## 3.2. CHARACTERISTIC AND DISTRIBUTION FUNCTIONS

Since probability distribution functions are uniquely characterized by their spectra (characteristic functions), it is reasonable to try and decide the extent to which the convergence of probability distribution functions, discussed in section 2.4, might be determined by the convergence properties of their corresponding characteristic functions. This line of reasoning turns out to be fruitful: the concept of *convergence in distribution can be translated into the easier property of pointwise convergence for characteristic functions*. This is the content of the next result, which will lead us ultimately to a proof of the Central Limit Theorem 2.5.1 in the next section.

**2.1 THEOREM : Convergence in Distribution via Characteristic Functions.** *Let $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ be a sequence of probability distribution functions, and let $\{\phi_n(\cdot)\}_{n\in\mathbf{N}}$ be the corresponding characteristic functions.*

(i) *If $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ converges to a probability distribution function $F(\cdot)$ at all continuity points $x$ of $F(\cdot)$, and if $\phi(\cdot)$ is the characteristic function of $F(\cdot)$, then $\phi_n(\cdot) \to \phi(\cdot)$ pointwise (i.e., $\lim_{n\to\infty} \phi_n(\xi) = \phi(\xi)$ for all $\xi \in \mathbf{R}$);*

(ii) *Conversely, suppose $\phi_n(\cdot) \to \phi(\cdot)$ pointwise and $\phi(\cdot)$ is continuous at $\xi = 0$; then there exists a probability distribution function $F(\cdot)$ with $\phi(\cdot)$ as its characteristic function, and $\lim_{n\to\infty} F_n(x) = F(x)$ at all continuity points $x$ of $F(\cdot)$.*

Part (i) is an immediate consequence of Theorem 2.4.3, since we can take $\Phi(y) = e^{i\xi y}$; then $\int \Phi \, dF$ is the characteristic function $\phi(\cdot)$ corresponding to $F(\cdot)$, evaluated at $\xi \in \mathbf{R}$ (the "harmonics" at "frequency" $\xi \in \mathbf{R}$). To prove (ii) we need a basic lemma, which says essentially that the space of distribution functions is "precompact".

**2.1 Lemma : Helly-Bray.** *Let $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ be any sequence of probability distribution functions. Then there exists a subsequence $\{F_{n_k}(\cdot)\}_{k\in\mathbf{N}}$ and a right-continuous, increasing function $F(\cdot)$, such that*

$$\lim_{k\to\infty} F_{n_k}(x) = F(x) \quad \textit{holds at all continuity points } x \textit{ of } F(\cdot)\,.$$

Note that the Helly-Bray Lemma does not guarantee $F(-\infty) = 0$ or $F(\infty) = 1$, so the function $F(\cdot)$ need not be a probability distribution function. For instance, if we define $F_n(x) := 1$, $x \geq n$ and $F_n(x) := 0$, $x < n$ for all $n \in \mathbf{N}$, then $\lim_{n\to\infty} F_n(x) = F(x) \equiv 0$, $\forall\, x \in \mathbf{R}$: all (probability) mass "leaks out at infinity".

*PROOF* : Let $\mathbf{D} = \{x_1, x_2, \cdots\}$ be a countable, dense set in $\mathbf{R}$. We can invoke the familiar **"diagonalization argument"**, to show that $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ *contains a subsequence* *that converges at every point of* $\mathbf{D}$. More precisely, since $0 \le F_n(x_1) \le 1$, we can select a convergent subsequence $F_{n_k}(\cdot) \equiv F_k^{(1)}(\cdot),\ k \in \mathbf{N}$ so that $\{F_k^{(1)}(x_1)\}_{k\in\mathbf{N}}$ converges. Repeating the argument with $\{F_n^{(1)}(\cdot)\}_{n\in\mathbf{N}}$, we see that we can select a subsequence, call it $F_{n_k}^{(1)}(\cdot) \equiv F_k^{(2)}(\cdot),\ k \in \mathbf{N}$, so that $\{F_k^{(2)}(x_2)\}_{k\in\mathbf{N}}$ converges. We arrive this way at a nested family

$$\{F_n^{(1)}(\cdot)\}_{n\in\mathbf{N}} \supseteq \{F_n^{(2)}(\cdot)\}_{n\in\mathbf{N}} \supseteq \cdots \supseteq \{F_n^{(m)}(\cdot)\}_{n\in\mathbf{N}} \supseteq \cdots \tag{2.1}$$

of subsequences, with the property that $\{F_n^{(m)}(x_m)\}_{n\in\mathbf{N}}$ converges at each $x_m \in \mathbf{D}$. Then the subsequence of functions $\{F_n^{(n)}(\cdot)\}_{n\in\mathbf{N}}$ satisfies the desired condition.

