1 Bochner’s Theorem

Before moving on to the proof of the Central Limit Theorem (CLT) using characteristic functions, we prove another important result, known as Bochner’s Theorem.

If $\phi$ is the characteristic function of a probability measure on $\mathbb{R}$, then $\phi$ is positive definite in the following sense:

$$
\sum_{i,j=1}^{n} \phi(t_i - t_j)\xi_i \xi_j \geq 0 \quad \text{for all } n \in \mathbb{N}, \xi_1, \ldots, \xi_n \in \mathbb{C}, \text{ and } t_1, \ldots, t_n \in \mathbb{R}. \quad (1.1)
$$

The proof is straightforward:

$$
\sum_{i,j=1}^{n} \phi(t_i - t_j)\xi_i \xi_j = \sum_{i,j=1}^{n} \xi_i \xi_j \int e^{it_i x} e^{-it_j x} \mu(dx) = \int \left| \sum_{i} \xi_i e^{it_i x} \right|^2 \mu(dx) \geq 0.
$$

The interesting fact is the converse, that every continuous positive definite function can be written as the characteristic function of a finite measure.

**Theorem 1.1 [Bochner’s Theorem]** Let $\phi$ be a positive definite function which is continuous at 0 with $\phi(0) = 1$. Then there exists a probability measure $\mu$ on $\mathbb{R}$ with $\phi(t) = \int e^{itx} \mu(dx)$.

**Remark 1.2** Positive definite functions appear in many contexts. For example, if $(X_s)_{s \in \mathbb{R}}$ is a stationary real-valued stochastic process (i.e., $(X_s)_{s \in \mathbb{R}}$ is a family of real-valued random variables indexed by $s \in \mathbb{R}$, with $(X_s)_{s \in \mathbb{R}} \overset{\text{dist}}{=} (X_{s+\theta})_{s \in \mathbb{R}}$ for all $\theta \in \mathbb{R}$), then the covariance function $C(t) = \mathbb{E}[X_0 X_t] - \mathbb{E}[X_0] \mathbb{E}[X_t]$ is positive definite, and hence $C(t) = \text{Var}(X_0) \int e^{itx} \mu(dx)$ for a symmetric probability measure $\mu$ on $\mathbb{R}$, symmetric because $C(t) = C(-t)$. Indeed, note that

$$
\sum_{i,j=1}^{n} \xi_i \xi_j C(t_i - t_j) = \sum_{i,j=1}^{n} \xi_i \xi_j \mathbb{E}[X'_i X'_{t_j}] = \mathbb{E} \left[ \left| \sum_{i=1}^{n} \xi_i X'_i \right|^2 \right] \geq 0,
$$

where $X'_i = X_t - \mathbb{E}[X_t]$, and we used the stationarity of $(X_s)_{s \in \mathbb{R}}$.

**Remark 1.3** The Hamburger moment problem is closely related to Bochner’s Theorem, with a moment sequence $(m_k)_{k \in \mathbb{N}}$ playing the role of the characteristic functions $(\phi(t))_{t \in \mathbb{R}}$. Indeed, if $m_k = \int x^k \mu(dx)$ for a probability measure $\mu$, then the sequence $(m_k)_{k \in \mathbb{N}}$ is Hankel positive definite in the sense that: for all $n \in \mathbb{N}$ and $\xi_1, \ldots, \xi_n \in \mathbb{R},$

$$
\sum_{i,j=1}^{n} \xi_i \xi_j m_{i+j} = \sum_{i,j=1}^{n} \int x^{i+j} \xi_i \xi_j \mu(dx) = \int \left( \sum_{i=1}^{n} \xi_i x^i \right)^2 \mu(dx) \geq 0.
$$

The main difference from the definition of positive definiteness of a characteristic function is that, $m_{i+j}$ depends on the indices $i$ and $j$ in terms of their sum, while $\phi(t_i - t_j)$ depends on the arguments $t_i$ and $t_j$ in terms of their difference, which can be called Toeplitz positive definite, in analogy with Toeplitz matrices and Hankel matrices. Hamburger proved the analogue of Bochner’s Theorem, that $(m_k)_{k \in \mathbb{N}}$ is a moment sequence of a finite measure if and only if it is Hankel positive definite. For an extensive discussion on representations of positive definite functions, see [1].
Proof of Theorem 1.1. We will construct a sequence of characteristic functions \( \phi_n \) such that \( \phi_n \to \phi \) pointwise, and then by Lévy’s Continuity Theorem, \( \phi \) must be a characteristic function as well.

