Chapter II
The Central Limit Theorem

In the preceding chapter we dealt with averages of random variables and showed that, in great generality, those averages converge almost surely or in probability to a constant. At last when all the random variables have the same distribution and have moments of all orders, one way of rationalizing this phenomenon is to recognize that the mean-value is conserved whereas all higher moments are driven to 0 unless that are infinite. Of course, the reason why it is easy to conserve the first moment is that mean of the sum is the sum of the means. Thus, if one is going to attempt to find a simple normalization procedure which conserves a quantity involving more than the mean-value, one should seek a quantity which shares this additivity property.

With this in mind, one is led ask what happens if one normalizes in a way which conserves the variance. For this purpose, suppose that $\{X_n : n \in \mathbb{Z}^+\}$ is a sequence of mutually independent, identically distributed random variables with mean-value 0 and variance 1, and set $S_n = \sum_1^n X_k$. Then $\breve{S}_n \equiv n^{-\frac{1}{2}} S_n$ again has mean-value 0 and variance 1. Because of Theorem 1.5.9, we know that, with probability 1, $\overline{\lim}_{n\to\infty} \breve{S}_n = \infty = -\underline{\lim}_{n\to\infty} \breve{S}_n$. Hence, from the point of view of either almost sure convergence or even convergence in probability, there is no hope that $\breve{S}_n$ will converge.

Nonetheless, the random variables $\{\breve{S}_n : n \geq 1\}$ possess remarkable stability when viewed from a distributional perspective. Indeed, if the X_n's are Gaussian, then so are the $\breve{S}_n$'s, and therefore $\breve{S}_n \in \mathcal{N}(0,1)$ for all $n \geq 1$. More generally, even if the X_n's are not Gaussian, controlling their mean-value and variance forces *all* their moments to stabalize. To be precise, trivially, $L_1 \equiv \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\breve{S}_n] = 0$ and $L_2 \equiv \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\breve{S}_n^2] = 1$. Next, assume that $L_\ell \equiv \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\breve{S}_n^\ell]$ exists for $1 \leq \ell \leq m$, where $m \geq 2$. We will show that that $L_{m+1} \equiv \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}[\breve{S}_n^{m+1}]$ exists and is equal to mL_{m-1}. To this end, first note that, since $\mathbb{E}^{\mathbb{P}}[X_n] = 0$,

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[S_n^{m+1}\big] &= n\mathbb{E}^{\mathbb{P}}\big[X_n\big(X_n + S_{n-1}\big)^m\big] = n\sum_{j=0}^m \binom{m}{j}\mathbb{E}^{\mathbb{P}}\big[X_n^{j+1}\big]\mathbb{E}^{\mathbb{P}}\big[S_{n-1}^{m-j}\big] \\ &= nm\mathbb{E}^{\mathbb{P}}\big[S_{n-1}^{m-1}\big] + n\sum_{j=2}^m \binom{m}{j}\mathbb{E}^{\mathbb{P}}\big[X_n^{j+1}\big]\mathbb{E}^{\mathbb{P}}\big[S_{n-1}^{m-j}\big].\end{aligned}$$

Thus, after dividing through by $n^{\frac{m+1}{2}}$, one gets the desired conclusion. Starting from $L_0 = 0$ and $L_1 = 1$, one can use induction to check that $L_{2m-1}0$ and $L_{2m} = \prod_{\ell=1}^m (2\ell - 1) = \frac{(2m)!}{2^m m!}$ for all $m \in \mathbb{Z}^+$. That is

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\breve{S}_n^{2m-1}\big] = 0 \quad \text{and} \quad \lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\breve{S}_n^{2m}\big] = \prod_{\ell=1}^{m}(2\ell-1) = \frac{(2m)!}{2^m m!},$$

for all $m \in \mathbb{Z}^+$. In other words, at least when X_n's has moments of all orders, $\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\breve{S}_n^m\big]$ exists and is independent of the particular choice of random variables. In partacular, since for the Gaussian case, $\mathbb{E}^{\mathbb{P}}[\breve{S}_n^m] = \mathbb{E}^{\mathbb{P}}[X_1^m]$, we conclude that all moments of the $\breve{S}_n$'s converge to the corresponding moments of a standard normal random variable.

In this chapter, we will see that the preceding stabaliztion result is just one manifestation of a general principle known as the *Central Limit phenomenon.*

§2.1 The Theorem of Lindeberg

Before beginning, we introduce the notation $\langle \varphi, \mu \rangle$ to denote the integral of a function φ against a measure μ.

Let $\{X_n : n \geq 1\}$ be a sequence of independent, square integrable random variables with mean-value 0, and set $\breve{S}_n = n^{-\frac{1}{2}} \sum_{m=1}^n X_m$. At least when the X_n's are identically distributed and have moments of all orders and variance 1, we just saw that (recall that γ_{m,σ^2} is the distribution of an $\mathcal{N}(m, \sigma^2)$-random variable)

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{S}_n)\big] = \langle \varphi, \gamma_{0,1} \rangle \tag{2.1.1}$$

for any polynomial $\varphi : \mathbb{R} \longrightarrow \mathbb{C}$. In this section, we will prove a result which shows that, under much more general conditions, (2.1.1) holds for all $\varphi \in C^3\big(\mathbb{R};\mathbb{C}\big)$ with bounded second and third order derivatives.

In the following statement,

$$\sigma_m = \sqrt{\operatorname{var}(X_m)} > 0, \quad \Sigma_n = \sqrt{\operatorname{var}(S_n)} = \sqrt{\sum_{m=1}^{n} \sigma_m^2}, \quad \text{and } \breve{S}_n \equiv \frac{S_n}{\Sigma_n}. \tag{2.1.2}$$

Notice that when the X_k's are identically distributed and have variance 1, the $\breve{S}_n$ in (2.1.2) is consistent with the notation used above. Finally, we set

$$r_n = \max_{1\leq m\leq n} \frac{\sigma_m}{\Sigma_n} \quad \text{and} \quad g_n(\epsilon) = \frac{1}{\Sigma_n^2} \sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\Big[X_m^2,\, \big|X_m\big| \geq \epsilon \Sigma_n\Big] \tag{2.1.3}$$

for $\epsilon > 0$. Clearly, in the identically distributed case, $r_n = n^{-\frac{1}{2}} \sigma_1^2$ and

$$g_n(\epsilon) = \sigma_1^{-2} \mathbb{E}^{\mathbb{P}}\Big[X_1^2,\, |X_1| \geq n^{\frac{1}{2}} \sigma_1 \epsilon\Big] \longrightarrow 0 \quad \text{as } n \to \infty \text{ for each } \epsilon > 0.$$

THEOREM 2.1.4 (**Lindeberg**). *Refer to the preceding, and let φ be an element of $C^3(\mathbb{R};\mathbb{R})$ with bounded second and third order derivatives. Then, for each $\epsilon > 0$,*

$$\left|\mathbb{E}^{\mathbb{P}}\Big[\varphi\big(\breve{S}_n\big)\Big] - \langle\varphi,\gamma_{0,1}\rangle\right| \le \left(\frac{\epsilon}{6} + \frac{r_n}{2}\right)\big\|\varphi'''\big\|_{\mathrm{u}} + g_n(\epsilon)\big\|\varphi''\big\|_{\mathrm{u}}. \tag{2.1.5}$$

In particular, because

$$r_n^2 \le \epsilon^2 + g_n(\epsilon), \quad \epsilon > 0, \tag{2.1.6}$$

(2.1.1) holds if $g_n(\epsilon) \longrightarrow 0$ as $n \to \infty$ for each $\epsilon > 0$.

PROOF: Choose $\mathcal{N}(0,1)$-random variables $Y_1, \dots, Y_n$ which are both mutually independent and independent of the X_m's. (After changing the probability spaces, if necessary, this can be done as an application of either Theorem 1.1.7 or Exercise 1.1.12.) Next, set

$$\breve{Y}_k = \frac{\sigma_k Y_k}{\Sigma_n} \quad \text{and} \quad \breve{T}_n = \sum_1^n \breve{Y}_k,$$

and observe that $\breve{T}_n$ is again an $\mathcal{N}(0,1)$-random variable and therefore that

$$\Delta \equiv \left|\mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{S}_n)\big] - \langle\varphi,\gamma_{0,1}\rangle\right| = \left|\mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{S}_n)\big] - \mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{T}_n)\big]\right|.$$

Further, set $\breve{X}_k = \frac{X_k}{\Sigma_n}$, and define

$$U_m = \sum_{1\le k\le m-1} \breve{X}_k + \sum_{m+1\le k\le n} \breve{Y}_k \quad \text{for } 1 \le m \le n,$$

where a sum over the empty set is taken to be 0. It is then clear that

$$\Delta \le \sum_1^n \Delta_m \quad \text{where } \Delta_m \equiv \left|\mathbb{E}^{\mathbb{P}}\big[\varphi\big(U_m + \breve{X}_m\big)\big] - \mathbb{E}^{\mathbb{P}}\big[\varphi\big(U_m + \breve{Y}_m\big)\big]\right|.$$

Moreover, if

$$R_m(\xi) \equiv \varphi\big(U_m + \xi\big) - \varphi(U_m) - \xi\varphi'(U_m) - \tfrac{\xi^2}{2}\varphi''(U_m), \quad \xi \in \mathbb{R},$$

then (because both $\breve{X}_m$ and $\breve{Y}_m$ are independent of U_m and have the same first two moments)

$$\Delta_m = \left|\mathbb{E}^{\mathbb{P}}\Big[R_m\big(\breve{X}_m\big)\Big] - \mathbb{E}^{\mathbb{P}}\Big[R_m\big(\breve{Y}_m\big)\Big]\right| \le \left|\mathbb{E}^{\mathbb{P}}\Big[R_m\big(\breve{X}_m\big)\Big]\right| + \left|\mathbb{E}^{\mathbb{P}}\Big[R_m\big(\breve{Y}_m\big)\Big]\right|.$$

In order to complete the derivation of (2.1.5), note that, by Taylor's Theorem,

$$\big|R_m(\xi)\big| \le \left(\big\|\varphi'''\big\|_{\mathrm{u}}\tfrac{|\xi|^3}{6}\right) \wedge \left(\big\|\varphi''\big\|_{\mathrm{u}}|\xi|^2\right);$$

and therefore, for each $\epsilon > 0$,

$$\begin{aligned}
&\sum_1^n \mathbb{E}^{\mathbb{P}}\Big[\big|R_m(\check{X}_m)\big|\Big] \\
&\quad \le \frac{\|\varphi'''\|_{\mathrm{u}}}{6}\sum_1^n \mathbb{E}^{\mathbb{P}}\Big[|\check{X}_m|^3,\ |X_m| \le \epsilon\Sigma_n\Big] + \|\varphi''\|_{\mathrm{u}}\sum_1^n \mathbb{E}^{\mathbb{P}}\Big[\check{X}_m^2,\ |X_m| \ge \epsilon\Sigma_n\Big] \\
&\quad \le \frac{\epsilon\|\varphi'''\|_{\mathrm{u}}}{6}\sum_1^n \frac{\sigma_m^2}{\Sigma_n^2} + \|\varphi''\|_{\mathrm{u}} g_n(\epsilon) = \frac{\epsilon\|\varphi'''\|_{\mathrm{u}}}{6} + \|\varphi''\|_{\mathrm{u}} g_n(\epsilon);
\end{aligned}$$

while

$$\sum_1^n \mathbb{E}^{\mathbb{P}}\Big[\big|R_m(\check{Y}_n)\big|\Big] \le \frac{\|\varphi'''\|_{\mathrm{u}}}{6}\,\mathbb{E}^{\mathbb{P}}\Big[\big|Y_1\big|^3\Big]\sum_1^n \frac{\sigma_m^3}{\Sigma_n^3} \le \frac{3^{\frac{3}{4}} r_n \|\varphi'''\|_{\mathrm{u}}}{6}.$$

Hence, (2.1.5) is now proved.

Given (2.1.5), all that remains is to prove (2.1.6). However, for any $1 \le m \le n$ and $\epsilon > 0$,

$$\sigma_m^2 = \mathbb{E}^{\mathbb{P}}\Big[X_m^2,\ |X_m| < \epsilon\Sigma_n\Big] + \mathbb{E}^{\mathbb{P}}\Big[X_m^2,\ |X_m| \ge \epsilon\Sigma_n\Big] \le \Sigma_n^2\big(\epsilon^2 + g_n(\epsilon)\big). \quad \square$$

The condition that $g_n(\epsilon) \longrightarrow 0$ for each $\epsilon > 0$ is often called *Lindeberg's condition*, because it was Lindeberg who introduced it and proved that it is a sufficient condition for (2.1.1) for $\varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{C})$. Later, Feller proved that (2.1.1) plus $r_n \to 0$ imply that Lindeberg's condition holds. Together, these two results are known as the **Lindeberg–Feller Theorem**. See Exercise 2.3.20 for a proof of Feller's part.

§2.1.1. The Central Limit Theorem. If one is not concerned about rates of convergence, then the differentiability requirement can be dropped from the last part of Theorem 2.1.4. In order to understand the reason for this, it is helpful to couch the statement of Theorem 2.1.4 entirely in terms of measures. Thus, let μ_n denote the distribution of $\check{S}_n$. Then, under the hypotheses there, Theorem 2.1.4 allows one to say that $\langle\varphi,\mu_n\rangle \longrightarrow \langle\varphi,\gamma_{0,1}\rangle$ for all $\varphi \in C^3(\mathbb{R}^N;\mathbb{C})$ with bounded second and third order derivatives. Because we are dealing with statements about integration and integration is a very forgiving operation, this sort of result self improves. To be more precise, we prove the following lemma.

LEMMA 2.1.7. *Suppose that* $\{\mu_n : n \geq 1\}$ *is a sequence of (non-negative) locally finite** *Borel measures on* $\mathbb{R}^N$ *and that* μ *is a locally finite Borel measure on* $\mathbb{R}^N$ *with the property that* $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ *for all* $\varphi \in C_c^\infty(\mathbb{R}^N;\mathbb{R})$. *Then, for any* $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$, $\langle \psi, \mu \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$. *Moreover, if* $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$ *is* μ_n*-integrable for each* $n \in \mathbb{Z}^+$ *and if* $\langle \psi, \mu_n \rangle \longrightarrow \langle \psi, \mu \rangle \in [0,\infty)$, *then for any sequence* $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ *which converges uniformly on compacts to a* $\varphi \in C(\mathbb{R}^N;\mathbb{C})$ *and satisfies* $|\varphi_n| \leq C\psi$ *for some* $C < \infty$ *and all* $n \geq 1$, $\langle \varphi_n, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$.

PROOF: Choose $\rho \in C_c^\infty\big(B(\mathbf{0},1);[0,\infty)\big)$ with total integral 1, and set $\rho_\epsilon(\mathbf{x}) = \epsilon^{-N}\rho(\epsilon \mathbf{x})$ for $\epsilon > 0$. Also, choose $\eta \in C_c^\infty\big(B(\mathbf{0},2);[0,1]\big)$ so that $\eta = 1$ on $\overline{B(\mathbf{0},1)}$, and set $\eta_R(\mathbf{x}) = \eta(R^{-1}\mathbf{x})$ for $R > 0$.

We begin by noting that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_c^\infty(\mathbb{R}^N;\mathbb{C})$. Next, suppose that $\varphi \in C_c(\mathbb{R}^N;\mathbb{C})$, and, for $\epsilon > 0$, set $\varphi_\epsilon = \rho_\epsilon \star \varphi$, the convolution

$$\int_{\mathbb{R}^N} \rho_\epsilon(\mathbf{x}-\mathbf{y})\varphi(\mathbf{y})\,d\mathbf{y}$$

of ρ_ϵ with φ. Then, for each $\epsilon > 0$, $\varphi_\epsilon \in C_c^\infty(\mathbb{R}^N;\mathbb{C})$ and therefore $\langle \varphi_\epsilon, \mu_n \rangle \longrightarrow \langle \varphi_\epsilon, \mu \rangle$. In addition, there is an $R > 0$ such that $\operatorname{supp}(\varphi_\epsilon) \subseteq B(0,R)$ for all $\epsilon \in (0,1]$. Hence,

$$\overline{\lim_{n\to\infty}} \big|\langle \varphi, \mu_n \rangle - \langle \varphi, \mu \rangle\big| \leq 2\langle \eta_R, \mu \rangle \|\varphi_\epsilon - \varphi\|_{\mathrm{u}}.$$

Since $\lim_{\epsilon \searrow 0} \|\varphi_\epsilon - \varphi\|_{\mathrm{u}} = 0$, we have now shown that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_c(\mathbb{R}^N;\mathbb{C})$.

Now suppose that $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$, and set $\psi_R = \eta_R\psi$, where η_R is as above. Then for each $R > 0$, $\langle \psi_R, \mu \rangle = \lim_{n\to\infty} \langle \psi_R, \mu_n \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$. Hence, by Fatou's Lemma, $\langle \psi, \mu \rangle \leq \underline{\lim}_{R\to\infty} \langle \psi_R, \mu \rangle \leq \underline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle$.

Finally, suppose that $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$ is μ_n-integrable for each $n \in \mathbb{Z}^+$ and that $\langle \psi, \mu_n \rangle \longrightarrow \langle \psi, \mu \rangle \in [0,\infty)$. Given $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ satisfying $|\varphi_n| \leq C\psi$ and converging uniformly on compacts to φ, one has

$$\big|\langle \varphi_n, \mu_n \rangle - \langle \varphi, \mu \rangle\big| \leq \big|\langle \varphi_n - \varphi, \mu_n \rangle + \big|\langle \varphi, \mu_n \rangle - \langle \varphi, \mu \rangle\big|.$$

Moreover, for each $R > 0$,

$$\begin{aligned}
&\overline{\lim_{n\to\infty}} \big|\langle (\varphi_n - \varphi), \mu_n \rangle\big| \\
&\qquad \leq \overline{\lim_{n\to\infty}} \sup_{\mathbf{x} \in B(\mathbf{0},2R)} |\varphi_n(\mathbf{x}) - \varphi(\mathbf{x})| \langle \eta_R, \mu_n \rangle + \overline{\lim_{n\to\infty}} \big|\langle (1-\eta_R)(\varphi_n - \varphi), \mu_n \rangle\big| \\
&\qquad \leq 2C \overline{\lim_{n\to\infty}} \langle (1-\eta_R)\psi, \mu_n \rangle = 2C\langle (1-\eta_R)\psi, \mu \rangle,
\end{aligned}$$

---

* A Borel measure on a topological space is locally finite if it gives finite measure to compacts.

and similarly

$$\begin{aligned}
&\overline{\lim_{n\to\infty}}\big|\langle\varphi,\mu_n\rangle-\langle\varphi,\mu\rangle\big|\\
&\qquad\le\overline{\lim_{n\to\infty}}\big|\langle\eta_R\varphi,\mu_n\rangle-\langle\eta_R\varphi,\mu\rangle\big|+C\,\overline{\lim_{n\to\infty}}\,\langle(1-\eta_R)\psi),\mu_n\rangle+\langle(\psi-\psi_R),\mu\rangle\\
&\qquad=2C\langle(1-\eta_R)\psi,\mu\rangle.
\end{aligned}$$

Finally, because ψ is μ-integrable, $\langle(1-\eta_R)\psi,\mu\rangle\longrightarrow 0$ as $R\to\infty$ by Lebesgue's Dominated Convergence Theorem, and so we are done. □

By combining Theorem 2.1.4 with the preceding, we have the following version of the famous **Central Limit Theorem**.

THEOREM 2.1.8 (**Central Limit Theorem**). *With the setting the same as it was in Theorem 2.1.4, assume that $g_n(\epsilon)\longrightarrow 0$ as $n\to\infty$ for each $\epsilon>0$. Then*

$$\lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\big[\varphi_n(\breve{S}_n)\big]=\langle\varphi,\gamma_{0,1}\rangle$$

whenever $\{\varphi_n:\,n\ge 1\}\subseteq C(\mathbb{R};\mathbb{C})$ satisfies

$$\sup_{n\ge1}\sup_{y\in\mathbb{R}}\frac{\big|\varphi_n(y)\big|}{1+|y|^2}<\infty$$

and tends to φ uniformly on compacts. Moreover, for every pair $-\infty\le a<b\le\infty$,

$$\lim_{n\to\infty}\mathbb{P}\Big(a\le\breve{S}_n\le b\Big)=\frac{1}{\sqrt{2\pi}}\int_a^b\exp\left[-\frac{y^2}{2}\right]\,dy. \tag{2.1.9}$$

(See Exercise 2.1.10 below for more information in the identically distributed case.)

PROOF: Take μ_n to be the distribution of $\breve{S}_n$. By Theorem 2.1.4, we know that $\langle\varphi,\mu_n\rangle\longrightarrow\langle\varphi,\gamma_{0,1}\rangle$ for all $\varphi\in C^\infty_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$. In addition, we know that $\langle\psi,\mu_n\rangle=2=\langle\psi,\gamma_{0,1}\rangle$ when $\psi(y)=1+y^2$. Hence, the first assertion is an application of Lemma 2.1.7.

Turning to the second assertion, let $a<b$ be given. To prove (2.1.9), choose $\{\varphi_k\}_1^\infty\subseteq C_{\mathrm{b}}(\mathbb{R};\mathbb{R})$ and $\{\psi_k\}_1^\infty\subseteq C_{\mathrm{b}}(\mathbb{R};\mathbb{R})$ so that $0\le\varphi_k\nearrow\mathbf{1}_{(a,b)}$ and $1\ge\psi_k\searrow\mathbf{1}_{[a,b]}$ as $k\to\infty$. Then,

$$\underline{\lim_{n\to\infty}}\,\mathbb{P}\Big(a<\breve{S}_n<b\Big)\ge\lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\Big[\varphi_k\big(\breve{S}_n\big)\Big]=\int_{\mathbb{R}}\varphi_k(y)\,\gamma_{0,1}(dy)\longrightarrow\gamma_{0,1}\big((a,b)\big)$$

as $k\to\infty$; and, similarly,

$$\overline{\lim_{n\to\infty}}\,\mathbb{P}\Big(a\le\breve{S}_n\le b\Big)\le\lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\Big[\psi_k\big(\breve{S}_n\big)\Big]=\int_{\mathbb{R}}\psi_k(y)\,\gamma_{0,1}(dy)\longrightarrow\gamma_{0,1}\big([a,b]\big).$$

Finally, note that $\gamma_{0,1}\big((a,b)\big)=\gamma_{0,1}\big([a,b]\big)$. □

Exercises for § 2.1

EXERCISE 2.1.10. Let $\{X_n\}_1^\infty$ be a sequence of independent, identically distributed random variables, define $\{\breve{S}_n : n \in \mathbb{Z}^+\}$ accordingly, and assume that

$$\varlimsup_{n\to\infty} \mathbb{E}^{\mathbb{P}}\Big[\breve{S}_n^2 \wedge R^2\Big] \le 1 \quad \text{for every } R \in [0,\infty).$$

In particular, this will certainly be the case whenever (2.1.1) holds for every $\varphi \in C_c(\mathbb{R};\mathbb{R})$. The purpose of this exercise is to show that the X_n's are square $\mathbb{P}$-integrable, have mean-value 0, and variance no more than 1; and the method which we will use is based on the same line of reasoning as was given in Exercise 1.5.13.

(**i**) Assuming that $X_1 \in L^2(\mathbb{P};\mathbb{R})$, show that $\mathbb{E}^{\mathbb{P}}\big[X_1\big] = 0$ and $\mathbb{E}^{\mathbb{P}}\big[X_1^2\big] \le 1$. In particular, use this together with the result in part (**i**) of Exercise 1.5.12 to see that it suffices to handle the case when the X_n's are symmetric.

(**ii**) In this, and the succeeding parts of this exercise, we will be assuming that the X_n's are symmetric. Following the same route as we took in (**ii**) of Exercise 1.5.12, we set

$$\breve{X}_n^t = X_n \mathbf{1}_{[0,t]}\big(|X_n|\big) - X_n \mathbf{1}_{(t,\infty)}\big(|X_n|\big), \quad n \in \mathbb{Z}^+,$$

and recall that

$$\Big(\breve{X}_1^t, \dots, \breve{X}_n^t, \dots\Big) \quad \text{and} \quad \Big(X_1, \dots, X_n, \dots\Big)$$

have the same distribution for each $t \in (0,\infty)$. Use this together with our basic assumption to see that

$$\lim_{R\to\infty} \sup_{\substack{n\in\mathbb{Z}^+\\ t\in(0,\infty)}} \mathbb{P}\Big(A_n(t,R)\Big) = 0,$$

where

$$A_n(t,R) \equiv \left\{ \left|\sum_1^n X_k\right| \vee \left|\sum_1^n \breve{X}_k^t\right| \ge n^{\frac{1}{2}} R \right\}.$$

(**iii**) Continuing in the setting of part (**ii**), set

$$\breve{S}_n^t = \frac{1}{n^{\frac{1}{2}}} \sum_1^n X_k \mathbf{1}_{[0,t]}\big(|X_k|\big).$$

After noting that the $X_n\mathbf{1}_{[0,t]}\big(|X_n|\big)$'s are symmetric, check (cf. the proof of Theorem 1.3.1) that

$$\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{S}_n^t\big|^4\Big] \le 3t^4.$$

In particular, conclude that, for each $t \in (0,\infty)$, there is an $R(t) \in (0,\infty)$ such that

$$\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{S}_n^t\big|^2,\, A_n\big(t,R(t)\big)\Big] \le 3^{\frac{1}{2}}t^2\mathbb{P}\Big(A_n\big(t,R(t)\big)\Big)^{\frac{1}{2}} \le 1$$

for all $n \in \mathbb{Z}^+$.

