## Laws of Large Numbers

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

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

**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

$$\frac{1}{b_n} \sum_{j=1}^{n} a_n \to 0 \text{ as } n \to \infty.$$

**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 4.14**: If $\{X_n\}$ ind such that $\sum_{n=1}^{\infty} \operatorname{E}|X_n|^{\alpha_n}/n^{\alpha_n} < \infty$ for $\alpha_n \in [1,2]$, then

$$\frac{S_n - \operatorname{E} S_n}{n} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \operatorname{E} X_i) \xrightarrow{\text{a.s.}} 0.$$

**Marcinkiewicz-Zygmund SLLN**: Let $\{X_n\}$ be iid, $S_n = \sum_{j=1}^{n} X_n$, and $p \in (0,2)$.

1. If $\frac{S_n - nc}{n^{1/p}} \xrightarrow{\text{a.s.}} 0$ for some $c \in \mathbb{R}$, then $\operatorname{E}|X_1|^p < \infty$.
2. If $\operatorname{E}|X_1|^p < \infty$, then (2) holds with $c = \operatorname{E} X_1$ if $p \in [1,2)$ and (2) holds for any $c \in \mathbb{R}$ if $p \in (0,1)$.

**Kolmogorov's SLLN**: If $\{X_n\}$ are iid, then

$$\bar{X}_n = \frac{S_n}{n} \xrightarrow{\text{a.s.}} \operatorname{E} X_1 \iff \operatorname{E}|X_1| < \infty \iff \frac{S_n - n \operatorname{E} X_1}{n} \xrightarrow{\text{a.s.}} 0.$$

**Useful Theorem**: For any r.v. $X$ and $r > 0$,

$$\sum_{n=1}^{\infty} P(|X| > n^{1/r}) \le \operatorname{E}|X|^r \le \sum_{n=0}^{\infty} P(|X| > n^{1/r}).$$

**Etemaldi's SLLN**: If $\{X_n\}$ are *pairwise* ind and identically distributed, then

$$\bar{X}_n = \frac{S_n}{n} \xrightarrow{\text{a.s.}} \operatorname{E} X_1 \iff \operatorname{E}|X_1| < \infty.$$

**Theorem 4.18** (general WLLN): $\{X_n\}$ ind and put $S_n = \sum_{j=1}^{n} X_j$. If

$$\sum_{j=1}^{n} P(|X_j| > n) \to 0 \quad \text{and} \quad \frac{1}{n^2} \sum_{j=1}^{n} EX_j^{(n)2} \to 0,$$

then

$$\frac{S_n - a_n}{n} \xrightarrow{\text{p}} 0,$$

where $a_n = \sum_{j=1}^{n} \operatorname{E} X_j^{(n)}$ and $X_j^{(n)} \equiv X_j I(|X_j| \le n)$.

**[Corollary] Feller's WLLN** (without a 1st moment hypothesis): if $\{X_n\}$ iid with $\lim_{n\to\infty} xP(|X_1| > x) = 0$, then

$$\frac{S_n}{n} - \operatorname{E} X_1^{(n)} \xrightarrow{\text{p}} 0.$$

## Empirical Distributions

The **empirical cdf** of $X_1, \ldots, X_n$ is the random cdf is the proportion of observations no larger than a fixed $x$:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x), \quad x \in \mathbb{R}.$$

- With $X_i$'s on $(\Omega, \mathcal{F}, P)$, for each $\omega \in \Omega$,
$$F_n(x, \omega) = \frac{1}{n} \sum_{i=1}^{n} I(X_i(\omega) \le x).$$
- $F_n(x)$ is a right-continuous, nondecreasing function of $x \in \mathbb{R}$,
- 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**: $\{X_n\}$ iid with cdf $F(\cdot)$. Let $F_n(\cdot)$ be the empirical cdf based on $X_1, \ldots, X_n$ and define

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F(x)|.$$

Then, (i) $D_n$ is a r.v. for any $n \ge 1$ and (ii) $D_n \xrightarrow{\text{a.s.}} 0$ as $n \to \infty$.