Assuming \( \phi \) is positive definite, we first establish some of its properties.

1. For any \( a \in \mathbb{R} \), \( \psi(t) = \phi(t)e^{ita} \) is also positive definite, which follows from the definition.

2. If \( \phi_1, \ldots, \phi_m \) are positive definite, and \( \lambda_1, \ldots, \lambda_m > 0 \), then \( \sum_{i=1}^{m} \lambda_i \phi_i \) is also positive definite, which also follows easily from the definition.

3. \( \phi(0) \geq 0 \), \( \phi(t) = \overline{\phi(-t)} \), and \( |\phi(t)| \leq \phi(0) \) for all \( t \in \mathbb{R} \). Indeed, if we let \( t_1 = 0 \) and \( t_2 = t \) in (1.1), then
   \[
   (|\xi_1|^2 + |\xi_2|^2)\phi(0) + \xi_1\xi_2\phi(-t) + \overline{\xi_1}\xi_2\phi(t) \geq 0 \text{ for all } \xi_1, \xi_2 \in \mathbb{C}.
   \]

   Letting \( \xi_1 = 1 \) and \( \xi_2 = 0 \) shows that \( \phi(0) \geq 0 \); letting \( \xi_1 = \xi_2 = 1 \) shows that \( \text{Im}[\phi(t)] = -\text{Im}[\phi(-t)] \); letting \( \xi_1 = 1 \) and \( \xi_2 = i \) shows that \( \text{Re}[\phi(t)] = \text{Re}[\phi(-t)] \); letting \( \xi_1 = \xi_2 = \sqrt{-\phi(t)} \) shows that \( |\phi(t)| \leq \phi(0) \).

4. For any \( s, t \in \mathbb{R} \), we have
   \[
   |\phi(t) - \phi(s)|^2 \leq 4\phi(0)|\phi(0) - \phi(t - s)|. \tag{1.2}
   \]

   To prove this, note that (3) and (1.1) imply that \((\phi(t_i - t_j))_{1 \leq i,j \leq n}\) is a Hermitian positive definite matrix. In particular, if we choose \( t_1 = t, t_2 = s \) and \( t_3 = 0 \), then the matrix
   \[
   \begin{pmatrix}
   \phi(0) & \phi(t-s) & \phi(t) \\
   \frac{\phi(0)}{\phi(t-s)} & \phi(0) & \phi(s) \\
   \phi(t) & \phi(s) & \phi(0)
   \end{pmatrix}
   \]

must have a non-negative determinant, which gives

\[
0 \leq \phi(0)^3 + \phi(s)\phi(t-s)\overline{\phi(t)} + \overline{\phi(s)}\phi(t-s)\phi(t) - \phi(0)
[|\phi(s)|^2 + |\phi(t)|^2 + |\phi(t-s)|^2]
= \phi(0)^3 - \phi(0)[|\phi(t) - \phi(s)|^2 + |\phi(t-s)|^2] - 2\text{Re}[\phi(s)\overline{\phi(t)}(\phi(0) - \phi(t-s))]
\leq \phi(0)^3 - \phi(0)[|\phi(t) - \phi(s)|^2 + |\phi(t-s)|^2] + 2\phi(0)^2|\phi(0) - \phi(t-s)|,
\]

where in the last line we used that \( |\phi(s)|, |\phi(t)| \leq \phi(0) \). Rearranging terms then gives

\[
|\phi(t) - \phi(s)|^2 \leq \phi(0)^2 - |\phi(t-s)|^2 + 2\phi(0)|\phi(0) - \phi(t-s)|
= (\phi(0) + |\phi(t-s)|)(\phi(0) - |\phi(t-s)|) + 2\phi(0)|\phi(0) - \phi(t-s)|
\leq 4\phi(0)|\phi(0) - \phi(t-s)|,
\]

where we again used that \( |\phi(t-s)| \leq \phi(0) \). This proves (1.2).

5. By (4), it follows that if \( \phi \) is continuous at 0, then in fact \( \phi \) is uniformly continuous on \( \mathbb{R} \). This is consistent with what we have proved for characteristic functions.

