Measure Theory: A Compact Note

Shuhei Watanabe

March 6, 2026

1 Foundations

Definition 1.1 (σ-algebra)**.** A collection $\mathcal{F}$ of subsets of a set Ω is a **σ-algebra** if:

1. $\Omega \in \mathcal{F}$,
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (closed under complementation),
3. $A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ (closed under countable unions).

Definition 1.2 (Measurable set)**.** A set A is **measurable** (with respect to $\mathcal{F}$) if $A \in \mathcal{F}$.

Definition 1.3 (Borel σ-algebra)**.** The **Borel σ-algebra** $\mathcal{B}(\mathbb{R})$ is the smallest σ-algebra on $\mathbb{R}$ containing all open sets.

Example 1.4 (Vitali set — a non-measurable set)**.** For each $r \in [0,1]$, consider the equivalence class $r + \mathbb{Q} = \{r + q : q \in \mathbb{Q}\}$. Two reals belong to the same class if and only if their difference is rational. By the axiom of choice, pick one representative from each equivalence class intersected with $[0,1]$ to form V. Let $\{r_k\}_{k=1}^{\infty}$ enumerate $\mathbb{Q} \cap [-1,1]$ and set $V_k = V + r_k$. The translates $\{V_k\}$ are pairwise disjoint (if $v_1 + r_j = v_2 + r_k$ then $v_1 - v_2 \in \mathbb{Q}$, so v_1 and v_2 lie in the same class, giving $v_1 = v_2$ and $r_j = r_k$) and satisfy

$$[0,1] \;\subseteq\; \bigcup_{k=1}^{\infty} V_k \;\subseteq\; [-1,2].$$

If V were Lebesgue measurable, translation invariance gives $\mu(V_k) = \mu(V)$ for all k, and σ-additivity yields

$$1 \;\leq\; \mu\left(\bigcup_{k=1}^{\infty} V_k\right) = \sum_{k=1}^{\infty} \mu(V_k) = \sum_{k=1}^{\infty} \mu(V) \;\leq\; 3 \;(\because V_i \cap V_j = \emptyset; i \neq j).$$

This is impossible: if $\mu(V) = 0$ the sum is 0, and if $\mu(V) > 0$ the sum is $+\infty$.

Definition 1.5 (Measurable space)**.** A pair $(\Omega, \mathcal{F})$ where Ω is a set and $\mathcal{F}$ is a σ-algebra on Ω.

Definition 1.6 (Lebesgue measure)**.** The Lebesgue measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is the unique measure satisfying $\mu([a,b]) = b - a$.

Remark 1.7. Lebesgue measure is just one example of a measure. In probability we use probability measures ($\mu(\Omega) = 1$); in counting problems, the counting measure; etc.

Definition 1.8 (Measurable function)**.** $f\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is **measurable** if $f^{-1}(B) \in \mathcal{F}$ for all $B \in \mathcal{B}(\mathbb{R})$.

Definition 1.9 (σ-additivity)**.** A set function $\mu\colon \mathcal{F} \to [0,\infty]$ is **σ-additive** if for any countable collection of pairwise disjoint sets $\{A_i\}_{i=1}^{\infty} \subset \mathcal{F}$:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i).$$

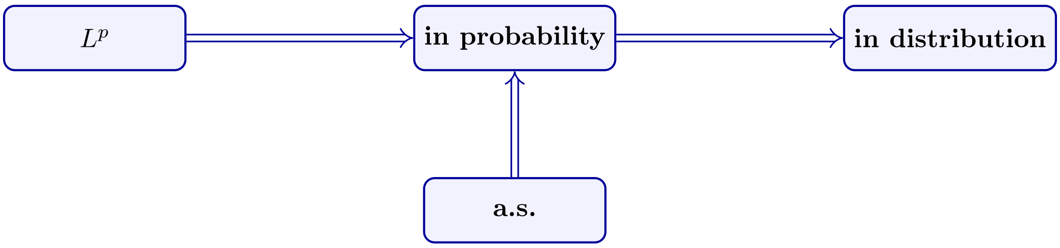
Definition 1.10 (Measure space)**.** A triple $(\Omega, \mathcal{F}, \mu)$ where $(\Omega, \mathcal{F})$ is a measurable space and $\mu\colon \mathcal{F} \to [0,\infty]$ is a σ-additive set function with $\mu(\emptyset) = 0$.

