Chapter IX
Convergence of Measures on a Polish Space

In Chapters II and III I introduced a notion of convergence on $\mathbf{M}_1(\mathbb{R}^N)$ which is appropriate when discussing either Central Limit phenomena or the sort of limits which arose in connection with infinitely divisible laws. In this chapter, I will give a systematic treatment of this sort of convergence and show how it extends to probability measures on any **Polish space**. That is, any complete, separable, metric space. Unfortunately, this extension will entail an excursion into territory which borders on *abstract nonsense*, although I hope to avoid crossing that border. In any case, just as Banach's great achievement was the ingenious use for infinite dimensional vector spaces of completeness to replace local compactness, so here we will have to learn how to substitute compactness by completeness in measure-theoretic arguments.

§ 9.1 Prohorov–Varadarajan Theory

The goal in this section is to generalize results like Lemma 2.1.7 and Theorem 3.1.1 to a very abstract setting.

§ 9.1.1. Some Background. When discussing the convergence of probability measures on a measurable space $(E, \mathcal{B})$, one always has at least two senses in which the convergence may take place, and (depending on additional structure that the space may possess) one may have more. To be more precise, let $B(E;\mathbb{R}) \equiv B\big((E,\mathcal{B});\mathbb{R}\big)$ be the space of bounded, $\mathbb{R}$-valued, $\mathcal{B}$-measurable functions on E, use $\mathbf{M}_1(E) \equiv \mathbf{M}_1(E,\mathcal{B})$ to denote the space of all probability measures on $(E,\mathcal{B})$, and define the **duality relation**

$$\langle \varphi, \mu \rangle = \int_E \varphi \, d\mu \quad \text{for } \varphi \in B(E;\mathbb{R}) \text{ and } \mu \in \mathbf{M}_1(E).$$

Next, again use $\|\varphi\|_{\mathrm{u}} \equiv \sup_{x\in E} |\varphi(x)|$ to denote the uniform norm of $\varphi \in B(E;\mathbb{R})$, and consider the neighborhood basis at $\mu \in \mathbf{M}_1(E)$ determined by the sets

$$U(\mu, r) = \big\{\nu \in \mathbf{M}_1(E) : \big|\langle \varphi, \nu\rangle - \langle \varphi, \mu\rangle\big| \le r \text{ for } \varphi \in B(E,\mathbb{R}) \text{ with } \|\varphi\|_{\mathrm{u}} \le 1\big\}$$

as r runs over $(0,\infty)$. For obvious reasons, the topology defined by these neighborhoods U is called the **uniform topology** on $\mathbf{M}_1(E)$. In order to develop some feeling for the uniform topology, I will begin by examining a few of its elementary properties.

LEMMA 9.1.1. *Define the variation distance between elements μ and ν of $\mathbf{M}_1(E)$ by*

$$\|\nu - \mu\|_{\mathrm{var}} = \sup\Big\{\big|\langle\varphi,\mu\rangle - \langle\varphi,\nu\rangle\big| : \varphi \in B(E;\mathbb{R}) \text{ with } \|\varphi\|_{\mathrm{u}} \le 1\Big\}.$$

Then $(\mu,\nu) \in \mathbf{M}_1(E)^2 \longmapsto \|\mu-\nu\|_{\mathrm{var}}$ is a metric on $\mathbf{M}_1(E)$ which is compatible with the uniform topology. Moreover, if $\mu,\, \nu \in \mathbf{M}_1(E)$ are two elements of $\mathbf{M}_1(E)$ and λ is any element of $\mathbf{M}_1(E)$ with respect to which both μ and ν are absolutely continuous (e.g., $\frac{\mu+\nu}{2}$), then

$$\|\mu - \nu\|_{\mathrm{var}} = \|g - f\|_{L^1(\lambda;\mathbb{R})} \quad \text{where} \quad f = \frac{d\mu}{d\lambda} \text{ and } g = \frac{\partial\nu}{\partial\lambda}. \tag{9.1.2}$$

In particular, $\|\mu-\nu\|_{\mathrm{var}} \le 2$, and equality holds precisely when $\nu \perp \mu$ (i.e., they are singular). Finally, the metric $(\mu,\nu) \in \mathbf{M}_1(E)^2 \longmapsto \|\mu - \nu\|_{\mathrm{var}}$ is complete.