We are now ready to prove the theorem. The basic idea is that if \( \phi \) is furthermore assumed to be absolutely integrable, i.e., \( \int |\phi(t)|dt < \infty \), then we can actually take the inverse Fourier transform

\[
f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-ixt} \phi(t)dt. \tag{1.3}
\]

If Bochner’s Theorem was true, then we expect \( f \) to be in fact a probability density, which we will try to establish. Then \( \phi \) is the Fourier transform of \( f \), and hence is a characteristic
function. When $\phi$ is not assumed to be integrable, we will approximate $\phi$ by $\phi_n$ which are positive definite and absolutely integrable. Since $\phi_n$ would be characteristic functions, by Lévy’s Continuity Theorem, we can then conclude that $\phi$ is also a characteristic function.

Now we show that $f$ defined in (1.3) from an absolutely integrable $\phi$ is indeed a probability density. We first use the positive definiteness of $\phi$ to show that $f \geq 0$. Note that

$$0 \leq \lim_{T \to \infty} \frac{1}{2\pi T} \int_0^T \int_0^T e^{-itx} e^{isx} \phi(t-s) dt \, ds$$

$$= \lim_{T \to \infty} \frac{1}{2\pi T} \int_{-T}^T \phi(u) e^{-iux} \left( \int_{s \in [0,T], s+u \in [0,T]} ds \right) du$$

$$= \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^T \phi(u) e^{-iux} \left( 1 - \frac{|u|}{T} \right) du$$

$$= \frac{1}{2\pi} \int e^{-itx} \phi(t) dt = f(x),$$

where the inequality can be deduced from a Riemann sum approximation of the two-fold integral and the positive definiteness of $\phi$, while the last equality follows from the assumption that $\phi$ is integrable and the dominated convergence theorem.

Next we show that

$$\phi(t) = \int e^{itx} f(x) dx,$$  \hspace{1cm} (1.4)

which will imply that $\int f = \phi(0) = 1$, and hence $\phi$ is a characteristic function. Note that if $f = \hat{\phi}$ is integrable, then $\int e^{itx} f(x) dx$ is well-defined and is the inverse Fourier transform of $f$, and hence must equal $\phi$ by the same proof as Theorem 1.5 in Lecture 6. More generally, without assuming the integrability of $f$, we can approximate $f$ by

$$f_\sigma(x) = f(x) e^{-\frac{\sigma^2 x^2}{2}}, \quad \sigma > 0.$$

Note that by Fubini,

$$\int e^{itx} f_\sigma(x) = \int e^{itx} \left( \frac{1}{2\pi} \int e^{-isx} \phi(s) ds \right) e^{-\frac{\sigma^2 x^2}{2}} dx$$

$$= \frac{1}{\sqrt{2\pi} \sigma^2} \int \phi(s) \left( \frac{1}{\sqrt{2\pi} / \sigma^2} \int e^{i(t-s)x - \frac{\sigma^2 x^2}{2}} dx \right) ds$$

$$= \frac{1}{\sqrt{2\pi} \sigma^2} \int \phi(s) e^{-\frac{(t-s)^2}{2\sigma^2}} ds = \frac{1}{\sqrt{2\pi}} \int \phi(t + \sigma u) e^{-\frac{u^2}{2}} du. \hspace{1cm} (1.5)$$

In particular, let $t = 0$, then

$$\int f_\sigma(x) dx = \frac{1}{\sqrt{2\pi} \sigma^2} \int \phi(s) e^{-\frac{s^2}{2\sigma^2}} ds \leq \sup_{s \in \mathbb{R}} |\phi(s)| = \phi(0) = 1,$$

which by monotone convergence theorem implies $\int f(x) dx \leq 1$ as we send $\sigma \downarrow 0$. Therefore in (1.5), as we let $\sigma \downarrow 0$, the left hand side (LHS) converges to $\int e^{itx} f(x) dx$ by the dominated convergence theorem, while the right hand side (RHS) converges to $\phi(t)$ because $\phi$ is bounded and continuous. This proves (1.4), and hence the theorem for integrable $\phi$.

