ finite sub collections $\left\{A_{i_1}, ..., A_{i_k}\right\}$, the following holds

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

## §4 Random Variables and Probability Distributions

We tacitly fix some sample space $(\Omega, \mathcal{F})$.

↪**Definition 4.1** (Random Variable): A real-valued function $X : \Omega \to \mathbb{R}$ is called a *random variable* or *rv* if

$$X^{-1}(B) \in \mathcal{F}$$

for all $B \in \mathfrak{B}_{\mathbb{R}}$.

↪**Theorem 4.1**: $X$ an rv $\Leftrightarrow$ for all $x \in \mathbb{R}$,

$$\{X \leq x\} \in \mathcal{F}.$$

↪**Theorem 4.2**: If $X$ a rv, then so is $aX + b$ for all $a, b \in \mathbb{R}$.

↪**Theorem 4.3**: Fix an rv $X$ defined on a probability space $(\Omega, \mathcal{F}, P)$. Then, $X$ induces a measure on the sample space $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, denote $Q$ and given by

$$Q(B) := P\left(X^{-1}B\right)$$

for any Borel set $B$.

**Remark 4.1**: If $X$ a random variable, then the sets $\{X = x\}, \{a < x \leq b\}, \{X < x\}$, etc are all events.

↪**Definition 4.2** (Distribution Function): An $\mathbb{R}$-valued function $F$ that is non-decreasing, right-continuous and satisfies

$$F(-\infty) = 0, F(+\infty) = 1$$

is called a *distribution function* or *df*.

↪**Theorem 4.4**: $\{x \mid F \text{ discontinuous}\}$ is at most countable.

↪**Definition 4.3**: Given a random variable $X$ and a probability space $(\Omega, \mathcal{F}, P)$, we define the df of $X$ as

$$F(x) = P(X \leq x).$$

**Remark 4.2**: It is not obvious a priori that this is indeed a df.

↪**Theorem 4.5**: If $Q$ a probability on $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$, then there exists a df $F$ where

$$F(x) = Q(-\infty, x],$$

and conversely, given a df $F$, there exists a unique probability on $(\mathbb{R}, \mathfrak{B}_{\mathbb{R}})$.

## §4.1 Discrete and Continuous Random Variables

↪**Definition 4.4**: $X$ called "discrete" if $\exists$ countable set $\mathrm{E} \subset \mathbb{R}$ such that $P(X \in \mathrm{E}) = 1$.

↪**Proposition 4.1**: Suppose $\mathrm{E} = \{x_n\}_{n=1}^{\infty}$ and put $p_n := P(X = x_n)$. Then,

$$\sum_{n=1}^{\infty} p_n = 1,$$

where $\{p_n\}$ defines a non-negative sequence.

↪**Definition 4.5** (PMF): Such a sequence $\{p_n\}$ satisying $0 \leq p_n = P(X = x_n)$ for a sequence $\{x_n\}$ and $\sum p_n = 1$ is called a *probability mass function* (pmf) of $X$. Then,

$$F_X(x) = P_X((-\infty, x]) = \sum_{n : x_n \leq x} p_n$$

and

$$X(\omega) = \sum_{n=1}^{\infty} x_n \mathbb{1}_{\{X = x_n\}}(\omega).$$

↪**Definition 4.6**: $X$ called *continuous* if $F$ induced by $X$ is absolutely continuous, i.e. if there exists a non-negative function $f(t)$ such that

$$F(x) = \int_{-\infty}^{x} f(t) \,\mathrm{d}t$$

for all $x \in \mathbb{R}$. Such a function $f$ is called the *probability density function* (pdf) of $X$.

↪**Theorem 4.6**: Let $X$ continuous with pdf $f$. Then

$$P(B) = \int_B f(t) \,\mathrm{d}t$$

for every $B \in \mathfrak{B}_{\mathbb{R}}$.

↪**Theorem 4.7**: Every nonnegative real function $f$ that is integral over $\mathbb{R}$ and such that $\int_{-\infty}^{\infty} f(x) \,\mathrm{d}x = 1$ is the PDF of some continuous $X$.