Definition 1.11 (σ-finiteness)**.** A measure μ on $(\Omega, \mathcal{F})$ is **σ-finite** if $\Omega = \bigcup_{i=1}^{\infty} A_i$ with $\mu(A_i) < \infty$ for all i. This condition is required by Fubini's theorem and Radon–Nikodym.

2 Modes of Convergence

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{X_n\}, X$ random variables.

Mode	**Definition**
L^p convergence	$\lim_{n\to\infty} E[\lvert X_n - X\rvert^p] = 0$
Almost sure (a.s.)	$P\left(\lim_{n\to\infty} X_n = X\right) = 1$
In probability	$\forall\,\varepsilon > 0,\ P(\lvert X_n - X\rvert > \varepsilon) \to 0$
In distribution	$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ at all continuity points of F_X



- $L^p \not\Rightarrow$ a.s. and a.s. $\not\Rightarrow L^p$ in general.
- In distribution $\not\Rightarrow$ in probability (unless the limit is a constant).

3 Core Theorems and Inequalities

3.1 Inequalities

Theorem 3.1 (Markov's inequality)**.** *For a non-negative random variable X and $a > 0$: $P(X \geq a) \leq E[X]/a$.*

Usage: Crude tail bound from the mean; foundation for Chebyshev.

Theorem 3.2 (Chebyshev's inequality)**.** *If $E[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$: $P(|X - \mu| \geq k\sigma) \leq 1/k^2$.*

Usage: Distribution-free tail bound; used to prove the weak LLN.

Theorem 3.3 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent with $X_i \in [a_i, b_i]$ a.s. Then for $\bar{X} = \frac{1}{n}\sum X_i$:*

$$P(|\bar{X} - E[\bar{X}]| \geq t) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Usage: Exponential concentration bound; widely used in ML generalization theory (PAC learning).

Theorem 3.4 (Monotone Convergence Theorem (MCT))**.** *If $0 \leq f_1 \leq f_2 \leq \cdots$ are measurable and $f_n \uparrow f$ pointwise, then:*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Proof. We prove the statement step by step:

1. **Upper bound.** Since $f_n \leq f$ for all n, monotonicity of the integral gives $\int f_n \, d\mu \leq \int f \, d\mu$ for every n. The sequence $\{\int f_n \, d\mu\}$ is non-decreasing (because $f_n \leq f_{n+1}$), so its limit exists in $[0, \infty]$ and satisfies

$$\lim_{n\to\infty} \int f_n \, d\mu \leq \int f \, d\mu.$$

2. **Lower bound — setup.** It remains to show $\lim_{n\to\infty} \int f_n \, d\mu \geq \int f \, d\mu$. Recall that by definition of the Lebesgue integral for non-negative measurable functions,

$$\int f \, d\mu = \sup\left\{ \int s \, d\mu : s \text{ simple}, \ 0 \leq s \leq f \right\}.$$

So it suffices to show $\lim \int f_n \, d\mu \geq \int s \, d\mu$ for every simple function s with $0 \leq s \leq f$.

3. **Lower bound — core argument.** Fix such a simple function s and fix $0 < c < 1$. Define $A_n = \{x \in \Omega : f_n(x) \geq c \cdot s(x)\}$. Since $f_n \uparrow f \geq s > c \cdot s$, every $x \in \Omega$ eventually belongs to A_n, so $A_n \uparrow \Omega$. Now:

$$\int f_n \, d\mu \geq \int_{A_n} f_n \, d\mu \geq c \int_{A_n} s \, d\mu.$$

The first inequality holds because $f_n \geq 0$; the second because $f_n \geq c \cdot s$ on A_n.

4. **Taking limits.** Write $s = \sum_{j=1}^{k} a_j \mathbf{1}_{E_j}$ where $E_j \in \mathcal{F}$. Then $\int_{A_n} s \, d\mu = \sum_j a_j \, \mu(A_n \cap E_j)$. Since $A_n \uparrow \Omega$, continuity of measure from below gives $\mu(A_n \cap E_j) \to \mu(E_j)$ for each j. Therefore $\int_{A_n} s \, d\mu \to \int s \, d\mu$, and so

$$\lim_{n\to\infty} \int f_n \, d\mu \geq c \int s \, d\mu.$$

Letting $c \uparrow 1$ gives $\lim \int f_n \, d\mu \geq \int s \, d\mu$. Since this holds for every simple $s \leq f$, taking the supremum yields $\lim \int f_n \, d\mu \geq \int f \, d\mu$.