**Quantile function**: $\phi(u) = \inf\{x \in \mathbb{R} : F(x) \ge u\} \equiv F^{-1}(u),\ u \in (0,1)$ which implies

$$F(x) \ge u \iff x \ge \phi(u) \quad \text{and} \quad F(\phi(u)-) \le u \le F(\phi(u)).$$

## Convergence in Distribution

: Let $\mu_n, n \ge 0$ be probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ for some $1 \le k < \infty$.

1. For $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$, the **cdf** of $\mu_n$ is
$$F_n(\mathbf{x}) = \mu_n((-\infty, x_1] \times \cdots \times (-\infty, x_n]).$$
If a random vector $X_n$ has probability distribution $\mu_n$ [i.e., $P(X_n \in A) = \mu_n(A),\ A \in \mathcal{B}(\mathbb{R}^k)$], then $F_n$ is also called the cdf of $X_n$.
2. A sequence of probability measures $\mu_n$ (or corresponding cdfs $F_n$) **converges weakly** to $\mu_o$ (to $F_0$), denoted as $\mu_n \Rightarrow \mu_0$ (or as $F_n \Rightarrow F_0$), if
$$\lim_{n\to\infty} F_n(\mathbf{x}) = F_0(\mathbf{x}) \quad \forall \mathbf{x} \in C(F_0),$$
where $C(F_0) = \left\{\mathbf{x} \in \mathbb{R}^k : F_0 \text{ is continuous at } \mathbf{x}\right\}$.
3. A sequence of random vectors $X_n$ in $R^k$ (with distributions $\mu_n$) **converges in distribution (law)** to a random variable $X_0$ (with distribution $\mu_0$) if $\mu_n \Rightarrow \mu_0$, denoted by $X_n \xrightarrow{\text{d}} X_0$.
That is, if $X_n = (X_{n,1}, \ldots, X_{n,i})$ has cdf $F_n, n \ge 0$, then
$$\begin{aligned}\lim_{n\to\infty} F_n(\mathbf{x}) &= \lim_{n\to\infty} P(X_{n,1} \le x_1, \ldots, X_{n,k} \le x_k) \\ &= P(X_{0,1} \le x_1, \ldots, X_{0,k} \le x_k) \\ &= F_0(\mathbf{x}) \quad \forall \mathbf{x} \in C(F_0).\end{aligned}$$
Note that
   - $\mathbf{x} = (x_1, \ldots, x_k) \in C(F_0) \iff F_0(\mathbf{x}) = P(X_{0,1} < x_1, \ldots, X_{0,k} < x_k) = F_0(\mathbf{x}-)$ i.e., if also left continuous.
   - $C(F_0)^c$ is at most countable.
   - $X_n \xrightarrow{\text{p}} X_0 \implies X_n \xrightarrow{\text{d}} X_0$ but not the other direction, unless $X_0$ is degenerate.

**Skorohod's Embedding Theorem**: If $\mu_n, n \ge 0$ are probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ for some $1 \le k < \infty$ such that $\mu_n \Rightarrow \mu_0$, then $\exists$ random vectors $\{Y_n\}_{n\ge 0}$ on a *common* probability space such that $Y_n$ has probability distribution $\mu_n$ for all $n \ge 0$ and $Y_n \xrightarrow{\text{a.s.}} Y_0$. That is, $P(Y_n \in A) = \mu_n(A),\ A \in \mathcal{B}(\mathbb{R}^k), n \ge 0$.

**Continuous Mapping Theorem**:

Version (a): Let $\mu_n, n \ge 0$ be probability measures on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ for $1 \le k < \infty$ and let $h : \mathbb{R}^k \to \mathbb{R}^m$ for $1 \le m < \infty$ be a $\langle \mathcal{B}(\mathbb{R}^k), \mathcal{B}(\mathbb{R}^m) \rangle$-mble function such that $\mu_0(D_h) = 0$, where $D_n \in \mathcal{B}(\mathbb{R}^k)$ denotes the set of all points of discontinuities of the function $h$. If $\mu_n \Rightarrow \mu_0$, then the induced measures converge weakly:

$$\mu_n h^{-1} \Rightarrow \mu_0 h^{-1}.$$

Version (b): Let $X_n, n \ge 0$ be $R^k$-valued random vectors and let measurable $h : \mathbb{R}^k \to \mathbb{R}^m$ 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

$$h(X_n) \xrightarrow{\text{d}} h(X_0).$$

Corollary: If $X_n, Y_n, n \ge 0$ be r.v.'s such that $(X_n, Y_n) \xrightarrow{\text{d}} (X_0, Y_0)$, then

$$X_n + Y_n \xrightarrow{\text{d}} X_0 + Y_0, \quad X_n Y_n \xrightarrow{\text{d}} X_0 Y_0, \quad X_n/Y_n \xrightarrow{\text{d}} X_0/Y_0 \text{ if } P(Y_0 = 0) = 0$$

Corollary (Slutsky's Theorem): If $X_n, Y_n, n \ge 1$ be r.v.'s such that $X_n \xrightarrow{\text{d}} X$ and $Y_n \xrightarrow{\text{p}} a$ for some $a \in \mathbb{R}$, then

$$X_n + Y_n \xrightarrow{\text{d}} X + a, \quad X_n Y_n \xrightarrow{\text{d}} aX, \quad X_n/Y_n \xrightarrow{\text{d}} X/a \text{ if } a \ne 0.$$

## Characterizations of Convergence in Distribution

For a probability measure $\mu$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, a set $A \in \mathcal{B}(\mathbb{R}^k)$ is called a $\mu$-**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 "Portmanteau" Theorem**: If $\mu_n, n \ge 0$ are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then

1. $\mu_n \Rightarrow \mu_0 \iff \mu_n(A) \to \mu_0(A)\ \forall A \in \mathcal{B}(\mathbb{R}) \ni \mu_0(\partial A) = 0$.
2. $\mu_n \Rightarrow \mu_0 \iff \int f d\mu_n \to \int f d\mu_0$ for all bounded continuous functions $f : \mathbb{R} \to \mathbb{R}$.

Remarks:

- $\mu_n \Rightarrow \mu_0 \not\Longleftarrow \mu_n(A) \to \mu_0(A) \forall A \in \mathcal{B}(\mathbb{R}^k)$.
- Holds for $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$.
- Can generalize to a metric space $(S, d)$.

**Lemma 5.8**: 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 function with $\mu_0(D_f)$ (where $D_f \in \mathcal{B}(\mathbb{R})$ is the set of discontinuity points of $f$), then

$$\int f d\mu_n \to \int f d\mu_0 \text{ as } n \to \infty.$$

A sequence of probability measures $\{\mu_n\}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is **tight** if $\forall \varepsilon > 0, \exists M_\varepsilon > 0$ such that

$$\sup_{n\ge 1} \mu_n\left(\left\{x \in \mathbb{R}^k : \|x\| > M_\varepsilon\right\}\right) < \varepsilon.$$

For a single probability measure on $\mathbb{R}$, given $\varepsilon$, we can find $M_\varepsilon$ such that $\mu([-M_\varepsilon, M_\varepsilon]) < \varepsilon$ and note $\mu([-M_\varepsilon, M_\varepsilon]) \uparrow \mu(\mathbb{R}) = 1$.

A sequence of $R^k$-valued random vectors $\{X_n\}$ is **tight** or **stochastically bounded** if their corresponding $\{\mu_n\}$ is tight. That is, $\forall \varepsilon > 0, \exists M_\varepsilon > 0$ such that

$$\sup_{n\ge 1} P(\|X_n\| > M_\varepsilon) = \sup_{n\ge 1} \mu_n\left(\left\{x \in \mathbb{R}^k : \|x\| > M_\varepsilon\right\}\right) < \varepsilon,$$

where $\mu_n(A) = P(X_n \in A),\ A \in \mathcal{B}(\mathbb{R}^k)$.