(**iv**) Given $t \in (0,\infty)$, choose $R(t) \in (0,\infty)$ as in the preceding. Taking into account the identity

$$\breve{S}_n^t = \frac{\sum_1^n X_k + \sum_1^n \check{X}_k^t}{2n^{\frac{1}{2}}},$$

show that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\Big[X_1^2,\, |X_1| \le t\Big] = \mathbb{E}^{\mathbb{P}}\Big[\big|\breve{S}_n^t\big|^2\Big] &\le \mathbb{E}^{\mathbb{P}}\Big[\big|\breve{S}_n^t\big|^2,\, A_n\big(t,R(t)\big)\complement\Big] + 1\\ &\le \mathbb{E}^{\mathbb{P}}\Big[\breve{S}_n^2 \wedge R(t)^2\Big] + 1\end{aligned}$$

for all $n \in \mathbb{Z}^+$ and $t \in (0,\infty)$. In particular, use this and our basic hypothesis to conclude first that

$$\mathbb{E}^{\mathbb{P}}\Big[X_1^2,\, |X_1| \le t\Big] \le 2$$

for all $t \in (0,\infty)$ and then that X_1 is square $\mathbb{P}$-integrable.

(**v**) After combining the preceding with the Central Limit Theorem, we see that, in the case of independent, identically distributed random variables, X_1 is square $\mathbb{P}$-integrable with $\mathbb{E}^{\mathbb{P}}[X_1] = 0$ and $\mathbb{E}^{\mathbb{P}}\big[X_1^2\big] \le \sigma^2$ if and only if

$$\varlimsup_{n\to\infty} \mathbb{E}^{\mathbb{P}}\Big[\breve{S}_n^2 \wedge R^2\Big] \le \sigma^2 \quad \text{for all } R \in (0,\infty).$$

Exercise 2.1.11. An interesting way in which to interpret The Central Limit Theorem is as the solution to a certain *fixed point problem*. Namely, let $\mathcal{P}$ denote the set of probability measures μ on $\big(\mathbb{R};\mathcal{B}_{\mathbb{R}}\big)$ with the properties that

$$\int_{\mathbb{R}} x^2\,\mu(dx) = 1 \quad \text{and} \quad \int_{\mathbb{R}} x\,\mu(dx) = 0.$$

Next, define $T\mu$ for $\mu \in \mathcal{P}$ to be the probability measure on $\big(\mathbb{R},\mathcal{B}_{\mathbb{R}}\big)$ given by

$$T\mu(\Gamma) = \iint\limits_{\mathbb{R}^2} \mathbf{1}_\Gamma\left(\frac{x+y}{\sqrt{2}}\right)\mu(dx)\mu(dy) \quad \text{for} \quad \Gamma \in \mathcal{B}_{\mathbb{R}}.$$

After checking that T maps $\mathcal{P}$ into itself, use The Central Limit Theorem to show that, for every $\mu \in \mathcal{P}$,

$$\lim_{n\to\infty} \int_{\mathbb{R}} \varphi\, d\, T^n\mu = \int_{\mathbb{R}} \varphi\, d\gamma_{0,1}, \quad \varphi \in C_{\mathrm{b}}(\mathbb{R};\mathbb{C}).$$

Conclude, in particular, that $\gamma_{0,1}$ is the one and only element μ of $\mathcal{P}$ with the property that $T\mu = \mu$ and that this fixed point is *attracting*. (See Exercise 2.2.24 below for more information.)

EXERCISE 2.1.12. Here is another indication of the remarkable stability of normal random variables. Namely, we outline below a derivation* of the **Lévy–Cramér Theorem** which says that if X and Y are independent random variables whose sum is normal (with some mean and variance), then both X and Y are normal.

(**i**) Assume that $X + Y \in \mathcal{N}(a, \sigma^2)$, and, by subtracting a from X, reduce to the case in which $X + Y \in \mathcal{N}(0, \sigma^2)$. Next, show that there is nothing more to do when $\sigma = 0$ and that one can always reduce to the case $\sigma = 1$ when $\sigma > 0$. Thus, from now on, assume that $X + Y \in \mathcal{N}(0, 1)$.

(**ii**) Choose $r \in (0,\infty)$ so that $\mathbb{P}\big(|X| \vee |Y| \geq r\big) \leq \frac{1}{2}$, and conclude (cf. (2.2.11)) that

$$\mathbb{P}\big(|X| \geq r + R\big) \vee \mathbb{P}\big(|Y| \geq r + R\big) \leq 4\exp\left[-\frac{R^2}{2}\right], \quad R \in (0,\infty).$$

In particular, show that the moment generating functions $z \in \mathbb{C} \longmapsto M(z) = \mathbb{E}^{\mathbb{P}}\left[e^{zX}\right] \in \mathbb{C}$ and $z \in \mathbb{C} \longmapsto N(z) = \mathbb{E}^{\mathbb{P}}\left[e^{zY}\right] \in \mathbb{C}$ exist and are entire functions. Further, note that $M(z)N(z) = \exp\left[\frac{z^2}{2}\right]$, and conclude that M and N never vanish. Finally, from the fact that $X + Y$ has mean 0, show that one can reduce to the case in which both X and Y have mean 0. Thus, from now on, we assume that $M'(0) = 0 = N'(0)$.

(**iii**) Because M never vanishes and $M(0) = 1$, elementary complex analysis (cf. Lemma 3.2.3) guarantees that there is a unique entire function $g : \mathbb{C} \longrightarrow \mathbb{C}$ such that $g(0) = 0$ and $M(z) = e^{g(z)}$ for all $z \in \mathbb{C}$. Further, from $M'(0) = 0$, note that $g'(0) = 0$. Thus,

$$g(z) = \sum_{n=2}^{\infty} c_n z^n \quad \text{where } n!c_n = \frac{d^n}{dx^n} \log\left(\mathbb{E}^{\mathbb{P}}\left[e^{xX}\right]\right)\bigg|_{x=0} \in \mathbb{R}.$$

Finally, note that $N(z) = \exp\left[\frac{z^2}{2} - g(z)\right]$.

---

* This derivation is based on a note by Z. Sasvári, who himself barrowed some of the ideas from A. Rényi. I know of no derivation which does not rely on complex analysis and would be very interested in learning one.

(**iv**) As an application of Hölder's inequality, observe that $x \in \mathbb{R} \longmapsto g(x) \in \mathbb{R}$ and $x \in \mathbb{R} \longmapsto \frac{x^2}{2} - g(x) \in \mathbb{R}$ are both convex. Thus, since $g'(0) = 0$, both these functions are nonincreasing on $(-\infty, 0]$ and nondecreasing on $[0, \infty)$. Use this observation to first check that

$$g(x) \geq 0 \leq \frac{x^2}{2} - g(x) \quad \text{for all } x \in \mathbb{R}.$$

Next, use the preceding in conjunction with the trivial remarks

$$\exp\Big[\mathfrak{Re}\big(g(z)\big)\Big] = \big|\mathbb{E}^{\mathbb{P}}\left[e^{zX}\right]\big| \leq e^{g(x)}$$

and

$$\exp\Big[\mathfrak{Re}\big(\tfrac{z^2}{2} - g(z)\big)\Big] = \big|\mathbb{E}^{\mathbb{P}}\left[e^{zY}\right]\big| \leq \exp\left[\tfrac{x^2}{2} - g(x)\right],$$

to arrive at

$$-y^2 \leq 2\mathfrak{Re}\big(g(z)\big) \leq x^2 \quad \text{for } z = x + \sqrt{-1}\,y \in \mathbb{C}.$$

In particular, this means that

$$\Big|\mathfrak{Re}\big(g(z)\big)\Big| \leq \frac{|z|^2}{2}, \quad z \in \mathbb{C}.$$

(**v**) To complete the program, observe that, for each $n \in \mathbb{Z}^+$ and $r > 0$, on the one hand

$$c_n r^n = \frac{1}{2\pi}\int_0^{2\pi} g\left(re^{\sqrt{-1}\,\theta}\right) e^{-\sqrt{-1}\,n\theta}\,d\theta, \quad r > 0,$$

while, on the other hand (since $\overline{g(z)} = g\big(\bar{z}\big)$),

$$0 = \int_0^{2\pi} \overline{g\left(re^{\sqrt{-1}\,\theta}\right)}\, e^{-\sqrt{-1}\,n\theta}\,d\theta.$$

Hence,

$$c_n r^n = \frac{1}{\pi}\int_0^{2\pi} \mathfrak{Re}\left(g\left(re^{\sqrt{-1}\,\theta}\right)\right) e^{-\sqrt{-1}\,n\theta}\,d\theta, \quad n \in \mathbb{Z}^+ \text{ and } r > 0.$$

Finally, in combination with the estimate obtained in (**ii**) and $c_0 = c_1 = 0$, this leads to the conclusion that $c_n = 0$ for $n \neq 2$ and therefore that $g(z) = c_2 z^2$ with $0 \leq c_2 \leq \frac{1}{2}$.

EXERCISE 2.1.13. An important result which is closely related to The Central Limit Theorem is the following observation, which occupies a central position in the development of classical statistical mechanics.*

For each $n \in \mathbb{Z}^+$, let λ_n denote the normalized surface measure on the $n-1$ dimensional sphere

$$\mathbb{S}^{n-1}\big(\sqrt{n}\big) = \big\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| = n^{\frac{1}{2}}\big\},$$

and denote by $\lambda_n^{(1)}$ the distribution of the coordinate x_1 under λ_n. Check that, when $n \geq 2$, $\mu_n(dt) = f_n(t)\,dt$, where

$$f_n(t) = \frac{\boldsymbol{\omega}_{n-2}}{n^{\frac{1}{2}}\boldsymbol{\omega}_{n-1}}\left(1 - \frac{t^2}{n}\right)^{\frac{n-3}{2}} \mathbf{1}_{(-1,1)}\big(n^{-\frac{1}{2}}t\big),$$

and $\boldsymbol{\omega}_{k-1}$ denotes the surface area of the $(k-1)$-dimensional unit sphere in $\mathbb{R}^k$. Using polar coordinates to compute the right-hand side of

$$(2\pi)^{\frac{k}{2}} = \int_{\mathbb{R}^k} e^{-\frac{|\mathbf{x}|^2}{2}}\,dx,$$

first check that

$$\boldsymbol{\omega}_{k-1} = \frac{2\pi^{\frac{k}{2}}}{\Gamma\big(\frac{k}{2}\big)},$$

where $\Gamma(t)$ is Euler's Γ-function (cf. (1.3.20), and then apply Stirling's formula (cf. (1.3.21)) to see that

$$\frac{\boldsymbol{\omega}_{n-2}}{n^{\frac{1}{2}}\boldsymbol{\omega}_{n-1}} \longrightarrow \frac{1}{\sqrt{2\pi}} \quad \text{as} \quad n \to \infty.$$

Now, using g to denote the density for the standard Gauss distribution (i.e., the Gauss kernel in (1.3.5)), apply these computations to show that

$$\sup_{n\geq 3}\sup_{t\in\mathbb{R}} \frac{f_n(t)}{g(t)} < \infty \quad \text{and that} \quad \frac{f_n(t)}{g(t)} \longrightarrow 1 \text{ uniformly on compacts.}$$

In particular, conclude that, for any $\varphi \in L^1(\gamma_{0,1};\mathbb{R})$:

$$\int_{\mathbb{R}} \varphi\,d\lambda_n^{(1)} \longrightarrow \int_{\mathbb{R}} \varphi\,d\gamma_{0,1}. \tag{2.1.14}$$

* Although E. Borel seems to have thought he was the first to discover this result and rhapsodizes about it a bit in "Sur les principes de la cinétique des gaz," *Ann. l'École Norm. sup.*, 3[e] **t. 23**, and probably was the first one to see its significance for statistical mechanics, it appears already in the 1866 article "Über die Entwicklungen einer Funktion von beliebig vielen Variabeln nach Laplaceshen Funktionen höherer Ordnung," *J. Reine u. Angewandte Math.* by F. Mehler. Actually, the preceding is only a small part of what Mehler discovered.

A less computational approach to the same question is the following. Let $X_1, \dots, X_n, \dots$ be a sequence of independent $\mathcal{N}(0,1)$ random variables, and set $R_n = \sqrt{X_1^2 + \cdots + X_n^2}$. First note that $\mathbb{P}(R_n = 0) = 0$ and then that the distribution of

$$\boldsymbol{\theta}_n \equiv \frac{n^{\frac{1}{2}}\big(X_1, \dots, X_n\big)}{R_n}$$

is λ_n. Next, use the Strong Law of Large Numbers to see that $\frac{R_n^2}{n} \longrightarrow 1$ (a.s., $\mathbb{P}$) and conclude that, for any $N \in \mathbb{Z}^+$,

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\Big[\varphi\big(\boldsymbol{\theta}_n^{(N)}\big)\Big] = \mathbb{E}^{\mathbb{P}}\Big[\varphi\big(X_1, \dots, X_N\big)\Big], \quad \varphi \in C_{\mathrm{c}}\big(\mathbb{R}^N; \mathbb{R}\big),$$

where, for $n \geq N$, $\boldsymbol{\theta}_n^{(N)} \in \mathbb{R}^N$ denotes the projection of $\boldsymbol{\theta}_n \in \mathbb{R}^n$ onto its first N coordinates. Conclude that if $\lambda_n^{(N)}$ on $\big(\mathbb{R}^N, \mathcal{B}_{\mathbb{R}^N}\big)$ denotes the distribution of $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n \longmapsto \mathbf{x}^{(N)} \equiv \big(x_1, \dots, x_N\big) \in \mathbb{R}^N$ under λ_n, then

$$\lim_{n\to\infty} \int_{\mathbb{R}^N} \varphi \, d\lambda_n^{(N)} = \int_{\mathbb{R}^N} \varphi \, d\gamma_{0,1}^N \quad \text{for all} \quad \varphi \in C_{\mathrm{b}}\big(\mathbb{R}^N; \mathbb{C}\big).$$

In particular, by considering the case when $N = 2$, show that, for any $\varphi \in C_{\mathrm{b}}(\mathbb{R}; \mathbb{R})$,

$$\lim_{n\to\infty} \int\limits_{\mathbf{S}^{n-1}(\sqrt{n})} \left(\frac{1}{n} \sum_{k=1}^n \varphi\big(x_k\big) - \int_{\mathbb{R}} \varphi \, d\gamma_{0,1} \right)^2 \lambda_n(d\mathbf{x}) = 0. \tag{2.1.15}$$

Notice that the noncomputational argument has the advantage that it immediately generalizes the earlier result to cover $\lambda_n^{(N)}$ for all $N \in \mathbb{Z}^+$, not just $N = 1$ (cf. Exercise 2.3.24). On the other hand, the conclusion is weaker in the sense that convergence of the densities has been replaced by convergence of integrals with bounded continuous integrands and that no estimate on the rate of convergence is provided. More work is required to restore the stronger statements.

When couched in terms of statistical mechanics, this result can be interpreted as a derivation of the Maxwell distribution of velocities for an **ideal gas** of free particles of mass 2 and having average energy 1.

Exercise 2.1.16. The most frequently encountered applications of Stirling's formula (cf. (1.3.21)) are to cases when $t \in \mathbb{Z}^+$. That is, one is usually interested in the formula

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n. \tag{2.1.17}$$

Here is a derivation of (2.1.17) as an application of the Central Limit Theorem. Namely, take $\{X_n\}_1^\infty$ to be a sequence of independent, random variables with $\mathbb{P}\big(X_n > x\big) = \exp\big(-(x+1)^+\big)$, $x \in \mathbb{R}$ for all $n \in \mathbb{Z}^+$. For $n \geq 1$, note that

$$\begin{aligned}\mathbb{P}\left(\breve{S}_{n+1} \in \left[0, \tfrac{1}{4}\right]\right) &= \frac{1}{n!}\int_{1+n}^{1+4^{-1}\sqrt{n}+n} x^n e^{-x}\,dx \\ &= \frac{n^{n+\frac{1}{2}}e^{-n}}{n!}\int_{n^{-\frac{1}{2}}}^{n^{-\frac{1}{2}}+\frac{1}{4}\sqrt{1+n^{-1}}} \big(1+n^{-\frac{1}{2}}y\big)^n e^{-\sqrt{n}\,y}\,dy.\end{aligned}$$

By the Central Limit Theorem,

$$\mathbb{P}\left(\breve{S}_n \in \left[0, \tfrac{1}{4}\right]\right) \longrightarrow \frac{1}{\sqrt{2\pi}}\int_0^{\frac{1}{4}} e^{-\frac{x^2}{2}}\,dx.$$

At the same time, an elementary computation shows that

$$\int_{n^{-\frac{1}{2}}}^{n^{-\frac{1}{2}}+\frac{1}{4}\sqrt{1+n^{-1}}} \big(1+n^{-\frac{1}{2}}y\big)^n e^{-\sqrt{n}\,y}\,dy \longrightarrow \int_0^{\frac{1}{4}} e^{-\frac{x^2}{2}}\,dx,$$

and clearly (2.1.17) follows from these. In fact, if one applies the Berry–Esseen estimate proved in the next section, one finds that

$$\frac{\sqrt{2\pi n}\,\left(\frac{n}{e}\right)^n}{n!} = 1 + \mathcal{O}\big(n^{-\frac{1}{2}}\big).$$

However, this last observation is not very interesting since we saw in Exercise 1.3.19 that the true correction term is of order t^{-1}.*

§ 2.2 The Berry–Esseen Theorem via Stein's Method

As we will see in the next section, the principles underlying the passage from Theorem 2.1.4 to Theorem 2.1.8 are very general. In fact, as we will see in Chapter V, some of these principles can be formulated in such a way that they extend to a very abstract setting. However, before we start delving into such extensions, we will devote this and the following sections to a closer examination of the situation at hand. Specifically, we are going to see how to make the final part of Theorem 2.1.8 quantitative.

From (2.1.5), we get a rate of convergence in terms of the second and third derivatives of φ. In fact, if we assume that

$$\tau_k \equiv \left(\mathbb{E}^{\mathbb{P}}\left[|X_k|^3\right]\right)^{\frac{1}{3}} < \infty, \quad 1 \leq k \leq n, \tag{2.2.1}$$

---

* For more information, see, for example, Wm. Feller's discussion of Stirling's formula in his *Introduction to Probability Theory and Its Applications, Vol. I*, J. Wiley Series in Probability and Math. Stat. (1968).

then (cf. the proof of Theorem 2.1.4) by using the estimates

$$\big|R_m(\xi)\big| \le \frac{\|\varphi'''\|_{\mathrm{u}}|\xi|^3}{6} \quad \text{and} \quad \sigma_k \le \tau_k,$$

one sees that (2.1.5) can be replaced by

$$\left|\mathbb{E}^{\mathbb{P}}\Big[\varphi\big(\breve{S}_n\big)\Big] - \int_{\mathbb{R}} \varphi\, d\gamma\right| \le \frac{2\|\varphi'''\|_{\mathrm{u}}}{3} \frac{\sum_1^n \tau_k^3}{\Sigma_n^3} \tag{2.2.2}$$

when the X_k's have third moments.

Although both (2.1.5) and (2.2.2) are interesting, neither one of them can be used to get very much information about the rate at which the distribution functions

$$x \in \mathbb{R} \longmapsto F_n(x) \equiv P\big(\breve{S}_n \le x\big) \in [0,1] \tag{2.2.3}$$

are tending to the **error function**

$$G(x) \equiv \gamma_{0,1}\big((-\infty, x]\big) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt. \tag{2.2.4}$$

To see how (2.1.5) and (2.2.2) must be modified in order to gain such information, first observe that

$$\begin{aligned} &\int_{\mathbb{R}} \varphi'(x)\big(F_n(x) - G(x)\big)\, dx \\ &\qquad = \mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{S}_n)\big] - \int_{\mathbb{R}} \varphi(y)\, \gamma(dy), \quad \varphi \in C_{\mathrm{b}}^1(\mathbb{R};\mathbb{R}). \end{aligned} \tag{2.2.5}$$

(To see (2.2.5), reduce to the case in which $\varphi \in C_{\mathrm{c}}^1(\mathbb{R};\mathbb{R})$ and $\varphi(0) = 0$; and for this case apply either Fubini's Theorem or integration by parts over the intervals $(-\infty, 0]$ and $[0, \infty)$ separately.) Hence, in order to get information about the distance between F_n and G, we will have to learn how to replace the right-hand sides of (2.1.5) and (2.2.2) with expressions which depend only on the first derivative of φ. For example, if the dependence is on $\|\varphi'\|_{\mathrm{u}}$, then we get information about the $L^1(\mathbb{R};\mathbb{R})$ distance between F_n and G; whereas if the dependence is on $\|\varphi'\|_{L^1(\mathbb{R})}$, then the information will be about the uniform distance between F_n and G.

The basic idea which we will use to get our estimates in terms of φ' was introduced by C. Stein and is an example of a procedure known as **Stein's method**.* In the case at hand, Stein's method rests on the simple observation that if $\psi \in C(\mathbb{R};\mathbb{R})$ has no more than linear growth at infinity, then the only obstruction to finding a boundedly differentiable solution f to the equation $f'(x) - xf(x) = \psi(x)$ is that $\langle \psi, \gamma_{0,1}\rangle = 0$. More precisely, we will use the following.

---

* Stein provided an introduction, by way of examples, to his own method in *Approximate Computation of Expectations*, IMS Lec. Notes & Monograph Series **7** (1986).

Lemma 2.2.6. *Let $\varphi \in C^1(\mathbb{R};\mathbb{R})$, assume that $\|\varphi'\|_{\mathrm{u}} < \infty$, set $\tilde{\varphi} = \varphi - \langle \varphi, \gamma_{0,1}\rangle$, and define*

$$x \in \mathbb{R} \longmapsto f(x) \equiv e^{\frac{x^2}{2}} \int_{-\infty}^{x} \tilde{\varphi}(t) e^{-\frac{t^2}{2}}\, dt. \tag{2.2.7}$$

Then $f \in C^2_{\mathrm{b}}(\mathbb{R};\mathbb{R})$,

$$\|f\|_{\mathrm{u}} \le 2\|\varphi'\|_{\mathrm{u}}, \quad \|f'\|_{\mathrm{u}} \le 3\sqrt{\tfrac{\pi}{2}}\|\varphi'\|_{\mathrm{u}}, \quad \|f''\|_{\mathrm{u}} \le 6\|\varphi'\|_{\mathrm{u}}, \tag{2.2.8}$$

and

$$f'(x) - xf(x) = \tilde{\varphi}(x), \quad x \in \mathbb{R}. \tag{2.2.9}$$

Proof: The facts that $f \in C^1(\mathbb{R};\mathbb{R})$ and that (2.2.9) holds are elementary applications of the Fundamental Theorem of Calculus. Moreover, knowing that $f \in C^1(\mathbb{R};\mathbb{R})$ and using (2.2.9), we see that $f \in C^2(\mathbb{R};\mathbb{R})$ and, in fact, that

$$f''(x) - xf'(x) = f(x) + \varphi'(x), \quad x \in \mathbb{R}. \tag{2.2.10}$$

To prove the estimates in (2.2.8), first note that, because $\tilde{\varphi}$ and therefore f are unchanged when φ is replaced by $\varphi - \varphi(0)$, we may and will assume that $\varphi(0) = 0$ and therefore that $|\varphi(t)| \le \|\varphi'\|_{\mathrm{u}}|t|$. In particular, this means that

$$\left| \int_{\mathbb{R}} \varphi \, d\gamma_{0,1} \right| \le \|\varphi'\|_{\mathrm{u}} \int_{\mathbb{R}} |t|\, \gamma_{0,1}(dt) = \|\varphi'\|_{\mathrm{u}} \sqrt{\tfrac{2}{\pi}}.$$

Next, observe that, because $\int_{\mathbb{R}} \tilde{\varphi}(t) e^{-\frac{t^2}{2}}\, dt = 0$, an alternative expression for f is

$$f(x) = -e^{\frac{x^2}{2}} \int_x^\infty \tilde{\varphi}(t) e^{-\frac{t^2}{2}}\, dt, \quad x \in \mathbb{R}.$$

Thus, by using the original expression for $f(x)$ when $x \in (-\infty, 0)$ and the alternative one when $x \in [0,\infty)$, we see first that

$$|f(x)| \le e^{\frac{x^2}{2}} \int_{|x|}^\infty \bigl|\tilde{\varphi}\bigl(-t\,\mathrm{sgn}(x)\bigr)\bigr| e^{-\frac{t^2}{2}}\, dt, \quad x \in \mathbb{R},$$

and then that

$$|f(x)| \le \|\varphi'\|_{\mathrm{u}} e^{\frac{x^2}{2}} \int_{|x|}^\infty \left(t + \sqrt{\tfrac{2}{\pi}}\right) e^{-\frac{t^2}{2}}\, dt.$$

But, since

$$\frac{d}{dx}\left(e^{\frac{x^2}{2}} \int_x^\infty e^{-\frac{t^2}{2}}\, dt \right) \le e^{\frac{x^2}{2}} \int_x^\infty t\, e^{-\frac{t^2}{2}}\, dt - 1 = 0 \quad \text{for } x \in [0,\infty),$$

we have that

$$e^{\frac{x^2}{2}}\int_{|x|}^{\infty} te^{-\frac{t^2}{2}}\,dt = 1 \text{ and } e^{\frac{x^2}{2}}\int_{|x|}^{\infty} e^{-\frac{t^2}{2}}\,dt \le \sqrt{\tfrac{\pi}{2}}, \quad x\in\mathbb{R}; \tag{2.2.11}$$

which means that we have now proved the first estimate in (2.2.8). To prove the other two estimates there, derive from (*) that

$$\frac{d}{dx}\Big(e^{-\frac{x^2}{2}}f'(x)\Big) = e^{-\frac{x^2}{2}}\big(f(x)+\varphi'(x)\big)$$

and therefore that

$$\begin{aligned} f'(x) &= e^{\frac{x^2}{2}}\int_{-\infty}^{x}\big(f(t)+\varphi'(t)\big)e^{-\frac{t^2}{2}}\,dt \\ &= -e^{\frac{x^2}{2}}\int_{x}^{\infty}\big(f(t)+\varphi'(t)\big)e^{-\frac{t^2}{2}}\,dt, \quad x\in\mathbb{R}. \end{aligned}$$

Thus, reasoning as we did above and using the first estimate in (2.2.8) and the identities in (2.2.11), (2.2.9), and (2.2.9), we arrive at the second and third estimates in (2.2.8). □

We now have the ingredients needed to apply Stein's method to the following example of a Berry–Esseen sort of estimate.

