11. Characteristic functions.

Given a random variable X, we define its *characteristic function* (or *Fourier transform*) by

$$\phi_X(t) \;=\; \mathbf{E}(e^{itX}) \;=\; \mathbf{E}[\cos(tX)] + i\,\mathbf{E}[\sin(tX)]\,, \qquad t \in \mathbf{R}\,.$$

The characteristic function is thus a function from the real numbers to the complex numbers. Of course, by the Change of Variable Theorem (Theorem 6.1.1), $\phi_X(t)$ depends only on the *distribution* of X. We sometimes write $\phi_X(t)$ as $\phi(t)$.

The characteristic function is clearly very similar to the moment generating function introduced earlier; the only difference is the appearance of the imaginary number $i = \sqrt{-1}$ in the exponent. However, this change is significant; since $|e^{itX}| = 1$ for any (real) t and X, the triangle inequality implies that $|\phi_X(t)| \le 1 < \infty$ for all t and all random variables X. This is quite a contrast to the case for moment generating functions, which could be infinite for any $s \neq 0$.

Like for moment generating functions, we have $\phi_X(0) = 1$ for any X, and if X and Y are independent then $\phi_{X+Y}(t) = \phi_X(t)\,\phi_Y(t)$ by (4.2.7). We further note that, with $\mu = \mathcal{L}(X)$, we have

$$|\phi_X(t+h) - \phi_X(t)| \;=\; \left|\int \left(e^{i(t+h)x} - e^{itx}\right)\mu(dx)\right|$$

$$\le\; \int \left|e^{i(t+h)x} - e^{itx}\right|\mu(dx) \;=\; \int \left|e^{itx}\right|\,\left|e^{ihx} - 1\right|\mu(dx)$$

$$=\; \int \left|e^{ihx} - 1\right|\mu(dx)\,.$$

Now, as $h \to 0$, this last quantity decreases to 0 by the bounded convergence theorem (since $|e^{ihx} - 1| \le 2$). We conclude that ϕ_X is always a (uniformly) continuous function.

The derivatives of ϕ_X are also straightforward. The following proposition is somewhat similar to the corresponding result for $M_X(s)$ (Theorem 9.3.3), except that here we do not require a severe condition like "$M_X(s) < \infty$ for all $|s| < s_0$".

Proposition 11.0.1. *Suppose X is a random variable with $\mathbf{E}\left(|X|^k\right) < \infty$. Then for $0 \le j \le k$, ϕ_X has finite j^{th} derivative, given by $\phi_X^{(j)}(t) = \mathbf{E}\left[(iX)^j e^{itX}\right]$. In particular, $\phi_X^{(j)}(0) = i^j\mathbf{E}(X^j)$.*

Proof. We proceed by induction on j. The case $j = 0$ is trivial. Assume now that the statement is true for $j-1$. For $t \in \mathbf{R}$, let $F_t = (iX)^{j-1}e^{itX}$, so that $|F_t'| = |(iX)^j e^{itX}| = |X|^j$. Since $E(|X|^k) < \infty$, therefore also $\mathbf{E}(|X|^j) < \infty$. It thus follows from Proposition 9.2.1 that

$$\phi_X^{(j)}(t) = \frac{d}{dt}\phi_X^{(j-1)}(t) = \frac{d}{dt}\mathbf{E}[(iX)^{j-1}e^{itX}]$$

$$= \mathbf{E}[\frac{\partial}{\partial t}(iX)^{j-1}e^{itX}] = \mathbf{E}[(iX)^j e^{itX}]. \qquad \blacksquare$$

11.1. The continuity theorem.

In this subsection we shall prove the continuity theorem for characteristic functions (Theorem 11.1.14), which says that if characteristic functions converge pointwise, then the corresponding distributions converge weakly: $\mu_n \Rightarrow \mu$ if and only if $\phi_n(t) \to \phi(t)$ for all t. This is a very important result; for example, it is used to prove the central limit theorem in the next subsection. Unfortunately, the proof is somewhat technical; we must show that characteristic functions completely determine the corresponding distribution (Theorem 11.1.1 and Corollary 11.1.7 below), and must also establish a simple criterion for weak convergence of "tight" measures (Corollary 11.1.11).

We begin with an inversion theorem, which tells how to recover information about a probability distribution from its characteristic function.

Theorem 11.1.1. *(Fourier inversion theorem) Let μ be a Borel probability measure on $\mathbf{R}$, with characteristic function $\phi(t) = \int_{\mathbf{R}} e^{itx}\mu(dx)$. Then if $a < b$ and $\mu\{a\} = \mu\{b\} = 0$, then*

$$\mu([a,b]) = \lim_{T\to\infty} \frac{1}{2\pi}\int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it}\phi(t)dt.$$

To prove Theorem 11.1.1, we use two computational lemmas.

Lemma 11.1.2. *For $T \geq 0$ and $a < b$,*

$$\int_{\mathbf{R}}\int_{-T}^{T}\left|\frac{e^{-ita} - e^{-itb}}{it}e^{itx}\right| dt\,\mu(dx) \;\leq\; 2T(b-a) \;<\; \infty.$$

Proof. We first note by the triangle inequality that

$$\left|\frac{e^{-ita} - e^{-itb}}{it}e^{itx}\right| = \left|\frac{e^{-ita} - e^{-itb}}{it}\right| = \left|\int_a^b e^{-itr}dr\right|$$

$$\leq \int_a^b \left|e^{-itr}\right| dr = \int_a^b 1\, dr = b - a\,.$$

Hence,

$$\int_{\mathbf{R}} \int_{-T}^{T} \left|\frac{e^{-ita} - e^{-itb}}{it} e^{itx}\right| dt\, \mu(dx) \;\leq\; \int_{\mathbf{R}} \int_{-T}^{T} (b-a)\, dt\, \mu(dx)$$

$$= \int_{\mathbf{R}} 2\,T\,(b-a)\,\mu(dx) \;=\; 2\,T\,(b-a)\,. \qquad \blacksquare$$

Lemma 11.1.3. *For $T \geq 0$ and $\theta \in \mathbf{R}$,*

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{\sin(\theta t)}{t} dt \;=\; \pi \operatorname{sign}(\theta)\,, \tag{11.1.4}$$

where $\operatorname{sign}(\theta) = 1$ *for* $\theta > 0$, $\operatorname{sign}(\theta) = -1$ *for* $\theta < 0$, *and* $\operatorname{sign}(0) = 0$. *Furthermore, there is* $M < \infty$ *such that* $|\int_{-T}^{T} [\sin(\theta t)/t]\, dt| \leq M$ *for all* $T \geq 0$ *and* $\theta \in \mathbf{R}$.