A sequence of random vectors $\{X_n\}$ is **uniformly integrable** if $\forall \varepsilon > 0, \exists t_\varepsilon > 0$ such that

$$\sup_{n\ge 1} \operatorname{E}\|X_n\| I(\|X_n\| > t_\varepsilon) = \sup_{n\ge 1} \int_{\|x\| > t_\varepsilon} \|x\| d\mu_n < \varepsilon,$$

where $\mu_n(A) = P(X_n \in A),\ A \in \mathcal{B}(\mathbb{R}^k)$.

**Proposition 5.9**: Let $\{X_n\}$ be r.v.'s.

1. If $X_n \xrightarrow{\text{d}} X_0$, then $\{X_n\}$ is tight.
2. If $\{X_n\}$ is tight and $Y_n \xrightarrow{\text{p}} 0$ for $X_n, Y_n$ defined on $(\Omega_n, \mathcal{F}_n, P_n)$, then $X_n Y_n \xrightarrow{\text{p}} 0$.
3. But, weak convergence "almost" implies tightness - see Prokhorov's theorem.

**Theorem 5.10**: A sequence of r.v.'s $\{X_n\}$ (or probability measures $\{\mu_n\}$) is tight iff for any subsequence $X_{n_k}$ of $X_n$ there exists a further subsequence $X_{n_{k_j}}$ of $X_{n_k}$ and a r.v. (or probability measure $\mu_0$) such that $X_{n_{k_j}} \xrightarrow{\text{d}} X_0$ (or $\mu_{n_{k_j}} \Rightarrow \mu_0$). Note: $X_0$ (or $\mu_0$) depends on the particular subsequence $X_{n_k}$.

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

**Theorem 5.12** (conv in dist + UI $\implies$ conv in mean): If $\{X_n\}, n \ge 1$ is UI and $X_n \xrightarrow{\text{d}} X_0$, then $\operatorname{E}|X_0| < \infty$ and $\operatorname{E} X_n \to \operatorname{E} X_0$.

**Corollary 5.13** If $X_n \xrightarrow{\text{d}} X_0$ and $\sup_{n\ge 1} \operatorname{E}|X_n|^{r+\delta} < \infty$ for some integer $r \ge 1$ and real $\delta > 0$, then $\operatorname{E}|X_0|^r < \infty$ and $\operatorname{E} X_n^r \to \operatorname{E} X_0^r$ (recall that $\sup_{n\ge 1} \operatorname{E}|Z_n|^{1+\delta} \implies \{Z_n\}$ is UI).

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

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

$$\operatorname{E} X^r = \frac{d^r}{dt^r} M_X(t)\Big|_{t=0}.$$

## Characteristic Functions

A **complex** number is $a + ib$, where $a, b \in \mathbb{R}$ and $i = \sqrt{-1}$. 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 $|a + bi| = \sqrt{a^2 + b^2} = \sqrt{(a + bi)(a - bi)}$. IMPORTANT! For any $b \in \mathbb{R}$,

$$e^{bi} = \cos(b) + i\sin(b)$$

and $\left|e^{bi}\right| = \sqrt{\cos^2(b) + \sin^2(b)} = 1$ and $e^{ai}e^{bi} = e^{(a+b)i}$ for $a, b \in \mathbb{R}$. ALSO, for fixed $b \in \mathbb{R}$, the function $g(t) = e^{tbi} : \mathbb{R} \to \mathbb{C}$ is infinitely differentiable in $t$ with $n$th derivative $(bi)^n e^{tbi}$.

For a random vector $X$ in $\mathbb{R}^k$, the **characteristic function** (CF) is defined as

$$\phi_X(t) = \mathrm{E}\, e^{it'X} = \mathrm{E}\cos(t'X) + i\,\mathrm{E}\sin(t'X), \quad t \in \mathbb{R}^k, i = \sqrt{-1}.$$

It allows for easy convolutions (same as MGF), always exists, uniquely identifies distribution, and p.w. convergence implies weak convergence.