Theorem 2.2.12 (**$\mathbf{L^1}$-Berry–Esseen Estimate**). *Continuing in the setting of Theorem 2.1.4, one has that for all $\epsilon>0$ (cf. (2.1.3), (2.2.3), and (2.2.4))*

$$\big\|F_n - G\big\|_{L^1(\mathbb{R};\mathbb{R})} \le 6(r_n+\epsilon) + 3\sqrt{2\pi}\,g_n(2\epsilon). \tag{2.2.13}$$

Moreover, if (cf. (2.2.1)) $\tau_m<\infty$ for each $1\le m\le n$, then

$$\big\|F_n - G\big\|_{L^1(\mathbb{R};\mathbb{R})} \le \left(6r_n + \frac{3\sum_{m=1}^n \tau_m^3}{\Sigma_n^3}\right)\wedge\left(\frac{9\sum_{m=1}^n\tau_m^3}{\Sigma_n^3}\right). \tag{2.2.14}$$

In particular, if $\sigma_m^2=1$ and $\tau_m\le\tau<\infty$ for each $1\le m\le n$, then

$$\big\|F_n - G\big\|_{L^1(\mathbb{R};\mathbb{R})} \le \frac{6+2\tau^3}{\sqrt{n}} \le \frac{8\tau^3}{\sqrt{n}}.$$

Proof: Let $\varphi\in C^1(\mathbb{R};\mathbb{R})$ with bounded first derivative be given, and define f accordingly, as in (2.2.7). Everything turns on the equality in (2.2.9). Indeed, because of that equality, we know that the right-hand side of (2.2.5) is equal to

$$\mathbb{E}^{\mathbb{P}}\big[f'(\check{S}_n)\big] - \mathbb{E}^{\mathbb{P}}\big[\check{S}_n f(\check{S}_n)\big] = \sum_{m=1}^{n}\Big(\check{\sigma}_m^2\mathbb{E}^{\mathbb{P}}\big[f'(\check{S}_n)\big] - \mathbb{E}^{\mathbb{P}}\big[\check{X}_m f(\check{S}_n)\big]\Big),$$

where we have set $\breve{\sigma}_m = \frac{\sigma_m}{\Sigma_n}$ and $\breve{X}_m = \frac{X_m}{\Sigma_n}$. Next, define

$$\breve{T}_{n,m}(t) = \breve{S}_n + (t-1)\breve{X}_m \quad \text{for } t \in [0,1],$$

note that $\breve{T}_{n,m}(0)$ is independent of $\breve{X}_m$, and conclude that

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m f(\breve{S}_n)\Big] &= \int_0^1 \mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^2 f'\big(\breve{T}_{n,m}(t)\big)\Big]\,dt \\ &= \breve{\sigma}_m^2 \mathbb{E}^{\mathbb{P}}\Big[f'\big(\breve{T}_{n,m}(0)\big)\Big] + \int_0^1 \mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^2\big(f'\big(\breve{T}_{n,m}(t)\big) - f'\big(\breve{T}_{n,m}(0)\big)\Big]\,dt\end{aligned}$$

for each $1 \le m \le n$. Hence, we now see that

$$\mathbb{E}^{\mathbb{P}}\big[\varphi(\breve{S}_n)\big] - \int_{\mathbb{R}} \varphi\,d\gamma = \sum_{m=1}^n \breve{\sigma}_m^2 A_m - \sum_{m=1}^n \int_0^1 B_m(t)\,dt \tag{2.2.15}$$

where

$$A_m \equiv \mathbb{E}^{\mathbb{P}}\Big[f'\big(\breve{S}_n\big) - f'\big(\breve{T}_{n,m}(0)\big)\Big]$$

and

$$B_m(t) \equiv \mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^2\Big(f'(\breve{T}_{n,m}(t)) - f'\big(\breve{T}_{n,m}(0)\big)\Big)\Big].$$

Obviously, by Taylor's Theorem and Hölder's inequality, for each $1 \le m \le n$,

$$|A_m| \le \breve{\sigma}_m \|f''\|_{\mathrm{u}} \le \left(r_n \wedge \frac{\tau_m}{\Sigma_n}\right) \|f''\|_{\mathrm{u}} \tag{*}$$

while, for each $t \in [0,1]$ and $\epsilon > 0$,

$$\big|B_m(t)\big| \le 2\epsilon t \breve{\sigma}_m^2 \|f''\|_{\mathrm{u}} + 2\frac{\|f'\|_{\mathrm{u}}}{\Sigma_n^2}\mathbb{E}^{\mathbb{P}}\Big[X_m^2,\,|X_m| \ge 2\epsilon\Sigma_n\Big].$$

Thus, after summing over $1 \le m \le n$, integrating with respect to $t \in [0,1]$, and using (2.2.5), (2.2.15), and (*), we arrive at

$$\left|\int_{\mathbb{R}} \varphi'(x)\big(F_n(x) - G(x)\big)\,dx\right| \le \big(r_n + \epsilon\big)\|f''\|_{\mathrm{u}} + 2g_n(2\epsilon)\|f'\|_{\mathrm{u}},$$

which, in conjunction with the estimates in (2.2.8), leads immediately to the one in (2.2.13). In order to get (2.2.14), simply note that

$$\big|B_m(t)\big| \le t\int_0^1 \mathbb{E}^{\mathbb{P}}\Big[|\breve{X}_m|^3\big|f''\big(\breve{T}_{n,m}(st)\big)\big|\Big]\,ds \le t\|f''\|_{\mathrm{u}}\frac{\tau_m^3}{\Sigma_n^3},$$

and again use (2.2.15), (2.2.8), and (*). □

§ **2.2.1. The Classical Berry–Esseen Theorem.** The result in Theorem 2.2.12 is already significant. However, it is not the classical Berry–Esseen Theorem, which is the analogous statement about $\|F_n - G\|_{\mathrm{u}}$.

In order to prove the classical result via Stein's method, we must learn how how to replace the $\|\varphi'''\|_{\mathrm{u}}$ in Lindeberg's Theorem by $\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}$. It turns out that this replacement is far more challenging than replacing $\|\varphi'''\|_{\mathrm{u}}$ by $\|\varphi'\|_{\mathrm{u}}$, which was the replacement needed to prove Theorem 2.2.12. The argument which we will use is a clever inductive procedure which was introduced into this context by E. Bolthausen.* But, before we can apply Bolthausen's argument, we will need the following variation on Lemma 2.2.6.

LEMMA 2.2.16. *Let $\varphi \in C^1(\mathbb{R};\mathbb{R})$, and define f accordingly, as in (2.2.7). Then $\|f\|_{\mathrm{u}} \le \sqrt{\frac{\pi}{8}}\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}$ and $\|f'\|_{\mathrm{u}} \le \|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}$.*

PROOF: We will assume, throughout, that $\|\varphi'\|_1 = 1$. Next, observe that, by the Fundamental Theorem of Calculus,

$$\tilde\varphi(x) = -\int_{\mathbb{R}} \tilde\varphi_y(x)\,\varphi'(y)\,dy, \quad \text{where } \varphi_y = \mathbf{1}_{(-\infty,y]},$$

and so (cf. (2.2.4))

$$f(x) = -\int_{\mathbb{R}} \psi_y(x)\,\varphi'(y)\,dy, \quad \text{where } \psi_y(x) = \sqrt{2\pi}e^{\frac{x^2}{2}}\big(G(x\wedge y) - G(x)G(y)\big) \ge 0.$$

At the same time, these, together with (2.2.9), give

$$f'(x) = -\int_{\mathbb{R}} \Big(x\psi_y(x) + \tilde\varphi_y(x)\Big)\,\varphi'(y)\,dy.$$

Hence, the desired estimates come down to checking that

$$e^{\frac{x^2}{2}}\big(G(x\wedge y) - G(x)G(y)\big) \le \frac{1}{4},$$

and

$$\Big|\sqrt{2\pi}xe^{\frac{x^2}{2}}\Big(G(x\wedge y) - G(x)G(y)\Big) + \mathbf{1}_{(-\infty,y]}(x) - G(y)\Big| \le 1$$

for all $(x,y) \in \mathbb{R}\times\mathbb{R}$. But

$$G(x\wedge y) - G(x)G(y) \le G(x) - G(x)^2 = \frac{1}{4}\Big(1 - 4\big(G(x) - \tfrac{1}{2}\big)^2\Big)$$

* The Berry–Esseen Theorem appears as a warm-up exercise in Bolthausen's "An estimate of the remainder term in a combinatorial central limit theorem," *Z. Wahr. Gebiete* **66**.

and

$$\begin{aligned}\left(G(x)-\tfrac{1}{2}\right)^2 &= \frac{1}{2\pi}\left(\int_0^{|x|} e^{-\frac{\xi^2}{2}}\,d\xi\right)^2\\ &\geq \frac{1}{8\pi}\iint\limits_{\xi^2+\eta^2\leq x^2} e^{-\frac{\xi^2+\eta^2}{2}}\,d\xi d\eta = \frac{1}{4}\Big(1-e^{-\frac{x^2}{2}}\Big),\end{aligned}$$

which proves the first inequality. To get the second one, it suffices to do consider each of the four cases $0\leq x\leq y$, $x\geq 0$ & $y<x$, $y<x<0$, and $x<0$ & $y\geq x$ separately and note that, from the first part of (2.2.11),

$$x\geq 0 \implies \sqrt{2\pi}xe^{\frac{x^2}{2}}\big(1-G(x)\big)\leq 1 \quad\text{and}\quad x<0\implies \sqrt{2\pi}|x|e^{\frac{x^2}{2}}G(x)\leq 1. \qquad \square$$

THEOREM 2.2.17 (**Classical Berry–Esseen Estimate**). *Let everything be as in Theorem 2.1.4, and assume that (cf. (2.2.1)) $\tau_m<\infty$ for each $1\leq m\leq n$. Then (cf. (2.2.3) and (2.2.4))*

$$\|F_n-G\|_{\mathrm{u}}\leq 10\,\frac{\sum_1^n \tau_m^3}{\Sigma_n^3}. \tag{2.2.18}$$

In particular, if $\sigma_m=1$ for all $1\leq m\leq n$, then (2.2.14) can be replaced by

$$\|F_n-G\|_{\mathrm{u}}\leq 10\,\frac{\sum_1^n \tau_m^3}{n^{\frac{3}{2}}}\leq 10\,\frac{\max\limits_{1\leq m\leq n}\tau_m^3}{\sqrt{n}}. \tag{2.2.19}$$

PROOF: For each $n\in\mathbb{Z}^+$, let β_n denote the smallest number β with the property that

$$\|F_n-G\|_{\mathrm{u}}\leq \beta\frac{\sum_1^n\tau_m^3}{\Sigma_n^3}$$

for all choices of random variables satisfying the hypotheses under which (2.2.18) is to be proved. Our goal is to give an inductive proof that $\beta_n\leq 10$ for all $n\in\mathbb{Z}^+$; and, because $\Sigma_1\leq\tau_1$ and therefore $\beta_1\leq 1$, we need only be concerned with $n\geq 2$.

Given $n\geq 2$ and $X_1,\dots,X_n$, define $\breve{X}_m$, $\breve{\sigma}_m$, and $\breve{T}_{n,m}(t)$ for $1\leq m\leq n$ and $t\in[0,1]$ as in the proof of Theorem 2.2.12. Next, for each $1\leq m\leq n$, set

$$\Sigma_{n,m}=\sqrt{\Sigma_n^2-\sigma_m^2},\quad \breve{\tau}_m=\frac{\tau_m}{\Sigma_n},\quad \rho_n=\sum_1^n\breve{\tau}_m^3,\quad\text{and } \rho_{n,m}=\sum_{\substack{1\leq\ell\leq n\\ \ell\neq m}}\left(\frac{\tau_\ell}{\Sigma_{n,m}}\right)^3.$$

Finally, set

$$S_{n,m} = \sum_{\substack{1\le \ell\le n\\ \ell\ne m}} X_\ell \quad \text{and} \quad \check{S}_{n,m} = \frac{S_{n,m}}{\Sigma_{n,m}},$$

and let $x \in \mathbb{R} \longmapsto F_{n,m}(x) \equiv P\big(\check{S}_{n,m} \le x\big) \in [0,1]$ denote the distribution function for $\check{S}_{n,m}$. Notice that, by definition, $\|F_{n,m} - G\|_{\mathrm{u}} \le \beta_{n-1}\rho_{n,m}$ for each $1 \le m \le n$. Furthermore, because (cf. (2.1.3))

$$\frac{\Sigma_{n,m}^2}{\Sigma_n^2} = 1 - \check{\sigma}_m^2 \ge 1 - r_n^2 \quad \text{and} \quad \rho_{n,m} \le \left(\frac{\Sigma_n}{\Sigma_{n,m}}\right)^3 \rho_n,$$

we see first that

$$\rho_{n,m} \le \frac{\rho_n}{(1-r_n^2)^{\frac{3}{2}}}, \quad 1 \le m \le n,$$

and therefore that

$$\max_{1\le m\le n} \|F_{n,m} - G\|_{\mathrm{u}} \le \frac{\rho_n \beta_{n-1}}{(1-r_n^2)^{\frac{3}{2}}}. \tag{2.2.20}$$

Now let $\varphi \in C_{\mathrm{b}}^2(\mathbb{R};\mathbb{R})$ with $\|\varphi''\|_{L^1(\mathbb{R})} < \infty$ be given, define f accordingly as in (2.2.7), and let

$$\{A_m : 1 \le m \le n\} \quad \text{and} \quad \{B_m(t) : 1 \le m \le n \;\&\; t \in [0,1]\}$$

be the associated quantities appearing in (2.2.15). By (2.2.9), we have that

$$\begin{aligned}
|A_m| &\le \Big|\mathbb{E}^{\mathbb{P}}\Big[\check{X}_m f(\check{S}_n)\Big]\Big| + \Big|\mathbb{E}^{\mathbb{P}}\Big[\check{T}_{n,m}(0)\Big(f(\check{S}_n) - f\big(\check{T}_{n,m}(0)\big)\Big)\Big]\Big| \\
&\qquad + \Big|\mathbb{E}^{\mathbb{P}}\Big[\varphi(\check{S}_n) - \varphi\big(\check{T}_{n,m}(0)\big)\Big]\Big| \\
&\le \mathbb{E}^{\mathbb{P}}\Big[\big|\check{X}_m\big|\Big]\|f\|_{\mathrm{u}} + \mathbb{E}^{\mathbb{P}}\Big[\big|\check{X}_m\,\check{T}_{n,m}(0)\big|\Big]\|f'\|_{\mathrm{u}} \\
&\qquad + \int_0^1 \Big|\mathbb{E}^{\mathbb{P}}\Big[\check{X}_m \varphi'\big(\check{T}_{n,m}(\xi)\big)\Big]\Big|\,d\xi \\
&\le \check{\sigma}_m\left(\|f\|_{\mathrm{u}} + \frac{\Sigma_{n,m}}{\Sigma_n}\|f'\|_{\mathrm{u}}\right) + \max_{\xi\in[0,1]} \Big|\mathbb{E}^{\mathbb{P}}\Big[\check{X}_m \varphi'\big(\check{T}_{n,m}(\xi)\big)\Big]\Big| \\
&\le \check{\sigma}_m\Big(\|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}}\Big) + \max_{\xi\in[0,1]} \Big|\mathbb{E}^{\mathbb{P}}\Big[\check{X}_m \varphi'\big(\check{T}_{n,m}(\xi)\big)\Big]\Big|.
\end{aligned}$$

Similarly, from (2.2.9)), one sees that $|B_m(t)|$ is dominated by:

$$\begin{aligned}
&t\Big|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^3 f\big(\breve{T}_{n,m}(t)\big)\Big]\Big| + \Big|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^2 \breve{T}_{n,m}(0)\Big(f\big(\breve{T}_{n,m}(t)\big) - f\big(\breve{T}_{n,m}(0)\big)\Big)\Big]\Big| \\
&\qquad + \Big|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^2\Big(\varphi\big(\breve{T}_{n,m}(t)\big) - \varphi\big(\breve{T}_{n,m}(0)\big)\Big)\Big]\Big| \\
&\quad \le t\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{X}_m\big|^3\Big]\|f\|_{\mathrm{u}} + t\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{X}_m\big|^3\Big]\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{T}_{n,m}(0)\big|\Big]\|f'\|_{\mathrm{u}} \\
&\qquad + t\int_0^1 \Big|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^3 \varphi'\big(\breve{T}_{n,m}(t\xi)\big)\Big]\Big|\,d\xi \\
&\quad \le t\breve{\tau}_m^3\Big(\|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}}\Big) + t\max_{\xi\in[0,1]}\Big|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^3 \varphi'\big(\breve{T}_{n,m}(\xi)\big)\Big]\Big|
\end{aligned}$$

In order to handle the second term in the last line of each of these calculations, we introduce the function

$$(\xi,\omega,y) \in [0,1]\times\Omega\times\mathbb{R} \longmapsto \psi(\xi,\omega,y) \equiv \varphi'\left(\xi\breve{X}_m(\omega) + \frac{\Sigma_{n,m}}{\Sigma_n}\,y\right).$$

Next, because $\breve{X}_m$ is independent of $\breve{T}_{n,m}(0)$,

$$\begin{aligned}
&\left|\mathbb{E}^{\mathbb{P}}\Big[\breve{X}_m^k \varphi'\big(\breve{T}_{n,m}(\xi)\big)\Big] - \int_\Omega \breve{X}_m(\omega)^k\left(\int_{\mathbb{R}}\psi(\xi,\omega,y)\,\gamma_{0,1}(dy)\right)\mathbb{P}(d\omega)\right| \\
&\quad \le \int_\Omega \big|\breve{X}_m(\omega)\big|^k\left|\int_{\mathbb{R}}\psi(\xi,\omega,y)\,dF_{n,m}(y) - \int_{\mathbb{R}}\psi(\xi,\omega,y)\,dG(y)\right|\mathbb{P}(d\omega) \\
&\quad = \int_\Omega \big|\breve{X}_m(\omega)\big|^k\left|\int_{\mathbb{R}}\psi'(t,\omega,y)\big(G(y) - F_{n,m}(y)\big)\,dy\right|\mathbb{P}(d\omega) \\
&\quad \le \frac{\beta_{n-1}\rho_n}{(1-r_n^2)^{\frac{3}{2}}}\mathbb{E}^{\mathbb{P}}\Big[\big|\breve{X}_m\big|^k\Big]\|\varphi''\|_{L^1(\mathbb{R};\mathbb{R})} \le \frac{\breve{\tau}_m^k\beta_{n-1}\|\varphi''\|_{L^1(\mathbb{R};\mathbb{R})}\rho_n}{(1-r_n^2)^{\frac{3}{2}}}, \quad k\in\{1,3\},
\end{aligned}$$

where we have used $\psi'(t,\omega,y)$ to denote the first derivative of $y \in \mathbb{R} \longmapsto \psi(\xi,\omega,y)$, applied (2.2.5) and (2.2.20), and noted that, for all $(\xi,\omega)\in[0,1]\times\Omega$, $\|\psi'(\xi,\omega,\cdot)\|_{L^1(\mathbb{R})} = \|\varphi''\|_{L^1(\mathbb{R})}$. At the same time, because

$$\|\psi(\xi,\omega,\,\cdot\,)\|_{L^1(\mathbb{R};\mathbb{R})} = \frac{\Sigma_n}{\Sigma_{n,m}}\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})} \quad \text{for all } (\xi,\omega)\in[0,1]\times\Omega,$$

we have that, for each $\xi\in[0,1]$,

$$\left|\int_\Omega \breve{X}_m(\omega)^k\left(\int_{\mathbb{R}}\psi(\xi,\omega,y)\,\gamma(dy)\right)\mathbb{P}(d\omega)\right| \le \frac{\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}\,\breve{\tau}_m^k}{\big(2\pi(1-r_n^2)\big)^{\frac{1}{2}}}.$$

Hence, by combining these estimates, we arrive at

$$|A_m| \le \breve{\tau}_m \left(\|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}} + \frac{\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}}{\big(2\pi(1-r_n^2)\big)^{\frac12}} + \frac{\beta_{n-1}\rho_n}{(1-r_n^2)^{\frac32}}\|\varphi''\|_{L^1(\mathbb{R};\mathbb{R})} \right)$$

and

$$|B_m(t)| \le t\breve{\tau}_m^3 \left(\|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}} + \frac{\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}}{\big(2\pi(1-r_n^2)\big)^{\frac12}} + \frac{\beta_{n-1}\rho_n}{(1-r_n^2)^{\frac32}}\|\varphi''\|_{L^1(\mathbb{R};\mathbb{R})} \right)$$

for all $1 \le m \le n$ and $t \in [0,1]$. After putting these together with (2.2.5) and (2.2.15), we conclude that

$$\begin{aligned} &\left| \int_{\mathbb{R}} \varphi'(y)\big(G(y) - F_n(y)\big)\,dy \right| \\ &\qquad \le \frac32 \Bigg(\|f\|_{\mathrm{u}} + \|f'\|_{\mathrm{u}} \\ &\qquad\qquad + \frac{\|\varphi'\|_{L^1(\mathbb{R};\mathbb{R})}}{\big(2\pi(1-r_n^2)\big)^{\frac12}} + \frac{\beta_{n-1}\|\varphi''\|_{L^1(\mathbb{R};\mathbb{R})}\rho_n}{(1-r_n^2)^{\frac32}} \Bigg)\rho_n. \end{aligned} \tag{2.2.21}$$

We next apply (2.2.21) to a special class of φ's. Namely, set

$$h(x) = \begin{cases} 1 & \text{if} \quad x < 0 \\ 1 - x & \text{if} \quad x \in [0,1] \\ 0 & \text{if} \quad x > 1, \end{cases}$$

and define

$$h_\epsilon(x) = \epsilon^{-1} \int_{\mathbb{R}} \eta\big(\epsilon^{-1}y\big)h(x-y)\,dy \quad \text{for } \epsilon > 0 \text{ and } x \in \mathbb{R},$$

where $\eta \in C_{\mathrm{c}}^\infty\big(\mathbb{R};[0,\infty)\big)$ satisfies $\int_{\mathbb{R}} \eta(y)\,dy = 1$. Finally, let $a \in \mathbb{R}$ be given, and set

$$\varphi_{\epsilon,L}(x) = h_\epsilon\left(\tfrac{x-a}{L\rho_n}\right), \quad x \in \mathbb{R} \text{ and } \epsilon,\ L > 0.$$