Proof (optional). When $\theta = 0$ both sides of (11.1.4) vanish, so assume $\theta \neq 0$. Making the substitution $s = |\theta| t$, $dt = ds/|\theta|$ gives

$$\int_{-T}^{T} \frac{\sin(\theta t)}{t} dt \;=\; \operatorname{sign}(\theta) \int_{-T}^{T} \frac{\sin(|\theta| t)}{t} dt \;=\; \operatorname{sign}(\theta) \int_{-|\theta|T}^{|\theta|T} \frac{\sin s}{s} ds\,, \tag{11.1.5}$$

and hence

$$\lim_{T\to\infty} \int_{-T}^{T} \frac{\sin(\theta s)}{s} ds \;=\; 2 \operatorname{sign}(\theta) \int_0^\infty \frac{\sin s}{s} ds\,. \tag{11.1.6}$$

Furthermore,

$$\int_0^\infty \frac{\sin s}{s} ds \;=\; \int_0^\infty (\sin s) \left(\int_0^\infty e^{-us} du\right) ds$$

$$= \int_0^\infty \left(\int_0^\infty (\sin s) e^{-us} ds\right) du \;=\; \int_0^\infty \left(\int_0^\infty (\sin s) e^{-us} ds\right) du\,.$$

Now, for $u > 0$, integrating by parts twice,

$$I_u \;\equiv\; \int_0^\infty (\sin s) e^{-us} ds \;=\; (-\cos s) e^{-us} \Big|_{s=0}^{\infty} - \int_0^\infty (-\cos s)(-u) e^{-us} ds$$

$$= \; 0 - (-1) - u \left((\sin s) e^{-us} \Big|_{s=0}^{\infty} + \int_0^\infty (\sin s)(-u) e^{-us} ds \right)$$

$$= \; 1 + 0 - 0 - u^2 \int_0^\infty (\sin s) e^{-us} ds\,.$$

Hence, $I_u = 1 - u^2 I_u$, so $I_u = 1/(1+u^2)$.

We then compute that

$$\int_0^\infty \frac{\sin s}{s}\, ds \;=\; \int_0^\infty I_u\, du \;=\; \int_0^\infty \frac{1}{1+u^2}\, du \;=\; \arctan(u)\,\Big|_{u=0}^\infty$$

$$= \; \arctan(\infty) - \arctan(0) \;=\; \pi/2 - 0 \;=\; \pi/2\,.$$

Combining this with (11.1.6) gives (11.1.4).

Finally, since convergent sequences are bounded, it follows from (11.1.4) that the set $\left\{ \int_{-T}^T [\sin(t)/t]\, dt \right\}_{T \ge 0}$ is bounded. It then follows from (11.1.5) that the set $\left\{ \int_{-T}^T [\sin(\theta t)/t]\, dt \right\}_{T \ge 0,\, \theta \in \mathbf{R}}$ is bounded as well. ∎

Proof of Theorem 11.1.1. We compute that

$$\begin{aligned}
&\tfrac{1}{2\pi} \int_{-T}^T \tfrac{e^{-ita} - e^{-itb}}{it}\, \phi(t)\, dt \\
&= \tfrac{1}{2\pi} \int_{-T}^T \tfrac{e^{-ita} - e^{-itb}}{it} \left(\int_{\mathbf{R}} e^{itx} \mu(dx) \right) dt \qquad [\text{by definition of } \phi(t)] \\
&= \tfrac{1}{2\pi} \int_{\mathbf{R}} \int_{-T}^T \tfrac{e^{it(x-a)} - e^{it(x-b)}}{it} dt\; \mu(dx) \qquad [\text{by Fubini and Lemma 11.1.2}] \\
&= \tfrac{1}{2\pi} \int_{\mathbf{R}} \int_{-T}^T \tfrac{\sin(t(x-a)) - \sin(t(x-b))}{t} dt\; \mu(dx) \qquad [\text{since } \tfrac{\cos(ct)}{t} \text{ is odd}]\,.
\end{aligned}$$

Hence, we may use Lemma 11.1.3, together with the bounded convergence theorem and Remark 9.1.9, to conclude that

$$\lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi(t) dt$$

$$= \; \frac{1}{2\pi} \int_{-\infty}^\infty \pi \big[\operatorname{sign}(x-a) - \operatorname{sign}(x-b)\big]\, \mu(dx)$$

$$= \; \int_{-\infty}^\infty \frac{1}{2} \big[\operatorname{sign}(x-a) - \operatorname{sign}(x-b)\big]\, \mu(dx)$$

$$= \frac{1}{2} \mu\{a\} + \mu\left((a,b)\right) + \frac{1}{2} \mu\{b\}\,.$$

(The last equality follows because $(1/2)[\operatorname{sign}(x-a) - \operatorname{sign}(x-b)]$ is equal to 0 if $x < a$ or $x > b$; is equal to $1/2$ if $x = a$ or $x = b$; and is equal to 1 if $a < x < b$.) But if $\mu\{a\} = \mu\{b\} = 0$, then this is precisely equal to $\mu\left([a,b]\right)$, as claimed. ∎

From this theorem easily follows the important

Corollary 11.1.7. *(Fourier uniqueness theorem) Let X and Y be random variables. Then $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbf{R}$ if and only if $\mathcal{L}(X) = \mathcal{L}(Y)$, i.e. if and only if X and Y have the same distribution.*

Proof. Suppose $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbf{R}$. From the theorem, we know that $\mathbf{P}(a \leq X \leq b) = \mathbf{P}(a \leq Y \leq b)$ provided that $\mathbf{P}(X = a) = \mathbf{P}(X = b) = \mathbf{P}(Y = a) = \mathbf{P}(Y = b) = 0$, i.e. for all but countably many choices of a and b. But then by taking limits and using continuity of probabilities, we see that $\mathbf{P}(X \in I) = \mathbf{P}(Y \in I)$ for all intervals $I \subseteq \mathbf{R}$. It then follows from uniqueness of extensions (Proposition 2.5.8) that $\mathcal{L}(X) = \mathcal{L}(Y)$.

Conversely, if $\mathcal{L}(X) = \mathcal{L}(Y)$, then Corollary 6.1.3 implies that $\mathbf{E}(e^{itX}) = \mathbf{E}(e^{itY})$, i.e. $\phi_X(t) = \phi_Y(t)$ for all $t \in \mathbf{R}$. ∎

This last result makes the continuity theorem at least plausible. However, to prove the continuity theorem we require some further results.