If $\phi$ is not assumed to be integrable, then we approximate $\phi$ by

$$\phi_\sigma(t) := \phi(t) e^{-\frac{\sigma^2 t^2}{2}} = \frac{1}{\sqrt{2\pi} \sigma^2} \int \phi(t) e^{ity} e^{-\frac{y^2}{2\sigma^2}} dy$$

where we note that $\phi_\sigma$ is continuous and integrable with $\phi_\sigma(0) = 1$, and $\phi_\sigma$ is a convex combination of the functions $\phi(t) e^{ity}, y \in \mathbb{R}$, each of which is positive definite, therefore $\phi_\sigma$ is also positive definite. In particular, by our arguments above, $\phi_\sigma$ is a characteristic function. Since $\phi_\sigma \to \phi$ as $\sigma \downarrow 0$, $\phi$ is also a characteristic function by Lévy’s Continuity Theorem.
2 Central Limit Theorem

We now use characteristic functions to prove the CLT for sums of i.i.d. random variables.

**Theorem 2.1 [Central Limit Theorem]** Let \((X_n)_{n \in \mathbb{N}}\) be i.i.d. random variables with \(\mathbb{E}[X_1] = 0\) and \(\text{Var}(X_1) = \sigma^2 \in (0, \infty)\). Let \(S_n := \sum_{i=1}^{n} X_i\). Then \(S_n/\sqrt{n}\) converges in distribution to \(Z\), a Gaussian (or normal) random variable with mean 0 and variance \(\sigma^2\), i.e.,

\[
\mathbb{P}(Z \leq x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} e^{-y^2/(2\sigma^2)} dy \quad \text{for } x \in \mathbb{R}.
\]

**Proof.** By Lévy’s continuity theorem, it suffices to show that the characteristic function of \(S_n/\sqrt{n}\) converges to that of \(Z\). Let \(\phi(t) := \mathbb{E}[e^{itX_1}]\). Then \(\phi'(0) = i\mathbb{E}[X_1] = 0\) and \(\phi''(0) = -\mathbb{E}[X_1^2] = -\sigma^2\). For each \(t \in \mathbb{R}\),

\[
\mathbb{E}[e^{it\frac{S_n}{\sqrt{n}}}] = \phi\left(\frac{t}{\sqrt{n}}\right)^n = \left(1 + \frac{\phi''(0) t^2}{2} + o\left(\frac{t^2}{n}\right)\right)^n \to \left(1 - \frac{\sigma^2 t^2}{2n} + o\left(\frac{t^2}{n}\right)\right)^n \quad \text{as } n \to \infty,
\]

which is precisely the characteristic function of \(Z\), as can be seen by

\[
\mathbb{E}[e^{itZ}] = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} e^{ity - \frac{y^2}{2\sigma^2}} dy = \frac{1}{\sqrt{2\pi}\sigma^2} \int_{\mathbb{R}} e^{-\frac{(y-it\sigma^2)^2}{2\sigma^2}} dy = e^{-\frac{t^2\sigma^2}{2}}.
\]

Therefore \(S_n/\sqrt{n}\) converges in distribution to \(Z\).

3 Lindeberg’s Theorem

The central limit theorem can be regarded as a result about a triangular array of random variables. Namely if we let \(X_{n,i} = X_i/\sqrt{n}\), then

\[
X_{1,1};
X_{2,1}, X_{2,2};
\vdots
X_{n,1}, \ldots, X_{n,n};
\]

forms a triangular array. For \(\frac{S_n}{\sqrt{n}} = X_{1,1} + X_{n,2} + \cdots + X_{n,n}\) to converge to a Gaussian distribution, it is actually not necessary to assume \(X_{n,1}, \ldots, X_{n,n}\) to be i.i.d. It suffices to assume that \(X_{n,1}, \ldots, X_{n,n}\) are independent and uniformly small in a suitable sense.

**Theorem 3.1 [Lindeberg’s Theorem]** For each \(n \in \mathbb{N}\), let \((X_{n,j})_{1 \leq j \leq n}\) be independent random variables with \(\mathbb{E}[X_{n,j}] = 0\) and \(\sum_{j=1}^{n} \mathbb{E}[X_{n,j}^2] = 1\). If \(X\) satisfies Lindeberg’s condition, namely that for each \(\epsilon > 0\),

\[
\sum_{j=1}^{n} \mathbb{E}[X_{n,j}^2 1_{\{|X_{n,j}| > \epsilon\}}] \to 0, \quad \text{as } n \to \infty.
\]

then \(Z_n := X_{n,1} + \cdots + X_{n,n}\) converges in distribution to a standard normal (or Gaussian) random variable, i.e., a normal random variable with mean 0 and variance 1.