It is then an easy matter to check that $\|\varphi'_{\epsilon,L}\|_{L^1(\mathbb{R})} = 1$ while $\|\varphi''_{\epsilon,L}\|_{L^1(\mathbb{R})} \le \frac{2}{L\rho_n}$. Hence, by plugging the estimates from Lemma 2.2.16 into (2.2.21) and then letting $\epsilon \searrow 0$, we find that, for each $L > 0$,

$$\begin{aligned} &\sup_{a\in\mathbb{R}} \left| \frac{1}{L\rho_n} \int_a^{a+L\rho_n} \big(G(y) - F_n(y)\big)\,dy \right| \\ &\qquad \le \frac32 \left(1 + \sqrt{\frac{\pi}{8}} + \frac{1}{\big(2\pi(1-r_n^2)\big)^{\frac12}} + \frac{2\beta_{n-1}}{(1-r_n^2)^{\frac32}L} \right)\rho_n. \end{aligned} \tag{2.2.22}$$

But

$$\frac{1}{L\rho_n}\int_{a-L\rho_n}^{a} F_n(y)\,dy \le F_n(a) \le \frac{1}{L\rho_n}\int_{a}^{a+L\rho_n} F_n(y)\,dy,$$

while

$$0 \le \frac{1}{L\rho_n}\int_{a}^{a+L\rho_n} G(y)\,dy - G(a) = \frac{1}{L\rho_n}\int_{a}^{a+L\rho_n} (a+L\rho_n-y)\,\gamma_{0,1}(dy) \le \frac{L\rho_n}{\sqrt{8\pi}},$$

and, similarly,

$$0 \le G(a) - \frac{1}{L\rho_n}\int_{a-L\rho_n}^{a} G(y)\,dy \le \frac{L\rho_n}{\sqrt{8\pi}}.$$

Thus, from (2.2.22), we first obtain, for each $L \in (0,\infty)$,

$$\|F_n - G\|_{\mathrm{u}} \le \left(\frac{3}{2} + \sqrt{\frac{9\pi}{32}} + \frac{3}{\bigl(8\pi(1-r_n^2)\bigr)^{\frac12}} + \frac{3\beta_{n-1}}{(1-r_n^2)^{\frac32}L} + \frac{L}{(8\pi)^{\frac12}}\right)\rho_n,$$

and then, after minimizing with respect to $L \in (0,\infty)$,

$$\begin{aligned}\|F_n - G\|_{\mathrm{u}} \le \Biggl(\frac{3}{2} + \sqrt{\frac{9\pi}{32}} + \sqrt{\frac{9}{8\pi}}\,\bigl(1-r_n^2\bigr)^{-\frac12} \\ + \sqrt[4]{\frac{18}{\pi}}\,\beta_{n-1}^{\frac12}\bigl(1-r_n^2\bigr)^{-\frac34}\Biggr)\rho_n.\end{aligned} \tag{2.2.23}$$

In order to complete the proof starting from (2.2.23), we have to consider the two cases determined by whether $\rho_n \ge \frac{1}{10}$ or $\rho_n < \frac{1}{10}$. Because $\|F_n - G\|_{\mathrm{u}} \le 1$, it is obvious that we can take $\beta_n \le 10$ in the first case. On the other hand, if $\rho_n \le \frac{1}{10}$ and we assume that $\beta_{n-1} \le 10$, then, because

$$\rho_n = \frac{1}{\Sigma_n^3}\sum_1^n \mathbb{E}^{\mathbb{P}}\bigl[|X_m|^3\bigr] \ge \frac{1}{\Sigma_n^3}\sum_1^n \mathbb{E}^{\mathbb{P}}\bigl[X_m^2\bigr]^{\frac32} = \sum_1^n \breve{\sigma}_m^3 \ge r_n^3,$$

(2.2.23) says that $\|F_n - G\|_{\mathrm{u}} \le 10\rho_n$. Hence, in either case, $\beta_{n-1} \le 10 \implies \beta_n \le 10$. $\square$

It is clear from the preceding derivation (in particular, the final step) that the constant 10 appearing in (2.2.18) and (2.2.19) can be replaced by the smallest $\beta > 1$ which satisfies the equation

$$\beta = \frac{3}{2} + \sqrt{\frac{9\pi}{32}} + \sqrt{\frac{9}{8\pi}}\,\bigl(1-\beta^{-\frac23}\bigr)^{-\frac12} + \sqrt[4]{\frac{18}{\pi}}\,\beta^{\frac12}\bigl(1-\beta^{-\frac23}\bigr)^{-\frac34}.$$

Numerical experimentation indicates that 10 is quite a good approximation to the actual solution to this equation. However, it should be recognized that, with sufficient diligence and entirely different techniques, one can show that the 10 in (2.2.18) can be replaced by a number which is less than 1. Thus, we do not claim that Stein's method gives the best result, only that it gives whatever it gives with relatively little pain.

Exercises for § 2.1

Exercise 2.2.24. It is important to know that, at least qualitatively, one can do better than Berry-Esseen. To see this, consider independent, standard Bernoulli random variables, and define F_n accordingly. Next, observe that when $t_n = -(2n+1)^{-\frac{1}{2}}$,

$$F_{2n+1}(t_n) - G(t_n) = \frac{1}{\sqrt{2\pi}} \int_{t_n}^{0} e^{-\frac{x^2}{2}}\, dx$$

and therefore that $\overline{\lim}_{n\to\infty} n^{\frac{1}{2}} \|F_n - G\|_{\mathrm{u}} \geq \frac{1}{\sqrt{2\pi}}$. In particular, since $\tau_m = 1$ for these Bernoulli random variables, we conclude that the constant in the Berry–Esseen estimate cannot be smaller than $(2\pi)^{-\frac{1}{2}}$.

Exercise 2.2.25. Because the derivation of Theorem 2.2.12 is so elegant and simple, one cannot help wondering whether (2.2.14) cannot be used as the starting point for a proof of (2.2.19). Unfortunately, the following naïve idea falls considerably short of the mark.

Let $X_1, \dots, X_n$ satisfy the hypotheses of Theorem 2.2.17. Starting from (2.2.14) and proceeding as we did in the passage from (2.2.22) to (2.2.23), show that for every $L > 0$

$$\|F_n - G\|_{\mathrm{u}} \leq \frac{6\sum_1^n \tau_m^3}{L\Sigma_n^3} + \frac{L}{\sqrt{8\pi}},$$

and conclude that

$$\|F_n - G\|_{\mathrm{u}} \leq \left(\frac{72}{\pi}\right)^{\frac{1}{4}} \left(\frac{\sum_1^n \tau_m^3}{\Sigma_n^3}\right)^{\frac{1}{2}}.$$

Obviously, this is unacceptably poor when $\Sigma_n^{-3}\sum_1^n \tau_m^3$ is small.

§2.3 Some Extensions of The Central Limit Theorem

In most modern treatments of the Central Limit Theorem, Fourier analysis plays a central role. Indeed, the Fourier transform makes the argument so simple that it can mask what is really happening. However, now that we know Lindeberg's argument, it is time to introduce Fourier techniques and see how they facilitate reasoning involving independent random variables.

§2.3.1. The Fourier Transform. The *Fourier transform* of finite, Borel $\mathbb{C}$-valued measure μ on $\mathbb{R}^N$ is the function $\hat{\mu} : \mathbb{R}^N \longrightarrow \mathbb{C}$ given by

$$\hat{\mu}(\boldsymbol{\xi}) = \int_{\mathbb{R}^N} \exp\left[\sqrt{-1}\,(\boldsymbol{\xi}, \mathbf{x})_{\mathbb{R}^N}\right] \mu(d\mathbf{x}) \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^N. \tag{2.3.1}$$

When μ is a probability measure which is the distribution of an $\mathbb{R}^N$-valued randomvariable $\mathbf{X}$, probabilists usual call its Fourier transform the **characteristic**

function of $\mathbf{X}$, and when μ admits a density φ with respect to Lebesgue's measure $\lambda_{\mathbb{R}^N}$, one uses

$$\hat{\varphi}(\boldsymbol{\xi}) = \int_{\mathbb{R}^N} \exp\Big[\sqrt{-1}\,(\boldsymbol{\xi},\mathbf{x})_{\mathbb{R}^N}\Big]\,\varphi(\mathbf{x})\,d\mathbf{x} \quad \text{for} \quad \boldsymbol{\xi}\in\mathbb{R}^N \tag{2.3.2}$$

in place of $\hat{\mu}$ to denote its Fourier transform.

Obviously, $\hat{\mu}$ is a continuous function which is bounded by the total variation $\|\mu\|_{\text{var}}$ of μ; and only slightly less obvious* is the fact that, for $\varphi \in C_{\text{c}}^{\infty}\big(\mathbb{R}^N;\mathbb{C}\big)$, $\hat{\varphi}\in C^{\infty}\big(\mathbb{R}^N;\mathbb{C}\big)$ and that $\hat{\varphi}$ as well as all its derivatives are **rapidly decreasing** (i.e., they tend to 0 at infinity faster than $\big(1+|\boldsymbol{\xi}|^2\big)^{-1}$ to any power).

Lemma 2.3.3. *Let μ be a finite Borel measure on $\mathbb{R}^N$. Then, for every $\varphi \in C_{\text{b}}\big(\mathbb{R}^N;\mathbb{C}\big)\cap L^1\big(\mathbb{R}^N;\mathbb{C}\big)$ with $\hat{\varphi}\in L^1(\mathbb{R}^N;\mathbb{C})$,*

$$\langle\varphi,\mu\rangle = \int_{\mathbb{R}^N}\varphi\,d\mu = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N}\hat{\varphi}(\boldsymbol{\xi})\overline{\hat{\mu}(\boldsymbol{\xi})}\,d\boldsymbol{\xi}. \tag{2.3.4}$$

Moreover, given a sequence $\big\{\mu_n : n\in\mathbb{Z}^+\big\}$ of Borel probability measures and a Borel probability measure μ on $\mathbb{R}^N$, $\widehat{\mu_n}\longrightarrow\hat{\mu}$ uniformly on compacts if $\langle\varphi,\mu_n\rangle\longrightarrow\langle\varphi,\mu\rangle$ for every $\varphi\in C_{\text{c}}\big(\mathbb{R}^N,\mathbb{R}\big)$. Conversely, if $\widehat{\mu_n}(\boldsymbol{\xi})\longrightarrow\hat{\mu}(\boldsymbol{\xi})$ point-wise, then $\langle\varphi_n,\mu_n\rangle\longrightarrow\langle\varphi,\mu\rangle$ whenever $\{\varphi_n : n\geq 1\}$ is a uniformly bounded sequence in $C_{\text{b}}(\mathbb{R}^N;\mathbb{C})$ which tends to φ uniformly on compacts. (Cf. Theorem 3.1.8 below for more information on this subject.)

Proof: Choose $\rho\in C_{\text{c}}^{\infty}\big(\mathbb{R}^N;[0,\infty)\big)$ to be an even function which satisfies $\int_{\mathbb{R}^N}\rho\,d\mathbf{x}=1$, and set $\rho_\epsilon(\mathbf{x})=\epsilon^{-N}\rho(\epsilon^{-1}\mathbf{x})$ for $\epsilon\in(0,\infty)$. Next, define ψ_ϵ for $\epsilon\in(0,\infty)$ to be the convolution $\rho_\epsilon\star\mu$ of ρ_ϵ with μ. That is,

$$\psi_\epsilon(\mathbf{x}) = \int_{\mathbb{R}^N}\rho_\epsilon(\mathbf{x}-\mathbf{y})\,\mu(d\mathbf{y}) \quad \text{for} \quad \mathbf{x}\in\mathbb{R}^N.$$

It is then an easy matter to check that $\psi_\epsilon\in C_{\text{b}}\big(\mathbb{R}^N;\mathbb{C}\big)$ and $\|\psi_\epsilon\|_{L^1(\mathbb{R}^N)}\leq\|\mu\|_{\text{var}}$ for every $\epsilon\in(0,\infty)$. In addition, one sees (by Fubini's Theorem) that $\hat{\psi}_\epsilon(\boldsymbol{\xi}) = \hat{\rho}(\epsilon\,\boldsymbol{\xi})\hat{\mu}(\boldsymbol{\xi})$. Thus, for any $\varphi\in C_{\text{b}}(\mathbb{R}^N;\mathbb{C})\cap L^1\big(\mathbb{R}^N;\mathbb{C}\big)$, Fubini's Theorem followed by the classical Parseval identity (cf. Exercise 2.3.23 below) yields

$$\int_{\mathbb{R}^N}\varphi_\epsilon\,d\mu = \int_{\mathbb{R}^N}\varphi(\mathbf{x})\,\psi_\epsilon(\mathbf{x})\,d\mathbf{x} = \frac{1}{(2\pi)^N}\int_{\mathbb{R}^N}\hat{\rho}(\epsilon\,\boldsymbol{\xi})\,\hat{\varphi}(\boldsymbol{\xi})\,\hat{\mu}(-\boldsymbol{\xi})\,d\boldsymbol{\xi},$$

where $\varphi_\epsilon\equiv\rho_\epsilon\star\varphi$ is the convolution of ρ_ϵ with φ. Since, as $\epsilon\searrow 0$, $\varphi_\epsilon\longrightarrow\varphi$ while $\hat{\rho}(\epsilon\,\boldsymbol{\xi})\longrightarrow 1$ boundedly and pointwise, (2.3.4) now follows from Lebesgue's Dominated Convergence Theorem.

---

* One uses integration by parts to check that $\widehat{\partial^\alpha\varphi}(\boldsymbol{\xi}) = (-\sqrt{-1}\boldsymbol{\xi})^\alpha\hat{\varphi}(\boldsymbol{\xi})$ and concludes that $|\boldsymbol{\xi}|^n|\hat{\varphi}(\boldsymbol{\xi})|$ is bounded by $\sum_{\|\alpha\|=n}\|\partial^\alpha\varphi\|_{L^1(\mathbb{R}^N)}$.

Turning to the second part of the theorem, first suppose that $\langle\varphi,\mu_n\rangle \longrightarrow \langle\varphi,\mu\rangle$ for every $\varphi \in C_{\mathrm{c}}(\mathbb{R}^N;\mathbb{R})$, and let $\boldsymbol{\xi}_n \longrightarrow \boldsymbol{\xi}$ in $\mathbb{C}$. Then, by the last part of Lemma 2.1.7 applied to $\varphi_n(\mathbf{x}) = e^{\sqrt{-1}\,(\boldsymbol{\xi}_n,\mathbf{x})_{\mathbb{R}^N}}$ and $\varphi(\mathbf{x}) = e^{\sqrt{-1}\,(\boldsymbol{\xi},\mathbf{x})_{\mathbb{R}^N}}$, $\widehat{\mu_n}(\boldsymbol{\xi}_n) \longrightarrow \hat{\mu}(\boldsymbol{\xi})$. Hence, $\widehat{\mu_n} \longrightarrow \hat{\mu}$ uniformly on compacts. Conversely, suppose that $\widehat{\mu_n} \longrightarrow \hat{\mu}$ point-wise. Again by Lemma 2.1.7, we need only check that $\langle\varphi,\mu_n\rangle \longrightarrow \langle\varphi,\mu\rangle$ when $\varphi \in C^\infty_{\mathrm{c}}\big(\mathbb{R}^N;\mathbb{C}\big)$. But, for such a φ, $\hat{\varphi}$ is smooth and rapidly decreasing, and therefore the result follows immediately from the first part of the present lemma together with Lebesgue's Dominated Convergence Theorem. $\square$

REMARK 2.3.5. Although it may seem too obvious to mention, an important, and rather amazing, consequence of Lemma 2.3.3 is that *a finite Borel measure on $\mathbb{R}^N$ is completely determined by its 1-dimensional marginals.* To understand this remark, recall that for a linear subspace L of $\mathbb{R}^N$, the **marginal distribution** of μ on L is the measure $\mu\circ(\Pi_L)^{-1}$, where Π_L denotes orthogonal projection onto L. In particular, if $\mathbf{e} \in \mathbb{S}^{N-1}$ and $\mu_{\mathbf{e}}$ is the marginal distribution of μ on the 1-dimensional subspace spanned by $\mathbf{e}$, then $\hat{\mu}(\xi\mathbf{e}) = \widehat{\mu_{\mathbf{e}}}(\xi)$. Hence, the Fourier transform of μ is determined by the Fourier transforms of $\{\mu_{\mathbf{e}} : \mathbf{e} \in \mathbb{S}^{N-1}\}$, and therefore, by Lemma 2.3.3, μ can be recovered from its 1-dimensional marginals. Of course, one should be careful when applying this observation. For instance, when applied to an $\mathbb{R}^N$-valued random variable $\mathbf{X} = (X_1,\dots,X_N)$, it says that the distribution of $\mathbf{X}$ can be recovered from a knowledge of the distributions of $(\mathbf{e},\mathbf{X})_{\mathbb{R}^N}$ for all $\mathbf{e} \in \mathbb{S}^{N-1}$, but it does not say that the distributions the coordinates X_i, $1 \le i \le N$, determine the distribution of $\mathbf{X}$.

§2.3.2. Multidimensional Central Limit Theorem. The great virtue of the Fourier transform is that it behaves so well under operations built out translation. In applications to probability theory, this virtue is of particular importance when adding independent random variables. Specifically, if $\mathbf{X}$ and $\mathbf{Y}$ are independent, then the characteristic function of $\mathbf{X}+\mathbf{Y}$ is the product of the characteristic functions of $\mathbf{X}$ and $\mathbf{Y}$. This observation combined with Lemma 2.3.3 leads to the following easy proof of the Central Limit Theorem for independent, identically distributed random variables $\{X_n : n \ge 1\}$ with mean-value 0 and variance 1. Namely, if μ_n is the distribution of $\check{S}_n$, then

$$\hat{\mu}_n(\xi) = \left(\hat{\mu}\left(\tfrac{\xi}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{\xi^2}{2n} + o\big(\tfrac{1}{n}\big)\right)^n \longrightarrow e^{-\frac{\xi^2}{2}} = \hat{\gamma}_{0,1}(\xi)$$

for every $\xi \in \mathbb{R}$.

Actually, as we are about to see, a slight variation on the preceding will allow us to lift the results which we already have for $\mathbb{R}$-valued random variables to random variables with values in $\mathbb{R}^N$. However, before we can state this result, we must introduce the analogs of the mean-value and variance for vector-valued random variables. Thus, given a $\mathbb{P}$-integrable, $\mathbb{R}^N$-valued random variable $\mathbf{X}$ on the probability space $(\Omega,\mathcal{F},\mathbb{P})$, the **mean-value** $\mathbb{E}^{\mathbb{P}}\big[\mathbf{X}\big]$ of $\mathbf{X}$ is that $\mathbf{m} \in \mathbb{R}^N$

which is determined by the property that

$$(\boldsymbol{\xi}, \mathbf{m})_{\mathbb{R}^N} = \mathbb{E}^{\mathbb{P}}\Big[\big(\boldsymbol{\xi}, \mathbf{X}\big)_{\mathbb{R}^N}\Big] \quad \text{for all} \quad \boldsymbol{\xi} \in \mathbb{R}^N.$$

Similarly, if $\mathbf{X}$ is square $\mathbb{P}$-integrable, then the **covariance** $\mathbf{cov}(\mathbf{X})$ of $\mathbf{X}$ is the symmetric linear transformation $\mathbf{C}$ on $\mathbb{R}^N$ determined by

$$\big(\boldsymbol{\xi}, \mathbf{C}\,\boldsymbol{\eta}\big)_{\mathbb{R}^N} = \mathbb{E}^{\mathbb{P}}\Big[\Big(\boldsymbol{\xi}, \mathbf{X} - \mathbb{E}^{\mathbb{P}}[\mathbf{X}]\Big)_{\mathbb{R}^N}\Big(\boldsymbol{\eta}, \mathbf{X} - \mathbb{E}^{\mathbb{P}}[\mathbf{X}]\Big)_{\mathbb{R}^N}\Big] \quad \text{for } \boldsymbol{\xi},\, \boldsymbol{\eta} \in \mathbb{R}^N.$$

Notice that $\mathbf{cov}(\mathbf{X})$ is not only symmetric but also nonnegative definite, since for each $\boldsymbol{\xi} \in \mathbb{R}^N$, $\big(\boldsymbol{\xi}, \mathbf{cov}(\mathbf{X})\,\boldsymbol{\xi}\big)_{\mathbb{R}^N}$ is nothing but the variance of $(\boldsymbol{\xi}, \mathbf{X})_{\mathbb{R}^N}$. Finally, given $\mathbf{m} \in \mathbb{R}^N$ and a symmetric, nonnegative $\mathbf{C} \in \mathbb{R}^N \otimes \mathbb{R}^N$, we use $\gamma_{\mathbf{m},\mathbf{C}}$ to denote the Borel probability measure on $\mathbb{R}^N$ which is determined by the property that

$$\int_{\mathbb{R}^N} \varphi\, d\gamma_{\mathbf{m},\mathbf{C}} = \int_{\mathbb{R}^N} \varphi\big(\mathbf{m} + \mathbf{C}^{\frac{1}{2}}\mathbf{y}\big)\, \gamma^N_{0,1}(d\mathbf{y}), \quad \varphi \in C_{\mathrm{b}}(\mathbb{R}^N;\mathbb{R}), \tag{2.3.6}$$

where $\mathbf{C}^{\frac{1}{2}}$ is the non-negative definite, symmetric square root of $\mathbf{C}$

Clearly, an $\mathbb{R}^N$-valued random variable $\mathbf{Y}$ has distribution $\gamma_{\mathbf{m},\mathbf{C}}$ if and only if, for each $\boldsymbol{\xi} \in \mathbb{R}^N$, $(\boldsymbol{\xi}, \mathbf{Y})_{\mathbb{R}^N}$ is a normal random variable with mean-value $(\boldsymbol{\xi}, \mathbf{m})_{\mathbb{R}^N}$ and variance $(\boldsymbol{\xi}, \mathbf{C}\,\boldsymbol{\xi})_{\mathbb{R}^N}$. For this reason, $\gamma_{\mathbf{m},\mathbf{C}}$ is called the **normal or Gaussian distribution** with mean-value $\mathbf{m}$ and covariance $\mathbf{C}$. For the same reason, a random variable with $\gamma_{\mathbf{m},\mathbf{C}}$ as its distribution is called a **normal or Gaussian random variable** with mean-value $\mathbf{m}$ and covariance $\mathbf{C}$, or, more briefly, an $\mathcal{N}(\mathbf{m}, \mathbf{C})$-random variable. Finally, one can use this characterization to see that

$$\widehat{\gamma_{\mathbf{m},\mathbf{C}}}(\boldsymbol{\xi}) = \exp\Big[\sqrt{-1}\big(\boldsymbol{\xi}, \mathbf{m}\big) - \tfrac{1}{2}\big(\boldsymbol{\xi}, \mathbf{C}\boldsymbol{\xi}\big)_{\mathbb{R}^N}\Big]. \tag{2.3.7}$$

In the following statements, we will be assuming that $\{\mathbf{X}_n : n \in \mathbb{Z}^+\}$ is a sequence of mutually independent, square $\mathbb{P}$-integrable, $\mathbb{R}^N$-valued random variables on the probability space $(\Omega, \mathcal{F}, P)$. Further, we will assume that, for each $n \in \mathbb{Z}^+$, $\mathbf{X}_n$ has mean-value $\mathbf{0}$ and strictly positive covariance $\mathbf{cov}(\mathbf{X}_n)$. Finally, for $n \in \mathbb{Z}^+$, we set

$$\mathbf{S}_n = \sum_{m=1}^{n} \mathbf{X}_m, \quad \mathbf{C}_n \equiv \mathbf{cov}(\mathbf{S}_n) = \sum_{m=1}^{n} \mathbf{cov}(\mathbf{X}_m),$$
$$\Sigma_n = \big(\det(\mathbf{C}_n)\big)^{\frac{1}{2N}} \text{ and } \breve{\mathbf{S}}_n = \frac{\mathbf{S}_n}{\Sigma_n}.$$

Notice that when $N = 1$, the above use of the notation Σ_n and $\breve{\mathbf{S}}_n$ is consistent with that of Section II.1.

With these preparations, we are ready to prove the following multidimensional generalization of Theorem 2.1.8.

THEOREM 2.3.8. *Referring to the preceding, assume that the limit*

$$\mathbf{A} \equiv \lim_{n\to\infty} \frac{\mathbf{C}_n}{\Sigma_n^2} \tag{2.3.9}$$

exists and that

$$\lim_{n\to\infty} \frac{1}{\Sigma_n^2} \sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\Big[|\mathbf{X}_m|^2,\, |\mathbf{X}_m| \geq \epsilon\Sigma_n\Big] = 0 \quad \textit{for each } \epsilon > 0. \tag{2.3.10}$$

Then, for every sequence $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ which satisfies

$$\sup_{n\geq 1} \sup_{\mathbf{y}\in\mathbb{R}^N} \frac{|\varphi_n(\mathbf{y})|}{1+|\mathbf{y}|^2} < \infty \tag{2.3.11}$$

and converges uniformly on compacts to φ,

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\Big[\varphi_n\big(\check{\mathbf{S}}_n\big)\Big] = \int_{\mathbb{R}^N} \varphi\, d\gamma_{\mathbf{0},\mathbf{A}}. \tag{2.3.12}$$

In particular, when the $\mathbf{X}_n$ are uniformly square $\mathbb{P}$-integrable random variables with mean-value $\mathbf{0}$ and common covariance $\mathbf{C}$,

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\left[\varphi_n\left(\frac{\mathbf{S}_n}{\sqrt{n}}\right)\right] = \int_{\mathbb{R}^N} \varphi\, d\gamma_{\mathbf{0},\mathbf{C}}$$

whenever $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ satisfies (2.3.11) and converges to φ uniformly on compacts.