Lemma 11.1.8. *(Helly Selection Principle) Let $\{F_n\}$ be a sequence of cumulative distribution functions (i.e. $F_n(x) = \mu_n((-\infty, x])$ for some probability distribution μ_n). Then there is a subsequence $\{F_{n_k}\}$, and a non-decreasing right-continuous function F with $0 \leq F \leq 1$, such that $\lim_{k\to\infty} F_{n_k}(x) = F(x)$ for all $x \in \mathbf{R}$ such that F is continuous at x. [On the other hand, we might not have $\lim_{x\to-\infty} F(x) = 0$ or $\lim_{x\to\infty} F(x) = 1$.]*

Proof. Since the rationals are countable, we can write them as $\mathbf{Q} = \{q_1, q_2, \ldots\}$. Since $0 \leq F_n(q_1) \leq 1$ for all n, the Bolzano-Weierstrass theorem (see page 204) says there is at least one subsequence $\{\ell_k^{(1)}\}$ such that $\lim_{k\to\infty} F_{\ell_k{}^{(1)}}(q_1)$ exists. Then, there is a further subsequence $\{\ell_k^{(2)}\}$ (i.e., $\{\ell_k^{(2)}\}$ is a subsequence of $\{\ell_k^{(1)}\}$) such that $\lim_{k\to\infty} F_{\ell_k{}^{(2)}}(q_2)$ exists (but also $\lim_{k\to\infty} F_{\ell_k{}^{(2)}}(q_1)$ exists, since $\{\ell_k^{(2)}\}$ is a subsequence of $\{\ell_k^{(1)}\}$). Continuing, for each $m \in \mathbf{N}$ there is a further subsequence $\{\ell_k^{(m)}\}$ such that $\lim_{k\to\infty} F_{\ell_k{}^{(m)}}(q_j)$ exists for $j \leq m$.

We now define the subsequence we want by $n_k = \ell_k^{(k)}$, i.e. we take the k^{th} element of the k^{th} subsequence. (This trick is called the *diagonalisation method.*) Since $\{n_k\}$ is a subsequence of $\{\ell_k\}$ from the k^{th} point onwards, this ensures that $\lim_{k\to\infty} F_{n_k}(q) \equiv G(q)$ exists for each $q \in \mathbf{Q}$. Since each F_{n_k} is non-decreasing, therefore G is also non-decreasing.

To continue, we set $F(x) = \inf\{G(q);\, q \in \mathbf{Q},\, q > x\}$. Then F is easily seen to be non-decreasing, with $0 \leq F(x) \leq 1$. Furthermore, F is right-continuous, since if $\{x_n\} \searrow x$ then $\big\{\{q \in \mathbf{Q} : q > x_n\}\big\} \nearrow \{q \in \mathbf{Q} : q > x\}$, and hence $F(x_n) \to F(x)$ as in Exercise 3.6.4. Also, since G is non-decreasing, we have $F(q) \geq G(q)$ for all $q \in \mathbf{Q}$.

Now, if F is continuous at x, then given $\epsilon > 0$ we can find rational numbers r, s, and u with $r < u < x < s$, and with $F(s) - F(r) < \epsilon$. We then note that

$$\begin{aligned}
F(x) - \epsilon &\leq F(r) \\
&= \inf_{q>r} G(q) \\
&= \inf_{q>r} \lim_k F_{n_k}(q) \\
&= \inf_{q>r} \liminf_k F_{n_k}(q) \\
&\leq \liminf_k F_{n_k}(u) \qquad \text{since } u \in \mathbf{Q},\ u > r \\
&\leq \liminf_k F_{n_k}(x) \qquad \text{since } x > u \\
&\leq \limsup_k F_{n_k}(x) \\
&\leq \limsup_k F_{n_k}(s) \qquad \text{since } s > x \\
&= G(s) \\
&\leq F(s) \\
&\leq F(x) + \epsilon\,.
\end{aligned}$$

This is true for any $\epsilon > 0$, hence we must have

$$\liminf_k F_{n_k}(x) = \limsup_k F_{n_k}(x) = F(x)\,,$$

so that $\lim_k F_{n_k}(x) = F(x)$, as claimed. ∎

Unfortunately, Lemma 11.1.8 does not ensure that $\lim_{x\to\infty} F(x) = 1$ or $\lim_{x\to-\infty} F(x) = 0$ (see e.g. Exercise 11.5.1). To rectify this, we require a new notion. We say that a collection $\{\mu_n\}$ of probability measures on $\mathbf{R}$ is *tight* if for all $\epsilon > 0$, there are $a < b$ with $\mu_n([a,b]) \geq 1 - \epsilon$ for all n. That is, all of the measures give most of their mass to the same finite interval; mass does not "escape off to infinity".

Exercise 11.1.9. Prove that:
(a) any *finite* collection of probability measures is tight.
(b) the *union* of two tight collections of probability measures is tight.
(c) any sub-collection of a tight collection is tight.

We then have the following.

Theorem 11.1.10. *If $\{\mu_n\}$ is a tight sequence of probability measures, then there is a subsequence $\{\mu_{n_k}\}$ and a probability measure μ, such that $\mu_{n_k} \Rightarrow \mu$, i.e. $\{\mu_{n_k}\}$ converges weakly to μ.*

Proof. Let $F_n(x) = \mu_n((-\infty, x])$. Then by Lemma 11.1.8, there is a subsequence F_{n_k} and a function F such that $F_{n_k}(x) \to F(x)$ at all continuity points of F. Furthermore $0 \le F \le 1$.

We now claim that F is actually a probability distribution function, i.e. that $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. Indeed, let $\epsilon > 0$. Then using tightness, we can find points $a < b$ which are continuity points of F, such that $\mu_n((a, b]) \ge 1 - \epsilon$ for all n. But then

$$\lim_{x\to\infty} F(x) \; - \; \lim_{x\to-\infty} F(x) \ge F(b) - F(a)$$

$$= \lim_k [F_{n_k}(b) - F_{n_k}(a)] = \lim_k \mu_{n_k}((a, b]) \ge 1 - \epsilon\,.$$

This is true for all $\epsilon > 0$, so we must have $\lim_{x\to\infty} F(x) - \lim_{x\to-\infty} F(x) = 1$, proving the claim.

Hence, F is indeed a probability distribution function. Thus, we can define the probability measure μ by $\mu((a, b]) = F(b) - F(a)$ for $a < b$. Then $\mu_{n_k} \Rightarrow \mu$ by Theorem 10.1.1, and we are done. ■

A main use of this theorem comes from the following corollary.

Corollary 11.1.11. *Let $\{\mu_n\}$ be a tight sequence of probability distributions on $\mathbf{R}$. Suppose that μ is the only possible weak limit of $\{\mu_n\}$, in the sense that whenever $\mu_{n_k} \Rightarrow \nu$ then $\nu = \mu$ (that is, whenever a subsequence of the $\{\mu_n\}$ converges weakly to some probability measure, then that probability measure must be μ). Then $\mu_n \Rightarrow \mu$, i.e. the full sequence converges weakly to μ.*

Proof. If $\mu_n \not\Rightarrow \mu$, then by Theorem 10.1.1, it is *not* the case that $\mu_n(\infty, x] \to \mu(-\infty, x]$ for all $x \in \mathbf{R}$ with $\mu\{x\} = 0$. Hence, we can find $x \in \mathbf{R}$, $\epsilon > 0$, and a subsequence $\{n_k\}$, with $\mu\{x\} = 0$, but with

$$\left|\mu_{n_k}((-\infty, x]) \; - \; \mu((-\infty, x])\right| \; \ge \; \epsilon\,, \qquad k \in \mathbf{N}\,. \tag{11.1.12}$$

On the other hand, $\{\mu_{n_k}\}$ is a subcollection of $\{\mu_n\}$ and hence tight, so by Theorem 11.1.10 there is a further subsequence $\{\mu_{n_{k_j}}\}$ which converges weakly to some probability measure, say ν. But then by hypothesis we must have $\nu = \mu$, which is a contradiction to (11.1.12). ■

Corollary 11.1.11 is nearly the last thing we need to prove the continuity theorem. We require just one further result, concerning a sufficient condition for a sequence of measures to be tight.