Indeed, for every $x_m \in \mathbf{D}$, we have that $\{F_n^{(n)}(x_m)\}_{n\ge m}$ is a subsequence of $\{F_n^{(m)}(x_m)\}_{n\ge m}$, and thus converges. We can introduce already a preliminary candidate $\widetilde{F}(\cdot)$, by setting

$$\widetilde{F}(x) := \lim_{n\to\infty} F_n^{(n)}(x) \ \text{ for } \ x \in \mathbf{D}, \qquad \widetilde{F}(x) := \sup_{\substack{x_k \le x \\ x_k \in \mathbf{D}}} \widetilde{F}(x_k) \ \text{ for } \ x \notin \mathbf{D}. \tag{2.2}$$

The resulting function $\widetilde{F}(\cdot)$ clearly takes values in $[0,1]$, and is easily seen to be increasing.

We show now that $\lim_{n\to\infty} F_n^{(n)}(x) = \widetilde{F}(x)$ holds *for every* $x$ *in the set* $\mathcal{C}$ *of continuity points of* $\widetilde{F}(\cdot)$. Choose $y, \xi \in \mathbf{D}$ with $y < x < \xi$, so that $F_n^{(n)}(y) \le F_n^{(n)}(x) \le F_n^{(n)}(\xi)$, and let $n \to \infty$ to obtain

$$\widetilde{F}(y) \le \liminf_{n\to\infty} F_n^{(n)}(x) \le \limsup_{n\to\infty} F_n^{(n)}(x) \le \widetilde{F}(\xi)\,. \tag{2.3}$$

Since $\widetilde{F}(\cdot)$ is continuous at $x$, we can let $y \in \mathbf{D}$, $\xi \in \mathbf{D}$ tend to $x$ (from below and above, respectively), to get $\lim_n F_n^{(n)}(x) = \widetilde{F}(x)$, as asserted.

Finally, the desired function $F(\cdot)$ can be constructed by setting $F(x) := \widetilde{F}(x)$ for $x \in \mathcal{C}$, and $F(x) := \lim_{y\downarrow x,\ y\in\mathcal{C}} \widetilde{F}(y)$ otherwise. This $F(\cdot)$ is still increasing, and is now right-continuous. Furthermore, the points of discontinuity of $F(\cdot)$ are the same as those of $\widetilde{F}(\cdot)$. $\diamondsuit$

The key additional property of the sequence $\{F_n(\cdot)\}_{n\in\mathbf{N}}$, which will ensure both $F(-\infty) = 0$ and $F(\infty) = 1$ in (and thus prevents the kind of "leakage of probability mass" we witnessed in the example immediately following) Lemma 2.1, is *tightness*.

More precisely, a sequence of probability measures $\{\mu_n\}_{n\in\mathbf{N}}$ is said to be **tight**, if for every $\varepsilon > 0$ there exists $K \in (0,\infty)$ such that

$$\mu_n((-K,K]) > 1-\varepsilon \quad \text{holds for all } \ n \in \mathbf{N}\,. \tag{2.4}$$

If $\{\mu_n\}$ are the Lebesgue-Stieltjes measures induced by the probability distribution functions $\{F_n(\cdot)\}$, then (2.4) is clearly equivalent to

$$F_n(K) - F_n(-K) \ > \ 1-\varepsilon \quad \text{holds for all } \ n \in \mathbf{N}\,. \tag{2.4$'$}$$

A sequence of probability distribution functions $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ for which (2.4)′ holds, will also be called **tight**.

