Measurable Transformations

$T : \Omega_1 \to \Omega_2$ is $\langle \mathcal{F}_1, \mathcal{F}_2 \rangle$-mble if $T^{-1}(A) \equiv \{\omega \in \Omega_1 : T(\omega) \in A\} \in \mathcal{F}_1$.

$T : \mathbb{R} \to \mathbb{R}$ is $\langle \mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}) \rangle$-mble $\iff$ $T^{-1}(-\infty, r) = \{\omega \in \mathbb{R} : T(\omega) < r\} \in \mathcal{B}(\mathbb{R})\ \forall r \in \mathbb{R}$.

Induced Measures & Distribution Functions

The **distribution** of X (denoted P_X), is the induced measure of X under P on $\mathcal{B}(\mathbb{R})$, i.e.,
$P_X(A) = P(X^{-1}(A)) = P(\{\omega \in \Omega : X(\omega) \in A\}) = P(X \in A), \quad A \in \mathcal{B}(\mathbb{R})$.

The **cumulative distribution func** (cdf) of a r.v. X is
$F(x) = P_X((-\infty, x]) = P(X \le x), \quad x \in \mathbb{R}$.
(1) F is right continuous: if $x_n \downarrow x_0$ and $x_n \ge x$ then
$F(x_n) = P_X((-\infty, x_n]) \downarrow P_X((-\infty, x_0]) = F(x_0)$ by mcfa.
(2) F is monotone nondecreasing: if $x \le y \implies (-\infty, x] \subset (-\infty, y]$ then
$F(x) = P_X((-\infty, x]) \le P_X((-\infty, y]) = F(y)$ by monotonicity
(3) $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$: show using argument similar to (1).

Integrals

For **disjoint** $A_1, A_2, \dots \in \mathcal{F}$, mble μ-int $f : \Omega \to \mathbb{R}$, by the DCT
$\int_\Omega f \mathbb{I}_{\bigcup_{n=1}^\infty A_n} d\mu = \sum_{n=1}^\infty \int_{A_i} f d\mu$.

Convergence Theorems

MCT: If $f_n : \Omega \to \overline{\mathbb{R}}$ is an increasing seq of nonneg mble funcs and $f_n(\omega) \uparrow f(\omega)$ a.e.(μ), then $\int_\Omega f_n d\mu \uparrow \int_\Omega f d\mu$. That is, $\int_\Omega f d\mu = \int_\Omega \lim_{n\to\infty} f_n d\mu = \lim_{n\to\infty} \int_\Omega f_n d\mu$.

Fatou's Lemma: If $f_n : \Omega \to \overline{\mathbb{R}}$ is a sequence of nonneg funcs, then $\int_\Omega \underline{\lim} f_n d\mu \le \underline{\lim} \int_\Omega f_n d\mu$.

DCT: Suppose (1) $g : \Omega \to \overline{\mathbb{R}}$ is a nonneg, μ-int func; (2) $|f_n| \le g$ a.e.(μ) $\forall n \ge 1$; and (3) $f_n \to f$ a.e.(μ). Then, f is μ-int and $\lim_{n\to\infty} \int_\Omega |f_n - f| d\mu = 0$ and $\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu$.

weaker conditions: If $\mu(\Omega) < \infty$ and $f, f_n : \Omega \to \overline{\mathbb{R}}$ are mble such that $f_n \to f$ a.e.(μ) and $\{f_n : n \ge 1\}$ is UI (see next section), then f is μ-int and $\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu$.

Scheffe's Theorem: Let $\nu_n(A) = \int_A f_n d\mu\ \forall A \in \mathcal{F}$ be finite measures with densities $f_n \ge 0$ for all $n \ge 0$. If $\nu_n(\Omega) = \nu_0(\Omega) < \infty$ for all $n \ge 1$ and $f_n \to f$ a.e.(μ), then
$\lim_{n\to\infty} \int_\Omega |f_n - f_0| d\mu = 0$.
Also, $\sup_{A\in\mathcal{F}} |\nu_n(A) - \nu_0(A)| = \frac{1}{2} \int_\Omega |f_n - f_0| d\mu \to 0$ as $n \to \infty$.

Uniform Integrability

If $f : \Omega \to \mathbb{R}$ is μ-int, by the DCT
$\lim_{n\to\infty} \int_{|f|>n} |f| d\mu = \lim_{n\to\infty} \int_\Omega \mathbb{I}_{|f|>n} |f| d\mu = 0$.

A family of μ-int funcs $\{f_\lambda : \lambda \in \Lambda\}$ on a msp $(\Omega, \mathcal{F}, \mu)$ is **uniformly integrable** (UI) w.r.t. μ if $\sup_{\lambda\in\Lambda} \int_{|f_\lambda|>t} |f_\lambda| d\mu \to 0$ as $t \to \infty$.

Suppose $\mathcal{A} \equiv \{f_\lambda : \lambda \in \Lambda\}$ is a collection of μ-int funcs on a msp $(\Omega, \mathcal{F}, \mu)$. Then, (1) if Λ is a finite set, then $\mathcal{A}$ is UI; (2) if $\exists \varepsilon > 0$ such that $\sup \left\{ \int |f_\lambda|^{1+\varepsilon} d\mu : \lambda \in \Lambda \right\} < \infty$, then $\mathcal{A}$ is UI; (3) if $|f_\lambda| \le f$ a.e.(μ) and $\int f d\mu < \infty$, then $\mathcal{A}$ is UI; (4) if $\mathcal{A}$ is UI and $\mu(\Omega) < \infty$, then $\exists M > 0$ such that $\sup \left\{ \int |f_\lambda| d\mu : \lambda \in \Lambda \right\} \le M$; (5) if $\{f_\lambda : \lambda \in \Lambda\}$ and $\{g_\lambda : \lambda \in \Lambda\}$ are both UI, then $\{f_\lambda + g_\lambda : \lambda \in \Lambda\}$ is also UI.

Independence

$\{A_i : i \in I\} \subset \mathcal{F}$ are **indep** if $\forall i_1, \dots, i_n \in I$ distinct indices and fixed $n \in \mathbb{N}$,
$P\left(\bigcap_{j=1}^n A_{i_j}\right) = \prod_{j=1}^n P\left(A_{i_j}\right)$, totalling $\sum_{k=2}^n \binom{n}{k} = 2^n - n - 1$ indep conditions.

$\{\mathcal{G}_i : i \in I\}$ are **indep** if any possible collection $\{A_i : A_i \in \mathcal{G}_i, i \in I\}$ of sets are indep.

$\{X_i : i \in I\}$ are **indep** if $\{\sigma\langle X_i\rangle : i \in I\}$ is indep, where
$\sigma\langle X_i\rangle = \left\{X_i^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\right\} = X_i^{-1}(\mathcal{B}(\mathbb{R}))$ is the σ-algebra generated by X_i. That is, $\forall i_1, \dots, i_n \in I$ distinct indices, fixed $n \in \mathbb{N}$ and $\forall B_{i_1}, \dots, B_{i_n} \in \mathcal{B}(\mathbb{R})$,
$P\left(X_{i_1} \in B_{i_1}, \dots, X_{i_n} \in B_{i_n}\right) = \prod_{j=1}^n P\left(X_{i_j} \in B_{i_j}\right)$. Equivalently,
$P\left(X_{i_1} \le x_1, \dots, X_{i_n} \le x_n\right) = \prod_{j=1}^n P\left(X_{i_j} \le x_j\right)\ \forall x_1, \dots, x_n \in \mathbb{R}$.