Combining Steps 1 and 4: $\lim\limits_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu$. □

Lemma 3.5 (Fatou's Lemma). *If $\{f_n\}$ are non-negative measurable functions, then:*

$$\int \liminf_{n\to\infty} f_n \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu.$$

Usage: "The integral of the limit $\leq$ the limit of the integrals." Used when we lack a dominating function for DCT.

Theorem 3.6 (Lebesgue's Dominated Convergence Theorem (DCT)). *If $f_n \to f$ pointwise a.e., and there exists an integrable g with $|f_n| \leq g$ a.e. for all n, then:*

$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

Proof. We prove the statement step by step:

1. **Verifying Fatou's lemma applies.** Since $|f_n| \leq g$ a.e., the functions $g + f_n \geq 0$ and $g - f_n \geq 0$ a.e., so Fatou's lemma is applicable to both sequences. Also, $f_n \to f$ pointwise a.e. and $|f| \leq g$ a.e., so f is integrable (i.e., $\int |f| \, d\mu \leq \int g \, d\mu < \infty$).

2. **Lower bound on** $\liminf$**.** Apply Fatou's lemma to the non-negative sequence $g + f_n$:

$$\int \liminf_{n\to\infty} (g + f_n) \, d\mu \leq \liminf_{n\to\infty} \int (g + f_n) \, d\mu.$$

The left side equals $\int (g + f) \, d\mu = \int g \, d\mu + \int f \, d\mu$ ($\because g + f_n \to g + f$ pointwise a.e.). The right side equals $\int g \, d\mu + \liminf \int f_n \, d\mu$ ($\because \int g \, d\mu < \infty$ is a constant). Subtracting $\int g \, d\mu$ (which is finite) from both sides:

$$\int f \, d\mu \leq \liminf_{n\to\infty} \int f_n \, d\mu.$$

3. **Upper bound on** $\limsup$**.** Apply Fatou's lemma to the non-negative sequence $g - f_n$:

$$\int \liminf_{n\to\infty}(g - f_n)\, d\mu \le \liminf_{n\to\infty} \int (g - f_n)\, d\mu.$$

The left side equals $\int (g-f)\, d\mu = \int g\, d\mu - \int f\, d\mu$. The right side equals $\int g\, d\mu + \liminf(-\int f_n\, d\mu) = \int g\, d\mu - \limsup \int f_n\, d\mu$ (using $\liminf(-a_n) = -\limsup(a_n)$). Subtracting $\int g\, d\mu$ from both sides and multiplying by -1:

$$\limsup_{n\to\infty} \int f_n\, d\mu \le \int f\, d\mu.$$

Combining Steps 2 and 3:

$$\int f\, d\mu \le \liminf_{n\to\infty} \int f_n\, d\mu \le \limsup_{n\to\infty} \int f_n\, d\mu \le \int f\, d\mu.$$

All inequalities are equalities, so the limit exists and equals $\int f\, d\mu$. □

4 Limit & Measure Theorems

Theorem 4.1 (Fubini's Theorem)**.** *Let* $(X, \mathcal{A}, \mu)$ *and* $(Y, \mathcal{B}, \nu)$ *be* σ*-finite measure spaces. If* $f\colon X \times Y \to \mathbb{R}$ *is integrable w.r.t. the product measure, then:*

$$\int_{X\times Y} f\, d(\mu \times \nu) = \int_X \left(\int_Y f(x, y)\, d\nu \right) d\mu = \int_Y \left(\int_X f(x, y)\, d\mu \right) d\nu.$$

Usage: Justifies swapping integration order. In ML, used for computing expectations over joint distributions.

Theorem 4.2 (Radon–Nikodym Theorem)**.** *Let* μ, ν *be* σ*-finite on* $(\Omega, \mathcal{F})$ *with* $\nu \ll \mu$ *(**absolute continuity**:* $\nu \ll \mu$ *means that for every* $A \in \mathcal{F}$*,* $\mu(A) = 0 \implies \nu(A) = 0$*). Then there exists measurable* $f \ge 0$ *such that* $\nu(A) = \int_A f\, d\mu$ *for all* $A \in \mathcal{F}$*. The function* $f = d\nu/d\mu$ *is the **Radon–Nikodym derivative**.*

Usage: Foundation for probability densities, likelihood ratios, and importance sampling.