PROOF: The first assertion needing comment is the one in (9.1.2). But, for every $\varphi \in B(E;\mathbb{R})$ with $\|\varphi\|_{\mathrm{u}} \le 1$,

$$\Big|\langle\varphi,\nu\rangle - \langle\varphi,\mu\rangle\Big| = \left|\int_E \varphi(g-f)\,d\lambda\right| \le \|g-f\|_{L^1(\lambda;\mathbb{R})},$$

and equality holds when $\varphi = \operatorname{sgn}\circ(g-f)$. To prove the assertion which follows (9.1.2), note that

$$\|g-f\|_{L^1(\lambda;\mathbb{R})} \le \|f\|_{L^1(\lambda;\mathbb{R})} + \|g\|_{L^1(\lambda;\mathbb{R})} = 2$$

and that the inequality is strict if and only if $fg > 0$ on a set of strictly positive λ-measure or, equivalently, if and only if $\mu \not\perp \nu$. Thus, all that remains is to check the completeness assertion. To this end, let $\{\mu_n : n \ge 1\} \subseteq \mathbf{M}_1(E)$ satisfying

$$\lim_{m\to\infty} \sup_{n\ge m} \|\mu_n - \mu_m\|_{\mathrm{var}} = 0$$

be given, and set $\lambda = \sum_{n=1}^\infty 2^{-n}\mu_n$. Clearly, λ is an element of $\mathbf{M}_1(E)$ with respect to which each μ_n is absolutely continuous. Moreover, if $f_n = \frac{d\mu_n}{d\lambda}$, then, by (9.1.2), $\{f_n : n \ge 1\}$ is a Cauchy convergent sequence in $L^1(\lambda;\mathbb{R})$. Hence, since $L^1(\lambda;\mathbb{R})$ is complete, there is an $f \in L^1(\lambda;\mathbb{R})$ to which the f_n's converge in $L^1(\lambda;\mathbb{R})$. Obviously, we may choose f to be non-negative, and certainly it has λ-integral 1. Thus, the measure μ given by $d\mu = f\,d\lambda$ is an element of $\mathbf{M}_1(E)$, and, by (9.1.2), $\|\mu_n - \mu\|_{\mathrm{var}} \longrightarrow 0$. $\square$

As a consequence of Lemma 9.1.1, we see that the uniform topology on $\mathbf{M}_1(E)$ admits a complete metric and that convergence in this topology is intimately related to L^1-convergence in the L^1-space of an appropriate element of $\mathbf{M}_1(E)$.

In fact, $\mathbf{M}_1(E)$ looks in the uniform topology like a *galaxy* which is broken into many *constellations*, each constellation consisting of measures which are all absolutely continuous with respect to some fixed measure. In particular, there will usually be too many *constellations* for $\mathbf{M}_1(E)$ in the uniform topology to be separable. To wit, if E is uncountable and $\{x\} \in \mathcal{B}$ for every $x \in E$, then the **point masses** δ_x, $x \in E$, (i.e., $\delta_x(\Gamma) = \mathbf{1}_\Gamma(x)$) form an uncountable subset of $\mathbf{M}_1(E)$ and $\|\delta_y - \delta_x\|_{\text{var}} = 2$ for $y \neq x$. Hence, in this case, $\mathbf{M}_1(E)$ cannot be covered by a countable collection of open $\|\cdot\|_{\text{var}}$-balls of radius 1.

As I said at the beginning of this section, the uniform topology is not the only one available. Indeed, for many purposes and, in particular, for probability theory, it is *too rigid* a topology to be useful. For this reason, it is often convenient to consider a more lenient topology on $\mathbf{M}_1(E)$. The first one which comes to mind is the one which results from eliminating the *uniformity* in the uniform topology. That is, given a $\mu \in \mathbf{M}_1(E)$, define

$$S\big(\mu, \delta; \varphi_1, \dots, \varphi_n\big) \equiv \Big\{\nu \in \mathbf{M}_1(E) : \max_{1 \le k \le n} \big|\langle \varphi_k, \nu\rangle - \langle \varphi_k, \mu\rangle\big| < \delta\Big\} \tag{9.1.3}$$