Independence of generated σ-algebras: If $\mathcal{G}_i \subset \mathcal{F}$ is a π-class $\forall i \in I$ and $\{\mathcal{G}_i : i \in I\}$ is indep, then $\{\sigma\langle\mathcal{G}_i\rangle : i \in I\}$ is indep.

Borel-Cantelli Lemmas

If $A_1, A_2, \dots \in \mathcal{F}$, then $\overline{\lim} A_n = \bigcap_{k=1}^\infty \bigcup_{n=k}^\infty A_n \in \mathcal{F}$ and
$\underline{\lim} A_n = \bigcup_{k=1}^\infty \bigcap_{n=k}^\infty A_n \in \mathcal{F}$. Also, $\underline{\lim} A_n \subset \overline{\lim} A_n$,
$(\text{“}A_n \text{ i.o”})^c = (\overline{\lim} A_n)^c = \underline{\lim} A_n^c = \text{“}A_n^c \text{ eventually,”}$, and
$(\text{“}A_n \text{ eventually”})^c = (\underline{\lim} A_n)^c = \overline{\lim} A_n^c = \text{“}A_n^c \text{ i.o.”}$ by De Morgan's laws.

Borel-Cantelli Lemma: For a psp $(\Omega, \mathcal{F}, P)$ with $A_1, A_2, \dots \in \mathcal{F}$:
(1) If $\sum_n P(A_n) < \infty$, then $P\left(\overline{\lim} A_n\right) = P(A_n \text{ occurs i.o.}) = 0$.
(2) If $\{A_n\}$ are indep and $\sum_n P(A_n) = \infty$, then $P\left(\overline{\lim} A_n\right) = P(A_n \text{ occurs i.o.}) = 1$.

Borel 0-1 Law: If $A_1, A_2, \dots$ are indep events, then
$P(\overline{\lim} A_n) = P(A_n \text{ occurs i.o.}) = \begin{cases} 0 \iff \sum_n P(A_n) < \infty, \\ 1 \iff \sum_n P(A_n) = \infty. \end{cases}$

Tail Events & Kolmogorov's 0-1 Law

The **tail σ-algebra** of $\{X_n\}_{n\ge1}$ is $\mathcal{T} = \bigcap_{n=1}^\infty \sigma\left\langle\left\{X_j : j \ge n\right\}\right\rangle$, where
$\sigma\left\langle\left\{X_j : j \ge n\right\}\right\rangle = \sigma\left\langle\left\{X_j^{-1} : B \in \mathcal{B}(\mathbb{R}), j \ge n\right\}\right\rangle$ is the σ-algebra generated by $X_j, j \ge n$. Any set (event) $A \in \mathcal{T}$ is a **tail event**.

A r.v. $T : \Omega \to \overline{\mathbb{R}}$ is a **tail r.v.** if T is $\langle \mathcal{F}, \mathcal{B}(\overline{\mathbb{R}})\rangle$-mble, e.g., $T^{-1}(B) \in \mathcal{F}\ \forall B \in \mathcal{B}(\mathbb{R})$.

Kolmogorov's 0-1 Law: If $\{X_n\}_{n\ge1}$ are indep and $A \in \mathcal{T}$, then $P(A) \in \{0, 1\}$.

Corollary: All tail r.v.'s are degenerate. That is, $T \in \mathcal{T} \implies \exists c \in \overline{\mathbb{R}}$ such that $P(T = c) = 1$. E.g, $\overline{\lim} S_n / a_n$ and $\underline{\lim} S_n / a_n$ are degenerate if $a_n \to \infty$.

Convergence

$X_1, X_2, \dots$ **converge almost surely** to X_0 on $(\Omega, \mathcal{F}, P)$ if
$P(\{\lim_{n\to\infty} X_n(\omega) = X_0(\omega)\}) = 1$.
TFAE:
(1) $X_n \xrightarrow{\text{as}} 0$,
(2) $P(|X_n| > \varepsilon \text{ i.o}) = 0, \quad \forall \varepsilon > 0$,
(3) $P(|X_n| > 1/k \text{ i.o.}) = 0, \quad \forall k \in \mathbb{N}$.
TFAE:
(1') $X_n \xrightarrow{\text{as}} X_0$,
(2') $\sup_{j\ge n} \left|X_j - X_0\right| \xrightarrow{\text{p}} 0$ as $n \to \infty$,
(3') $\lim_{n\to\infty} P\left(\bigcap_{j=n}^\infty \left[\left|X_j - X_0\right| \le \varepsilon\right]\right) = 1, \quad \forall \varepsilon > 0$.

$X_1, X_2, \dots$ **converge in probability** to X_0 on $(\Omega, \mathcal{F}, P)$ if
$\lim_{n\to\infty} P(|X_n - X_0| > \varepsilon) = 0, \quad \forall \varepsilon > 0$.
TFAE:
(1) $X_n \xrightarrow{\text{p}} X_0$,
(2) $\sup_{m\ge n} (|X_m - X_n| > \varepsilon) \to 0$ as $n \to \infty, \quad \forall \varepsilon > 0$,
(3) $\forall \left\{n_j\right\}$ of $\{X_n\}$, $\exists \{n_{j_k}\}$ such that $X_{n_{j_k}} \xrightarrow{\text{as}} X_0$.

Continuous functions preserve convergence. If $g : \mathbb{R} \to \mathbb{R}$ is continuous, then
(1) $X_n \xrightarrow{\text{as}} X_0 \implies g(X_n) \xrightarrow{\text{as}} g(X_0)$,
(2) $X_n \xrightarrow{\text{p}} X_0 \implies g(X_n) \xrightarrow{\text{p}} g(X_0)$.

$X_1, X_2, \dots \in \mathcal{L}_r(\Omega, \mathcal{F}, P) \equiv \left\{\text{mble } X \in \mathbb{R} : \int_\Omega |X|^r dP < \infty\right\}$ **converges in $\mathcal{L}_r$** to X_0 if $\lim_{n\to\infty} \int_\Omega |X_n - X_0|^r dP = 0$.

If $X \in \mathcal{L}_r$, then $t^r P(|X| > t) \to 0$ as $t \to \infty$, that is, $\uparrow r \implies$ faster convergence. If $\exists p \in (0, \infty)$ such that $t^p P(|X| > t) \to 0$, then $X \in \mathcal{L}_r\ \forall r \in (0, p)$. Also,
$X_n \xrightarrow{\mathcal{L}_r} X_0 \implies X_n \xrightarrow{\mathcal{L}_p} X_0\ \forall p \in (0, r)$.

If $\{X_n\}_{n\ge1} \subset \mathcal{L}_r$, then $\exists X_0 \in \mathcal{L}_r$ such that
$X_n \xrightarrow{\mathcal{L}_r} X_0 \iff \sup_{m\ge n} \mathrm{E}\,|X_m - X_n|^r \to 0$ as $n \to \infty$.