Lemma 11.1.13. *Let $\{\mu_n\}$ be a sequence of probability measures on $\mathbf{R}$, with characteristic functions $\phi_n(t) = \int e^{itx}\mu_n(dx)$. Suppose there is a* *function g which is continuous at 0, such that $\lim_n \phi_n(t) = g(t)$ for each $|t| < t_0$ for some $t_0 > 0$. Then $\{\mu_n\}$ is tight.*

Proof. We first note that $g(0) = \lim_n \phi_n(0) = \lim_n 1 = 1$. We then compute that, for $y > 0$,

$$\begin{aligned}
\mu_n\left((-\infty, -\tfrac{2}{y}] \cup [\tfrac{2}{y}, \infty)\right) &= \int_{|x| \ge 2/y} 1\ \mu_n(dx) \\
&\le 2\int_{|x| \ge 2/y} \left(1 - \tfrac{1}{y|x|}\right) \mu_n(dx) \\
&\le 2\int_{|x| \ge 2/y} \left(1 - \tfrac{\sin(yx)}{yx}\right) \mu_n(dx) \\
&= \int_{|x| \ge 2/y} (1/y) \int_{-y}^{y} \left(1 - \cos(tx)\right) dt\ \mu_n(dx) \\
&\le \int_{x \in \mathbf{R}} (1/y) \int_{-y}^{y} \left(1 - \cos(tx)\right) dt\ \mu_n(dx) \\
&= \int_{x \in \mathbf{R}} (1/y) \int_{-y}^{y} (1 - e^{itx})\, dt\ \mu_n(dx) \\
&= \tfrac{1}{y} \int_{-y}^{y} (1 - \phi_n(t))\, dt\,.
\end{aligned}$$

Here the first inequality uses that $1 - \frac{1}{y|x|} \ge \frac{1}{2}$ whenever $|x| \ge 2/y$, the second inequality uses that $\frac{\sin(yx)}{yx} = \frac{\sin(y|x|)}{y|x|} \le \frac{1}{y|x|}$, the second equality uses that $\int_{-y}^{y} \cos(tx)\, dt = \frac{2\sin(yx)}{x}$, the final inequality uses that $1 - \cos(tx) \ge 0$, the third equality uses that $\int_{-y}^{y} \sin(tx)\, dt = 0$, and the final equality uses Fubini's theorem (which is justified since the function is bounded and hence has finite double-integral).

To finish the proof, let $\epsilon > 0$. Since $g(0) = 1$ and g is continuous at 0, we can find y_0 with $0 < y_0 < t_0$ such that $|1 - g(t)| \le \epsilon/4$ whenever $|t| \le y_0$. Then $\left|\frac{1}{y_0} \int_{-y_0}^{y_0} (1 - g(t))dt\right| < \epsilon/2$. Now, $\phi_n(t) \to g(t)$ for all $|t| \le y_0$, and $|\phi_n(t)| \le 1$. Hence, by the bounded convergence theorem, we can find $n_0 \in \mathbf{N}$ such that $\left|\frac{1}{y_0} \int_{-y_0}^{y_0} (1 - \phi_n(t))dt\right| < \epsilon$ for all $n \ge n_0$.

Hence, $\mu_n\left(-\frac{2}{y_0}, \frac{2}{y_0}\right) = 1 - \mu_n\left((-\infty, -\frac{2}{y_0}] \cup [\frac{2}{y_0}, \infty)\right) > 1 - \epsilon$ for all $n \ge n_0$. It follows from the definition that $\{\mu_n\}$ is tight. ∎

We are now, finally, in a position to prove the continuity theorem.

Theorem 11.1.14. *(Continuity Theorem) Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures, with corresponding characteristic functions $\phi, \phi_1, \phi_2, \ldots$. Then $\mu_n \Rightarrow \mu$ if and only if $\phi_n(t) \to \phi(t)$ for all $t \in \mathbf{R}$. In words, the probability measures $\{\mu_n\}$ converge weakly to μ if and only if their characteristic functions converge pointwise to that of μ.*

Proof. First, suppose that $\mu_n \Rightarrow \mu$. Then, since $\cos(tx)$ and $\sin(tx)$ are bounded continuous functions, we have as $n \to \infty$ for each $t \in \mathbf{R}$ that

$$\begin{aligned}
\phi_n(t) &= \int \cos(tx)\mu_n(dx) + i\int \sin(tx)\mu_n(dx) \\
&\to \int \cos(tx)\mu(dx) + i\int \sin(tx)\mu(dx) \\
&= \phi(t)\,.
\end{aligned}$$

Conversely, suppose that $\phi_n(t) \to \phi(t)$ for each $t \in \mathbf{R}$. Then by Lemma 11.1.13 (with $g = \phi$), the $\{\mu_n\}$ are tight. Now, suppose that we have $\mu_{n_k} \Rightarrow \nu$ for some subsequence $\{\mu_{n_k}\}$ and some measure ν. Then, from the previous paragraph we must have $\phi_{n_k}(t) \to \phi_\nu(t)$ for all t, where $\phi_\nu(t) = \int e^{itx}\nu(dx)$. On the other hand, we know that $\phi_{n_k}(t) \to \phi(t)$ for all t; hence, we must have $\phi_\nu = \phi$. But from Fourier uniqueness (Corollary 11.1.7), this implies that $\nu = \mu$.

Hence, we have shown that μ is the only possible weak limit of the $\{\mu_n\}$. Therefore, from Corollary 11.1.11, we must have $\mu_n \Rightarrow \mu$, as claimed. ∎

11.2. The Central Limit Theorem.

Now that we have proved the continuity theorem (Theorem 11.1.14), it is very easy to prove the classical central limit theorem.

First, we compute the characteristic function for the standard normal distribution $N(0,1)$, i.e. for a random variable X having density with respect to Lebesgue measure given by $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. That is, we wish to compute

$$\phi_X(t) \;=\; \int_{-\infty}^{\infty} e^{itx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx\,.$$

Comparing with the computation leading to (9.3.2), we might expect that $\phi_X(t) = M_X(it) = e^{(it)^2/2} = e^{-t^2/2}$. This is in fact correct, and can be justified using theory of complex analysis. But to avoid such technicalities, we instead resort to a trick.