We will give two proofs of Theorem 3.1. The first proof is based on the convergence of characteristic functions, while the second proof is based on Lindeberg’s original proof, which has recently attracted renewed interest because Lindeberg’s method also applies to more
complex (non-linear) functionals of independent random variables, where the distribution of the functional need not be Gaussian asymptotically. One recent example in the literature is the study of eigenvalues of a random matrix with independent entries. As the matrix size tends to infinity, it turns out that many statistics of the eigenvalues are asymptotically independent of the distribution of the individual matrix entries, similar in spirit to the central limit theorem.

**Proof by characteristic functions.** Let \( \phi_{n,j}(t) = \mathbb{E}[e^{itX_{n,j}}] \), and \( \phi_n(t) = \mathbb{E}[e^{itZ_n}] = \prod_{j=1}^n \phi_{n,j}(t) \). It suffices to show that

\[
\phi_n(t) \to e^{-\frac{t^2}{2}} \quad \text{for each } t \in \mathbb{R}. \tag{3.7}
\]

Let \( \sigma^2_{n,j} := \mathbb{E}[X^2_{n,j}] \). We need to Taylor expand \( \phi_{n,j}(t) = 1 - \frac{\sigma^2_{n,j}t^2}{2} + R_{n,j}(t) \) with a good control on the remainder \( R_{n,j}(t) \), so that we can approximate \( \phi_{n,j}(t) \) by \( e^{-(\frac{\sigma^2_{n,j}}{2})t^2} \), which will then imply (3.7).

Note that by Taylor expansion with remainder, we have

\[
e^{ix} = 1 + ix - \frac{x^2}{2} - \frac{i}{2} \int_0^x e^{is}(x-s)^2 ds. \tag{3.8}
\]

Therefore

\[
|e^{ix} - 1 - ix + \frac{x^2}{2}| \leq \frac{1}{6}|x|^3. \tag{3.9}
\]

On the other hand, using integration by parts,

\[
\int_0^x e^{is}(x-s)^2 ds = i^{-1}e^{is}(x-s)^2\bigg|_0^x + 2i^{-1} \int_0^x e^{is}(x-s) ds = -i^{-1}x^2 + 2i^{-1} \int_0^x e^{is}(x-s) ds.
\]

Therefore

\[
|e^{ix} - 1 - ix + \frac{x^2}{2}| \leq \frac{1}{2} \int_0^x e^{is}(x-s)^2 ds \leq |x|^2. \tag{3.10}
\]

Combining (3.9) and (3.10) then shows that for any random variable \( Y \) with \( \mathbb{E}[Y] = 0 \) and \( \mathbb{E}[Y^2] = \sigma^2_Y \),

\[
|\mathbb{E}[e^{itY}] - 1 + \frac{\sigma^2_Y t^2}{2}| = \left|\mathbb{E}\left[e^{itY} - 1 - itY + \frac{t^2Y^2}{2}\right]\right| \leq \mathbb{E}[\min\{Y^2, |Y|^3/6\}]. \tag{3.11}
\]

Applying this bound to \( X_{n,j} \) then gives

\[
|R_{n,j}(t)| = |\phi_{n,j}(t) - 1 + \frac{\sigma^2_{n,j}t^2}{2}| \leq \mathbb{E}[\min\{X^2_{n,j}, |X_{n,j}|^3\}] \leq \mathbb{E}[X^2_{n,j}1_{\{|X_{n,j}|>\epsilon\}}] + \epsilon \mathbb{E}[X^2_{n,j}],
\]

Therefore

\[
\sum_{j=1}^n |R_{n,j}(t)| \leq \sum_{j=1}^n \mathbb{E}[X^2_{n,j}1_{\{|X_{n,j}|>\epsilon\}}] + \epsilon.
\]