Fixed $m(X) \in \mathbb{R}$ is a **median** if $P(X \ge m(X)) \ge 1/2$ and $P(X \le m(X)) \ge 1/2$. Can be defined as $\inf\{x \in \mathbb{R} : P(X \le x) \ge 1/2\}$. If $P(|X| \ge c) < \varepsilon \le 1/2$ then $|m(X)| \le c$.

Levy's Inequality: If $\{X_n\}_{n\ge1}$ are independent, then $\forall \varepsilon > 0$,
(1) $P\left(\max_{1\le j\le n}\left(S_j - m(S_j - S_n)\right) \ge \varepsilon\right) \le 2P(S_n \ge \varepsilon)$,
(2) $P\left(\max_{1\le j\le n}\left|S_j - m(S_j - S_n)\right| \ge \varepsilon\right) \le 2P(|S_n| \ge \varepsilon)$.

Levy's Theorem: If $\{X_n\}_{n\ge1}$ are indep, then $S_n \xrightarrow{\text{as}} S \iff S_n \xrightarrow{\text{p}} S$.

Khintchine-Kolmogorov Convergence Theorem: If $\{X_n\}_{n\ge1}$ are indep with
$\mathrm{E}(X_n) = 0, \mathrm{E}(X_n^2) < \infty$ for all $n \ge 1$ and $\sum_n \mathrm{E}(X_n^2) < \infty$, then S_n converges a.s.(P) and in $\mathcal{L}_2$ to $S = \sum_n X_n$. Also, $\mathrm{E}(S) = 0, \mathrm{E}(S^2) = \sum_n E(X_n^2)$.

$\{X_n\}$ and $\{Y_n\}$ are **tail equivalent** if $\sum_n P(X_n \ne Y_n) < \infty$. If $\{X_n\}$ and $\{Y_n\}$ are tail equivalent, then
(1) By Borel-Cantelli, $P\left(\overline{\lim}(X_n \ne Y_n)\right) = 0 \implies P\left(X_n = Y_n \text{ for large } n\right) = 1$,
(2) $S_n = \sum_{j=1}^n X_j \xrightarrow{\text{as}} S \iff S_n' = \sum_{j=1}^n Y_j \xrightarrow{\text{as}} S'$,
(3) If $b_n \to \infty$, then $\frac{\sum_{j=1}^n X_j}{b_n} \xrightarrow{\text{as}} 0 \iff \frac{\sum_{j=1}^n Y_j}{b_n} \xrightarrow{\text{as}} 0$.

Berry-Esseen Lemma: If $X_1, \dots, X_n$ are indep with $\mathrm{E}(X_i) = 0$ and $\mathrm{E}\left|X_i\right|^3 < \infty, 1 \le i \le n$, then $\forall n \ge 4$, $\sup_{x\in\mathbb{R}} \left|P\left(\frac{S_n}{\sigma_n} \le x\right) - \Phi(x)\right| \le \frac{2.75}{\sigma_n^3} \sum_{i=1}^n \mathrm{E}\left|X_i\right|^3$, where
$S_n = \sum_{j=1}^n X_j,\ \sigma_n^2 = \mathrm{Var}(S_n) = \sum_{i=1}^n \mathrm{E}(X_i^2)$, and $\Phi(\cdot)$ is the $\mathcal{N}(0, 1)$ cdf.

Kolmogorov's 3-Series Theorem: If $\{X_n\}_{n\ge1}$ are indep, define for fixed $c > 0$,
$\sum_n P(|X_n| > c), \quad \sum_n \mathrm{E}(X_n^{(c)}), \quad \sum_n \mathrm{Var}(X_n^{(c)})$, where $X_n^{(c)} = X_n \mathbb{I}_{|X_n|\le c}$. Then,
(1) if the 3 series conv for *some* $c > 0$, then $S_n \xrightarrow{\text{as}} S$,
(2) if $S_n \xrightarrow{\text{as}} S$, then the 3 series converge for *all* $c > 0$.

Corollary: If $\{X_n\}_{n\ge1}$ are indep with $\mathrm{E}(X_n) = 0$, then
(1) if $\sum_n \left[\mathrm{E}(X_n^{(c)})^2 + \mathrm{E}\,|X_n| \mathbb{I}_{|X_n|>c}\right] < \infty$ for *some* $c > 0$, then $S_n \xrightarrow{\text{as}} S$,
(2) if $\sum_n \mathrm{E}\,|X_n|^{\alpha_n} < \infty$ for *some* $\{\alpha_n\} \subset [1, 2]$, then $S_n \xrightarrow{\text{as}} S$.

Laws of Large Numbers

$\{X_n\}_{n\ge1}$ obeys the LLN if $\exists \{b_n\} \subset \mathbb{R}$ and $0 < a_n \uparrow$ such that

$$\textbf{SLLN: } \frac{S_n - b_n}{a_n} \xrightarrow{\text{as}} 0, \quad \textbf{WLLN: } \frac{S_n - b_n}{a_n} \xrightarrow{\text{p}} 0.$$

Kronecker's Lemma: If $\{a_n\}, \{b_n\} \subset \mathbb{R}$ such that $0 < b_n \uparrow \infty$ and $\sum_{n=1}^\infty a_n / b_n$ converges, then $\frac{1}{b_n} \sum_{j=1}^n a_n \to 0$ as $n \to \infty$.

Cesaro's Mean Summability Theorem: If $\{x_n\} \subset \mathbb{R}$ such that $\lim_{n\to\infty} x_n = x < \infty$, then $\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^n x_j = x$.

Theorem: If $\{X_n\}$ indep such that $\sum_{n=1}^\infty \mathrm{E}\,|X_n|^{\alpha_n} / n^{\alpha_n} < \infty$ for $\alpha_n \in [1, 2]$, then $\frac{S_n - \mathrm{E}\,S_n}{n} = \frac{1}{n} \sum_{i=1}^n (X_i - \mathrm{E}\,X_i) \xrightarrow{\text{as}} 0$.

Marcinkiewicz-Zygmund SLLN: If $\{X_n\}_{n\ge1}$ are iid and $p \in (0, 2)$,
(1) if $\exists c \in \mathbb{R}$ s.t. $\frac{S_n - nc}{n^{1/p}} \xrightarrow{\text{as}} 0$, then $\mathrm{E}\,|X_1|^p < \infty$.
(2) if $\mathrm{E}\,|X_1|^p < \infty$, then (2) holds with $c = \mathrm{E}\,X_1$ if $p \in [1, 2)$ and (2) holds $\forall c \in \mathbb{R}$ if $p \in (0, 1)$.

Kolmogorov's SLLN: If $\{X_n\}$ are iid, then
$\bar{X}_n = \frac{S_n}{n} \xrightarrow{\text{as}} \mathrm{E}\,X_1 \iff \mathrm{E}\,|X_1| < \infty \iff \frac{S_n - n\,\mathrm{E}\,X_1}{n} \xrightarrow{\text{as}} 0$.

Useful Theorem: For any r.v. X and $r > 0$,
$\sum_{n=1}^\infty P(|X| > n^{1/r}) \le \mathrm{E}\,|X|^r \le \sum_{n=0}^\infty P(|X| > n^{1/r})$.