Theorem 4.3 (Law of Large Numbers (LLN))**.** *Let* $X_1, X_2, \dots$ *be i.i.d. with* $E[X_1] = \mu$*.*
- ***Weak:*** $\bar{X}_n \xrightarrow{P} \mu$*.*
- ***Strong:*** $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$ *(assuming* $E[X_1^2] < \infty$*).*

Usage: Justifies sample means as estimators; underpins Monte Carlo methods and ERM.

Theorem 4.4 (Central Limit Theorem (CLT))**.** *Let* $X_1, X_2, \dots$ *be i.i.d. with* $E[X_1] = \mu$*,* $\mathrm{Var}(X_1) = \sigma^2 \in (0, \infty)$*. Then:*

$$\sqrt{n}\,(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

Proof. We prove the statement step by step:

1. **Standardization.** Let $Y_i = (X_i - \mu)/\sigma$ so that $E[Y_i] = 0$ and $E[Y_i^2] = 1$. Define $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$. Then $S_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$, so it suffices to show $S_n \xrightarrow{d} N(0, 1)$.

2. **Characteristic function of** Y_1**.** The characteristic function of Y_1 is $\varphi_Y(t) = E[e^{itY_1}]$. By Taylor expansion (using $E[Y_1] = 0$ and $E[Y_1^2] = 1$):

$$\varphi_Y(t) = E\left[1 + itY_1 + \frac{(it)^2}{2} Y_1^2 + o(t^2)\right] = 1 + it \cdot 0 - \frac{t^2}{2} \cdot 1 + o(t^2) = 1 - \frac{t^2}{2} + o(t^2)$$

as $t \to 0$.

3. **Characteristic function of S_n.** Since $Y_1, \dots, Y_n$ are independent, the characteristic function of S_n factorizes:

$$\varphi_{S_n}(t) = E\Big[e^{it\cdot \frac{1}{\sqrt{n}}\sum_{j=1}^n Y_j}\Big] = \prod_{j=1}^{n} \varphi_Y\left(\frac{t}{\sqrt{n}}\right) = \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

4. **Taking the limit.** Substituting the expansion from Step 2 with t replaced by $t/\sqrt{n}$:

$$\varphi_{S_n}(t) = \left[1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right]^n.$$

   Using the standard limit $\lim_{n\to\infty}(1 + a/n)^n = e^a$ (here $a = -t^2/2$, and the $o(1/n)$ remainder vanishes in the limit):

$$\lim_{n\to\infty} \varphi_{S_n}(t) = e^{-t^2/2}.$$

   This is the characteristic function of $N(0, 1)$.

5. **Conclusion via Lévy's continuity theorem.** Lévy's continuity theorem states: if $\varphi_{S_n}(t) \to \varphi(t)$ pointwise for all $t \in \mathbb{R}$, and φ is continuous at $t = 0$, then S_n converges in distribution to the distribution with characteristic function φ.

Since $e^{-t^2/2}$ is continuous (everywhere, in particular at 0), we conclude $S_n \xrightarrow{d} N(0, 1)$, which gives $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$. □

Theorem 4.5 (Continuous Mapping Theorem)**.** *If $X_n \xrightarrow{d} X$ and g is continuous a.e. (w.r.t. the distribution of X), then $g(X_n) \xrightarrow{d} g(X)$. Analogous results hold for convergence in probability and a.s.*

Usage: Extends convergence through transformations, e.g. $X_n \xrightarrow{d} X \Rightarrow X_n^2 \xrightarrow{d} X^2$.

Theorem 4.6 (Cramér–Wold Theorem)**.** *$\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$ in $\mathbb{R}^d$ if and only if $\boldsymbol{t}^\top \boldsymbol{X}_n \xrightarrow{d} \boldsymbol{t}^\top \boldsymbol{X}$ for every $\boldsymbol{t} \in \mathbb{R}^d$.*

Usage: Reduces multivariate convergence in distribution to univariate. Essential for proving the multivariate CLT.

Theorem 4.7 (Glivenko–Cantelli Theorem)**.** *Let F_n be the empirical CDF of i.i.d. samples from distribution F. Then:*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{a.s.}} 0.$$

Usage: The empirical CDF converges uniformly to the true CDF. Foundational for the Kolmogorov–Smirnov test and uniform convergence in statistical learning theory.