PROOF: Given $\mathbf{e} \in \mathbb{S}^{N-1}$, set

$$\Sigma_n(\mathbf{e}) = \sqrt{\big(\mathbf{e}, \mathbf{C}_n e\big)_{\mathbb{R}^N}} \quad \text{and} \quad \rho_n(\mathbf{e}) = \frac{\Sigma_n(\mathbf{e})}{\Sigma_n}.$$

Then, $\rho(\mathbf{e}) \equiv \inf_{n\geq 1} \rho_n(\mathbf{e}) > 0$ and $\rho_n(\mathbf{e}) \longrightarrow \sqrt{(\mathbf{e}, \mathbf{A}\mathbf{e})_{\mathbb{R}^N}}$ as $n \to \infty$. In particular, if $(\mathbf{e}_1, \dots, \mathbf{e}_N)$ is an orthonormal basis in $\mathbb{R}^N$, then

$$\begin{aligned}\mathbb{E}^{\mathbb{P}}\big[|\check{\mathbf{S}}_n|^2\big] &= \sum_{i=1}^{N} \mathbb{E}^{\mathbb{P}}\big[\big(\mathbf{e}_i, \check{\mathbf{S}}_n\big)^2_{\mathbb{R}^N}\big] = \sum_{i=1}^{N} \rho_n(\mathbf{e}_i)^2 \\ &\longrightarrow \sum_{i=1}^{N} \big(\mathbf{e}_i, \mathbf{A}\mathbf{e}_i\big)_{\mathbb{R}^N} = \int_{\mathbb{R}^N} |\mathbf{y}|^2\, \gamma_{\mathbf{0},\mathbf{A}}(dy).\end{aligned}$$

Hence, by Lemmas 2.1.7 and 2.3.3 plus (2.3.7), all that we have to do is check that

$$(*)\qquad f_n(\boldsymbol{\xi}) \equiv \mathbb{E}^{\mathbb{P}}\Big[e^{\sqrt{-1}\,(\boldsymbol{\xi},\breve{\mathbf{S}}_n)_{\mathbb{R}^N}}\Big] \longrightarrow e^{-\frac{1}{2}(\boldsymbol{\xi},\mathbf{A}\boldsymbol{\xi})_{\mathbb{R}^N}}$$

for each $\boldsymbol{\xi} \in \mathbb{R}^N$.

When $\boldsymbol{\xi} = \mathbf{0}$, (*) is trivial. Thus, assume that $\boldsymbol{\xi} \neq \mathbf{0}$, set $\mathbf{e} = \frac{\boldsymbol{\xi}}{|\boldsymbol{\xi}|}$, and take $\breve{S}_n(\mathbf{e}) = \frac{(\mathbf{e},\mathbf{S}_n)_{\mathbb{R}^N}}{\Sigma_n(\mathbf{e})}$. Because

$$\frac{1}{\Sigma_n(\mathbf{e})^2}\sum_{m=1}^n \mathbb{E}^{\mathbb{P}}\big[(\mathbf{e},\mathbf{X}_m)_{\mathbb{R}^N}^2,\, \big|(\mathbf{e},\mathbf{X}_m)_{\mathbb{R}^N}\big| \geq \epsilon\Sigma_n(\mathbf{e})\big]$$
$$\leq \frac{1}{\rho(\mathbf{e})^{-1}\Sigma_n}\sum_{m=1}^n \mathbb{E}^{\mathbb{P}}\big[(\mathbf{e},\mathbf{X}_m)_{\mathbb{R}^N}^2,\, \big|(\mathbf{e},\mathbf{X}_m)_{\mathbb{R}^N}\big| \geq \rho(\mathbf{e})\epsilon\Sigma_n(\mathbf{e})\big]$$

tends to 0 for each $\epsilon > 0$, Theorem 2.1.8 combined with Lemma 2.3.3 guarantees that, for any $\eta \in \mathbb{R}$,

$$\mathbb{E}^{\mathbb{P}}\big[e^{\sqrt{-1}\,\eta_n\breve{S}_n(\mathbf{e})}\big] \longrightarrow e^{-\frac{1}{2}|\eta|^2}$$

for any $\{\eta_n : n \geq 1\} \subseteq \mathbb{R}$ which tends to η. In particular, if $\eta = \sqrt{(\boldsymbol{\xi},\mathbf{A}\boldsymbol{\xi})_{\mathbb{R}^N}}$ and $\eta_n = \rho_n(\mathbf{e})|\boldsymbol{\xi}|$, we find that

$$f_n(\boldsymbol{\xi}) = \mathbb{E}^{\mathbb{P}}\big[e^{\sqrt{-1}\,\eta_n\breve{S}_n(\mathbf{e})}\big] \longrightarrow e^{-\frac{1}{2}(\boldsymbol{\xi},\mathbf{A}\boldsymbol{\xi})_{\mathbb{R}^N}}. \quad \square$$

§2.3.3. Higher Moments. In this subsection we will show that when the X_n's possess higher moments, then (2.1.1) remains true for φ's which can grow faster than $1+|y|^2$. As an intitial step in this direction, we give the following simple example.

LEMMA 2.3.13. *Suppose that $\{X_n : n \geq 1\}$ is a sequence of independent, indentically distributed random variables with mean-value 0 and variance 1. If $\mathbb{E}^{\mathbb{P}}[X_1^{2\ell}] < \infty$ for some $\ell \in \mathbb{Z}^+$, then (2.1.1) holds for any $\varphi \in C(\mathbb{R}^N;\mathbb{C})$ which satisfies*

$$(2.3.14)\qquad \sup_{y\in\mathbb{R}} \frac{|\varphi(y)}{1+|y|^{2\ell}} < \infty.$$

PROOF: Refer to the discussion in the introduction to this chapter and observe that the argument there shows that

$$\lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\big[\breve{S}_n^{2\ell}\big] = \frac{(2\ell)!}{2^\ell\ell!} = \int_{\mathbb{R}} y^{2\ell}\,\gamma_{0,1}(dy)$$

just so long as the 2ℓth moment of X_1 is finite. Hence the desired conclusion is an application of the last part of Lemma 2.1.7 with $\psi(y) = 1+|y|^{2\ell}$.

In most situations, one cannot carry out the computations to give a proof that the last part of Lemma 2.1.7 applies, and for this reason the following lemma is often useful.

Lemma 2.3.15. *Suppose that $\{\mu_n : n \geq 1\}$ is a sequence of finite (non-negative) Borel measures on $\mathbb{R}^N$, and assume μ is a finite Borel measure with the property that $\langle \varphi, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ for all $\varphi \in C_{\mathrm{b}}^\infty(\mathbb{R}^N;\mathbb{R})$. If for some $\psi \in C\big(\mathbb{R}^N;[0,\infty)\big)$ and $p \in (1,\infty)$*

$$\sup_{n\geq 1} \langle \psi^p, \mu_n \rangle < \infty, \tag{2.3.16}$$

then $\langle \varphi_n, \mu_n \rangle \longrightarrow \langle \varphi, \mu \rangle$ whenever $\{\varphi_n : n \geq 1\} \subseteq C(\mathbb{R}^N;\mathbb{C})$ is a sequence which satisfies $|\varphi_n| \leq \psi$ for all $n \in \mathbb{Z}^+$ and converges to φ uniformly on compacts.

Proof: By Lemma 2.1.7, all that we have to prove is that $\langle \psi, \mu_n \rangle \longrightarrow \langle \psi, \mu \rangle$. For this purpose, note that, under our present hypotheses, Lemma 2.1.7 shows that $\overline{\lim}_{n\to\infty} \langle \psi, \mu_n \rangle \leq \langle \psi, \mu \rangle$ and that, for each $R > 0$, $\langle \psi \wedge R, \mu_n \rangle \longrightarrow \langle \psi \wedge R, \mu \rangle \leq \langle \psi, \mu \rangle$. Thus, it suffices to observe that

$$\sup_{n\geq 1} \langle (\psi - \psi \wedge R), \mu_n \rangle = \sup_{n\geq 1} \int_{\{\psi > R\}} \psi \, d\mu_n \leq R^{1-p} \sup_{n\geq 1} \langle \psi^p, \mu_n \rangle \longrightarrow 0$$

as $R \to \infty$. $\square$

Knowing Lemma 2.3.15, ones problem is to find conditions under which one can show that $\sup_{n\geq 1} \mathbb{E}^{\mathbb{P}}[\psi(\breve{S}_n)] < \infty$ for an interesting class of non-negative ψ's. One such class is provided by the notion of a subGaussian random variable. Given $\beta \in [0,\infty)$, an $\mathbb{R}$-valued random variable X is said to be **β-subGaussian** if

$$\mathbb{E}^{\mathbb{P}}\big[e^{\xi X}\big] \leq e^{\frac{\beta^2 \xi^2}{2}}, \quad \xi \in \mathbb{R}. \tag{2.3.17}$$

The origin of this terminology should be clear: if $X \in \mathcal{N}(0,\sigma^2)$, then equality holds in (2.3.17) with $\beta = \sigma$.

Lemma 2.3.18. *Let X be an $\mathbb{R}$-valued random varianble. If X is a β-subGaussian, then, $\mathbb{E}^{\mathbb{P}}[X] = 0$, $\mathbb{E}^P[X^2] \leq \beta^2$,*

$$\mathbb{P}\big(|X| \geq R\big) \leq 2e^{-\frac{R^2}{2\beta^2}}, \quad R > 0.$$

and, for each $\alpha \in [0, \beta^{-1})$,

$$\mathbb{E}^{\mathbb{P}}\big[e^{\frac{\alpha^2 X^2}{2}}\big] \leq \big(1 - (\alpha\beta)^2\big)^{-\frac{1}{2}}.$$

Conversely, if $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{\alpha^2 X^2}{2}}\big] < \infty$ for some $\alpha \in (0,\infty)$ and $\mathbb{E}^{\mathbb{P}}[X] = 0$, then X is β-subGaussian for some $\beta = \beta(\alpha) \in (0,\infty)$, where $\beta : (0,\infty) \longrightarrow (0,\infty)$ is a non-decreasing function. Moreover, if $X_1, \dots, X_n$ are independent random variables, and, for each $1 \leq m \leq n$, X_m is β_m-subGaussian, then for any $a_1, \dots, a_n \in \mathbb{R}$, $\sum_{m=1}^n a_m X_m$ is β-subGaussian when $\beta = \sqrt{\sum_{m=1}^n (a_m \beta_m)^2}$.

PROOF: Since the moment generating function of the sum of independent random variables is the product of the moment generating functions of the summands, the final assertion is essentially trivial.

To prove the first assertion, use Lebesgue's Dominated Convergence Theorem to justify

$$\pm\mathbb{E}^{\mathbb{P}}[X] = \lim_{\xi\searrow 0}\xi^{-1}\Big(\mathbb{E}^{\mathbb{P}}\big[e^{\pm\xi X}\big]-1\Big) \le \lim_{\xi\searrow 0}\frac{e^{\frac{\beta^2\xi^2}{2}}-1}{\xi} = 0$$

and

$$\mathbb{E}^{\mathbb{P}}\big[X^2\big] = \lim_{\xi\searrow 0}\xi^{-2}\Big(\mathbb{E}^{P}\big[e^{\xi X}\big]+\mathbb{E}^{\mathbb{P}}\big[e^{-\xi X}\big]-2\Big)\le 2\lim_{\xi\searrow 0}\frac{e^{\frac{\beta^2\xi^2}{2}}-1}{\xi^2} = \beta^2.$$

Next, from

$$\mathbb{P}(X\ge R)\le e^{-\xi R}\mathbb{E}^{\mathbb{P}}\big[e^{\xi R}\big]\le \exp\left(-\xi R+\frac{\beta^2\xi^2}{2}\right)$$

for all $\xi\ge 0$, one gets $\mathbb{P}(X\ge R)\le e^{-\frac{R^2}{2\beta^2}}$ by minimizing over $\xi\ge 0$. Since the same estimate holds for $-X$, the estimate for $\mathbb{P}(|X|\ge R)$ follows. To get the estimate on $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{\alpha^2X^2}{2}}\big]$, use Tonelli's Theorem to see that

$$\mathbb{E}^{\mathbb{P}}\big[e^{\frac{\alpha^2X^2}{2}}\big] = \int_{\mathbb{R}}\mathbb{E}^{\mathbb{P}}\big[e^{\xi X}\big]\,\gamma_{0,\alpha}(d\xi)\le\int_{\mathbb{R}}e^{\frac{\beta^2\xi^2}{2}}\,\gamma_{0,\alpha^2}(d\xi) = \big(1-(\alpha\beta)^2\big)^{-\frac{1}{2}}.$$

Finally, assume that $\mathbb{E}^{\mathbb{P}}\big[e^{\frac{\alpha^2X^2}{2}}\big]<\infty$ for some $\alpha\in(0,\infty)$ and $\mathbb{E}^{\mathbb{P}}[X]=0$. Then, for any $a\in[0,\infty)$,

$$\mathbb{E}^{\mathbb{P}}\big[e^{\xi X}\big] = 1+\xi^2\int_0^1(1-t)\mathbb{E}^{\mathbb{P}}\big[X^2e^{t\xi X}\big]\,dt\le 1+\frac{\xi^2e^{\frac{a^2\xi^2}{2}}}{2}\mathbb{E}^{\mathbb{P}}\big[X^2e^{\frac{X^2}{2a^2}}\big],$$

since $\xi X-\frac{X^2}{2a^2}\le\frac{a^2\xi^2}{2}$. Now determine a by the equation $a^2=\mathbb{E}^{\mathbb{P}}\big[X^2e^{\frac{X^2}{2a^2}}\big]$, and set $\beta=e^{\frac{e}{2}}a$. Then,

$$\mathbb{E}^{\mathbb{P}}\big[e^{\xi X}\big]\le 1+\frac{a^2\xi^2}{2}e^{\frac{a^2\xi^2}{2}} = 1+\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{a^2\xi^2}{2}\right)^{n+1}\le e^{\frac{\beta^2\xi^2}{2}},$$

since $(n+1)e^{-(n+1)\frac{e}{2}}\le 1$. $\square$

By combining Lemmas 2.3.15 and 2.3.18 with Theorem 2.3.8, we get the following.

THEOREM 2.3.19. *Working with the setting and notation in Theorem 2.3.8, assume that, for each* $n \in \mathbb{Z}^+$,

$$\mathbb{E}\mathbb{P}\big[e^{(\boldsymbol{\xi},\mathbf{X}_n)_{\mathbb{R}^N}}\big] \le e^{\beta_n|\boldsymbol{\xi}|^2}, \quad \boldsymbol{\xi} \in \mathbb{R}^N$$

where $\beta_n \in (0,\infty)$. *If*

$$\beta \equiv \sup_{n\ge1} \frac{\sqrt{\sum_{m=1}^n \beta_m^2}}{\Sigma_n} < \infty,$$

then (2.3.12) holds for any $\varphi \in C(\mathbb{R}^N;\mathbb{C})$ *satisfying*

$$|\varphi(\mathbf{y})| \le Ce^{\frac{\alpha^2|\mathbf{y}|^2}{2}}, \quad \mathbf{y} \in \mathbb{R}^N,$$

for some $C < \infty$ *and* $\alpha \in \big(0,\frac{1}{\beta}\big)$. *In particular, if the* $\mathbf{X}_n$*'s are identically distributed and* $\mathbb{E}^{\mathbb{P}}\big[e^{\alpha^2|\mathbf{X}_1|^2}\big] < \infty$ *for some* $\alpha \in (0,\infty)$, *then, for any* $\varphi \in C(\mathbb{R};\mathbb{C})$,

$$\varlimsup_{|\mathbf{y}|\to\infty} |\mathbf{y}|^{-2}\log\big(1+|\varphi(\mathbf{y})|\big) = 0 \implies \lim_{n\to\infty}\mathbb{E}^{\mathbb{P}}\left[\frac{\mathbf{S}_n}{n^{\frac12}}\right] = \langle\varphi,\gamma_{\mathbf{0},\mathbf{C}}\rangle.$$

Exercises for § 2.3

EXERCISE 2.3.20. Here is a proof of Feller's part of the Lindeberg–Feller Theorem. Referring to Theorem 2.1.4 and the discussion proceeding it, assume that $r_n \longrightarrow 0$ and that

$$\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\frac{\xi X_m}{\Sigma_n}}\right] \longrightarrow e^{-\frac{\xi^2}{2}} \quad \text{for all}\xi \in \mathbb{R}.$$

(**i**) Show that

$$\max_{1\le m\le n}\left|1-\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\frac{\xi X_m}{\Sigma_n}}\right]\right| \le \frac{\xi^2 r_n^2}{2},$$

and conclude that, for each $R>0$ there is an N_R such that

$$\max_{1\le m\le n}\left|1-\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\frac{\xi X_m}{\Sigma_n}}\right]\right| \le \frac12 \quad \text{for } n \ge N_R \text{ and } |\xi|\le R.$$

(**ii**) Define $\log\zeta = -\sum_{k=1}^\infty \frac{(1-\zeta)^k}{k}$ for $\zeta\in\mathbb{C}$ with $|1-\zeta|<1$, and check that $\big|(1-\zeta)+\log\zeta\big| \le |1-\zeta|^2$ for $|1-\zeta|\le\frac12$. Conclude first that

$$\left|\sum_{m=1}^n \mathbb{E}^{\mathbb{P}}\left[1-e^{\sqrt{-1}\frac{\xi X_m}{\Sigma_n}}\right] + \sum_{m=1}^n \log\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}\frac{\xi X_m}{\Sigma_n}}\right]\right| \le \frac{R^2r_n^2}{2}$$

for $n\ge N_R$ and $|\xi|\le R$, and then that

$$\Delta_n(\xi) \equiv \frac{\xi^2}{2} - \sum_{m=1}^n \mathbb{E}^{\mathbb{P}}\left[1-\cos\frac{\xi X_m}{\Sigma_n}\right] \longrightarrow 0$$

uniformly for ξ's in compacts.

(**iii**) Given $\epsilon > 0$, show that

$$\sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\left[1-\cos\frac{\xi X_m}{\Sigma_n},\, |X_m| < \epsilon\Sigma_n\right] \le \frac{\xi^2}{2\Sigma_n}\sum_{m=1}^{n}\mathbb{E}^{\mathbb{P}}\big[X_m^2,\, |X_m|<\epsilon\Sigma_n\big]$$
$$\le \frac{\xi^2}{2} - \frac{\xi^2}{2} g_n(\epsilon)$$

and that

$$\sum_{m=1}^{n} \mathbb{E}^{\mathbb{P}}\left[1-\cos\frac{\xi X_m}{\Sigma_n},\, |X_m| \ge \epsilon\Sigma_n\right] \le \epsilon^{-2}.$$

Finally, combine these and apply (**ii**) to get $\overline{\lim}_{n\to\infty} \xi^2 g_n(\epsilon) \le \epsilon^{-2}$ for all $\xi \in \mathbb{R}$.

Exercise 2.3.21. It is of some interest to note that the second moment assumption can be removed from hypotheses in Exercise 2.1.11. To explain what we have in mind, first use that exercise to see that if $\sigma^2 = \int_{\mathbb{R}} x^2\,\mu(dx) < \infty$, then $\mu = T\mu \implies \mu \in \mathcal{N}(0,\sigma^2)$. What we want to do now is remove the *a priori* assumption that $\int_{\mathbb{R}} x^2\,\mu(dx) < \infty$. That is, we want to show that, for *any* probability measure μ on $\mathbb{R}$, $\mu = T\mu \iff \mu \in \mathcal{N}(0,\sigma^2)$ for some $\sigma \in [0,\infty)$. Since the "$\Longleftarrow$" direction is obvious, and, by the above discussion, the "$\implies$" direction is already covered when $\int_{\mathbb{R}} x^2\,\mu(dx) < \infty$, all that remains is to show that

$$\mu = T\mu \implies \int_{\mathbb{R}} x^2\,\mu(dx) < \infty. \tag{2.3.22}$$

(**i**) We check (2.3.22) first under the condition that μ is symmetric (i.e., $\mu(-\Gamma) = \mu(\Gamma)$ for all $\Gamma \in \mathcal{B}_{\mathbb{R}}$). But, if μ is symmetric, show that

$$\hat{\mu}(\xi) = \int_{\mathbb{R}} \cos(\xi x)\,\mu(dx), \quad \xi \in \mathbb{R}.$$

At the same time, show that

$$\mu = T\mu \implies \hat{\mu}\big(2^{-\frac{1}{2}}\xi\big) = \hat{\mu}(\xi)^{\frac{1}{2}}, \quad \xi \in \mathbb{R}.$$

Conclude from these two that $\hat{\mu} > 0$ everywhere and that

$$\int_{\mathbb{R}} \cos\big(2^{-\frac{n}{2}}\xi x\big)\,\mu(dx) = \hat{\mu}(\xi)^{2^{-n}}, \quad n \in \mathbb{N} \text{ and } \xi \in \mathbb{R}.$$

Finally, note that $1 - x \le -\log x$ for $x \in (0,1]$, apply this to the preceding to get

$$2^n \int_{\mathbb{R}} \Big(1-\cos\big(2^{-\frac{n}{2}}x\big)\Big)\,\mu(dx) \le -\log\big(\hat{\mu}(1)\big) < \infty, \quad n \in \mathbb{N},$$

and arrive at

$$\int_{\mathbb{R}} x^2\,\mu(dx) \le -2\log\big(\hat{\mu}(1)\big)$$

after an application of Fatou's Lemma.

(**ii**) To complete the program, let μ be any solution to $\mu = T\mu$, and define ν by

$$\nu(\Gamma) = \iint_{\mathbb{R}^2} \mathbf{1}_\Gamma(x-y)\,\mu(dx)\mu(dy).$$

Check that ν is symmetric and that $\nu = T\nu$. Hence, by (**i**), $\int_{\mathbb{R}} x^2\,\nu(dx) < \infty$ (in fact, ν is centered normal). Finally, use this and part (**i**) of Exercise 1.5.12 to deduce that $\int_{\mathbb{R}} x^2\,\mu(dx) < \infty$.

EXERCISE 2.3.23. In connection with the preceding exercise, define $T_\alpha\mu$, for $\alpha \in (0,\infty)$ and probability measures μ on $\mathbb{R}$, so that

$$T_\alpha\mu(\Gamma) = \iint_{\mathbb{R}^2} \mathbf{1}_\Gamma\big(2^{-\frac{1}{\alpha}}(x+y)\big)\,\mu(dx)\mu(dy), \quad \Gamma \in \mathcal{B}_{\mathbb{R}}.$$

The problem under consideration here is that of determining for which α's there exist nontrivial (i.e., $\mu \neq \delta_0$) solutions to the fixed point equation $\mu = T_\alpha\mu$. Begin by repeating the argument given in part (**ii**) above to see that there is some solution if and only if there is one which is symmetric. Next, assuming that μ is a nontrivial, symmetric solution, use the reasoning in part (**i**) there to see that

$$\int_{\mathbb{R}} x^2\,\mu(dx) = \begin{cases} \infty & \text{if } \alpha \in (0,2) \\ 0 & \text{if } \alpha \in (2,\infty). \end{cases}$$

In particular, when $\alpha \in (2,\infty)$, there are no nontrivial solutions to $\mu = T_\alpha\mu$. (See Exercise 3.2.25 for more on this topic.)

EXERCISE 2.3.24. Return to the setting of Exercise 2.1.13. After noting that, so long as $\mathbf{e} \in \mathbf{S}^{n-1}$, the distribution of

$$\mathbf{x} \in \mathbf{S}^{n-1}\big(\sqrt{n}\big) \longmapsto (\mathbf{e},\mathbf{x})_{\mathbb{R}^n} \in \mathbb{R}$$

is independent of $\mathbf{e}$, use Lemma 2.3.3 to prove that the assertion in (2.1.15) follows as a consequence of the one in (2.1.14).

EXERCISE 2.3.25. Begin by checking the identity (cf. (1.3.20))

$$\int_0^\infty t^s e^{-\frac{t^2}{2\beta^2}}\,dt = 2^{\frac{s-1}{2}}\beta^{s+1}\Gamma\left(\frac{s+1}{2}\right)$$

for all $\beta \in (0,\infty)$ and $s \in (-1,\eta)$. Use the preceding to see that, for each $p \in (0,\infty)$

$$\mathbb{E}^{\mathbb{P}}\big[|X|^p\big] = \sqrt{\frac{2^p}{\pi}}\,\Gamma\left(\frac{p+1}{2}\right)\sigma^p \quad \text{if } X \in \mathcal{N}(0,\sigma^2). \tag{2.3.26}$$

The goal of the exercise is to show that the moments of subGaussian random variable display similar behavior.