Etemaldi's SLLN: If $\{X_n\}_{n\ge1}$ are *pairwise* indep and identically distributed, then
$\bar{X}_n = \frac{S_n}{n} \xrightarrow{\text{as}} \mathrm{E}\,X_1 \iff \mathrm{E}\,|X_1| < \infty$.

General WLLN: If $\{X_n\}_{n\ge1}$ are indep, $\sum_{j=1}^n P\left(\left|X_j\right| > n\right) \to 0$, and
$\frac{1}{n^2} \sum_{j=1}^n E X_j^{(n)2} \to 0$, then $\frac{S_n - a_n}{n} \xrightarrow{\text{p}} 0$, where $a_n = \sum_{j=1}^n \mathrm{E}\,X_j^{(n)}$ and $X_j^{(n)} \equiv X_j I(\left|X_j\right| \le n)$.

Feller's WLLN: If $\{X_n\}$ iid with $\lim_{n\to\infty} xP(|X_1| > x) = 0$, then $\frac{S_n}{n} - \mathrm{E}\,X_1^{(n)} \xrightarrow{\text{p}} 0$.

Empirical Distributions

The **empirical cdf** of $X_1, \dots, X_n$ is the random cdf: $F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \le x), \quad x \in \mathbb{R}$.
(1) With X_i's on $(\Omega, \mathcal{F}, P)$, for each $\omega \in \Omega$, $F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^n I(X_i(\omega) \le x)$.
(2) $F_n(x)$ is a right-continuous, nondecreasing func of $x \in \mathbb{R}$,
(3) For any $x \in \mathbb{R}$, $F_n(x)$ is a r.v., i.e., is $\langle \mathcal{F}, \mathcal{B}(\mathbb{R})\rangle$-mble:
$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i^{-1}(-\infty, x])(\omega)$.

Glivenko-Cantelli Theorem: If $\{X_n\}_{n\ge1} \overset{\text{iid}}{\sim} F$, then $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow{\text{as}} 0$.

Quantile func: $\phi(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\} \equiv F^{-1}(u),\ u \in (0, 1)$ which implies $F(x) \ge u \iff x \ge \phi(u)$ and $F(\phi(u)-) \le u \le F(\phi(u))$.

Convergence in Distribution

The **cdf** of μ_n is $F_n(\mathbf{x}) = \mu_n\left((-\infty, x_1] \times \dots \times (-\infty, x_n]\right), \quad \mathbf{x} \in \mathbb{R}^k$.
If X_n has prob dist μ_n [i.e., $P(X_n \in A) = \mu_n(A),\ A \in \mathcal{B}(\mathbb{R}^k)$], then F_n is the **cdf** of X_n.

$\{\mu_n\}_{n\ge1}$ ($\{F_n\}_{n\ge1}$) **converges weakly** to μ_0 (F_0), denoted $\mu_n \Rightarrow \mu_0$ ($F_n \Rightarrow F_0$), if $\lim_{n\to\infty} F_n(\mathbf{x}) = F_0(\mathbf{x})\ \forall \mathbf{x} \in C(F_0)$, where $C(F_0) = \left\{\mathbf{x} \in \mathbb{R}^k : F_0 \text{ is continuous at } \mathbf{x}\right\}$.

$\{X_n\}_{n\ge1} \subset R^k$ **converges in distribution** to a r.v. X_0 if $\mu_n \Rightarrow \mu_0$, denoted by $X_n \xrightarrow{\text{d}} X_0$.
That is, if $X_n = (X_{n,1}, \dots, X_{n,k})$ has cdf F_n, $n \ge 0$, then

$$\lim_{n\to\infty} F_n(\mathbf{x}) = \lim_{n\to\infty} P(X_{n,1} \le x_1, \dots, X_{n,k} \le x_k)$$
$$= P(X_{0,1} \le x_1, \dots, X_{0,k} \le x_k) = F_0(\mathbf{x}) \quad \forall \mathbf{x} \in C(F_0).$$

NB: (1) $\mathbf{x} = (x_1, \dots, x_k) \in C(F_0) \iff F_0(\mathbf{x}) = P(X_{0,1} < x_1, \dots, X_{0,k} < x_k) = F_0(\mathbf{x}-)$ i.e., if also left continuous; (2) $C(F_0)^c$ is at most countable; (3) $X_n \xrightarrow{\text{p}} X_0 \implies X_n \xrightarrow{\text{d}} X_0$ but not the other direction, unless X_0 is degenerate.

Skorohod's Embedding Theorem: If $\{\mu_n\}_{n\ge0}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ such that $\mu_n \Rightarrow \mu_0$, then $\exists$ random vectors $\{Y_n\}_{n\ge0}$ on a *common* psp such that Y_n has distribution μ_n for all $n \ge 0$ and $Y_n \xrightarrow{\text{as}} Y_0$. That is, $P(Y_n \in A) = \mu_n(A),\ A \in \mathcal{B}(\mathbb{R}^k), n \ge 0$.

Continuous Mapping Theorem:
(a): Let $\{\mu_n\}_{n\ge0}$ are p.m.'s on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ and $h : \mathbb{R}^k \to \mathbb{R}^m$ is a $\langle \mathcal{B}(\mathbb{R}^k), \mathcal{B}(\mathbb{R}^m)\rangle$-mble func such that $\mu_0(D_h) = 0$, where $D_h \in \mathcal{B}(\mathbb{R}^k)$ denotes the set of all points of discontinuities of h. If $\mu_n \Rightarrow \mu_0$, then the induced measures converge weakly: $\mu_n h^{-1} \Rightarrow \mu_0 h^{-1}$.
(b): Let $X_{n\,n\ge0} \subset R^k$ and mble $h : \mathbb{R}^k \to \mathbb{R}^m$ be such that $P(X_0 \in D_h) = 0$, where D_h is as above. If $X_n \xrightarrow{\text{d}} X_0$, then $h(X_n) \xrightarrow{\text{d}} h(X_0)$.

Slutsky's Theorem: If $\{X_n\}_{n\ge1}, \{Y_n\}_{n\ge1}$ are such that $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{p}} a$ for some $a \in \mathbb{R}$, then (1) $X_n + Y_n \xrightarrow{\text{d}} X + a$, (2) $X_n Y_n \xrightarrow{\text{d}} aX$, (3) $X_n / Y_n \xrightarrow{\text{d}} X/a$ if $a \ne 0$.

Characterizations of Convergence in Distribution

For a p.m. μ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, a set $A \in \mathcal{B}(\mathbb{R}^k)$ is called a **μ-continuity** set if $\mu(\partial A) = 0$, where $\partial A = \overline{A} \setminus \text{int}A$. E.g., $\partial(-\infty, x] = (-\infty, x] \setminus (-\infty, x) = \{x\}$.

Helly-Bray Theorem: If $\{\mu_n\}_{n\geq 0}$ are p.m.'s on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then
(1) $\mu_n \Rightarrow \mu_0 \iff \mu_n(A) \to \mu_0(A)\ \forall A \in \mathcal{B}(\mathbb{R})$ with $\mu_0(\partial A) = 0$.
(2) $\mu_n \Rightarrow \mu_0 \iff \int f d\mu_n \to \int f d\mu_0$ for all bounded cont. func. $f: \mathbb{R} \to \mathbb{R}$.