**2.2 Lemma : Tightness.** *Let $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ be a sequence of probability distribution functions, and let $\{\phi_n(\cdot)\}_{n\in\mathbf{N}}$ be the corresponding characteristic functions.*

(i) *Suppose there exists an increasing function $F : \mathbf{R} \to \mathbf{R}$, for which $\lim_{n\to\infty} F_n(x) = F(x)$ holds at all continuity points $x$ of $F(\cdot)$.*

*Then $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ is tight, if and only if $F(-\infty) = 0$ and $F(\infty) = 1$.*

(ii) *Suppose that $\{\phi_n(\cdot)\}_{n\in\mathbf{N}}$ converges pointwise to a function $\phi(\cdot)$ which is continuous at the origin $\xi = 0$; then the sequence $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ is tight.*

*Proof* : (i) Assume first that $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ is tight. For any $\varepsilon > 0$, let $K$ be so large that $F_n(K) - F_n(-K) > 1 - \varepsilon$ holds for all $n \in \mathbf{N}$, and take an arbitrary point of continuity $x \geq K$ of $F(\cdot)$. Then $F_n(x) - F_n(-x) > 1 - \varepsilon$, hence $F_n(x) \geq F_n(K) > 1 - \varepsilon$. We may now let $x \to \infty$ and invoke the arbitrariness of $\varepsilon > 0$, to obtain $\lim_{x\to\infty} F(x) = 1$. Similarly, $F(-x) \leq \varepsilon$, whence $\lim_{x\to-\infty} F(x) = 0$.

Conversely, assume that $F(-\infty) = 0$, $F(\infty) = 1$. Suppose $\{F_n(\cdot)\}_{n\in\mathbf{N}}$ is not tight; then there exist an $\varepsilon > 0$ and a subsequence $\{F_{n_k}(\cdot)\}_{k\in\mathbf{N}}$, so that $F_{n_k}(k) - F_{n_k}(-k) \leq 1 - \varepsilon$ holds for all $k \in \mathbf{N}$. Let $a, b$ be *any* two points of continuity of $F(\cdot)$ with $a < b$. For $k$ large enough, we have $(a, b] \subset (-k, k]$ and thus $F_{n_k}(b) - F_{n_k}(a) \leq 1 - \varepsilon$. Letting $k \to \infty$ gives $F(b) - F(a) \leq 1 - \varepsilon$, and hence $F(\infty) - F(-\infty) \leq 1 - \varepsilon$ by letting $b \to \infty$, $a \to -\infty$ through the set of continuity points of $F(\cdot)$. This is a contradiction, and (i) is proved.

• To prove (ii), we observe that the function $\xi \mapsto \phi(\xi) + \phi(-\xi)$ is real-valued, takes the value 2 at $\xi = 0$, and satisfies $|\phi(\xi) + \phi(-\xi)| \leq 2$. The continuity of $\phi(\cdot)$ at $\xi = 0$ implies that for any $\varepsilon > 0$, there is a $\delta > 0$ so that

$$0 \leq \frac{1}{\delta} \int_0^\delta \left[ \left(2 - (\phi(\xi) + \phi(-\xi)\right) \right] d\xi \leq \varepsilon.$$

In view of the Lebesgue Dominated Convergence Theorem, this implies

$$\frac{1}{\delta} \int_0^\delta \left[ \left(2 - (\phi_n(\xi) + \phi_n(-\xi)\right) \right] d\xi \leq 2\,\varepsilon, \quad \text{for all large enough } n \in \mathbf{N}.$$

If we write $\phi_n(\cdot)$ in terms of the measure $\mu_n$ induced by $F_n(\cdot)$ and apply Tonelli's Theorem, we obtain the inequality

$$2\varepsilon \geq \frac{2}{\delta} \int_{\mathbf{R}} \left( \int_0^\delta \left(1 - \cos(\xi x)\right) d\xi \right) d\mu(x) = 2 \int_{\mathbf{R}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) d\mu_n(x).$$

The integrand $\left(1 - \frac{\sin(\delta x)}{\delta x}\right)$ is positive because $|\sin(u)| \le |u|$ for all $u \in \mathbf{R}$, and we have the bounds

$$\varepsilon \ge \int_{\{\delta|x|>2\}} \left(1 - \frac{\sin(\delta x)}{\delta x}\right) d\mu_n(x) \ge \int_{\{\delta|x|>2\}} \left(1 - \frac{1}{|\delta x|}\right) d\mu_n(x) \ge \frac{1}{2} \cdot \mu_n\big(\{\delta|x| > 2\}\big)$$

or equivalently $\mu_n\left(\left[-\frac{2}{\delta}, \frac{2}{\delta}\right]\right) \ge 1 - 2\varepsilon$, and (ii) is proved. $\diamondsuit$

*PROOF OF THEOREM 2.1* : We have noted already that part (i) is an immediate consequence of Theorem 2.4.3. For part (ii) note that, by the Helly-Bray Lemma, we can find a subsequence $\{F_{n_k}(\cdot)\}_{k\in\mathbf{N}}$ that converges to an increasing, right-continuous function $F(\cdot)$ at all its continuity points. Of course, the corresponding characteristic functions $\{\phi_{n_k}(\cdot)\}_{k\in\mathbf{N}}$ converge pointwise to $\phi(\cdot)$, which is continuous at the origin $\xi = 0$.