(**i**) Suppose that X is β-subGaussian, and show that, for each $p \in (0,\infty)$,

$$\mathbb{E}^{\mathbb{P}}\big[|X|^p\big] \le K_p\beta^p \quad \text{where } K_p \equiv p2^{\frac{p}{2}}\Gamma\left(\frac{p}{2}\right) = 2^{\frac{p}{2}+1}\Gamma\left(\frac{p}{2}+1\right).$$

(**ii**) Again suppose that X is β-subGaussian, and let σ^2 be its variance. Show that

$$\mathbb{E}^{\mathbb{P}}\big[|X|^p\big] \ge K_4^{-(1-\frac{p}{2})^+}\left(\frac{\sigma}{\beta}\right)^{2+|p-2|}\beta^p$$

for each $p \in (0,\infty)$.

Hint: When $p \ge 2$, the inequality is trivial. To prove it when $p < 2$, and that, for any $q \in (1,\infty)$,

$$\sigma^2 \le \mathbb{E}^{\mathbb{P}}\big[|X|^p\big]^{\frac{1}{q}}\mathbb{E}^{\mathbb{P}}\big[|X|^{\frac{2q-p}{q-1}}\big]^{\frac{1}{q}'}$$

where $q' = \frac{q}{q-1}$ is the Hölder conjugate of q.

(**iii**) Suppose that $X_1,\dots,X_n$ are independent and that, for each $1 \le m \le n$, X_m is β_m-subGaussian and has variance σ_m^2. Given $\{a_1,\dots,a_n\} \subseteq \mathbb{R}$, set

$$S = \sum_{m=1}^n a_m X_m, \quad \Sigma = \sqrt{\sum_{m=1}^n (a_m\sigma_m)^2}, \quad \text{and} \quad B = \sqrt{\sum_{m=1}^n (a_m\beta_m)^2},$$

and show that, for each $p \in (0,\infty)$

$$K_4^{-(1-\frac{p}{2})^+}\left(\frac{\Sigma}{B}\right)^{2+|p-2|}B^p \le \mathbb{E}^{\mathbb{P}}\big[|S|^p\big] \le K_pB^p.$$

In particular, if $\beta_m = \beta$ and $\sigma_m = \sigma$ for all $1 \le m \le n$, then

$$K_4^{-(1-\frac{p}{2})^+}\left(\frac{\sigma}{\beta}\right)^{2+|p-2|}(\beta A)^p \le \mathbb{E}^{\mathbb{P}}\big[|S|^p\big] \le K_p(\beta A)^p \quad \text{where } A = \sqrt{\sum_{m=1}^n a_m^2}.$$

(**iv**) The most famous case of the situation discussed in (**iii**) is when the X_m's are symmetric Bernoulli (i.e., $\mathbb{P}(X_m = \pm 1) = \frac{1}{2}$). First use (**iii**) in Exercise 1.3.17 to check that X_m is 1-subGaussian, and then conclude that

$$K_4^{-(1-\frac{p}{2})^+}\left(\sum_{m=1}^n a_m\right)^{\frac{p}{2}} \le \mathbb{E}^P\left[\left|\sum_{m=1}^n a_m X_m\right|^p\right] \le K_p\left(\sum_{m=1}^n a_m\right)^{\frac{p}{2}} \tag{2.3.27}$$

for all $\{a_1,\dots,a_n\} \subseteq \mathbb{R}$. This fact is known as **Khinchine's inequality**.

EXERCISE 2.3.28. Let $X_1, \dots, X_n$ be independent, symmetric (Exercise 1.4.25) random variables, and set $S = \sum_1^n X_m$. Show that, for each $p \in (0,\infty)$ (cf. part (**ii**) in Exercise 2.3.25),

$$K_4^{-(1-\frac{p}{2})^+}\mathbb{E}^{\mathbb{P}}\left[\left(\sum_1^n X_m^2\right)^{\frac{p}{2}}\right] \le \mathbb{E}^{\mathbb{P}}\Big[|S|^p\Big] \le K_p\mathbb{E}^{\mathbb{P}}\left[\left(\sum_1^n X_m^2\right)^{\frac{p}{2}}\right].$$

Hint: Refer to the beginning of the proof of Lemma 1.1.6, and let $R_1, \dots, R_n$ be the Rademacher functions on $[0,1)$, set $\mathbb{Q} = \lambda_{[0,1)} \times \mathbb{P}$ on $\big([0,1) \times \Omega, \mathcal{B}_{[0,1)} \times \mathcal{F}\big)$, and observe that

$$\omega \in \Omega \longmapsto S_n(\omega) \equiv \sum_1^n X_m(\omega)$$

has the same distribution under $\mathbb{P}$ as

$$(t,\omega) \in [0,1) \times \Omega \longmapsto T_n(t,\omega) \equiv \sum_1^n R_m(t)X_m(\omega)$$

does under $\mathbb{Q}$. Next, apply Khinchine's inequality to see that, for each $\omega \in \Omega$,

$$K_4^{-(1-\frac{p}{2})^+}\left(\sum_1^n X_m(\omega)^2\right)^{\frac{p}{2}} \le \int_{[0,1)} \big|T_n(t,\omega)\big|^p\,dt \le K_p\left(\sum_1^n X_m(\omega)^2\right)^{\frac{p}{2}},$$

and complete the proof by taking the $\mathbb{P}$-integral of this with respect to ω.

At least when $p \in (1,\infty)$, we will show later that this sort of inequality holds in much greater generality. Specifically, see Burkholder's inequality in (?).

EXERCISE 2.3.29. Suppose that $\mathbf{X}$ is an $\mathbb{R}^N$-valued Gaussian random variable with mean-value $\mathbf{0}$ and covariance $\mathbf{C}$. Given a linear subspace L of $\mathbb{R}^N$, let $\mathcal{F}_L$ be the σ-algebra generated by $\{(\boldsymbol{\xi},\mathbf{X})_{\mathbb{R}^N} : \xi \in L\}$, and let $L^{\perp_{\mathbf{C}}}$ be the subspace of $\boldsymbol{\eta}$ such that $(\boldsymbol{\eta},\mathbf{C}\boldsymbol{\xi})_{\mathbb{R}^N} = 0$ for all $\boldsymbol{\xi} \in L$.

(**i**) Show that if $\mathbf{A} : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ is a linear transformation, then $\mathbf{AX}$ is an $\mathcal{N}(\mathbf{0},\mathbf{ACA}^\top)$ random variable, where $\mathbf{A}^\top$ is the adjoint transformation.

(**ii**) Show that $\mathcal{F}_L$ is independent of $\mathcal{F}_{L^{\perp_{\mathbf{C}}}}$.

Hint: Show that, because of linearity, it suffices to check that

$$\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}(\boldsymbol{\xi},\mathbf{X})_{\mathbb{R}^N}}e^{\sqrt{-1}(\boldsymbol{\eta},\mathbf{X})_{\mathbb{R}^N}}\right] = \mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}(\boldsymbol{\xi},\mathbf{X})_{\mathbb{R}^N}}\right]\mathbb{E}^{\mathbb{P}}\left[e^{\sqrt{-1}(\boldsymbol{\eta},\mathbf{X})_{\mathbb{R}^N}}\right]$$

for all $\boldsymbol{\xi} \in L$ and $\boldsymbol{\eta} \in L^{\perp_{\mathbf{C}}}$.

(**iii**) Suppose that $N = N_1 + N_2$, where $N_i \in \mathbb{Z}^+$ for $i \in \{1,2\}$, write $\mathbb{R}^N \ni \mathbf{x} = \begin{pmatrix} \mathbf{x}_{(1)} \\ \mathbf{x}_{(2)} \end{pmatrix} \in \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$, and take $L = \{\mathbf{x} : \mathbf{x}_{(1)} = \mathbf{0}_{(1)}\}$. Show that if Π is a linear transformation taking $\mathbb{R}^N$ onto L and satisfies $\big(\boldsymbol{\xi} - \Pi\boldsymbol{\xi}, \mathbf{C}\boldsymbol{\eta}\big)_{\mathbb{R}^N} = 0$ for all $\boldsymbol{\xi} \in \mathbb{R}^N$ and $\boldsymbol{\eta} \in L$, then $\Pi^\top \mathbf{X}$ is independent of $(\mathbf{I} - \Pi^\top)\mathbf{X}$.

(**iv**) Write

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{(11)} & \mathbf{C}_{(12)} \\ \mathbf{C}_{(21)} & \mathbf{C}_{(22)} \end{pmatrix},$$

where the block structure corresponds to $\mathbb{R}^N = \mathbb{R}^{N_1} \times \mathbb{R}^{N_2}$, and assume that $\mathbf{C}_{(22)}$ is non-degenerate. Show that the one and only transformation Π of the sort in part (**iii**) is given by

$$\Pi = \begin{pmatrix} \mathbf{0}_{(11)} & \mathbf{0}_{(12)} \\ \mathbf{C}_{(22)}^{-1}\mathbf{C}_{(21)} & \mathbf{I}_{(22)} \end{pmatrix},$$

and therefore that

$$\Pi^\top = \begin{pmatrix} \mathbf{0}_{(11)} & \mathbf{C}_{(12)}\mathbf{C}_{(22)}^{-1} \\ \mathbf{0}_{(21)} & \mathbf{I}_{(22)} \end{pmatrix}.$$

Hint: Note that $\Pi\boldsymbol{\xi} = \mathbf{0}$ if $\boldsymbol{\xi}_{(2)} = \mathbf{0}_{(2)}$, $\Pi\boldsymbol{\xi} = \boldsymbol{\xi}$ if $\boldsymbol{\xi}_{(1)} = \mathbf{0}_{(1)}$, and that $\big(\mathbf{C}(\mathbf{I} - \Pi)\big)_{(21)} = \mathbf{0}_{(21)}$.

(**v**) Continuing with the assumption that $\mathbf{C}_{(22)}$ is non-degenerate, show that

$$\mathbf{X} = \begin{pmatrix} \mathbf{C}_{(12)}\mathbf{C}_{(22)}^{-1}\mathbf{Y} \\ \mathbf{Y} \end{pmatrix} + \begin{pmatrix} \mathbf{Z} \\ \mathbf{0} \end{pmatrix},$$

where $\mathbf{Y}$ is an $\mathbb{R}^{N_2}$-valued $\mathcal{N}(\mathbf{0}, \mathbf{C}_{(22)}$ random variable, $\mathbf{Z}$ is an $\mathbb{R}^{N_1}$-valued $\mathcal{N}(\mathbf{0}, \mathbf{B})$ random variable with $\mathbf{B} = \mathbf{C}_{(11)} - \mathbf{C}_{(12)}\mathbf{C}_{(22)}^{-1}\mathbf{C}_{(21)}\big)$, and $\mathbf{Y}$ is independent of $\mathbf{Z}$. Conclude that, for any measurable $F : \mathbb{R}^{N_1} \times \mathbb{R}^{N_2} \longrightarrow \mathbb{R}$ which is bounded below, then $\mathbb{E}^{\mathbb{P}}\Big[F\big(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}\big)\Big]$ equals

$$\int_{\mathbb{R}^{N_2}} \left(\int_{\mathbb{R}^{N_1}} F\big(\mathbf{x}_{(1)}, \mathbf{x}_{(2)}\big)\, \gamma_{\mathbf{C}_{(12)}\mathbf{C}_{(22)}^{-1}\mathbf{x}_{(2)}, \mathbf{B}}(d\mathbf{x}_{(1)}) \right) \gamma_{\mathbf{0}, C_{(22)}}(d\mathbf{x}_{(2)}).$$

Exercise 2.3.30. Given $h \in L^2(\mathbb{R}^N;\mathbb{C})$, recall that the convolution $h^{\star(n+2)}$ is a bounded continuous function for each $n \in \mathbb{N}$. Next, assume that $h(-\mathbf{x}) = \overline{h(\mathbf{x})}$ for almost every $\mathbf{x} \in \mathbb{R}^N$ and that $h \equiv 0$ off of $B_{\mathbb{R}^N}(\mathbf{0}, 1)$. As an application of part (**iii**) in Exercise 1.3.22, show that

$$\big|h^{\star(n+2)}(\mathbf{x})\big| \le 2\|h\|_{L^2(\mathbb{R}^N;\mathbb{C})}^2 \|h\|_{L^1(\mathbb{R}^N;\mathbb{C})}^n \exp\left[-\frac{\big((|\mathbf{x}| - 2)^+\big)^2}{2n} \right].$$

Hint: Note that $h \in L^1(\mathbb{R}^N;\mathbb{C})$, assume that $M \equiv \|h\|_{L^1(\mathbb{R}^N;\mathbb{C})} > 0$, and define $Af = M^{-1}h \star f$ for $f \in L^2(\mathbb{R}^N;\mathbb{C})$. Show that A is a self-adjoint contraction on $L^2(\mathbb{R}^N;\mathbb{C})$, check that

$$h^{\star(n+2)}(\mathbf{x}) = M^n\big(\tau_{\mathbf{x}}h, A^n h\big)_{L^2(\mathbb{R}^N;\mathbb{C})},$$

where $\tau_{\mathbf{x}}h \equiv h(\,\cdot + \mathbf{x})$, and note that

$$\big(\tau_{\mathbf{x}}h, A^\ell h\big)_{L^2(\mathbb{R}^N;\mathbb{C})} = 0 \quad \text{if } \ell \le |\mathbf{x}| - 2.$$

§2.4 An Application to Hermite Multipliers

This section does not really belong here and should probably be skipped by those readers who want to restrict their attention to purely probabilistic matters. On the other hand, for those who want to see how probability theory interacts with other aspects of mathematical analysis, the present section may come as something of a revelation.

§2.4.1. Hermite Multipliers. The topic of this section will be a class of linear operators called Hermite multipliers, and what will be discussed are certain boundedness properties of these operators. The setting is as follows. For $n \in \mathbb{N}$, define

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{d^n}{dx^n}\left(e^{-\frac{x^2}{2}}\right), \quad x \in \mathbb{R}. \tag{2.4.1}$$

Clearly, H_n is an nth order, real, monic (i.e., 1 is the coefficient of the highest order term) polynomial. Moreover, if we define the **raising operator** A_+ on $C^1(\mathbb{R};\mathbb{C})$ by

$$\big[A_+\varphi\big](x) = -e^{\frac{x^2}{2}} \frac{d}{dx}\left(e^{-\frac{x^2}{2}}\varphi(x)\right) = -\frac{d\varphi}{dx}(x) + x\varphi(x), \quad x \in \mathbb{R},$$

then

$$H_{n+1} = A_+ H_n \quad \text{for all } n \in \mathbb{N}. \tag{2.4.2}$$

At the same time, if φ and ψ are continuously differentiable functions whose first derivatives are **tempered** (i.e., have at most polynomial growth at infinity), then

$$\big(\varphi, A_+\psi\big)_{L^2(\gamma_{0,1};\mathbb{C})} = \big(A_-\varphi, \psi\big)_{L^2(\gamma_{0,1};\mathbb{C})}, \tag{2.4.3}$$

where A_- is the **lowering operator** given by $A_-\varphi = \frac{d\varphi}{dx}$. After combining (2.4.2) with (2.4.3), we see that, for all $0 \le m \le n$,

$$\big(H_m, H_n\big)_{L^2(\gamma_{0,1};\mathbb{C})} = \big(H_m, A_+^n H_0\big)_{L^2(\gamma_{0,1};\mathbb{C})} = \big(A_-^n H_m, H_0\big)_{L^2(\gamma_{0,1};\mathbb{C})} = m!\,\delta_{m,n},$$

where, at the last step, we have used the fact that H_m is a monic mth order polynomial. Hence, the (normalized) **Hermite polynomials**

$$\overline{H}_n(x) = \frac{H_n(x)}{\sqrt{n!}} = \frac{(-1)^n}{\sqrt{n!}} e^{\frac{x^2}{2}} \frac{d^n}{dx^n}\left(e^{-\frac{x^2}{2}}\right), \quad x \in \mathbb{R}$$

form an orthonormal set in $L^2(\gamma_{0,1};\mathbb{C})$. (Indeed, they are one choice of the orthogonal polynomials relative to the Gauss weight.)

LEMMA 2.4.4. *For each $\lambda \in \mathbb{C}$, set*

$$H(x;\lambda) = \exp\left[\lambda x - \frac{\lambda^2}{2}\right], \quad x \in \mathbb{R}.$$

Then

$$H(x;\lambda) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} H_n(x), \quad x \in \mathbb{R}, \tag{2.4.5}$$

where the convergence is both uniform on compact subsets of $\mathbb{R}\times\mathbb{C}$ and, for λ's in compact subsets of $\mathbb{C}$, uniform in $L^2(\gamma_{0,1};\mathbb{C})$. In particular, $\{\overline{H}_n : n \in \mathbb{N}\}$ is an orthonormal basis in $L^2(\gamma_{0,1};\mathbb{C})$.

PROOF: By (2.4.1) and Taylor's expansion for the function $e^{-\frac{x^2}{2}}$, it is clear that (2.4.5) holds for each (x,λ) and that the convergence is uniform on compact subsets of $\mathbb{R}\times\mathbb{C}$. Furthermore, because the H_n's are orthogonal, the asserted uniform convergence in $L^2(\gamma_{0,1};\mathbb{C})$ comes down to checking that

$$\lim_{m\to\infty} \sup_{|\lambda|\le R} \sum_{n=m}^{\infty} \left|\frac{\lambda^n}{n!}\right|^2 \|H_n\|^2_{L^2(\gamma_{0,1})} = 0$$

for every $R \in (0,\infty)$, and obviously this follows from our earlier calculation that $\|H_n\|^2_{L^2(\gamma_{0,1})} = n!$.

To prove the assertion that $\{\overline{H}_n : n \in \mathbb{N}\}$ forms an orthonormal basis, it suffices to check that any $\varphi \in L^2(\gamma_{0,1};\mathbb{C})$ which is orthogonal to all of the H_n's must be 0. But, because of the $L^2(\gamma_{0,1};\mathbb{C})$-convergence in (2.4.5), we would have that

$$\int_{\mathbb{R}} \varphi(x)\, e^{\lambda x}\, \gamma_{0,1}(dx) = 0, \quad \lambda \in \mathbb{C},$$

for such a φ. Hence, if we set

$$\psi(x) = \frac{e^{-\frac{x^2}{2}}\varphi(x)}{\sqrt{2\pi}}, \quad x \in \mathbb{R},$$

then $\|\psi\|_{L^1(\mathbb{R};\mathbb{C})} = \|\varphi\|_{L^1(\gamma_{0,1};\mathbb{C})} \le \|\varphi\|_{L^2(\gamma_{0,1};\mathbb{C})} < \infty$, and (cf. (2.3.2)) $\hat{\psi} \equiv 0$; which, by the $L^1(\mathbb{R};\mathbb{C})$-Fourier inversion formula

$$\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\alpha|\xi|}\, e^{-\sqrt{-1}\,x\xi}\hat{\psi}(\xi)\,d\xi \xrightarrow{\alpha\searrow 0} \psi \quad \text{in } L^1(\mathbb{R};\mathbb{C}),$$

means that ψ and therefore φ vanish Lebesgue-almost everywhere. □

Now that we know $\{\overline{H}_n : n \in \mathbb{N}\}$ is an orthonormal basis, we can uniquely determine a normal operator $\mathcal{H}_\theta$ for each $\theta \in \mathbb{C}$ by specifying that

$$\mathcal{H}_\theta H_n = \theta^n H_n \quad \text{for each } n \in \mathbb{N}.$$

The operator $\mathcal{H}_\theta$ is called the **Hermite multiplier** with parameter θ, and clearly

$$\mathrm{Dom}(\mathcal{H}_\theta) = \left\{ \varphi \in L^2(\gamma_{0,1};\mathbb{C}) : \sum_{n=1}^{\infty} |\theta|^{2n} \left| \left(\varphi, \overline{H}_n\right)_{L^2(\gamma_{0,1};\mathbb{C})} \right|^2 < \infty \right\}$$

$$\mathcal{H}_\theta \varphi = \sum_{n=0}^{\infty} \theta^n \left(\varphi, \overline{H}_N\right)_{L^2(\gamma_{0,1};\mathbb{C})} \overline{H}_n, \quad \varphi \in \mathrm{Dom}(\mathcal{H}_\theta).$$

In particular, $\mathcal{H}_\theta$ is a contraction if and only if θ is an element of the closed unit disk $\mathbb{D}$ in $\mathbb{C}$, and it is unitary precisely when $\theta \in \mathbb{S}^1 \equiv \partial\mathbb{D}$. Also, the adjoint of $\mathcal{H}_\theta$ is $\mathcal{H}_{\overline{\theta}}$, and so it is self-adjoint if and only if $\theta \in \mathbb{R}$.

As we are about to see, there are special choices of θ for which the corresponding Hermite multiplier has interesting alternative interpretations and unexpected additional properties. For example, consider the **Mehler kernel***

$$M(x,y;\theta) = \frac{1}{\sqrt{1-\theta^2}} \exp\left[-\frac{(\theta x)^2 - 2\theta xy + (\theta y)^2}{2(1-\theta^2)} \right]$$

for $\theta \in (0,1)$ and $x, y \in \mathbb{R}$. By a straightforward Gaussian computation (i.e., "complete the square" in the exponential) one can easily check that

$$\int_{\mathbb{R}} H(y;\lambda)\, M(x,y;\theta)\, \gamma_{0,1}(dy) = H(x;\theta\lambda)$$

for all $\theta \in (0,1)$ and $(x,\lambda) \in \mathbb{R}\times\mathbb{C}$. In conjunction with (2.4.5), this means that

$$\text{(2.4.6)} \qquad \mathcal{H}_\theta \varphi = \int_{\mathbb{R}} M(\,\cdot\,, y;\theta)\, \varphi(y)\, \gamma_{0,1}(dy), \quad \theta \in (0,1) \text{ and } \varphi \in L^2(\gamma_{0,1};\mathbb{C}),$$

and from here it is not very difficult to prove the following properties of $\mathcal{H}_\theta$ for $\theta \in (0,1)$.

* This kernel appears in the 1866 article by Mehler referred to in the footnote following (2.1.14). It arises there as the generating function for spherical harmonics on the sphere $\mathbb{S}^\infty\left(\sqrt{\infty}\right)$.