Lemma: If $\mu_n \Rightarrow \mu_0$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $f: \mathbb{R} \to \mathbb{R}$ is a bounded, Borel-mble func with $\mu_0(D_f)$ (where $D_f \in \mathcal{B}(\mathbb{R})$ is the set of discontinuity points of f), then $\int f d\mu_n \to \int f d\mu_0$ as $n \to \infty$.

$\{\mu_n\}_{n\geq 1}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is **tight** if $\forall \varepsilon > 0, \exists M_\varepsilon > 0$ such that $\sup_{n\geq 1} \mu_n\left(\left\{x \in \mathbb{R}^k : \|x\| > M_\varepsilon\right\}\right) < \varepsilon$.

$\{X_n\}_{n\geq 1} \subset \mathbb{R}^k$ is **tight** if $\{\mu_n\}_{n\geq 1}$ is tight. That is, $\forall \varepsilon > 0, \exists M_\varepsilon > 0$ such that $\sup_{n\geq 1} P(\|X_n\| > M_\varepsilon) = \sup_{n\geq 1} \mu_n\left(\left\{x \in \mathbb{R}^k : \|x\| > M_\varepsilon\right\}\right) < \varepsilon$.

$\{X_n\}$ is **uniformly integrable** if $\forall \varepsilon > 0, \exists t_\varepsilon > 0$ such that $\sup_{n\geq 1} \mathrm{E}\,\|X_n\|\, I(\|X_n\| > t_\varepsilon) = \sup_{n\geq 1} \int_{\|x\|>t_\varepsilon} \|x\|\, d\mu_n < \varepsilon$.

Proposition:
(1) If $X_n \xrightarrow{d} X_0$, then $\{X_n\}$ is tight.
(2) If $\{X_n\}$ is tight and $Y_n \xrightarrow{p} 0$ for X_n, Y_n on $(\Omega_n, \mathcal{F}_n, P_n)$, then $X_n Y_n \xrightarrow{p} 0$.

Theorem: $\{X_n\}_{n\geq 1}$ ($\{\mu_n\}_{n\geq 1}$) is tight iff for any subseq X_{n_k} of X_n there $\exists$ a further subseq $X_{n_{k_j}}$ of X_{n_k} and a r.v. (p.m.) such that $X_{n_{k_j}} \xrightarrow{d} X_0$ ($\mu_{n_{k_j}} \Rightarrow \mu_0$).

Corollary: If $\{X_n\}$ is tight and its convergent subseq converge in law to the *same* r.v., then $X_n \xrightarrow{d} X_0$.

Theorem: If $\{X_n\}, n \geq 1$ is UI and $X_n \xrightarrow{d} X_0$, then $\mathrm{E}\,|X_0| < \infty$ and $\mathrm{E}\, X_n \to \mathrm{E}\, X_0$.

Corollary: If $X_n \xrightarrow{d} X_0$ and $\sup_{n\geq 1} \mathrm{E}\,|X_n|^{r+\delta} < \infty$ for some $r \in \mathbb{N}$ and $\delta > 0$, then $\mathrm{E}\,|X_0|^r < \infty$ and $\mathrm{E}\, X_n^r \to \mathrm{E}\, X_0^r$ (recall that $\sup_{n\geq 1} \mathrm{E}\,|Z_n|^{1+\delta} \implies \{Z_n\}$ is UI).

Fréchet-Shohat Theorem: If $\lim_{n\to\infty} \mathrm{E}\, X_n^r = \beta_r \in \mathbb{R}$ for all integers $r \geq 1$ and if $\{\beta_r : r \geq 1\}$ are the moments of a *unique* r.v. X_0, then $X_n \xrightarrow{d} X_0$.

Moments uniquely determine distribution when Cardeman's condition is met, $\sum_{r=1}^\infty \beta_{2r}^{-1/(2r)} = \infty$, or if the MGF $M_X(t) = \mathrm{E}\, e^{tX} < \infty\ \forall\, |t| < \varepsilon$ for some $\varepsilon > 0$. Recall: $\mathrm{E}\, X^r = \frac{d^r}{dt^r} M_X(t)\Big|_{t=0}$.

Characteristic Functions

If $a + bi$ and $c + di$ are complex, then their sum is $(a + b) + (c + d)i$, their product is $(ac - bd) + (ad + bc)i$, and the modulus is $|a + bi| = \sqrt{a^2 + b^2} = \sqrt{(a + bi)(a - bi)}$ For any $b \in \mathbb{R}$, $e^{bi} = \cos(b) + i\sin(b)$ and $\left|e^{bi}\right| = \sqrt{\cos^2(b) + \sin^2(b)} = 1$. For fixed $b \in \mathbb{R}$, $g(t) = e^{tbi} : \mathbb{R} \to \mathbb{C}$ is infinitely differentiable in t with nth derivative $(bi)^n e^{tbi}$.

The **cf** of $X \in \mathbb{R}^k$ is $\phi_X(t) = \mathrm{E}\, e^{it'X} = \mathrm{E}\cos(t'X) + i\,\mathrm{E}\sin(t'X), \quad t \in \mathbb{R}^k$.

Note that $\phi_X(0) = 1$ and $\phi_X(t)$ is uniformly continuous on $\mathbb{R}^k$: by the BCT, $\sup_{t\in\mathbb{R}^k} \left|\phi_X(t + h) - \phi_X(t)\right| \to 0$ as $|h| \to 0$.

Theorem: If $X \in \mathcal{L}_r$, then $\phi_X(t)$ is r-times diffble on $\mathbb{R}$ and $\phi_X^{(r)}(t) = \mathrm{E}(iX)^r e^{itX}, \quad t \in \mathbb{R}$.

Riemann-Lebesgue Lemma: If X has a density f w.r.t. m on $\mathbb{R}$, then $\phi_X(t) \to 0$ as $|t| \to \infty$.

Levy Continuity Theorem:
(1) If $X_n \xrightarrow{d} X_0$, then $\forall T > 0$, $\sup_{|t|<T} \left|\phi_{X_n}(t) - \phi_{X_0}(t)\right| \to 0$ as $n \to \infty$.
(2) If $\phi_{X_n}(t) \to g(t)$ as $n \to \infty\ \forall t \in \mathbb{R}$ and $g(\cdot)$ is continuous at zero, then $g(\cdot)$ is a cf and $X_n \xrightarrow{d} X_0$, where X_0 has cf $g(\cdot)$.

Corollary: $X_n \xrightarrow{d} X_0 \iff \phi_{X_n}(t) \to \phi_{X_0}(t)$ as $n \to \infty \quad \forall t \in \mathbb{R}$.

Levy Inversion Formula in $\mathbb{R}^k$: Let $X \in \mathbb{R}^k$ with cf $\phi_X(t)$ for $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$. Then, $\forall$ rectangle $A = (a_1, b_1] \times \cdots \times (a_k, b_k]$ with $P(X \in \partial A) = 0$,
$P(X \in A) =$
$\lim_{T\to\infty} \frac{1}{(2\pi)^k} \int_{-T}^T \cdots \int_{-T}^T \prod_{j=1}^k \frac{e^{-it_j a_j} - e^{-it_j b_j}}{it_j} \phi_X(t_1, \ldots, t_k) dt_1 \ldots dt_k$.