By Lindeberg’s condition (3.6) and the fact that \( \epsilon > 0 \) can be chosen arbitrarily, we have

\[
\sum_{j=1}^n |R_{n,j}(t)| \xrightarrow{n \to \infty} 0. \tag{3.12}
\]
Again by Lindeberg’s condition, we note that
\[
\max_{1 \leq j \leq n} \sigma_{n,j}^2 = \max_{1 \leq j \leq n} \mathbb{E}[X_{n,j}^2] \leq \max_{1 \leq j \leq n} \mathbb{E}[X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}| \leq \epsilon\}}] + \max_{1 \leq j \leq n} \mathbb{E}[X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}| > \epsilon\}}] \\
\leq \epsilon^2 + \sum_{j=1}^{n} \mathbb{E}[X_{n,j}^2 \mathbb{1}_{\{|X_{n,j}| > \epsilon\}}] \xrightarrow{n \to \infty} \epsilon^2.
\]
Since \(\epsilon > 0\) can be chosen arbitrarily, it follows that
\[
\max_{1 \leq j \leq n} \sigma_{n,j}^2 \xrightarrow{n \to \infty} 0. \tag{3.13}
\]
By (3.12) and (3.13), we can Taylor expand
\[
\log \phi_{n,j}(t) = \log \left(1 - \frac{\sigma_{n,j}^2 t^2}{2} + R_{n,j}(t)\right) = -\frac{\sigma_{n,j}^2 t^2}{2} + R_{n,j}(t) + O\left(\frac{\sigma_{n,j}^2 t^2}{2} + |R_{n,j}(t)|^2\right),
\]
and hence
\[
\log \phi_n(t) = \sum_{j=1}^{n} \log \phi_{n,j}(t) = -\frac{t^2}{2} + O\left(\sum_{j=1}^{n} |R_{n,j}(t)| + |R_{n,j}(t)|^2\right) + O\left(\max_{1 \leq j \leq n} \sigma_{n,j}^2 t^4\right) \xrightarrow{n \to \infty} -\frac{t^2}{2},
\]
which proves (3.7).

\textbf{Proof by Lindeberg’s method.} To prove that \(Z_n := \sum_{j=1}^{n} X_{n,j}\) converges in distribution to a standard normal random variable \(W\), by the definition of weak convergence, it suffices to show that \(\mathbb{E}[f(Z_n)] \xrightarrow{n \to \infty} \mathbb{E}[f(W)]\) for all bounded continuous \(f : \mathbb{R} \to \mathbb{R}\). In fact, since for any \(-\infty < a < b < \infty\), the indicator function \(\mathbb{1}_{(a,b)}(x)\) can be approximated from above and below by \(f \in C^2_b(\mathbb{R})\), the space of three-times continuously differentiable functions with compact support, it suffices to show that
\[
\mathbb{E}[f(Z_n)] \xrightarrow{n \to \infty} \mathbb{E}[f(W)] \quad \text{for} \quad f \in C^2_b(\mathbb{R}). \tag{3.14}
\]
Let us denote \(g_n(X_{n,1}, X_{n,2}, \ldots, X_{n,n}) = f(Z_n) = f(\sum_{i=1}^{n} X_{n,i})\). The basic idea is that under Lindeberg’s condition, the influence of each coordinate \(X_{n,j}\) on \(g_n\) is small. Therefore we will successively exchange \(X_{n,1}, X_{n,2}, \ldots\) with Gaussian random variables \(W_{n,1}, W_{n,2}, \ldots\) with the same mean and variance, and hope that the aggregate effect of these exchanges on \(g_n\) is still small. More precisely, let \(W_{n,j}, 1 \leq j \leq n\), be independent Gaussian random variables with mean 0 and variance \(\sigma_{n,j}^2\). Note that \(W := \sum_{j=1}^{n} W_{n,j}\) is a standard normal random variable.