Lemma 2.4.7. *For each $\varphi \in L^2(\gamma_{0,1};\mathbb{C})$, $(\theta,x) \in (0,1)\times\mathbb{R} \longmapsto \mathcal{H}_\theta\varphi(x) \in \mathbb{C}$ may be chosen to be a continuous function which is nonnegative if $\varphi \geq 0$ Lebesgue-almost everywhere. In addition, for each $\theta \in (0,1)$ and every $p \in [1,\infty]$,*

$$\big\|\mathcal{H}_\theta\varphi\big\|_{L^p(\gamma_{0,1};\mathbb{C})} \leq \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})}. \tag{2.4.8}$$

Proof: The first assertions are immediate consequences of the representation in (2.4.6). To prove the second assertion, observe that $\mathcal{H}_\theta\mathbf{1} = \mathbf{1}$ and therefore, as a special case of (2.4.6),

$$\int_{\mathbb{R}} M(x,y;\theta)\,\gamma_{0,1}(dy) = 1 \quad \text{for all} \quad \theta \in (0,1) \text{ and } x \in \mathbb{R}.$$

Hence, by (2.4.6) and Jensen's inequality, for any $p \in [1,\infty)$,

$$\big|\big[\mathcal{H}_\theta\varphi\big](x)\big|^p \leq \int_{\mathbb{R}} M(x,y;\theta)\,|\varphi(y)|^p\,\gamma_{0,1}(dy).$$

At the same time, by symmetry, $\int_{\mathbb{R}} M(x,y;\theta)\,\gamma_{0,1}(dx) = 1$ for all $(\theta,y) \in (0,1)\times\mathbb{R}$, and therefore

$$\int_{\mathbb{R}} \big|\big[\mathcal{H}_\theta\varphi\big](x)\big|^p\,\gamma_{0,1}(dx) \leq \iint_{\mathbb{R}\times\mathbb{R}} M(x,y;\theta)\,|\varphi(y)|^p\,\gamma_{0,1}(dx)\gamma_{0,1}(dy) = \int_{\mathbb{R}} |\varphi|^p\,d\gamma_{0,1}.$$

Hence, (2.4.8) is now proved for $p \in [1,\infty)$. The case when $p = \infty$ is even easier and is left to the reader. □

The consequences drawn in Lemma 2.4.7 from the Mehler representation in (2.4.6) are interesting but not very deep (cf. Exercise 2.3.23 below). A deeper fact is the relationship between Hermite multipliers and the Fourier transform. For the purposes of this analysis, it is best to define the **Fourier operator** $\mathcal{F}$ by

$$\big[\mathcal{F}f\big](\xi) = \int_{\mathbb{R}} e^{\sqrt{-1}\,2\pi\xi x}\,f(x)\,dx, \quad \xi \in \mathbb{R}, \tag{2.4.9}$$

for $f \in L^1(\mathbb{R};\mathbb{C})$. The advantage of this choice is that, without the introduction of any further factors of $\sqrt{2\pi}$, the Parseval identity (cf. Exercise 2.3.24) becomes the statement that $\mathcal{F}$ determines a unitary operator on $L^2(\mathbb{R};\mathbb{C})$. In order to relate $\mathcal{F}$ to Hermite multipliers, observe that, after analytically continuing the result of another simple Gaussian computation,

$$\frac{1}{\sqrt{2\pi p}}\int_{\mathbb{R}} \exp\left[\big(\lambda + \sqrt{-1}\,\eta\big)y - \frac{y^2}{2p}\right]\,dy = \exp\left[\frac{p}{2}\big(\lambda + \sqrt{-1}\,\eta\big)^2\right]$$

for all $p \in (1,\infty)$ and all real numbers λ and η. Hence, after making the change of variables $y = \sqrt{2\pi p}\,x$ and $\eta = \sqrt{\frac{2\pi}{p}}\,\xi$, we see from (2.4.5) that

$$\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\int_{\mathbb{R}} e^{\sqrt{-1}\,2\pi\xi x} H_n\big(\sqrt{2\pi p}\,x\big)e^{-\pi x^2}\,dx$$
$$= e^{-\pi\xi^2}\exp\left[\frac{(p-1)\lambda^2}{2} + \sqrt{-1}\,\lambda\sqrt{2\pi p}\,\xi\right] = e^{-\pi\xi^2}\sum_{n=0}^{\infty}\frac{\lambda^n}{n!}\theta_p^n H_n\big(\sqrt{2\pi p'}\,\xi\big),$$

where $p' = \frac{p}{p-1}$ is the **Hölder conjugate** of p and $\theta_p \equiv \sqrt{-1}\,(p-1)^{\frac{1}{2}}$. Thus, we have now proved that, for each $p \in (1,\infty)$ and $n \in \mathbb{N}$,

$$\int_{\mathbb{R}} e^{\sqrt{-1}\,2\pi\xi x} H_n\big(\sqrt{2\pi p}\,x\big)e^{-\pi x^2}\,dx = \theta_p^n\, H_n\big(\sqrt{2\pi p'}\,x\big)\,e^{-\pi\xi^2}. \tag{2.4.10}$$

In particular, when $p=2$, (2.4.10) says that

$$\mathcal{F}h_n = \big(\sqrt{-1}\,\big)^n h_n, \quad n \in \mathbb{N}, \tag{2.4.11}$$

where h_n is the nth (un-normalized) **Hermite function** given by

$$h_n(x) = H_n\big(2\pi^{\frac{1}{2}}x\big)e^{-\pi x^2}, \quad n \in \mathbb{N} \text{ and } x \in \mathbb{R}. \tag{2.4.12}$$

More generally, (2.4.10) leads to the following relationship between $\mathcal{F}$ and Hermite multipliers. Namely, for each $p \in (1,\infty)$, define $\mathcal{U}_p$ on $L^p(\gamma_{0,1};\mathbb{C})$ by

$$\big[\mathcal{U}_p\varphi\big](x) = p^{\frac{1}{2p}}\varphi\big((2\pi p)^{\frac{1}{2}}x\big)e^{-\pi x^2}, \quad x \in \mathbb{R}.$$

It is then an easy matter to check that $\mathcal{U}_p$ is an isometric surjection from $L^p(\gamma_{0,1};\mathbb{C})$ onto $L^p(\mathbb{R};\mathbb{C})$. In addition, (2.4.10) can now be interpreted as the statement that, for every $p \in (1,\infty)$ and every polynomial φ,

$$\mathcal{U}_{p'}^{-1}\circ\mathcal{F}\circ\mathcal{U}_p\varphi = A_p\mathcal{H}_{\theta_p}\varphi \quad \text{where} \quad A_p \equiv \left(\frac{p^{\frac{1}{p}}}{(p')^{\frac{1}{p'}}}\right)^{\frac{1}{2}}. \tag{2.4.13}$$

See Exercise 2.3.21 below to see that $A_p < 1$ for $p \in (0,1)$.

§2.4.2. Beckner's Theorem. Having completed this brief introduction to Hermite multipliers, we will now address a problem to which The Central Limit Theorem has something to contribute. The problem is that of determining the set of $(\theta,p,q) \in \mathbb{D}\times(1,\infty)\times(0,\infty)$ with $p \le q$ for which $\mathcal{H}_\theta$ determines a contraction from $L^p(\gamma_{0,1};\mathbb{C})$ into $L^q(\gamma_{0,1};\mathbb{C})$. In view of the preceding discussion, when $\theta \in (0,1)$, a solution to this problem has implications for the Mehler transform; and, when $q = p'$, the solution tells us about the Fourier operator. The rôle that The Central Limit Theorem plays in this analysis is hidden in the following beautiful criterion, which was first discovered by Wm. Beckner.*

* See Beckner's "Inequalities in Fourier analysis," *Ann. Math.* **102**.

Theorem 2.4.14 (**Beckner**). *Let $\theta \in \mathbb{D}$ and $1 \le p \le q < \infty$ be given. Then*

$$\big\|\mathcal{H}_\theta\varphi\big\|_{L^q(\gamma_{0,1};\mathbb{C})} \le \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})} \quad \textit{for all} \quad \varphi \in L^2(\gamma_{0,1};\mathbb{C}) \tag{2.4.15}$$

if and only if

$$\left(\frac{|1-\theta\zeta|^q + |1+\theta\zeta|^q}{2}\right)^{\frac{1}{q}} \le \left(\frac{|1-\theta\zeta|^p + |1+\theta\zeta|^q}{2}\right)^{\frac{1}{p}} \tag{2.4.16}$$

for every $\zeta \in \mathbb{C}$. Hence

That (2.4.15) implies (2.4.15) is trivial: simply take

$$\varphi(x) = \begin{cases} 1-\zeta & \text{if} \quad x \in (-\infty,0) \\ 1+\zeta & \text{if} \quad x \in [0,\infty). \end{cases}$$

On the other hand, the opposite implication is remarkable! Indeed, it takes a problem in infinite dimensional analysis and reduces it to a calculus question about functions on the complex plane. Even though, as we will see later, this reduction leads to highly nontrivial problems in calculus, Theorem 2.4.14 has to be considered a major step toward understanding the contraction properties of Hermite multipliers.*

The first step in the proof of Theorem 2.4.14 is to interpret (2.4.15) in operator theoretic language. For this purpose, let β denote the standard Bernoulli probability measure on $(\mathbb{R},\mathcal{B}_{\mathbb{R}})$. That is, $\beta\big(\{\pm 1\}\big) = \frac{1}{2}$. Next, use $\chi_\emptyset$ to denote the function on $\mathbb{R}$ which is constantly equal to 1 and $\chi_{\{1\}}$ to stand for the identity function on $\mathbb{R}$ (i.e., $\chi_{\{1\}}(x) = x,\ x \in \mathbb{R}$). It is then clear that $\chi_\emptyset$ and $\chi_{\{1\}}$ constitute an orthonormal basis in $L^2(\beta;\mathbb{C})$; in fact, they are the orthogonal polynomials there. Hence, for each $\theta \in \mathbb{C}$, we can define the **Bernoulli multiplier** $\mathcal{K}_\theta$ as the unique normal operator on $L^2(\beta;\mathbb{C})$ prescribed by

$$\mathcal{K}_\theta\chi_F = \begin{cases} \chi_\emptyset & \text{if} \quad F = \emptyset \\ \theta\chi_{\{1\}} & \text{if} \quad F = \{1\}. \end{cases}$$

Furthermore, (2.4.15) is equivalent to the statement that

$$\big\|\mathcal{K}_\theta\varphi\big\|_{L^q(\beta;\mathbb{C})} \le \|\varphi\|_{L^p(\beta,\mathbb{C})} \quad \text{for all} \quad \varphi \in L^2(\beta;\mathbb{C}). \tag{2.4.17}$$

Indeed, it is obvious that (2.4.15) is equivalent to (2.4.17) restricted to φ's of the form $x \in \mathbb{R} \longmapsto 1 + \zeta x$ as ζ runs over $\mathbb{C}$; and from this, together with

* Later, in his article "Gaussian kernels have only Gaussian maximizers," *Invent. Math.* **12** (1990), E. Lieb has essentially killed this line of research. His argument, which is entirely different from the one discussed here, handles not only the Hermite multipliers but essentially every operator whose kernel can be represented as the exponential of a second order polynomial.

the observation that every element of $L^2(\beta;\mathbb{C})$ can be represented in the form $a\chi_\emptyset + b\chi_{\{1\}}$ as (a,b) runs over $\mathbb{C}^2$, one quickly concludes that (2.4.15) implies (2.4.17) for general $\varphi \in L^2(\beta;\mathbb{C})$.

We next want to show that (2.4.17) can be parlayed into a seemingly more general statement. To this end, we define the n-fold tensor product operator $\mathcal{K}_\theta^{\otimes n}$ on $L^2(\beta^n;\mathbb{C})$ as follows. For $F \subseteq \{1,\dots,n\}$ set $\chi_F \equiv 1$ if $F = \emptyset$ and define

$$\chi_F(\mathbf{x}) = \prod_{j\in F} \chi_{\{1\}}(x_j) \quad \text{for} \quad \mathbf{x} = \bigl(x_1,\dots,x_n\bigr) \in \mathbb{R}^n$$

if $F \neq \emptyset$. Note that $\bigl\{\chi_F : F \subseteq \{1,\dots,n\}\bigr\}$ is an orthonormal basis for $L^2(\beta^n;\mathbb{C})$, and define $\mathcal{K}_\theta^{\otimes n}$ to be the unique normal operator on $L^2(\beta^n;\mathbb{C})$ for which

$$\mathcal{K}_\theta^{\otimes n}\chi_F = \theta^{|F|}\chi_F, \quad F \subseteq \{1,\dots,n\}, \tag{2.4.18}$$

where $|F|$ is used here to denote the number of elements in the set F. Alternatively, one can describe $\mathcal{K}_\theta^{\otimes n}$ inductively on $n \in \mathbb{Z}^+$ by saying that $\mathcal{K}_\theta^{\otimes 1} = \mathcal{K}_\theta$ and that, for $\Phi \in C\bigl(\mathbb{R}^{n+1};\mathbb{C}\bigr)$ and $(\mathbf{x},y) \in \mathbb{R}^n \times \mathbb{R}$,

$$\bigl[\mathcal{K}_\theta^{\otimes(n+1)}\Phi\bigr](\mathbf{x},y) = \bigl[\mathcal{K}_\theta\Psi(\mathbf{x},\,\cdot\,)\bigr](y) \quad \text{where} \quad \Psi(\mathbf{x},y) = \bigl[\mathcal{K}_\theta^{\otimes n}\Phi(\,\cdot\,,y)\bigr](\mathbf{x}).$$

It is this alternative description which makes it easiest to see the extension of (2.4.17) alluded to above. Namely, what we will now show is that, for every $n \in \mathbb{Z}^+$,

$$(2.4.17) \implies \bigl\|\mathcal{K}_\theta^{\otimes n}\Phi\bigr\|_{L^q(\beta^n)} \le \|\Phi\|_{L^p(\beta^n)}, \quad \Phi \in L^2(\beta^n;\mathbb{C}). \tag{2.4.19}$$

Obviously, there is nothing to do when $n = 1$. Next, assume (2.4.19) for n, let $\Phi \in C\bigl(\mathbb{R}^{n+1};\mathbb{C}\bigr)$ be given, and define Ψ as in the second description of $\mathcal{K}_\theta^{\otimes(n+1)}\Phi$. Then, by (2.4.17) applied to $\Psi(\mathbf{x},\,\cdot\,)$ for each $\mathbf{x} \in \mathbb{R}^n$ and by the induction hypothesis applied to $\Phi(\,\cdot\,,y)$ for each $y \in \mathbb{R}$, we have that

$$\begin{aligned}
\bigl\|\mathcal{K}_\theta^{\otimes(n+1)}\Phi\bigr\|^q_{L^q(\beta^{n+1})} &= \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}}\Bigl|\bigl[\mathcal{K}_\theta\Psi(\mathbf{x},\,\cdot\,)\bigr](y)\Bigr|^q\,\beta(dy)\right)\beta^n(d\mathbf{x}) \\
&\le \int_{\mathbb{R}^n}\left(\int_{\mathbb{R}}|\Psi(\mathbf{x},y)|^p\,\beta(dy)\right)^{\frac{q}{p}}\beta^n(d\mathbf{x}) = \left\|\int_{\mathbb{R}^n}\bigl|\Psi(\,\cdot\,,y)\bigr|^p\,\beta(dy)\right\|^{\frac{q}{p}}_{L^{\frac{q}{p}}(\beta^n)} \\
&\le \left(\int_{\mathbb{R}^n}\bigl\|\,|\Psi(\,\cdot\,,y)|^p\,\bigr\|_{L^{\frac{q}{p}}(\beta^n)}\beta(dy)\right)^{\frac{q}{p}} = \left(\int_{\mathbb{R}}\bigl\|\Psi(\,\cdot\,,y)\bigr\|^p_{L^q(\beta^n)}\beta(dy)\right)^{\frac{q}{p}} \\
&\le \left(\int_{\mathbb{R}}\bigl\|\Phi(\,\cdot\,,y)\bigr\|^p_{L^p(\beta^n)}\beta(dy)\right)^{\frac{q}{p}} = \|\Phi\|^q_{L^p(\beta^{n+1})},
\end{aligned}$$

where, in the passage to the third line, we have used the continuous form of Minkowski's equality (it is at this point that the only essential use of the hypothesis $p \le q$ is made).

We are now ready to take the main step in the proof of Theorem 2.4.14.

Lemma 2.4.20. *Define* $\mathcal{A}_n : L^2(\beta;\mathbb{C}) \longrightarrow L^2\big(\beta^n;\mathbb{C}\big)$ *by*

$$\big[\mathcal{A}_n\varphi\big](\mathbf{x}) = \varphi\left(\frac{\sum_{\ell=1}^n x_\ell}{\sqrt{n}}\right) \quad \text{for} \quad \mathbf{x}\in\mathbb{R}^n.$$

Then, for every pair of tempered φ *and* ψ *from* $C(\mathbb{R};\mathbb{C})$,

$$\|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})} = \lim_{n\to\infty}\big\|\mathcal{A}_n\varphi\big\|_{L^p(\beta^n;\mathbb{C})} \quad \text{for every} \quad p\in[1,\infty) \tag{2.4.21}$$

and

$$\Big(\mathcal{H}_\theta\varphi,\psi\Big)_{L^2(\gamma_{0,1};\mathbb{C})} = \lim_{n\to\infty}\Big(\mathcal{K}_\theta^{\otimes n}\circ\mathcal{A}_n\varphi,\mathcal{A}_n\psi\Big)_{L^2(\beta^n;\mathbb{C})} \tag{2.4.22}$$

for every $\theta\in(0,1)$. *Moreover, if, in addition, either* φ *or* ψ *is a polynomial, then* (2.4.22) *continues to hold for all* $\theta\in\mathbb{C}$.

Proof: Let φ and ψ be tempered elements of $C(\mathbb{R};\mathbb{C})$, and define

$$f_n(\theta) = \Big(\mathcal{K}_\theta^{\otimes n}\circ\mathcal{A}_n\varphi,\mathcal{A}_n\psi\Big)_{L^2(\beta^n)} \quad \text{and} \quad f(\theta) = \Big(\mathcal{H}_\theta\varphi,\psi\Big)_{L^2(\gamma_{0,1})}$$

for $n\in\mathbb{Z}^+$ and $\theta\in\mathbb{C}$. We begin by showing that

$$\lim_{n\to\infty} f_n(\theta) = f(\theta), \quad \theta\in(0,1). \tag{2.4.23}$$

Notice that (2.4.23) is (2.4.22) for $\theta\in(0,1)$ and therefore that (2.4.21) follows immediately from (2.4.23) by replacing φ and ψ, respectively, with $\mathbf{1}$ and $|\varphi|^p$.

In order to prove (2.4.23), we will need to introduce other expressions for $f(\theta)$ and the $f_n(\theta)$'s. To this end, set

$$\mathbf{C}_\theta = \begin{pmatrix} 1 & \theta \\ \theta & 1 \end{pmatrix},$$

and, using (2.4.6), observe (cf. (2.3.6)) that

$$f(\theta) = \int_{\mathbb{R}^2} \varphi(x)\,\overline{\psi(\mathbf{y})}\,\gamma_{\mathbf{0},\mathbf{C}_\theta}(dx\times dy).$$

Next, let $\boldsymbol{\beta}_\theta$ be the probability measure on $\mathbb{R}^2$ determined by

$$\boldsymbol{\beta}_\theta\big(\{\pm1,\pm1\}\big) = \tfrac{1+\theta}{4} \quad \text{and} \quad \boldsymbol{\beta}_\theta\big(\{\pm1,\mp1\}\big) = \tfrac{1-\theta}{4};$$

and note that, because

$$\Big(\mathcal{K}_\theta\varphi,\psi\Big)_{L^2(\beta)} = \int_{\mathbb{R}^2}\varphi(x)\,\overline{\psi(y)}\,\boldsymbol{\beta}_\theta(dx\times dy),$$

one has that

$$\Big(\mathcal{K}_\theta^{\otimes n}\varPhi, \Psi\Big)_{L^2(\beta)} = \int_{\mathbb{R}^2}\cdots\int_{\mathbb{R}^2} \varPhi(\mathbf{x})\,\overline{\Psi(\mathbf{y})}\,\boldsymbol{\beta}_\theta(dx_1\times dy_1)\cdots\boldsymbol{\beta}_\theta(dx_n\times dy_n)$$

for all $\varPhi$, $\Psi \in C(\mathbb{R}^n;\mathbb{C})$. Hence, if (cf. Exercise 1.1.12) $\Omega = \big(\mathbb{R}^2\big)^{\mathbb{Z}^+}$, $\mathcal{F} = \mathcal{B}_\Omega$, and $\mathbb{P}_\theta = \big(\boldsymbol{\beta}_\theta\big)^{\mathbb{Z}^+}$, then

$$f_n(\theta) = \mathbb{E}^{P_\theta}\left[F\left(\frac{\sum_1^n \mathbf{Z}_n}{\sqrt{n}}\right)\right],$$

where $F(\mathbf{z}) \equiv \varphi(x)\,\overline{\psi(y)}$ for $\mathbf{z} = (x,y) \in \mathbb{R}^2$ and $\mathbf{Z}_n(\omega) = \mathbf{z}_n$, $n \in \mathbb{Z}^+$, when $\omega = (\mathbf{z}_1, \dots, \mathbf{z}_n, \dots) \in \Omega$. Note that, under $\mathbb{P}_\theta$, the $\mathbf{Z}_n$'s are mutually independent, identically distributed $\mathbb{R}^2$-valued random variables with mean-value $\mathbf{0}$ and covariance $\mathbf{C}_\theta$. In addition, $\mathbf{Z}_1$ is bounded, and therefore the last part of Theorem 2.3.19 applies and guarantees that (2.4.23) holds.

To complete the proof, suppose that φ is a polynomial of degree k. It is then an easy matter to check that

$$\big(\mathcal{A}_n\varphi, \chi_F\big)_{L^2(\beta^n;\mathbb{C})} = 0 \quad \text{if } |F| > k,$$

and therefore (cf. (2.4.18)) $\theta \in \mathbb{C} \longmapsto f_n(\theta) \in \mathbb{C}$ is also a polynomial of degree no more than k. Moreover, because

$$\big|f_n(\theta)\big| = \left|\sum_F \theta^{|F|}\big(\mathcal{A}_n\varphi, \chi_F\big)_{L^2(\beta^n)}\big(\chi_F, \mathcal{A}_n\psi\big)_{L^2(\beta^n)}\right|,$$

we also know that

$$\big|f_n(\theta)\big| \le \big(|\theta| \vee 1\big)^k \big\|\mathcal{A}_n\varphi\big\|_{L^2(\beta^n;\mathbb{C})}\,\big\|\mathcal{A}_n\psi\big\|_{L^2(\beta^n;\mathbb{C})}, \quad n \in \mathbb{Z}^+ \text{ and } \theta \in \mathbb{C}.$$

Hence, because of (2.4.21) with $p = 2$, $\{f_n : n \in \mathbb{Z}^+\}$ is a family of entire functions on $\mathbb{C}$ which are uniformly bounded on compact subsets. At the same time, because $(\varphi, H_m)_{L^2(\gamma_{0,1})} = 0$ for $m > k$, f is also a polynomial of degree $k \wedge n$; and therefore (2.4.23) already implies that the convergence extends to the whole of $\mathbb{C}$ and is uniform on compacts. Finally, in the case when ψ, instead of φ, is a polynomial, simply note that

$$\Big(\mathcal{K}_\theta^{\otimes n}\circ\mathcal{A}_n\varphi, \mathcal{A}_n\psi\Big)_{L^2(\beta^n)} = \overline{\Big(\mathcal{K}_{\bar\theta}^{\otimes n}\circ\mathcal{A}_n\psi, \mathcal{A}_n\varphi\Big)}_{L^2(\beta^n)},$$

and apply the preceding. $\square$

Proof of Theorem 2.4.14: Assume that (2.4.15) holds for a given pair $1 < p \le q < \infty$ and $\theta \in \mathbb{D}$. We then know that (2.4.19) holds for every $n \in \mathbb{Z}^+$. Hence, by Lemma 2.4.20, if φ and ψ are tempered elements of $C(\mathbb{R};\mathbb{C})$ and at least one of them is a polynomial, then

$$\begin{aligned}\Big|\big(\mathcal{H}_\theta\varphi,\psi\big)_{L^2(\gamma_{0,1};\mathbb{C})}\Big| &= \lim_{n\to\infty}\Big|\Big(\mathcal{K}_\theta^{\otimes n}\circ\mathcal{A}_n\varphi,\mathcal{A}_n\psi\Big)_{L^2(\beta^n;\mathbb{C})}\Big| \\ &\le \varliminf_{n\to\infty}\big\|\mathcal{A}_n\varphi\big\|_{L^p(\beta^n;\mathbb{C})}\big\|\mathcal{A}_n\psi\big\|_{L^{q'}(\beta^n;\mathbb{C})} = \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})}\|\psi\|_{L^{q'}(\gamma_{0,1};\mathbb{C})}.\end{aligned}$$

In other words, we now know that for all tempered φ and ψ from $C(\mathbb{R};\mathbb{C})$

$$\Big|\big(\mathcal{H}_\theta\varphi,\psi\big)_{L^2(\gamma_{0,1};\mathbb{C})}\Big| \le \|\varphi\|_{L^p(\gamma_{0,1})}\|\psi\|_{L^{q'}(\gamma_{0,1};\mathbb{C})} \tag{2.4.24}$$

so long as one or the other is a polynomial.

We next complete the proof in the case when $p \in (1,2]$. To this end, note that, for any fixed polynomial φ, (2.4.24) for every tempered $\psi \in C(\mathbb{R};\mathbb{C})$ guarantees that the inequality in (2.4.15) holds for that φ. At the same time, because $p \in (1,2]$ and the polynomials are dense in $L^2(\gamma_{0,1};\mathbb{C})$, (2.4.15) follows immediately from its own restriction to polynomials.

Finally, assume that $p \in [2,\infty)$ and therefore that $q' \in (1,2]$. Then, again because the polynomials are dense in $L^2(\gamma_{0,1};\mathbb{C})$, (2.4.24) for a fixed tempered $\varphi \in C(\mathbb{R};\mathbb{C})$ and all polynomials ψ implies (2.4.15) first for all tempered continuous ψ's and thence for all $\psi \in L^2(\gamma_{0,1};\mathbb{C})$. $\square$

§2.4.3. Applications of Beckner's Theorem. We will now apply Theorem 2.4.14 to two important examples. The first example involves the case when $\theta \in (0,1)$ and shows that the contraction property proved in Lemma 2.4.7 can be improved to say that, for each $p \in (1,\infty)$ and $\theta \in (0,1)$, there is a $q = q(p,\theta) \in (p,\infty)$ such that $\mathcal{H}_\theta$ is a contraction on $L^p(\gamma_{0,1};\mathbb{C})$ into $L^q(\gamma_{0,1};\mathbb{C})$. Such an operator is said to be **hypercontractive**, and the fact that $\mathcal{H}_\theta$ is hypercontractive was first proved by E. Nelson in connection with his renowned construction of a nontrivial, two-dimensional quantum field.* The proof which we will give is entirely different from Nelson's and is much closer to the ideas introduced by L. Gross† as they were developed by Beckner.