Also, if $\int_{\mathbb{R}^k} \left|\phi_X(t_1, \ldots, t_k)\right| dt_1 \ldots dt_k < \infty$, then X has a bounded, continuous density $f_X(x)$ w.r.t. m on $\mathbb{R}^k$ given by
$f_X(x) = \frac{1}{(2\pi)^k} \int_{\mathbb{R}^k} e^{-i\sum_{j=1}^k x_j t_j} \phi_X(t_1, \ldots, t_k) dt_1 \cdots dt_k, \quad x \in \mathbb{R}^k$.

Theorem: $X_1, \ldots, X_k$ are indep $\iff \forall t_1, \ldots, t_k \in \mathbb{R}$,
$\phi_{X_1,\ldots,X_k}(t_1, \ldots, t_k) \equiv \mathrm{E}\, e^{i\sum_{j=1}^k X_j t_j} = \prod_{j=1}^k \mathrm{E}\, e^{iX_j t_j} = \prod_{j=1}^k \phi_{X_j}(t_j)$.

Theorem: If $\{X_n\}, n \geq 0 \subset \mathbb{R}^k$,
(1) $X_n \xrightarrow{d} X_0 \iff \phi_{X_n}(t) \to \phi_{X_0}(t)\ \forall t \in \mathbb{R}^k$,
(2) (Cramer-Wold device) $X_n \xrightarrow{d} X_0 \iff t'X_n \xrightarrow{d} t'X_0\ \forall t \in \mathbb{R}^k$.

Central Limit Theorems

Let $\left\{X_{n,j} : 1 \leq j \leq r_n\right\}_{n\geq 1}$ be an independent triangular array with

$$\mathrm{E}\, X_{n,j} = 0, \quad 0 < \mathrm{E}\, X_{n,j}^2 = \sigma_{n,j}^2 < \infty, \quad \nu_n^2 = \sum_{j=1}^{r_n} \sigma_{n,j}^2, \tag{1}$$

Lindeberg Condition: $\forall \varepsilon > 0$, $\lim_{n\to\infty} \nu_n^{-2} \sum_{j=1}^{r_n} \mathrm{E}\, X_{n,j}^2 \mathbb{I}_{|X_{n,j}|>\varepsilon\nu_n} = 0$.

Lindeberg CLT: if $\left\{X_{n,j} : 1 \leq j \leq r_n\right\}_{n\geq 1}$ is a triangular array satisfying (1) and the Lindeberg condition, then $\frac{S_n}{\nu_n} = \frac{\sum_{j=1}^{r_n} X_{n,j}}{\nu_n} \xrightarrow{d} \mathcal{N}(0, 1)$.

Corollary: if $\left\{X_{n,j}\right\}$ is a null array, i.e., $\max_{1\leq j\leq r_n} P\left(\left|X_{n,j}\right| > \varepsilon\nu_n\right) \to 0$, then the CLT holds $\iff$ Lindeberg condition.

Lyapounov's Condition: $\exists \delta > 0$ such that $\lim_{n\to\infty} \nu_n^{-(d+\delta)} \sum_{j=1}^{r_n} \mathrm{E}\left|X_{n,j}\right|^{2+\delta} = 0$.

Lyapounov's CLT: if $\left\{X_{n,j} : 1 \leq j \leq r_n\right\}_{n\geq 1}$ is a triangular array satisfying (1) and the Lyapounov's condition, then $\frac{S_n}{\nu_n} = \frac{\sum_{j=1}^{r_n} X_{n,j}}{\nu_n} \xrightarrow{d} \mathcal{N}(0, 1)$.

Multivariate CLT: if $\{X_n\}_{n\geq 1} \subset \mathbb{R}^d$ is iid with $\mathrm{E}\,\|X_1\|^2 < \infty$ and nonsingular $\mathrm{Var}(X_1) = \Sigma$ ($|\Sigma| \neq 0$), then $\sqrt{n}(\bar{X}_n - \mathrm{E}\, X_1) \xrightarrow{d} \mathcal{N}(0, \Sigma)$ by the Cramer-Wold device and Slutsky's theorem.

Infinitely Divisible & Stable Distributions

X is **infinitely divisible** if $\forall n \geq 1$, $\exists$ cf ϕ_n such that $\phi_X(t) = [\phi_n(t)]^n, \ \forall t \in \mathbb{R}$.
Note $X \overset{d}{=} X_{n,1} + \cdots + X_{n,n}$ for iid $X_{n,j}$ with cf ϕ_n.

X is infinitely divisible $\iff \phi_X(t) = \exp\left[itb + \int_{\mathbb{R}} \left(\frac{e^{itx} - 1 - it\tau(x)}{x^2}\right) dM(x)\right]\ \forall t \in \mathbb{R}$,
where $b \in \mathbb{R}$, M is a "cannonical" measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ satisfying $M(I) < \infty\ \forall$ finite intervals $I \in \mathbb{R}$ and $\forall x > 0$, $\int_{-\infty}^{-x} |y|^{-2} dM(y) + \int_x^\infty y^2 dM(y) < \infty$, and

$$\tau(x) = \begin{cases} 1 & \text{if } x > 1 \\ x & \text{if } |x| \leq 1 \\ -1 & \text{if } x < -1 \end{cases}$$

Theorem: If $\left\{X_{n,j} : 1 \leq j \leq r_n\right\}_{n\geq 1}$ is a null triangular array, then $\exists b_n \in \mathbb{R}$ such that $S_n - b_n \xrightarrow{d} X$ where X is infinitely divisible $\iff$ (a), (b), (c) hold:
(a) $\lim_{n\to\infty} \left[\sum_{j=1}^{r_n} \mathrm{E}\,\tau(X_{n,j})\right] - b_n = b$,
(b) $\lim_{n\to\infty} \sum_{j=1}^{r_n} \mathrm{Var}\left[\tau(X_{n,j})\right] = M\big((-1, 1)\big) + \int_{|y|\geq 1} y^{-2} dM(y)$,
(c) $\forall x > 0$ with $M(\{\pm x\}) = 0$, $\lim_{n\to\infty} \sum_{j=1}^{r_n} P(X_{n,j} > x) = \int_x^\infty y^{-2} dM(y)$, and $\lim_{n\to\infty} \sum_{j=1}^{r_n} P(X_{n,j} < -x) = \int_{-\infty}^{-x} y^{-2} dM(y)$.

Weaker result: X is infinitely divisible $\iff \exists \{X_n\}_{n\geq 1}$ iid with $\sum_{j=1}^n X_j \xrightarrow{d} X$.

Nondegenerate X is **stable** if $\forall n \geq 1$, $\exists a_n > 0$, $b_n \in \mathbb{R}$ such that $\phi_X(t) = \left[\phi_X\left(\frac{t}{a_n}\right)\right]^n \exp\left[-\frac{itb_n}{a_n}\right], \ \forall t \in \mathbb{R}$.
Note $X \overset{d}{=} [X_{n,1} + \cdots + X_{n,n} - b_n]/a_n$ for iid $X_{n,j} \sim X$, i.e., $S_n \overset{d}{=} a_n X + b_n$, where $a_n = n^{1/\alpha}$ with $\alpha \in (0, 2]$, e.g., $\alpha = 1$ for Cauchy, $\alpha = 2$ for normal. $\alpha < 2 \implies$ infinite variance.