From the Tightness Lemma 2.2(ii), the sequence $\{F_{n_k}(\cdot)\}_{k\in\mathbf{N}}$ is tight; and Lemma 2.2(i) implies that $F(\cdot)$ is actually a probability distribution function. By part (i) of the Theorem, the sequence $\{\phi_{n_k}(\xi)\}_{k\in\mathbf{N}}$ converges for each $\xi \in \mathbf{R}$ to the value of the characteristic function of $F(\cdot)$ at $\xi$, which must then coincide with $\phi(\xi)$.

- Our last task is to show that the *entire sequence* $\{F_n(x)\}_{n\in\mathbf{N}}$ *converges to* $F(x)$, for every continuity point $x$ of $F(\cdot)$.

Assume otherwise. Then there exists a subsequence $\{F_{m_k}(\cdot)\}_{k\in\mathbf{N}} \subseteq \{F_n(\cdot)\}_{n\in\mathbf{N}}$ such that

$$|F_{m_k}(x) - F(x)| > \delta \quad \text{holds for all } k \in \mathbf{N} \text{ large enough,} \tag{2.5}$$

for some $\delta > 0$ and some continuity point $x$ of $F(\cdot)$. But we can apply again the Helly-Bray and Tightness Lemmata, and obtain a subsequence $\{F_{m_{k_\ell}}(\cdot)\}_{\ell\in\mathbf{N}}$ converging to some distribution function $G(\cdot)$ at all its continuity points, with corresponding characteristic functions $\{\phi_{m_{k_\ell}}(\cdot)\}_{\ell\in\mathbf{N}}$ converging to the characteristic function of $G(\cdot)$.

Since $\{\phi_n(\cdot)\}_{n\in\mathbf{N}}$ converges to $\phi(\cdot)$ pointwise, it follows (thanks to part (i) of the Theorem) that $\phi(\cdot)$ is also the characteristic function of $G(\cdot)$. But a distribution function is uniquely determined by its characteristic function (Fourier-Lévy Inversion Theorem 1.1), so $G(\cdot) \equiv F(\cdot)$. Thus $x$ is a continuity point for $G(\cdot)$, and we must have $|F_{m_{k_\ell}}(x) - F(x)| = |F_{m_{k_\ell}}(x) - G(x)| \le \delta$ for $\ell \in \mathbf{N}$ large enough; a contradiction to (2.5). $\diamondsuit$

**2.1 Exercise:** Let $\mu$, $\{\mu_n\}_{n\in\mathbf{N}}$ be probability measures on $\mathcal{B}(\mathbf{R})$ with corresponding characteristic functions $\phi \in \mathbf{L}^1(\mathbf{R})$, $\{\phi_n\}_{n\in\mathbf{N}} \subseteq \mathbf{L}^1(\mathbf{R})$, and assume that $\phi_n \longrightarrow \phi$ as $n \to \infty$ in $\mathbf{L}^1(\mathbf{R})$. Then the densities $f = d\mu/d\lambda$, $f_n = d\mu_n/d\lambda$ $(n \in \mathbf{N})$ exist and are bounded and continuous on $\mathbf{R}$, and we have $\lim_{n\to\infty} \sup_{x\in\mathbf{R}} |f_n(x) - f(x)| = 0$.

**2.2 Exercise:** If $\phi(\cdot)$ is a characteristic function, then so is $\exp[\lambda(\phi(\cdot) - 1)]$.

(*Hint:* Consider the limit of $\Big(1 + (\lambda\,/\,n)\,(\phi(\cdot) - 1)\Big)^n$ as $n \to \infty$, and recall Exercise 1.2 and Theorem 2.1(ii).

**2.3 Exercise:** If every subsequence of a sequence $\{\mu_n\}_{n\in\mathbf{N}}$ of probability measures converges vaguely to the *same* probability measure $\mu$, then the entire sequence converges vaguely to this measure.