We can rewrite (3.14) as
\[
\mathbb{E}[g_n(X_{n,1}, \ldots, X_{n,n})] - \mathbb{E}[g_n(W_{n,1}, \ldots, W_{n,n})] \xrightarrow{n \to \infty} 0. \tag{3.15}
\]
We rewrite the above difference as a telescopic sum:
\[
\mathbb{E}[g_n(X_{n,1}, \ldots, X_{n,n})] - \mathbb{E}[g_n(W_{n,1}, \ldots, W_{n,n})] = \sum_{j=1}^{n} \left( \mathbb{E}[g_n(X_{n,1}, \ldots, X_{n,j}, W_{n,j+1}, \ldots, W_{n,n})] - \mathbb{E}[g_n(X_{n,1}, \ldots, X_{n,j-1}, W_{n,j}, \ldots, W_{n,n})] \right).
\]
In the \(j\)-th term, we effectively switched the \(j\)-th argument of \(g_n\) from \(W_{n,j}\) to \(X_{n,j}\), and we can bound this difference by Taylor expansion as follows. Given \(x := (x_1, \ldots, x_n) \in \mathbb{R}^n\) and \(y \in \mathbb{R}\), let us denote \(h_{n,j}^x(y) = g_n(x_1, \ldots, x_{j-1}, y, x_{j+1}, \ldots, x_n)\). Then
\[
h_{n,j}^x(y) = h_{n,j}^x(0) + \frac{d h_{n,j}^x}{dy}(0)y + \frac{1}{2} \frac{d^2 h_{n,j}^x}{dy^2}(0)y^2 + R_{n,j}^x(y),
\]
where by the same arguments as in (3.8)–(3.10), the remainder
\[ R_{n,j}^x(y) = \frac{1}{2} \int_0^y \frac{d^3 h_{n,j}^x(t)}{d^3 y} (y-t)^2 dt \]
satisfies
\[ |R_{n,j}^x(y)| \leq \frac{1}{6} \left| \frac{d^3 h_{n,j}^x}{d^3 y} \right| \infty |y|^3 = \frac{1}{6} |f'''| \infty |y|^3, \quad (3.16) \]
\[ |R_{n,j}^x(y)| \leq \frac{2}{d^2 h_{n,j}^x}{d^2 y} \infty y^2 = |f'''| \infty y^2, \quad (3.17) \]
where $|\phi|_\infty := \sup_x |\phi(x)|$ denotes the supremum norm of a function $\phi$, and we can assume without loss of generality that $|f|_\infty, |f'|_\infty, |f''|_\infty$ and $|f'''|_\infty$ are all bounded by 1. Similar to (3.11), for any $x \in \mathbb{R}^n$ and any random variable $Y$ with $E[Y] = 0$ and $E[Y^2] = \sigma_2^2$, we have
\[ \left| E[h_{n,j}^x(Y)] - h_{n,j}^x(0) - \frac{\sigma_2^2}{2} \frac{d^2 h_{n,j}^x}{d^2 y}(0) \right| \leq E[\min\{Y^2, |Y|^3/6\}] \]
In particular, for each $1 \leq j \leq n$, by first conditioning on $X_{n,i} \text{ and } W_{n,i}$ with $i \neq j$, we obtain
\[ |E[g_n(X_{n,1}, \ldots, X_{n,j}, W_{n,j+1}, \ldots, W_{n,n})] - E[g_n(X_{n,1}, \ldots, X_{n,j-1}, W_{n,j}, \ldots, W_{n,n})]| \leq E[\min\{X_{n,j}^2, |X_{n,j}|^3\}] + E[\min\{W_{n,j}^2, |W_{n,j}|^3\}], \]
and hence
\[ \left| E[g_n(X_{n,1}, \ldots, X_{n,n})] - E[g_n(W_{n,1}, \ldots, W_{n,n})] \right| \leq \sum_{j=1}^n E[\min\{X_{n,j}^2, |X_{n,j}|^3\}] + \sum_{j=1}^n E[\min\{W_{n,j}^2, |W_{n,j}|^3\}], \quad (3.18) \]
As shown in the calculations leading to (3.12), Lindeberg’s condition (3.6) implies that the first sum tends to 0 as $n \to \infty$. Since $W_{n,j}$ are Gaussian random variables with the same mean and variance as $X_{n,j}$, using the bound
\[ E[W_{n,j}^2 1\{|W_{n,j}| > \epsilon\}] \leq \epsilon^{-2} E[W_{n,j}^4] = 3\epsilon^{-2} \sigma_{n,j}^4, \]
it is easily seen that $(W_{n,j})_{1 \leq j \leq n}$ also satisfies Lindeberg’s condition, and hence the second sum in (3.18) also tends to 0 as $n \to \infty$, which concludes the proof that $\sum_{j=1}^n X_{n,j}$ converges in distribution to a standard normal random variable.
\[ \blacksquare \]
As a last remark, we note that neither characteristic functions nor Lindeberg’s method can be easily adapted to prove CLT for sums of dependent random variables. There is however a method which is particularly suited for proving CLT for dependent random variables, known as Stein’s method, which also gives bounds on the rate of convergence. Stein’s method has been developed also for Poisson distribution, exponential distribution, etc. See e.g. [2] for a recent survey.

References