All stable distributions are infinitely divisible and have canonical measure $M(A) = \sigma^2 \mathbb{I}_{0\in A}$ for normal distribution. For non-normals, for $0 < \alpha < 2$ and $x > 0$, $M_\alpha\big((0, x]\big) = cpx^{2-\alpha}$ and $M_\alpha\big([-x, 0)\big) = cqx^{2-\alpha}$ where $c > 0$, $p, q \geq 0$ with $p + q = 1$.

Theorem: X is stable $\iff \exists \{X_n\}, \{a_n\}, \{b_n\}$ such that $\frac{S_n - b_n}{a_n} \xrightarrow{d} X$.

Conditional Expectation

The **conditional expectation** of Y given $\mathcal{G} \subset \mathcal{F}$ under P, denoted $\mathrm{E}(Y \mid \mathcal{G})$, is $g : \Omega \to \mathbb{R}$ satisfying
(1) g is $\langle\mathcal{G}, \mathcal{B}B(\mathbb{R})\rangle$-mble, i.e., is a r.v.: $\mathrm{E}(Y \mid \mathcal{G})^{-1}(B) \in \mathcal{G}\ \forall B \in \mathcal{B}(\mathbb{R})$,
(2) $\forall G \in \mathcal{G}$, $\int_G g dP = \int_G Y dP$.

The **conditional probability** of $A \in \mathcal{F}$ given $\mathcal{G} \subset \mathcal{F}$, denoted $P(A \mid \mathcal{G})$, is $P(A \mid \mathcal{G}) = \mathrm{E}(\mathbb{I}_A \mid \mathcal{G})$.

Examples:
- If $\mathcal{G} = \{\emptyset, \Omega\}$, then $\omega \mapsto \mathrm{E}(Y)$ is $\mathcal{G}$-mble (since $\mathrm{E}(Y)^{-1}(B) = \emptyset$ if $E(Y) \in B$ and Ω o.w.) and trivially $\int_A \mathrm{E}(Y) dP = \int_A Y dP\ \forall A \in \mathcal{G}$. Thus $\mathrm{E}(Y \mid \mathcal{G}) = \mathrm{E}(Y)$.
- If $\mathcal{G} = \mathcal{F}$, then clearly $\mathrm{E}(Y \mid \mathcal{G}) = Y$.
- If $\mathcal{F} = \{\emptyset, \Omega, B, B^c\}$ where $0 < P(B) < 1$, then
$\mathrm{E}(Y \mid \mathcal{G})(\omega) = \left(\frac{1}{P(B)} \int_B Y dP\right) \mathbb{I}_B(\omega) + \left(\frac{1}{P(B^c)} \int_{B^c} Y dP\right) \mathbb{I}_{B^c}(\omega)$.

Projection Theorem:
- *Hilbert space*: $H = \mathcal{L}_2(\Omega, \mathcal{F}, P)$,
- *Inner product*: $\langle X, Z\rangle = \mathrm{E}(XZ)\ \forall X, Z \in H$,
- *Orthogonality*: $X, Z \in H$ are orthogonal if $\langle X, Z\rangle = 0$,
- *Distance*: squared distance between $X, Z \in H$ is the mse: $\|X - Z\|^2 = \langle X - Z, X - Z\rangle = \mathrm{E}(X - Z)^2$,
- *Statement of the Theorem*: let $H_0 \subset H$ be the subspace of all funcs of X with finite second moment. Then,
(1) $\exists \hat{Y} \in H_0$ such that $\left\|Y - \hat{Y}\right\| = \min\left\{\|Y - h(X)\|^2 : h(X) \in H_0\right\}$, where $\hat{Y}$ is the conditional expectation of Y given X,
(2) Let $V \in H$. Then $V = \hat{Y} \iff$ (a) $V \in H_0$ (i.e., V is $\langle\mathcal{G} = \sigma\langle X\rangle, \mathcal{B}(\mathbb{R})\rangle$-mble & $\mathrm{E}\, V^2 < \infty$), and (b) residual $Y - V$ is orthogonal to any other $h(X) \in H_0$ (i.e., $\mathrm{E}[(Y - V)h(X)] = \langle Y - V, h(X)\rangle = 0$).

Existence & a.s. Uniqueness: $\mathrm{E}(Y \mid \mathcal{G})$ satisfying (1)-(2) of the definition exists and if g and h are two versions of it, then $g = h$ as(P). *Proof*: Define $\mu(A) = \int_A Y dP\ \forall A \in \mathcal{G}$. Then μ is finite and $\mu \ll P$ on the restricted psp $(\Omega, \mathcal{G}, P)$ since $A \in \mathcal{G}$ with $P(A) = 0 \implies \mu(A) = 0$. By the Radon-Nikodym theorem, $\exists$ density $g = \frac{d\mu}{dP} \geq 0$ and for all $G \in \mathcal{G}$, $\int_G g dP = \mu(G) = \int_G Y dP$.

Example: (cble partition) Let $\mathcal{G} \subset \mathcal{F}$ is generated by countable partition $\left\{B_i\right\}_{i\geq 1}$ of disjoints sets in $\mathcal{F}$.
(1) for any $X \in \mathcal{L}_1(P)$, $\mathrm{E}(X \mid \mathcal{G}) = \sum_{i=1}^\infty \mathrm{E}_{B_i}(X)\mathbb{I}_{B_i}$ where $\mathrm{E}_{B_i}(X) = \int_{B_i} X dP / P(B_i)\mathbb{I}_{P(B_i)>0}$.
(2) for any $A \in \mathcal{F}$, $P(A \mid \mathcal{G}) = \sum_{i=1}^\infty P(A \mid B_i)\mathbb{I}_{B_i}$ where $P(A \mid B_i) = P(A \cup B_i)/P(B_i)\mathbb{I}_{P(B_i)>0}$.
Proof: let $h(\omega) = \sum_{i=1}^\infty \mathrm{E}_{B_i}(X)\mathbb{I}_{B_i}(\omega)$, $\omega \in \Omega$ and note $\mathcal{G} = \left\{\cup_I B_i : I \subset \mathbb{N}\right\}$.

Example: (discrete case) Let X be a discrete r.v. with support $x_1, x_2, \ldots$, then for $A \in \mathcal{B}(\mathbb{R})$, $P(A \mid X) = P(A \mid \sigma\langle X\rangle) = \sum_{i=1}^\infty P(A \mid X = x_i)\mathbb{I}_{X=x_i}$.

Example: (absolutely continuous case) $P(Y \in C \mid X) = \phi(X) = \frac{\int_C f(X,t)dt}{f_X(x)} \mathbb{I}_{f_X(x)>0}$ by showing $\int_B \phi(X) dP = \int_{X^{-1}(A)} \phi(X) dP = \int_A \phi(x) PX^{-1}(dx) = P([Y \in C] \cap B)$.