Theorem 2.4.25 (**Nelson**). *Let $\theta \in (0,1)$ and $p \in (1,\infty)$ be given, and set*

$$q(p,\theta) = 1 + \frac{p-1}{\theta^2}.$$

---

* Nelson's own proof appeared in his "The free Markov field," *J. Fnal. Anal.* **12**.

† See Gross's "Logarithmic Sobolev inequalities," *Amer. J. Math.* **97**. In this paper, Gross introduced the idea of proving estimates on $\mathcal{H}_\theta$ from the corresponding estimates for $\mathcal{K}_\theta$. In this connection, have a look at Exercises 2.3.28 and 2.3.29 below.

Then

$$\left\|\mathcal{H}_\theta\varphi\right\|_{L^q(\gamma_{0,1};\mathbb{C})} \le \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})}, \quad \varphi \in L^2(\gamma_{0,1};\mathbb{C}), \tag{2.4.26}$$

for every $1 \le q \le q(p,\theta)$. *Moreover, if* $q > q(p,\theta)$, *then*

$$\sup\left\{\left\|\mathcal{H}_\theta\varphi\right\|_{L^q(\gamma_{0,1})} : \|\varphi\|_{L^q(\gamma_{0,1};\mathbb{C})} = 1\right\} = \infty. \tag{2.4.27}$$

Proof: We will leave the proof of (2.4.27) as an exercise. (Try taking φ's of the form $e^{\lambda x^2}$.) Also, because $\gamma_{0,1}$ is a probability measure and therefore the left-hand side of (2.4.26) is nondecreasing as a function of q, we will restrict our attention to the proof of (2.4.26) for $q = q(p,\theta)$. Hence, by Theorem 2.4.14, what we have to do is prove (2.4.15) for every $1 < p < q < \infty$ and $\theta \in (0,1)$ which are related by

$$\theta = \left(\frac{p-1}{q-1}\right)^{\frac{1}{2}}. \tag{2.4.28}$$

We begin with the case when $1 < p < q \le 2$; and we first consider $\zeta \in [0,1)$. Introducing the generalized binomial coefficients

$$\binom{r}{\ell} \equiv \frac{r(r-1)\cdots(r-\ell+1)}{\ell!} \quad \text{for} \quad r \in \mathbb{R} \text{ and } \ell \in \mathbb{N},$$

we can write

$$\frac{|1-\theta\zeta|^q + |1+\theta\zeta|^q}{2} = 1 + \sum_{k=1}^\infty \binom{q}{2k}(\theta\zeta)^{2k}$$

and

$$\frac{|1-\zeta|^p + |1+\zeta|^p}{2} = 1 + \sum_{k=1}^\infty \binom{p}{2k}\zeta^{2k}.$$

Noting that, because $q \le 2$, $\binom{q}{2k} \ge 0$ for every $k \in \mathbb{Z}^+$, and using the fact that, because $\frac{p}{q} \in (0,1)$, $(1+x)^{\frac{p}{q}} \le 1 + \frac{p}{q}x$ for all $x \ge 0$, we see that

$$\left(\frac{|1-\theta\zeta|^q + |1+\theta\zeta|^q}{2}\right)^{\frac{p}{q}} \le 1 + \frac{p}{q}\sum_{k=1}^\infty \binom{q}{2k}(\theta\zeta)^{2k}.$$

Hence, we will have completed the case under consideration once we check that

$$\frac{p}{q}\sum_{k=1}^\infty \binom{q}{2k}(\theta\zeta)^{2k} \le \sum_{k=1}^\infty \binom{p}{2k}\zeta^{2k};$$

and clearly this will follow if we show that

$$\frac{p}{q}\binom{q}{2k}\theta^{2k} \le \binom{p}{2k} \quad \text{for each} \quad k \in \mathbb{Z}^+.$$

But the choice of θ in (2.4.28) makes the preceding an equality when $k = 1$; and, when $k \ge 2$,

$$\frac{\frac{p}{q}\binom{q}{2k}\theta^{2k}}{\binom{p}{2k}} \le \prod_{j=2}^{2k-1} \frac{j-q}{j-p} \le 1,$$

since $1 < p < q \le 2$.

At this point, we have proved (2.4.15) for $1 < p < q \le 2$ and θ given by (2.4.28) when $\zeta \in (0,1)$. Continuing with this choice of p, q, and θ, note that (2.4.15) extends immediately to $\zeta \in [-1,1]$ by continuity and symmetry. Finally, for general $\zeta \in \mathbb{C}$, set

$$a = \frac{|1-\zeta| + |1+\zeta|}{2}, \quad b = \frac{|1-\zeta| - |1+\zeta|}{2}, \quad \text{and } c = \frac{b}{a}.$$

Then,

$$|1 \pm \theta\zeta| = \left|\tfrac{1+\theta}{2}(1 \pm \zeta) + \tfrac{1-\theta}{2}(1 \mp \zeta)\right| \le a \pm \theta b,$$

and therefore, by the preceding applied to $c \in [-1,1]$, we have that

$$\begin{aligned}
&\left(\frac{|1-\theta\zeta|^q + |1+\theta\zeta|^q}{2}\right)^{\frac{1}{q}} \le a\left(\frac{|1-\theta c|^q + |1+\theta c|^q}{2}\right)^{\frac{1}{q}} \\
&\le a\left(\frac{|1-c|^p + |1+c|^p}{2}\right)^{\frac{1}{p}} = \left(\frac{|a-b|^p + |a+b|^p}{2}\right)^{\frac{1}{p}} = \left(\frac{|1-\zeta|^p + |1+\zeta|^p}{2}\right)^{\frac{1}{p}}.
\end{aligned}$$

Hence, we have now completed the case when $1 < p < q \le 2$ and θ is given by (2.4.28).

To handle the other cases, we use the equivalence of (2.4.15) and (2.4.17). Thus, what we already know is that (2.4.17) holds for $1 < p < q \le 2$ and the θ in (2.4.28). Next, suppose that $2 \le p < q < \infty$. Then, since $1 < q' < p' \le 2$ and

$$\frac{p-1}{q-1} = \frac{q'-1}{p'-1},$$

an application to q' and p' of the result which we already have yields

$$\begin{aligned}
\|\mathcal{K}_\theta\varphi\|_{L^q(\beta)} &= \sup\left\{\left(\mathcal{K}_\theta\varphi, \psi\right)_{L^2(\beta)} : \psi \in L^2(\beta;\mathbb{C}) \text{ with } \|\psi\|_{L^{q'}(\beta)} = 1\right\} \\
&= \sup\left\{\left(\varphi, \mathcal{K}_\theta\psi\right)_{L^2(\beta)} : \psi \in L^2(\beta;\mathbb{C}) \text{ with } \|\psi\|_{L^{q'}(\beta)} = 1\right\} \\
&\le \|\varphi\|_{L^p(\beta)},
\end{aligned}$$

where the θ is the one given in (2.4.28). Thus, the only case which remains is the one when $1 < p \le 2 \le q < \infty$. But, in this case, set $\xi = (p-1)^{\frac{1}{2}}$, $\eta = (q-1)^{-\frac{1}{2}}$, and observe that, because the associated θ in (2.4.28) is the product of ξ with η, $\mathcal{K}_\theta = \mathcal{K}_\eta \circ \mathcal{K}_\xi$ and therefore

$$\left\|\mathcal{K}_\theta\varphi\right\|_{L^q(\beta)} \le \left\|\mathcal{K}_\xi\varphi\right\|_{L^2(\beta)} \le \|\varphi\|_{L^p(\beta)}. \quad \square$$

As our second, and final, application of Theorem 2.4.14, we present the theorem of Beckner for which Theorem 2.4.14 was concocted in the first place. The result was conjectured originally by H. Weyl, who guessed, on the basis of $\mathcal{F}h_0 =)\sqrt{-1})^n h_0$, that the norm $\|\mathcal{F}\|_{p\to p'}$ of $\mathcal{F}$ as an operator on $L^p(\mathbb{R};\mathbb{C})$ to $L^{p'}(\mathbb{R};\mathbb{C})$ should be achieved by h_0. Weyl's conjecture was partially verified by I. Babenko when p' is an even integer. In particular, when combined with Riesz–Thorin Interpolation Theorem, Babenko's result already shows (cf. Exercise 2.3.21) that $\|\mathcal{F}\|_{p\to p'} < 1$ for $p \in (0,1)$.

Theorem 2.4.29 (**Beckner**). *For each $p \in [1,2]$,*

$$\|\mathcal{F}f\|_{L^{p'}(\mathbb{R};\mathbb{C})} \le A_p\|f\|_{L^p(\mathbb{R};\mathbb{C})}, \quad f \in L^1(\mathbb{R};\mathbb{C}) \cap L^2(\mathbb{R};\mathbb{C}), \tag{2.4.30}$$

where $\mathcal{F}$ is the Fourier operator in (2.4.9) and A_p is the constant in (2.4.13). Moreover, if f is the Gauss kernel $e^{-\pi x^2}$, then (2.4.30) is an equality.

Proof: Because of (2.4.11), the second part is a straightforward computation which we leave to the reader. Also, we will only consider (2.4.30) when $p \in (1,2)$, the other cases being well-known (cf. Exercise 2.4.33).

Because of (2.4.13), the proof of (2.4.30) comes down to showing that

$$\left\|\mathcal{H}_{\theta_p}\varphi\right\|_{L^{p'}(\gamma_{0,1};\mathbb{C})} \le \|\varphi\|_{L^p(\gamma_{0,1};\mathbb{C})}, \quad \varphi \in L^2(\gamma_{0,1};\mathbb{C}),$$

where $\theta_p = \sqrt{-1}\,(p-1)^{\frac{1}{2}}$; and, by Theorem 2.4.14, this will follow as soon as we prove (2.4.16) for θ_p. For this purpose, write

$$\zeta = \xi + \sqrt{-1}\,(p-1)^{-\frac{1}{2}}\eta \quad \text{where} \quad \xi,\, \eta \in \mathbb{R}.$$

Then, proving (2.4.16) for θ_p becomes the problem of checking that

$$\left(\frac{\Big[\big(1-\eta\big)^2 + (p-1)\xi^2\Big]^{\frac{p'}{2}} + \Big[\big(1+\eta\big)^2 + (p-1)\xi^2\Big]^{\frac{p'}{2}}}{2}\right)^{\frac{1}{p'}}$$

$$\le \left(\frac{\Big[\big(1-\xi\big)^2 + (p-1)\eta^2\Big]^{\frac{p}{2}} + \Big[\big(1+\xi\big)^2 + (p-1)\eta^2\Big]^{\frac{p}{2}}}{2}\right)^{\frac{1}{p}} \tag{*}$$

for all $\xi,\, \eta \in \mathbb{R}$.

To prove (*), consider, for each $\alpha \in (0,\infty)$, the function $g_\alpha : [0,\infty)^2 \longrightarrow [0,\infty)$ defined by $g_\alpha(x,y) = \bigl[x^{\frac{1}{\alpha}} + y^{\frac{1}{\alpha}}\bigr]^\alpha$. It is an easy matter to check that g_α is concave or convex depending on whether $\alpha \in [1,\infty)$ or $\alpha \in (0,1)$. In particular, since $\frac{p'}{2} \in (1,\infty)$, when we set $\alpha = \frac{p'}{2}$, $x_\pm = |1 \pm \eta|^{p'}$, and $y = (p-1)^{\frac{p'}{2}}|\xi|^{p'}$, we get

$$\begin{aligned}
&\frac{\Bigl[\bigl(1-\eta\bigr)^2 + (p-1)\xi^2\Bigr]^{\frac{p'}{2}} + \Bigl[\bigl(1+\eta\bigr)^2 + (p-1)\xi^2\Bigr]^{\frac{p'}{2}}}{2} \\
&\qquad = \frac{g_\alpha\bigl(x_-,y\bigr) + g_\alpha\bigl(x_+,y\bigr)}{2} \le g_\alpha\left(\frac{x_- + x_+}{2}, y\right) \\
&\qquad = \left[\left(\frac{|1-\eta|^{p'} + |1+\eta|^{p'}}{2}\right)^{\frac{p}{2}} + (p-1)\xi^2\right]^{\frac{p'}{2}};
\end{aligned}$$

and similarly, because $\frac{p}{2} \in (0,1)$,

$$\begin{aligned}
&\frac{\Bigl[\bigl(1-\xi\bigr)^2 + (p-1)\eta^2\Bigr]^{\frac{p}{2}} + \Bigl[\bigl(1+\xi\bigr)^2 + (p-1)\eta^2\Bigr]^{\frac{p}{2}}}{2} \\
&\qquad \ge \left[\left(\frac{|1-\xi|^{p} + |1+\xi|^{p}}{2}\right)^{\frac{2}{p}} + (p-1)\eta^2\right]^{\frac{p}{2}}.
\end{aligned}$$

Thus, (*) will be proved if we show that

$$\text{(**)}\qquad \begin{aligned}
&\left(\frac{|1-\eta|^{p'} + |1+\eta|^{p'}}{2}\right)^{\frac{2}{p'}} + (p-1)\xi^2 \\
&\qquad \le \left(\frac{|1-\xi|^{p} + |1+\xi|^{p}}{2}\right)^{\frac{2}{p}} + (p-1)\eta^2.
\end{aligned}$$

But because (cf. Theorems 2.4.14 and 2.4.25) we know that (2.4.16) holds with p replaced by 2, $q = p'$, and $\theta = \bigl(p'-1\bigr)^{\frac{1}{2}}$, the left side of (**) is dominated by

$$(p-1)\xi^2 + \frac{\left(1 - \frac{\eta}{(p'-1)^{\frac{1}{2}}}\right)^2 + \left(1 + \frac{\eta}{(p'-1)^{\frac{1}{2}}}\right)^2}{2} = 1 + (p-1)\bigl(\xi^2 + \eta^2\bigr).$$

At the same time, again by (2.4.16), only this time with p, 2, and $\theta = (p-1)^{-\frac{1}{2}}$, we see that the right-hand side of (**) dominates

$$(p-1)\eta^2 + \frac{\bigl(1 - (p-1)^{\frac{1}{2}}\xi\bigr)^2 + \bigl(1 + (p-1)^{\frac{1}{2}}\xi\bigr)^2}{2} = 1 + (p-1)\bigl(\xi^2 + \eta^2\bigr). \quad \square$$

Exercises for § 2.4

Exercise 2.4.31. Because the Fourier operator $\mathcal{F}$ (cf. (2.4.9)) is a contraction from $L^1(\mathbb{R};\mathbb{C})$ to $L^\infty(\mathbb{R};\mathbb{C})$ as well as from $L^2(\mathbb{R};\mathbb{C})$ into $L^2(\mathbb{R};\mathbb{C})$, the Riesz–Thorin Interpolation Theorem guarantees that it is a contraction from $L^p(\mathbb{R})$ into $L^{p'}(\mathbb{R})$ for each $p \in (0,1)$. Hence, we know, from Theorem 2.4.29, that the number A_p in (2.4.13) must be less than or equal to 1. However, the preceding is a rather convoluted line of reasoning to what must be a far more elementary fact. Indeed, show that

$$t \in \left(\tfrac{1}{2}, 1\right) \longmapsto \log A_{\frac{1}{t}} \in \mathbb{R}$$

is a strictly convex function which tends to 0 at both end points and is therefore strictly negative. In particular, Beckner's result proves is that the Fourier operator is one for which interpolation fails to give the best result.

Exercise 2.4.32. The inequality in (2.4.8) is an example of a general principle. Namely, if $(E,\mathcal{B})$ is any measurable space, then a map $(x,\Gamma) \in E \times \mathcal{B} \longmapsto \Pi(x,\Gamma) \in [0,1]$ is called a **transition probability** whenever $x \in E \longmapsto \Pi(x,\Gamma)$ is $\mathcal{B}$-measurable for each $\Gamma \in \mathcal{B}$ and $\Gamma \in \mathcal{B} \longmapsto \Pi(x,\Gamma)$ is a probability measure on $(E,\mathcal{B})$ for each $x \in E$. Given a transition probability $\Pi(x,\,\cdot\,)$, we define the linear operator Π on $B(E;\mathbb{C})$ (the space of bounded, $\mathcal{B}$-measurable $\varphi : E \longrightarrow \mathbb{C}$) by

$$\left[\Pi\varphi\right](x) = \int_E \varphi(y)\,\Pi(x,dy), \quad x \in E, \quad \text{for} \quad \varphi \in B(E;\mathbb{C}).$$

Check that Π takes $B(E;\mathbb{C})$ into itself and that $\|\Pi\varphi\|_{\mathrm{u}} \le \|\varphi\|_{\mathrm{u}}$. Next, given a σ-finite measure μ on $(E,\mathcal{B})$, we say that μ is **Π-invariant** if

$$\mu(\Gamma) = \int_E \Pi(x,\Gamma)\,\mu(dx) \quad \text{for all} \quad \Gamma \in \mathcal{B}.$$

Using Jensen's inequality, first show that, for each $p \in [1,\infty)$,

$$\left|\left[\Pi\varphi\right](x)\right|^p \le \left[\Pi|\varphi|^p\right](x), \quad x \in E,$$

and then that, for any Π-invariant μ,

$$\|\Pi\varphi\|_{L^p(\mu;\mathbb{C})} \le \|\varphi\|_{L^p(\mu;\mathbb{C})}, \quad \varphi \in B(E;\mathbb{C}).$$

Finally, show that μ is Π-invariant if it is **Π-reversing** in the sense that

$$\int_{\Gamma_1} \Pi\big(x,\Gamma_2\big)\,\mu(dx) = \int_{\Gamma_2} \Pi\big(y,\Gamma_1\big)\,\mu(dy) \quad \text{for all} \quad \Gamma_1,\,\Gamma_2 \in \mathcal{B}.$$

EXERCISE 2.4.33. Recall the Hermite functions h_n, $n \in \mathbb{N}$, in (2.4.12) and define the **normalized Hermite functions** $\overline{h}_n$, $n \in \mathbb{N}$ by

$$\overline{h}_n = \frac{2^{\frac{1}{4}}}{(n!)^{\frac{1}{2}}}\, h_n, \quad n \in \mathbb{N}.$$

By noting that (cf. the discussion following (2.4.12)) $\overline{h}_n = \mathcal{U}_2\overline{H}_n$, show that $\{\overline{h}_n : n \in \mathbb{N}\}$ constitutes an orthonormal basis in $L^2(\mathbb{R};\mathbb{C})$; and from this together with (2.4.11), arrive at **Parseval's Identity**

$$\|\mathcal{F}f\|_{L^2(\mathbb{R};\mathbb{C})} = \|f\|_{L^2(\mathbb{R};\mathbb{C})}, \quad f \in L^1(\mathbb{R};\mathbb{C}) \cap L^2(\mathbb{R};\mathbb{C}),$$

and conclude that $\mathcal{F}$ determines a unique unitary operator $\overline{\mathcal{F}}$ on $L^2(\mathbb{R};\mathbb{C})$ such that $\overline{\mathcal{F}}f = \mathcal{F}f$ for $f \in L^1(\mathbb{R};\mathbb{C}) \cap L^2(\mathbb{R};\mathbb{C})$. Finally, use this to verify the L^2-Fourier inversion formula $\overline{\mathcal{F}}^{-1} = \tilde{\mathcal{F}}$, where $[\tilde{\mathcal{F}}f](x) \equiv [\mathcal{F}f](-x)$, $x \in \mathbb{R}$, for $f \in L^1(\mathbb{R};\mathbb{C}) \cap L^2(\mathbb{R};\mathbb{C})$.

EXERCISE 2.4.34. By the same reasoning as we used to prove Theorem 2.4.29, show that, for any pair $1 < p \le 2 \le q < \infty$ and any complex number $\theta = \xi + \sqrt{-1}\,\eta$, (2.4.16), and therefore (2.4.15), holds if and only if both $(q-1)\eta^2+\xi^2 \le 1$ and

$$(q-2)(\xi\eta)^2 \le \big[1-\xi^2-(q-1)\eta^2\big]\big[(p-1)-(q-1)\alpha^2-\beta^2\big].$$

EXERCISE 2.4.35. L. Gross had a somewhat different approach to the proof of (2.4.26). As in the proof which we have given, he reduced everything to checking (2.4.17). However, he did this in a different way. Namely, given $b \in (0,1)$ he set $f(x) = 1 + bx$ and introduced the functions

$$f_t(x) \equiv \big[\mathcal{K}_{e^{-t}}f\big](x) = \tfrac{1+e^{-t}}{2}f(x) + \tfrac{1-e^{-t}}{2}f(-x), \quad (t,x) \in [0,\infty)\times\mathbb{R},$$

and $q(t) = 1 + (p-1)e^{2t}$, $t \in [0,\infty)$, and proved that

$$(*) \qquad \frac{d}{dt}\big\|f_t\big\|_{L^{q(t)}(\beta;\mathbb{C})} \le 0.$$

Following the steps below, see if you can reproduce Gross's calculation.

(**i**) Set

$$F(t) = \|f_t\|_{L^{q(t)}(\beta);\mathbb{C}},$$

and, by somewhat tedious but completely elementary differential calculus, show that

$$\frac{dF}{dt}(t) = \frac{F(t)^{1-q(t)}}{q(t)^2}\Bigg[-\dot{q}(t)\int_{\mathbb{R}} f^{q(t)}\log\Big(\tfrac{f_t}{F(t)}\Big)^{q(t)}\,d\beta + \tfrac{q(t)^2}{2}\int_{\mathbb{R}} f_t(x)^{q(t)-1}\big(f_t(-x)-f_t(x)\big)\,\beta(dx)\Bigg].$$

Next, check that

$$\int_{\mathbb{R}} f_t(x)^{q(t)-1}\big(f_t(-x)-f_t(x)\big)\,\beta(dx)$$
$$= -\tfrac{1}{2}\int_{\mathbb{R}}\big(f_t(x)^{q(t)-1}-f_t(-x)^{q(t)-1}\big)\big(f_t(x)-f_t(-x)\big)\,\beta(dx),$$

and, after verifying that

$$\big(\xi^{q-1}-\eta^{q-1}\big)(\xi-\eta)\ge \frac{4(q-1)\big(\xi^{\frac{q}{2}}-\eta^{\frac{q}{2}}\big)^2}{q^2},\quad \xi,\,\eta\in(0,\infty)\text{ and } q\in(1,\infty),$$

conclude that

$$(**)\qquad \begin{aligned}\tfrac{dF}{dt}(t)\le \frac{F(t)^{1-q(t)}}{q(t)^2}\Bigg[&-\dot q(t)\int_{\mathbb{R}} f^{q(t)}\log\Big(\tfrac{f_t}{F(t)}\Big)^{q(t)}\,d\beta\\ &+\big(q(t)-1\big)\int_{\mathbb{R}}\big(f_t(x)^{\frac{q(t)}{2}}-f_t^{\frac{q(t)}{2}}(-x)\big)^2\,\beta(dx)\Bigg].\end{aligned}$$

(**ii**) Prove the **logarithmic Sobolev inequality**

$$(2.4.36)\qquad \int_{\mathbb{R}}\varphi^2\log\Big(\tfrac{\varphi}{\|\varphi\|_{L^2(\beta;\mathbb{C})}}\Big)^2\,d\beta\le 2\int_{\mathbb{R}}\big(\varphi(x)-\varphi(-x)\big)^2\,\beta(dx)$$

for strictly positive φ's on $\mathbb{R}$.

Hint: Reduce to the case when $\varphi(x)=1+bx$ for some $b\in(0,1)$, and, in this case, check that (2.4.36) is the elementary calculus inequality

$$(1+b)^2\log(1+b)+(1-b)^2\log(1-b)-(1+b^2)\log(1+b^2)\le 2b^2,\quad b\in(0,1).$$

(**iii**) By plugging (2.4.36) into (**), arrive at (*), and conclude that (2.4.17) holds for $\theta\in(0,1)$ and $q=1+\frac{p-1}{\theta^2}$.

Exercise 2.4.37. The major difference between Gross and Beckner's approach to proving Nelson's Theorem 2.4.25 is that Gross based his proof on the equivalence of contraction results like (2.4.17) and (2.4.15) to logarithmic Sobolev inequalities like (2.4.36). In Exercise 2.4.34, we outlined how one passes from a logarithmic Sobolev inequality to a contraction result. The object of this exercise is to go in the opposite direction. Specifically, starting from (2.4.26), show that

$$(2.4.38)\qquad \int_{\mathbb{R}}\varphi^2\log\Bigg(\frac{\varphi}{\|\varphi\|_{L^2(\gamma_{0,1};\mathbb{C})}}\Bigg)^2\,d\gamma_{0,1}\le 2\int_{\mathbb{R}}|\varphi'|^2\,\gamma_{0,1}(dx)$$

for non-negative, continuously differentiable $\varphi\in L^2(\gamma_{0,1};\mathbb{C})\setminus\{0\}$ with $\varphi'\in L^2(\gamma_{0,1};\mathbb{C})$