Proposition 11.2.1. *If $X \sim N(0,1)$, then $\phi_X(t) = e^{-t^2/2}$ for all $t \in \mathbf{R}$.*

Proof. By Proposition 9.2.1 (with $F_t = e^{itX}$ and $Y = |X|$, so that $\mathbf{E}(Y) < \infty$ and $|F_t'| = |(iX)e^{itX}| = |X| \le Y$ for all t), we can differentiate under the integral sign, to obtain that

$$\phi_X'(t) \;=\; \int_{-\infty}^{\infty} ix\,e^{itx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \;=\; \int_{-\infty}^{\infty} i\,e^{itx}\frac{1}{\sqrt{2\pi}}x\,e^{-x^2/2}dx\,.$$

Integrating by parts gives that

$$\phi_X'(t) \;=\; \int_{-\infty}^{\infty} i(it)e^{itx}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx \;=\; -t\,\phi_X(t)\,.$$

Hence, $\phi_X'(t) = -t\,\phi_X(t)$, so that $\frac{d}{dt}\log\phi_X(t) = -t$. Also, we know that $\log\phi_X(0) = \log 1 = 0$. Hence, we must have $\log\phi_X(t) = \int_0^t(-s)\,ds = -t^2/2$, whence $\phi_X(t) = e^{-t^2/2}$. ∎

We can now prove

Theorem 11.2.2. *(Central Limit Theorem) Let $X_1, X_2, \ldots$ be i.i.d. with finite mean m and finite variance v. Set $S_n = X_1 + \ldots + X_n$. Then as $n \to \infty$,*

$$\mathcal{L}\left(\frac{S_n - nm}{\sqrt{vn}}\right) \Rightarrow \mu_N\,,$$

where $\mu_N = N(0,1)$ is the standard normal distribution, i.e. the distribution having density $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ with respect to Lebesgue measure.

Proof. By replacing X_i by $\frac{X_i - m}{\sqrt{v}}$, we can (and do) assume that $m = 0$ and $v = 1$.

Let $\phi_n(t) = \mathbf{E}\left(e^{itS_n/\sqrt{n}}\right)$ be the characteristic function of $S_n/\sqrt{n}$. By the continuity theorem (Theorem 11.1.14), and by Proposition (11.2.1), it suffices to show that $\lim_n \phi_n(t) = e^{-t^2/2}$ for each fixed $t \in \mathbf{R}$.

To this end, set $\phi(t) = \mathbf{E}(e^{itX_1})$. Then as $n \to \infty$, using a Taylor expansion and Proposition 11.0.1,

$$\begin{aligned}
\phi_n(t) \quad &= \mathbf{E}\left(e^{it(X_1+\ldots+X_n)/\sqrt{n}}\right) \\
&= \phi(t/\sqrt{n})^n \\
&= \left(1 + \tfrac{it}{\sqrt{n}}\mathbf{E}(X_1) + \tfrac{1}{2!}\left(\tfrac{it}{\sqrt{n}}\right)^2 \mathbf{E}\left((X_1)^2\right) + o(1/n)\right)^n \\
&= \left(1 - \tfrac{t^2}{2n} + o(1/n)\right)^n \\
&\to e^{-t^2/2}\,,
\end{aligned}$$

as claimed. (Here $o(1/n)$ means a quantity q_n such that $q_n/(1/n) \to 0$ as $n \to \infty$. This implies that for any $\epsilon > 0$, for sufficiently large n we have $|nq_n| \le \epsilon$. Hence, as $n \to \infty$, the $\phi_n(t)$ limit is multiplied by e^{nq_n}, which becomes arbitrarily close to 1 and hence can be ignored. ∎

Since the normal distribution has no points of positive measure, this theorem immediately implies (by Theorem 10.1.1) the simpler-seeming

Corollary 11.2.3. *Let $\{X_n\}$, m, v, and S_n be as above. Then for each fixed $x \in \mathbf{R}$,*

$$\lim_{n\to\infty} \mathbf{P}\left(\frac{S_n - nm}{\sqrt{nv}} \le x\right) \;=\; \Phi(x)\,, \tag{11.2.4}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ is the cumulative distribution function for the standard normal distribution.

This can also be written as $\mathbf{P}(S_n \le nm + x\sqrt{nv}) \to \Phi(x)$. That is, $X_1+\ldots+X_n$ is approximately equal to nm, with deviations from this value of order $\sqrt{n}$. For example, suppose $X_1, X_2, \ldots$ each have the **Poisson**(5) distribution. This implies that $m = \mathbf{E}(X_i) = 5$ and $v = \mathbf{Var}(X_i) = 5$. Hence, for each fixed $x \in \mathbf{R}$, we see that

$$\mathbf{P}\left(X_1 + \ldots + X_n \le 5n + x\sqrt{5n}\right) \to \Phi(x)\,, \qquad n \to \infty\,.$$

Remarks.

1. It is not essential in the Central Limit Theorem to divide by $\sqrt{v}$. Without doing so, the theorem asserts instead that

$$\mathcal{L}\left(\frac{S_n - nm}{\sqrt{n}}\right) \Rightarrow N(0, v)\,.$$

2. Going backwards, the Central Limit Theorem in turn implies the WLLN, since if $y > m$, then as $n \to \infty$,

$$P[(S_n/n) \le y] = P[S_n \le ny] = P[(S_n - nm)/\sqrt{nv} \le (ny - nm)/\sqrt{nv}]$$
$$\approx \Phi[(ny - nm)/\sqrt{nv}] = \Phi[\sqrt{n}(y-m)/\sqrt{v}] \to \Phi(+\infty) = 1\,,$$

and similarly if $y < m$ then $P[(S_n/n) \le y] \to \Phi(-\infty) = 0$. Hence, $\mathcal{L}(S_n/n) \Rightarrow \delta_m$, and so S_n/n converges to m in probability.

11.3. Generalisations of the Central Limit Theorem.

The classical central limit theorem (Theorem 11.2.2 and Corollary 11.2.3) is extremely useful in many areas of science. However, it does have certain limitations. For example, it provides no quantitative bounds on the convergence in (11.2.4). Also, the insistence that the random variables be i.i.d. is sometimes too severe.

The first of these problems is solved by the Berry-Esseen Theorem, which states that if $X_1, X_2, \ldots$ are i.i.d. with finite mean m, finite positive variance v, and $\mathbf{E}\left(|X_i - m|^3\right) = \rho < \infty$, then

$$\left|\mathbf{P}\left(\frac{X_1 + \ldots + X_n - nm}{\sqrt{vn}} \le x\right) - \Phi(x)\right| \le \frac{3\rho}{\sqrt{nv^3}}\,.$$

This theorem thus provides a quantitative bound on the convergence in (11.2.4), depending only on the third moment. For a proof see e.g. Feller (1971, Section XVI.5). Note, however, that this error bound is absolute, not relative: as $x \to -\infty$, both $\mathbf{P}\left(\frac{X_1+\ldots+X_n-nm}{\sqrt{vn}} \le x\right)$ and $\Phi(x)$ get small, and the Berry-Esseen Theorem says less and less. In particular, the theorem does *not* assert that $\mathbf{P}\left(\frac{X_1+\ldots+X_n-nm}{\sqrt{vn}} \le x\right)$ decreases as $O(e^{-x^2/2})$ as $x \to -\infty$, even though $\Phi(x)$ does.