Properties:
(1) $\mathrm{E}(\mathrm{E}(Y \mid \mathcal{G})) = \mathrm{E}(Y)$ directly from the definition since $\Omega \in \mathcal{G}$,
(2) if Y is $\mathcal{G}$-mble, then $\mathrm{E}(Y \mid \mathcal{G}) = Y$ as(P) by taking $g = Y$,
(3) if $\mathcal{G}_1 \subset \mathcal{G}_2 \subset \mathcal{F}$, then $\mathrm{E}(Y \mid \mathcal{G}_1) = \mathrm{E}[\mathrm{E}(Y \mid \mathcal{G}_1) \mid \mathcal{G}_2] = \mathrm{E}[\mathrm{E}(Y \mid \mathcal{G}_2) \mid \mathcal{G}_1]$ as(P),
(4) if $Y \geq 0$ w.p.1, then $\mathrm{E}(Y \mid \mathcal{G}) \geq 0$ w.p.1,
(5) for $a, b \in \mathbb{R}$, $\mathrm{E}(aY_1 + bY_2 \mid \mathcal{G}) = a\,\mathrm{E}(Y_1 \mid \mathcal{G}) + b\,\mathrm{E}(Y_2 \mid \mathcal{G})$ as(P),
(6) if $Y_1 \geq Y_2$ w.p.1, then $\mathrm{E}(Y_1 \mid \mathcal{G}) \geq \mathrm{E}(Y_2 \mid \mathcal{G})$ w.p.1,
(7) if U is $\mathcal{G}$-mble and $\mathrm{E}\,|YU| < \infty$, then $\mathrm{E}(UY \mid \mathcal{G}) = U\,\mathrm{E}(Y \mid \mathcal{G})$ as(P),
(8) if $\phi : (a, b) \to \mathbb{R}$ is convex for $a, b \in \overline{\mathbb{R}}$, $P(Y \in (a, b)) = 1$, and $\mathrm{E}\,|\phi(Y)| < \infty$, then $\mathrm{E}(\phi(Y) \mid \mathcal{G}) \geq \phi(\mathrm{E}(Y \mid \mathcal{G}))$ as(P).

MCT for CE: If $0 \leq Y_n \leq Y_{n+1}$ w.p.1 $\forall n \geq 1$ and $Y_n \xrightarrow{as} Y$, then $\lim_{n\to\infty} \mathrm{E}(Y_n \mid \mathcal{G}) = \mathrm{E}(Y \mid \mathcal{G})$ w.p.1.

Fatou's Lemma for CE: If $0 \leq Y_n\ \forall n \geq 1$, then $\mathrm{E}(\underline{\lim} Y_n \mid \mathcal{G}) \leq \underline{\lim}\, \mathrm{E}(Y_n \mid \mathcal{G})$ w.p.1.

DCT for CE: If $Y_n \xrightarrow{as} Y$ and $|Y_n| \leq Z$ w.p.1 $\forall n \geq 1$ where $\mathrm{E}\,|Z| < \infty$, then $\lim_{n\to\infty} \mathrm{E}(Y_n \mid \mathcal{G}) = \mathrm{E}(Y \mid \mathcal{G})$ w.p.1.

Proposition: If $\mathcal{C}$ is a π-system such that $\mathcal{G} = \sigma\langle\mathcal{C}\rangle$, Ω is a countable union of disjoint $\mathcal{C}$-sets, and $\mathrm{E}\,|Y| < \infty$, then a P-integrable func g is a version of $\mathrm{E}(Y \mid \mathcal{G})$ if g is $\mathcal{G}$-measurable and $\mathrm{E}\, g\mathbb{I}_G = \int_G g dP = \int_G Y dP = \mathrm{E}\, Y\mathbb{I}_G\ \forall G \in \mathcal{C}$.

Conditional Distributions

The **conditional distribution** of Y given $\mathcal{G} \subset \mathcal{F}$ (or the regular conditional probability of μ_Y on $\mathbb{R}^d$ given $\mathcal{G}$) is a function $\mu : \Omega \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ satisfying
(1) $\forall \omega \in \Omega$, $\mu(\omega, \cdot)$ is a p.m. on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$,
(2) $\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\mu(\cdot, A)$ is $\mathcal{G}$-mble,
(3) $\forall A \in \mathcal{B}(\mathbb{R}^d)$, $\int_G \mu(\omega, A) dP(\omega) = P(G \cap \{Y \in A\})\ \forall G \in \mathcal{G}$.
Note: (2)-(3) $\iff \mu(\cdot, A)$ is a version of $P(Y \in A \mid \mathcal{G})$ as $\mu(\omega, A) = P(Y \in A \mid \mathcal{G})(\omega) = \mathrm{E}(\mathbb{I}_A(Y) \mid \mathcal{G})(\omega)$.

Sometimes $\mu(\cdot, \cdot)$ is denoted $\mathcal{L}(Y \mid \mathcal{G})$. For X, Y we write $\mathcal{L}(Y \mid X)$ for $\mathcal{L}(Y \mid \sigma\langle X\rangle)$. Define $\mu\big((x, y), C\big) = \mathbb{I}_{f_X(x)>0} \int_C f_{Y \mid X=x}(t) dt = \mathbb{I}_{f_X(x)>0} \int_C \frac{f(x,t)}{f_X(x)} dt, \ C \in \mathcal{B}(\mathbb{R})$.

If $\sigma\langle Y\rangle$ and $\mathcal{G}$ are independent, then $\mathrm{E}(Y \mid \mathcal{G} = \mathrm{E}(Y)$ since (1) $\mathrm{E}(Y)$ is a constant $\implies \mathrm{E}(Y)$ is $\mathcal{G}$-mble and (2) by independence $\int_G Y dP = \int_\Omega Y\mathbb{I}_G dP = \int_\Omega Y dP \int_\Omega \mathbb{I}_G dP = \mathrm{E}(Y)P(G) = \int_G \mathrm{E}(Y) dP\ \forall G \in \mathcal{G}$.

Theorem: For $Y \in \mathbb{R}^d$,
(a) $\exists \mu : \Omega \times \mathcal{B}(\mathbb{R}^d) \to [0, 1]$ satisfying (1)-(3) in the def of the conditional distribution of Y given $\mathcal{G}$,
(a) if $\phi : \mathbb{R}^d \to \mathbb{R}$ is Borel-mble with $\mathrm{E}\,|\phi(Y)| < \infty$ then $\forall \omega \in \Omega$, $E(\phi(Y) \mid \mathcal{G})(\omega) = \int_{\mathbb{R}^d} \phi(y)\mu(\omega, dy)$ as(P).

Inequalities & Miscellanea

For positive a, b, p, $(a + b)^p \leq 2^p(a^p + b^p)$.

Markov's ineq: if X is a nonneg r.v. and $a > 0$, then $P(X \geq a) \leq \mathrm{E}(X)/a$.

Hölder's ineq: If $1/p + 1/q = 1$, then for mble f, g, $\|fg\|_1 \leq \|f\|_p \|g\|_q$.

Jensen's ineq: $\forall r \in (0, q)$, $\phi(x) = x^{q/r}$ is convex $\implies \left[\mathrm{E}\,|X|^q\right]^{1/q} \geq \left[\mathrm{E}\,|X|^r\right]^{1/r}$.

Bayes rule: $P(X \in A \mid Y \in B) = \frac{P(Y \in B \mid X \in A)P(X \in A)}{\sum_x P(Y \in B \mid X = x)P(X = x)}$.