Note that $\phi_X(0) = 1$ and $\phi_X(t)$ is uniformly continuous on $\mathbb{R}^k$: by the BCT,

$$\begin{aligned}
\sup_{t \in \mathbb{R}^k} |\phi_X(t + h) - \phi_X(t)| &= \sup_{t \in \mathbb{R}^k} \left|\mathrm{E}\, e^{i(t+h)'X} - \mathrm{E}\, e^{it'X}\right| \\
&\le \mathrm{E}\left|e^{ih'X-1}\right| \\
&= \int_{R^k} \left|e^{ih'x} - 1\right| d\mu_X(x) \to 0 \text{ as } |h| \to 0.
\end{aligned}$$

**Theorem 5.15**: If $X$ is a r.v. with $\mathrm{E}\,|X|^r < \infty$ for some $r \ge 1$, then $\phi_X(t)$ is $r$-times differentiable on $\mathbb{R}$ and

$$\phi_X^{(r)}(t) = \mathrm{E}(iX)^r e^{itX}, \quad t \in \mathbb{R}.$$

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

**Levy Inversion Formula** (use CF to recover dist): If $X$ is a r.v. with CF $\phi_X$, then for any $a, b \in \mathbb{R}$ with $P(X = a) = 0 = P(X = b)$,

$$P(a < X \le b) = \lim_{T\to\infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t)dt.$$

**Corollary**: If we also assume $\int |\phi_X(t)|\, dt < \infty$ (which implies $C(F) = \mathbb{R}$), then $X$ has a pdf $f$ w.r.t. the Lebesgue measure on $\mathbb{R}$ given by

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \phi_X(t)dt.$$

**Levy Continuity Theorem**: suppose $\{X_n\}$ is a sequence of r.v.'s each with CF $\phi_{X_n}$

1. If $X_n \xrightarrow{\text{d}} X_o$, then for any $T > 0$,
$$\sup_{|t|<T} \left|\phi_{X_n}(t) - \phi_{X_0}(t)\right| \to 0 \text{ as } n \to \infty.$$
2. If $\phi_{X_n}(t) \to g(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$ and $g(\cdot)$ is continuous at zero, then $g(\cdot)$ is a CF and $X_n \xrightarrow{\text{d}} X_0$, where $X_0$ is the r.v. with CF $g(\cdot)$.

**Corollary 5.20**: $X_n \xrightarrow{\text{d}} X_0 \iff \phi_{X_n}(t) \to \phi_{X_0}(t)$ as $n \to \infty$ for all $t \in \mathbb{R}$.

**Inversion Formula in $\mathbb{R}^k$**: Let $X$ be a $\mathbb{R}^k$-valued random vector with CF $\phi_X(t)$ for $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$. Then, for any 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 $\int_{\mathbb{R}^k} |\phi_X(t_1, \ldots, t_k)|\, dt_1 \ldots dt_k < \infty$, then $X$ has a bounded, continuous density $f_X(x)$ w.r.t. the Lebesgue measure in $\mathbb{R}^k$ given by

$$f_X(x) = \frac{1}{(2\pi)^k} \int_{\mathbb{R}^k} e^{-i\sum_{j=1}^{k} x_j t_j} \phi_X(t_1, \ldots, t_k)dt_1 \cdots dt_k, \quad x = (x_1, \ldots, x_k) \in \mathbb{R}^k.$$

**Theorem 5.22**: On a psp $(\Omega, \mathcal{F}, P)$, r.v.'s $X_1, \ldots, X_k$ are ind iff for all $t_1, \ldots, t_k \in \mathbb{R}$,

$$\phi_{X_1,\ldots,X_k}(t_1, \ldots, t_k) \equiv \mathrm{E}\, e^{i\sum_{j=1}^{k} X_j t_j} = \prod_{j=1}^{k} \mathrm{E}\, e^{iX_j t_j} = \prod_{j=1}^{k} \phi_{X_j}(t_j).$$

**Theorem 5.23**: For a sequence $\{X_n\}, n \ge 0$ of $\mathbb{R}^k$-valued random vectors,

1. $X_n \xrightarrow{\text{d}} X_0 \iff \phi_{X_n}(t) \to \phi_{X_0}(t)\ \forall t \in \mathbb{R}^k$,
2. (Cramer-Wold device) $X_n \xrightarrow{\text{d}} X_0 \iff t'X_n \xrightarrow{\text{d}} t'X_0\ \forall t \in \mathbb{R}^k$,