Regarding the second problem, we mention just two of many results. To state them, we shall consider collections $\{Z_{nk};\ n \ge 1,\ 1 \le k \le r_n\}$ of random variables such that each row $\{Z_{nk}\}_{1\le k \le r_n}$ is independent, called *triangular arrays.* (If $r_n = n$ they form an actual triangle.) We shall assume for simplicity that $\mathbf{E}(Z_{nk}) = 0$ for each n and k. We shall further set $\sigma_{nk}^2 = \mathbf{E}(Z_{nk}^2)$ (assumed to be finite), $S_n = Z_{n1} + \ldots + Z_{nr_n}$, and $s_n^2 = \mathbf{Var}(S_n) = \sigma_{n1}^2 + \ldots + \sigma_{nr_n}^2$.

For such a triangular array, the Lindeberg Central Limit Theorem states that $\mathcal{L}(S_n/s_n) \Rightarrow N(0,1)$, provided that for each $\epsilon > 0$, we have

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \mathbf{E}[Z_{nk}^2 \mathbf{1}_{|Z_{nk}|\ge \epsilon s_n}] \;=\; 0\,. \tag{11.3.1}$$

This *Lindeberg condition* states, roughly, that as $n \to \infty$, the tails of the Z_{nk} contribute less and less to the variance of S_n.

Exercise 11.3.2. Consider the special case where $r_n = n$, with $Z_{nk} = \frac{1}{\sqrt{nv}} Y_k$ where $\{Y_k\}$ are i.i.d. with mean 0 and variance $v < \infty$ (so $s_n = 1$).
(a) Prove that the Lindeberg condition (11.3.1) is satisfied in this case. [Hint: Use the Dominated Convergence Theorem.]
(b) Prove that the Lindeberg CLT implies Theorem 11.2.2.

This raises the question that, if (11.3.1) is *not* satisfied, then what other limiting distributions may arise? Call a distribution μ a *possible limit* if there exists a triangular array as defined above, with $\sup_n s_n^2 < \infty$ and $\lim_{n\to\infty} \max_{1\le k\le r_n} \sigma_{nk}^2 = 0$ (so that no one term dominates the contribution to $\mathbf{Var}(S_n)$), such that $\mathcal{L}(S_n) \Rightarrow \mu$. Then we can ask, what distributions are possible limits? Obviously the normal distributions $N(0,v)$ are possible limits; indeed $\mathcal{L}(S_n) \Rightarrow N(0,v)$ whenever (11.3.1) is satisfied and $s_n^2 \to v$. But what else?

The answer is that the possible limits are precisely the infinitely divisible distributions having mean 0 and finite variance. Here a distribution μ is called *infinitely divisible* if for all $n \in \mathbf{N}$, there is a distribution ν_n such that the n-fold convolution of ν_n equals μ (in symbols: $\nu_n * \nu_n * \ldots * \nu_n = \mu$). Recall that this means that, if $X_1, X_2, \ldots, X_n \sim \nu_n$ are independent, then $X_1 + \ldots + X_n \sim \mu$.

Half of this theorem is obvious; indeed, if μ is infinitely divisible, then we can take $r_n = n$ and $\mathcal{L}(X_{nk}) = \nu_n$ in the triangular array, to get that

$\mathcal{L}(S_n) \Rightarrow \mu$. For a proof of the converse, see e.g. Billingsley (1995, Theorem 28.2).

11.4. Method of moments.

There is another way of proving weak convergence of probability measures, which does not explicitly use characteristic functions or the continuity theorem (though its proof of correctness does, through Corollary 11.1.11). Instead, it uses *moments*, as we now discuss.

Recall that a probability distribution μ on $\mathbf{R}$ has moments defined by $\alpha_k = \int x^k \mu(dx)$, for $k = 1, 2, 3, \ldots$. Suppose these moments all exist and are all finite. Then is μ the *only* distribution having precisely these moments? And, if a sequence $\{\mu_n\}$ of distributions have moments which *converge* to those of μ, then does it follow that $\mu_n \Rightarrow \mu$, i.e. that the μ_n converges weakly to μ? We shall see in this section that such conclusions hold sometimes, but not always.

We shall say that a distribution μ is *determined by its moments* if all its moments are finite, and if no other distribution has identical moments. (That is, we have $\int |x^k| \mu(dx) < \infty$ for all $k \in \mathbf{N}$, and furthermore whenever $\int x^k \mu(dx) = \int x^k \nu(dx)$ for all $k \in \mathbf{N}$, then we must have $\nu = \mu$.)

We first show that, for those distributions determined by their moments, convergence of moments implies weak convergence of distributions; this result thus reduces the second question above to the first question.

Theorem 11.4.1. *Suppose that μ is determined by its moments. Let $\{\mu_n\}$ be a sequence of distributions, such that $\int x^k \mu_n(dx)$ is finite for all $n, k \in \mathbf{N}$, and such that $\lim_{n\to\infty} \int x^k \mu_n(dx) = \int x^k \mu(dx)$ for each $k \in \mathbf{N}$. Then $\mu_n \Rightarrow \mu$, i.e. the μ_n converge weakly to μ.*

Proof. We first claim that $\{\mu_n\}$ is tight. Indeed, since the moments converge to finite quantities, we can find $K_k \in \mathbf{R}$ with $\int x^k \mu_n(dx) \le K_k$ for all $n \in \mathbf{N}$. But then, by Markov's inequality, letting $Y_n \sim \mu_n$, we have

$$\begin{aligned} \mu_n([-R, R]) &= \mathbf{P}(|Y_n| \le R) \\ &= 1 - \mathbf{P}(|Y_n| > R) \\ &= 1 - \mathbf{P}\left(Y_n^2 > R^2\right) \\ &\ge 1 - (\mathbf{E}[Y_n^2] / R^2) \\ &\ge 1 - (K_2 / R^2), \end{aligned}$$

which is $\ge 1 - \epsilon$ whenever $R \ge \sqrt{K_2/\epsilon}$, thus proving tightness.

We now claim that if any subsequence $\{\mu_{n_r}\}$ converges weakly to some distribution ν, then we must have $\nu = \mu$. The theorem will then follow from Corollary 11.1.11.