## §4.2 Functions of a Random Variable

↪**Theorem 4.8**: Let $X$ be an rv and $g$ a Borel-measurable function on $\mathbb{R}$. Then, $g(X)$ also an rv.

↪**Theorem 4.9**: Let $Y = g(X)$ as above. Then, $P(Y \leq y) = P\big(X \in g^{-1}(-\infty, y]\big)$.

⊛ **Example 4.1**: Let $X$ be an RV with Poisson distribution; we write $X \sim \text{Poisson}(\lambda)$; where

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

for $k \in \mathbb{N} \cup \{0\}$. Let $Y = X^2 + 3$. We say that $X$ has *support* $\{0, 1, 2, \text{dots}\}$ (more generally, where $X$ can take values), and so $Y$ has support on $\{3, 4, 7, \ldots\} =: B$. Then

$$P(Y = y) = P\left(X = \sqrt{y-3}\right) = \frac{e^{-\lambda}\lambda^{\sqrt{y-3}}}{\sqrt{y-3}!},$$

for $y \in B$ and $P(Y = y) = 0$ for $y \notin B$.

↪**Theorem 4.10**: Let $X$ cont. rv with pdf f_X. Let $Y = g(X)$ be differentiable for all $x$ and with either strictly positive or negative derivative. Then, $Y = g(X)$ also a continuous rv with pdf given by

$$h(y) = \begin{cases} f_x\left(g^{-1}(y)\right) \left|\frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y)\right| \text{ for } \alpha < y < \beta \\ 0 \text{ else} \end{cases},$$

where

$$\alpha := \min\{g(-\infty), g(\infty)\}, \beta := \max\{g(-\infty), g(\infty)\}.$$

↪**Theorem 4.11**: Let $X$ continuous rv with cdf $F_X(x)$. Let $Y = F_X(X)$. Then, $Y \sim \text{Unif}\,(0, 1)$.

Proof.

$$\begin{aligned} P(Y \leq y) &= P(F_X(X) \leq y) \\ &= P\left(X \leq F_X^{-1}(y)\right). \end{aligned}$$

■

↪**Theorem 4.12**: Let $X$ continuous rv with pdf $f_X$ and $y = g(x)$

## §5 Moments and Moment Generating Functions

↪**Definition 5.1** (Expected Value): Let $X$ be a discrete (continuous) rv with PMF (PDF) $p_k = P(X = x_k)$ $(f)$. If $\sum |x_k|\, p_k < \infty$ $(\int |x|\, f_X(x)\,\mathrm{d}x < \infty)$ then we say the *expected value* of $X$ exists, and write

$$\mathbb{E}(X) = \sum x_k p_k \left(= \int x \cdot f(x)\,\mathrm{d}x\right).$$

↪**Theorem 5.1**: If $X$ symmetric about $\alpha \in \mathbb{R}$, i.e. $P(X \geq \alpha + x) = P(X \leq \alpha - x)$ for all $x \in \mathbb{R}$ (or in the continuous case, $f(\alpha - x) = f(\alpha + x)$), then $\mathbb{E}(X) = \alpha$.

↪**Theorem 5.2**: Let $g$ Borel-measurable and $Y = g(X)$. Then,

$$\mathbb{E}(Y) = \sum_{j=1}^{\infty} g(x_j) P_X(X = x_j).$$

If $X$ continuous,

$$= \int g(x) f(x) \,\mathrm{d}x.$$

↪**Definition 5.2**: For $\alpha > 0$, we say $\mathbb{E}(|X|^\alpha)$ (if it exists) is the *$\alpha$-th moment* of $X$.

⊛ **Example 5.1**: Let $X$ such that $P(X = k) = \frac{1}{N}$, $k = 1, ..., N$, namely $X \sim \mathrm{Unif}_{\{1,...,N\}}$. Then

$$\mathbb{E}(X) = \sum_{k=1}^{N} \frac{k}{N} = \frac{N+1}{2}.$$

↪**Theorem 5.3**: If the $t$th moment of $X$ exists, so does the $s$th moment for $s < t$.

↪**Theorem 5.4**: If $\mathbb{E}\left(|X|^k\right) < \infty$ for some $k > 0$, then

$$n^k P(|X| > n) \to 0$$

as $n \to \infty$.