for $n \in \mathbb{Z}^+$, $\varphi_1, \dots, \varphi_n \in B(E;\mathbb{R})$, and $\delta > 0$. Clearly these sets S determine a Hausdorff topology on $\mathbf{M}_1(E)$ in which the net $\{\mu_\alpha : \alpha \in A\}$ converges to μ if and only if $\lim_\alpha \langle \varphi, \mu_\alpha\rangle = \langle \varphi, \mu\rangle$ for every $\varphi \in B(E;\mathbb{R})$. For historical reasons, in spite of the fact that it is obviously *weaker* than the uniform topology, this topology on $\mathbf{M}_1(E)$ is sometimes called the **strong topology**; although, in some of the statistics literature, it is also known as the $\boldsymbol{\tau}$**-topology**.

A good understanding of the relationship between the strong and uniform topologies is most easily gained through functional analytic considerations which will not be particularly important for what follows. Nonetheless, it will be useful to recognize that, except in very special circumstances, the strong topology is *strictly weaker* than the uniform topology. For example, take $E = [0,1]$ with its Borel field, and consider the probability measures $\mu_n(dt) = \big(1 + \sin(2n\pi t)\big)\,dt$ for $n \in \mathbb{Z}^+$. Noting that, since $|\sin(2n\pi t) - \sin(2m\pi t)| \le 2$ and therefore

$$\begin{aligned}\tfrac{1}{2}\|\mu_n - \mu_m\|_{\text{var}} &= \int_0^1 \frac{|\sin(2n\pi t) - \sin(2m\pi t)|}{2}\,dt \\ &\ge \tfrac{1}{4}\int_0^1 \big|\sin(2n\pi t) - \sin(2m\pi t)\big|^2\,dt = \frac{1}{4}\end{aligned}$$

for $m \neq n$, one sees that $\{\mu_n : n \ge 1\}$ not only fails to converge in the uniform topology, it does not even have any limit points as $n \to \infty$. On the other hand, because $\big\{2^{\frac{1}{2}}\sin(2n\pi t) : n \ge 1\big\}$ is orthonormal in $L^2\big(\lambda_{[0,1]};\mathbb{R}\big)$, Bessel's inequality says that

$$2\sum_{n=1}^\infty \left(\int_{[0,1]} \varphi(t)\,\sin(2n\pi t)\,dt\right)^2 \le \|\varphi\|^2_{L^2(\lambda_{[0,1]})} \le \|\varphi\|_{\mathrm{u}}^2 < \infty$$

and therefore $\langle\varphi,\mu_n\rangle \longrightarrow \langle\varphi,\lambda_{[0,1]}\rangle$ for every $\varphi \in B\big([0,1];\mathbb{R}\big)$. In other words, $\{\mu_n : n \geq 1\}$ *converges to* $\lambda_{[0,1]}$ *in the strong topology, but it converges to nothing at all in the uniform topology.*

§ **9.1.2. The Weak Topology.** Although the strong topology is weaker than the uniform and can be effectively used in various applications, it is still not weak enough for most probabilistic applications. Indeed, even when E possesses a good topological structure and $\mathcal{B} = \mathcal{B}_E$ is the Borel field over E, the strong topology on $\mathbf{M}_1(E)$ shows no respect for the topology on E. For example, suppose that E is a metric space and, for each $x \in E$, consider the point mass δ_x on $\mathcal{B}_E$. Then, no matter how *close* $x \in E \setminus \{x\}$ gets to y in the sense of the topology on E, δ_x is not getting *close* to δ_y in the strong topology on $\mathbf{M}_1(E)$. More generally (cf. Exercise 9.1.15 below), measures cannot be close in the strong topology unless their *sets of small measure* are essentially the same. Thus, for example, the convergence which is occurring in The Central Limit Theorem (cf. Theorem 2.1.8) cannot, in general, be taking place in the strong topology; and since The Central Limit Theorem is an archetypal example of the sort of convergence result at which probabilists look, it is only sensible for us to take a hint from the result which we got there.