Indeed, suppose $\mu_{n_r} \Rightarrow \nu$. By Skorohod's theorem, we can find random variables Y and $\{Y_r\}$ with $\mathcal{L}(Y) = \nu$ and $\mathcal{L}(Y_r) = \mu_{n_r}$, such that $Y_r \to Y$ with probability 1. But then also $Y_r^k \to Y^k$ with probability 1. Furthermore, for $k \in \mathbf{N}$ and $\alpha > 0$, we have

$$\mathbf{E}\left(|Y_r|^k \mathbf{1}_{|Y_r|^k \geq \alpha}\right) \leq \mathbf{E}\left(\frac{|Y_r|^{2k}}{\alpha} \mathbf{1}_{|Y_r|^k \geq \alpha}\right) \leq \frac{1}{\alpha}\mathbf{E}\left(|Y_r|^{2k}\right)$$

$$= \frac{1}{\alpha}\mathbf{E}\left((Y_r)^{2k}\right) \leq \frac{K_{2k}}{\alpha},$$

which is independent of r, and goes to 0 as $\alpha \to \infty$. Hence, the $\{Y_r^k\}$ are uniformly integrable. Thus, by Theorem 9.1.6, $\mathbf{E}(Y_r^k) \to \mathbf{E}(Y^k)$, i.e. $\int x^k \mu_{n_r}(dx) \to \int x^k \nu(dx)$.

But we already know that $\int x^k \mu_{n_r}(dx) \to \int x^k \mu(dx)$. Hence, the moments of ν and μ must coincide. And, since μ is determined by its moments, we must have $\nu = \mu$, as claimed. ∎

This theorem leads to the question of which distributions are determined by their moments. Unfortunately, not all distributions are, as the following exercise shows.

Exercise 11.4.2. Let $f(x) = \frac{1}{x\sqrt{2\pi}} e^{-(\log x)^2/2}$ for $x > 0$ (with $f(x) = 0$ for $x \leq 0$) be the density function for the random variable e^X, where $X \sim N(0,1)$. Let $g(x) = f(x)\,(1 + \sin(2\pi \log x))$. Show that $g(x) \geq 0$ and that $\int x^k g(x)dx = \int x^k f(x)dx$ for $k = 0, 1, 2, \ldots$. [Hint: Consider $\int x^k f(x) \sin(2\pi \log x)dx$, and make the substitution $x = e^s e^k$, $dx = e^s e^k ds$.] Show further that $\int |x|^k f(x)dx < \infty$ for all $k \in \mathbf{N}$. Conclude that g is a probability density function, and that g gives rise to the same (finite) moments as does f. Relate this to Theorem 11.4.1 above and Theorem 11.4.3 below.

On the other hand, if a distribution satisfies that its moment generating function is finite in a neighbourhood of the origin, then it will be determined by its moments, as we now show. (Unfortunately, the proof requires a result from complex function theory.)

Theorem 11.4.3. *Let $s_0 > 0$, and let X be a random variable with moment generating function $M_X(s)$ which is finite for $|s| < s_0$. Then $\mathcal{L}(X)$ is determined by its moments (and also by $M_X(s)$).*

Proof (optional). Let $f_X(z) = \mathbf{E}(e^{zX})$ for $z \in \mathbf{C}$. Since $|e^{zX}| = e^{X\,\Re e\, z}$, we see that $f_X(z)$ is finite whenever $|\Re e\, z| < s_0$. Furthermore, just like for $M_X(s)$, it follows that $f_X(z)$ is *analytic* on $\{z \in \mathbf{C};\ |\Re e\, z| < s_0\}$. Now, if

Y has the same moments as does X, then for $|s| < s_0$, we have by order-preserving and countable linearity that

$$\mathbf{E}\left(e^{sY}\right) \;\leq\; \mathbf{E}\left(e^{sY} + e^{-sY}\right)$$
$$=\; 2\,\mathbf{E}\left(1 + \frac{s^2Y^2}{2!} + \ldots\right) = 2\left(1 + \frac{s^2\mathbf{E}(Y^2)}{2!} + \ldots\right)$$
$$=\; 2\left(1 + \frac{s^2\mathbf{E}(X^2)}{2!} + \ldots\right) = M_X(s) + M_X(-s) < \infty\,.$$

Hence, $M_Y(s) < \infty$ for $|s| < s_0$. It now follows from Theorem 9.3.3 that $M_Y(s) = M_X(s)$ for $|s| < s_0$, i.e. that $f_X(s) = f_Y(s)$ for real $|s| < s_0$. By the uniqueness of analytic continuation, this implies that $f_Y(z) = f_X(z)$ for $|\Re e\, z| < s_0$. In particular, since $\phi_X(t) = f_X(it)$ and $\phi_Y(t) = f_Y(it)$, we have $\phi_X = \phi_Y$. Hence, by the uniqueness theorem for characteristic functions (Theorem 11.1.7), we must have $\mathcal{L}(Y) = \mathcal{L}(X)$, as claimed. ∎

Remark 11.4.4. Proving weak convergence by showing convergence of moments is called the *method of moments**. Indeed, it is possible to prove the central limit theorem in this manner, under appropriate assumptions.

Remark 11.4.5. By similar reasoning, it is possible to show that if $M_{X_n}(s) < \infty$ and $M_X(s) < \infty$ for all $n \in \mathbf{N}$ and $|s| < s_0$, and also $M_{X_n}(s) \to M_X(s)$ for all $|s| < s_0$, then we must have $\mathcal{L}(X_n) \Rightarrow \mathcal{L}(X)$.

11.5. Exercises.

Exercise 11.5.1. Let $\mu_n = \delta_n$ be a point mass at n (for $n = 1, 2, \ldots$).
(a) Is $\{\mu_n\}$ tight?
(b) Does there exist a subsequence $\{\mu_{n_k}\}$, and a Borel probability measure μ, such that $\mu_{n_k} \Rightarrow \mu$? (If so, then specify $\{n_k\}$ and μ.) Relate this to theorems from this section.
(c) Setting $F_n(x) = \mu_n\left((-\infty, x]\right)$, does there exist a non-decreasing, right-continuous function F such that $F_n(x) \to F(x)$ for all continuity points x of F? (If so, then specify F.) Relate this to the Helly Selection Principle.
(d) Repeat part (c) for the case where $\mu_n = \delta_{-n}$ is a point mass at $-n$.

Exercise 11.5.2. Let $\mu_n = \delta_{n \bmod 3}$ be a point mass at n mod 3. (Thus, $\mu_1 = \delta_1$, $\mu_2 = \delta_2$, $\mu_3 = \delta_0$, $\mu_4 = \delta_1$, $\mu_5 = \delta_2$, $\mu_6 = \delta_0$, etc.)

*This should not be confused with the statistical estimation procedure of the same name, which estimates unknown parameters by choosing them to make observed moments equal theoretical ones.

(a) Is $\{\mu_n\}$ tight?
(b) Does there exist a Borel probability measure μ, such that $\mu_n \Rightarrow \mu$? (If so, then specify μ.)
(c) Does there exist a subsequence $\{\mu_{n_k}\}$, and a Borel probability measure μ, such that $\mu_{n_k} \Rightarrow \mu$? (If so, then specify $\{n_k\}$ and μ.)
(d) Relate parts (b) and (c) to theorems from this section.

Exercise 11.5.3. Let $\{x_n\}$ be any sequence of points in the interval $[0, 1]$. Let $\mu_n = \delta_{x_n}$ be a point mass at x_n.
(a) Is $\{\mu_n\}$ tight?
(b) Does there exist a subsequence $\{\mu_{n_k}\}$, and a Borel probability measure μ, such that $\mu_{n_k} \Rightarrow \mu$? (Hint: by compactness, there must be a subsequence of points $\{x_{n_k}\}$ which converges, say to $y \in [0, 1]$. Then what does μ_{n_k} converge to?)

Exercise 11.5.4. Let $\mu_{2n} = \delta_0$, and let $\mu_{2n+1} = \delta_n$, for $n = 0, 1, 2, \ldots$.
(a) Does there exist a Borel probability measure μ, such that $\mu_n \Rightarrow \mu$?
(b) Suppose for some subsequence $\{\mu_{n_k}\}$ and some Borel probability measure ν, we have $\mu_{n_k} \Rightarrow \nu$. What must ν be?
(c) Relate parts (a) and (b) to Corollary 11.1.11. Why is there no contradiction?

Exercise 11.5.5. Let $\mu_n = \mathbf{Uniform}[0, n]$, so $\mu_n([a, b]) = (b - a)/n$ for $0 \le a \le b \le n$.
(a) Prove or disprove that $\{\mu_n\}$ is tight.
(b) Prove or disprove that there is some probabilty measure μ such that $\mu_n \Rightarrow \mu$.

Exercise 11.5.6. Suppose $\mu_n \Rightarrow \mu$. Prove or disprove that $\{\mu_n\}$ must be tight.

Exercise 11.5.7. Define the Borel probability measure μ_n by $\mu_n(\{x\}) = 1/n$, for $x = 0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}$. Let λ be Lebesgue measure on $[0, 1]$.
(a) Compute $\phi_n(t) = \int e^{itx} \mu_n(dx)$, the characteristic function of μ_n.
(b) Compute $\phi(t) = \int e^{itx} \lambda(dx)$, the characteristic function of λ.
(c) Does $\phi_n(t) \to \phi(t)$, for each $t \in \mathbf{R}$?
(d) What does the result in part (c) imply?

Exercise 11.5.8. Use characteristic functions to provide an alternative solution of Exercise 10.3.2.

Exercise 11.5.9. Use characteristic functions to provide an alternative solution of Exercise 10.3.3.

Exercise 11.5.10. Use characteristic functions to provide an alternative solution of Exercise 10.3.4.

Exercise 11.5.11. Compute the characteristic function $\phi_X(t)$, and also $\phi'_X(0) = i\,\mathbf{E}(X)$, where X follows
(a) the binomial distribution: $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, 1, 2, \ldots, n$.
(b) the Poisson distribution: $P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$, for $k = 0, 1, 2, \ldots$.
(c) the exponential distribution, with density with respect to Lebesgue measure given by $f_X(x) = \lambda e^{-\lambda x}$ for $x > 0$, and $f_X(x) = 0$ for $x < 0$.

Exercise 11.5.12. Suppose that for $n \in \mathbf{N}$, we have $\mathbf{P}[X_n = 5] = 1/n$ and $\mathbf{P}[X_n = 6] = 1 - (1/n)$.
(a) Compute the characteristic function $\phi_{X_n}(t)$, for all $n \in \mathbf{N}$ and $t \in \mathbf{R}$.
(b) Compute $\lim_{n\to\infty} \phi_{X_n}(t)$.
(c) Specify a distribution μ such that $\lim_{n\to\infty} \phi_{X_n}(t) = \int e^{itx}\,\mu(dx)$ for all $t \in \mathbf{R}$.
(d) Determine (with explanation) whether or not $\mathcal{L}(X_n) \Rightarrow \mu$.

Exercise 11.5.13. Let $\{X_n\}$ be i.i.d., each having mean 3 and variance 4. Let $S = X_1 + X_2 + \ldots + X_{10,000}$. In terms of $\Phi(x)$, give an approximate value for $\mathbf{P}[S \le 30,500]$.

Exercise 11.5.14. Let $X_1, X_2, \ldots$ be i.i.d. with mean 4 and variance 9. Find values $C(n, x)$, for $n \in \mathbf{N}$ and $x \in \mathbf{R}$, such that as $n \to \infty$, $\mathbf{P}[X_1 + X_2 + \ldots + X_n \le C(n, x)] \approx \Phi(x)$.

Exercise 11.5.15. Prove that the **Poisson**(λ) distribution, and the $N(m, v)$ (normal) distribution, are both infinitely divisible (for any $\lambda > 0$, $m \in \mathbf{R}$, and $v > 0$). [Hint: Use Exercises 9.5.15 and 9.5.16.]

Exercise 11.5.16. Let X be a random variable whose distribution $\mathcal{L}(X)$ is infinitely divisible. Let $a > 0$ and $b \in \mathbf{R}$, and set $Y = aX + b$. Prove that $\mathcal{L}(Y)$ is infinitely divisible.

Exercise 11.5.17. Prove that the **Poisson**(λ) distribution, the $N(m, v)$ distribution, and the **Exp**(λ) (exponential) distribution, are all determined by their moments, for any $\lambda > 0$, $m \in \mathbf{R}$, and $v > 0$.

Exercise 11.5.18. Let $X, X_1, X_2, \ldots$ be random variables which are uniformly bounded, i.e. there is $M \in \mathbf{R}$ with $|X| \le M$ and $|X_n| \le M$ for all n. Prove that $\{\mathcal{L}(X_n)\} \Rightarrow \mathcal{L}(X)$ if and only if $\mathbf{E}\left(X_n^k\right) \to \mathbf{E}\left(X^k\right)$ for all $k \in \mathbf{N}$.

11.6. Section summary.

This section introduced the characteristic function $\phi_X(t) = \mathbf{E}(e^{itX})$. After introducing its basic properties, it proved an Inversion Theorem (to recover the distribution of a random variable from its characteristic function) and a Uniqueness Theorem (which shows that if two random variables have the same characteristic function then they have the same distribution).

Then, using the Helly Selection Principle and the notion of tightness, it proved the important Continuity Theorem, which asserts the equivalence of weak convergence of distributions and pointwise convergence of characteristic functions. This important theorem was used to prove the Central Limit Theorem about weak convergence of averages of i.i.d. random variables to a normal distribution. Some generalisations of the Central Limit Theorem were briefly discussed.

The section ended with a discussion of the method of moments, an alternative way of proving weak convergence of random variables using only their moments, but not their characteristic functions.