That is, let E be a metric space, set $\mathcal{B} = \mathcal{B}_E$, and consider the neighborhood basis at $\mu \in \mathbf{M}_1(E)$ given by the sets $S(\mu,\delta;\,\varphi_1,\dots,\varphi_n)$ in (9.1.3) when the φ_k's are *restricted to be elements of* $C_{\mathrm{b}}(E;\mathbb{R})$. The topology which results is *much weaker* than the strong topology, and is therefore justifiably called the **weak topology** on $\mathbf{M}_1(E)$. (The reader who is familiar with the language of functional analysis will, with considerable justice, complain about this terminology. Indeed, if one thinks of $C_{\mathrm{b}}(E;\mathbb{R})$ as a Banach space and of $\mathbf{M}_1(E)$ as a subspace of its dual space $C_{\mathrm{b}}(E;\mathbb{R})^*$, then the topology which I am calling the weak topology is what a functional analyst would call the weak* topology. However, because it is the most commonly accepted choice of probabilists, I will continue to use the term *weak* instead of the more correct term *weak**.) In particular, the weak topology respects the topology on E: δ_y tends to δ_x in the weak topology on $\mathbf{M}_1(E)$ if and only if $y \longrightarrow x$ in E. Lemma 2.3.3 provides further evidence that the weak topology is well adapted to the sort of analysis encountered in probability theory, since, by that lemma, weak convergence of $\{\mu_n : n \geq 1\} \subseteq \mathbf{M}_1(\mathbb{R}^N)$ to μ is equivalent to pointwise convergence of $\widehat{\mu_N}(\boldsymbol{\xi})$ to $\hat{\mu}(\boldsymbol{\xi})$.

Besides being well adapted to probabilistic analysis, the weak topology turns out to have many intrinsic virtues which are not shared by either the uniform or strong topologies. In particular, as we will see shortly, when E is a separable metric space, the weak topology on $\mathbf{M}_1(E)$ is not only a metric topology, which (cf. Exercise 9.1.15) the strong topology seldom is, but it is even separable, which, as we have seen, the uniform topology seldom is. In order to check these properties, we will first have to review some elementary facts about separable metric spaces.

Given a metric ρ for a topological space E, I will use $U^\rho_{\mathrm{b}}(E;\mathbb{R})$ to denote the space of bounded, ρ-uniformly continuous $\mathbb{R}$-valued functions on E and will endow $U^\rho_{\mathrm{b}}(E;\mathbb{R})$ with the topology determined by the uniform metric. Thus, $U^\rho_{\mathrm{b}}(E;\mathbb{R})$ becomes in this way a closed subspace of $C_{\mathrm{b}}(E;\mathbb{R})$.

LEMMA 9.1.4. *Let E be a separable metric space. Then E is homeomorphic to a subset of $[0,1]^{\mathbb{Z}^+}$. In particular:*

(**i**) *If E is compact, then the space $C(E;\mathbb{R})$ is separable with respect to the uniform metric.*

(**ii**) *Even when E is not compact, it nonetheless admits a metric $\hat\rho$ with respect to which it becomes a totally bounded metric space.*

(**iii**) *If $\hat\rho$ is a totally bounded metric on E, then $U^{\hat\rho}_{\mathrm{b}}(E;\mathbb{R})$ is separable.*

PROOF: Let ρ be any metric on E, and choose $\{p_n : n \ge 1\}$ to be a countable, dense subset of E. Next, define $\mathbf{h} : E \longrightarrow [0,1]^{\mathbb{Z}^+}$ to be the mapping whose nth coordinate is given by

$$h_n(x) = \frac{\rho(x,p_n)}{1+\rho(x,p_n)}, \quad x \in E.$$

It is then an easy matter to check that $\mathbf{h}$ is homeomorphic onto a subset of $[0,1]^{\mathbb{Z}^+}$.

To prove (**i**), I will first check it for compact subsets K of $E = [0,1]^{\mathbb{Z}^+}$. To this end, denote by $\mathcal{P}$ the space of polynomials $p : [0,1]^{\mathbb{Z}^+} \longrightarrow \mathbb{R}$. That is, $\mathcal{P}$ consists of finite, $\mathbb{R}$-linear combinations of the monomials $\boldsymbol{\xi} \in [0,1]^{\mathbb{Z}^+} \longmapsto \xi_{k_1}^{n_1}\cdots\xi_{k_\ell}^{n_\ell}$, where $\ell \ge 1$, $1 \le k_1 < \cdots < k_\ell$, and $\{n_1,\dots,n_\ell\} \subseteq \mathbb{N}$. Clearly, if $\mathcal{P}_0$ is the subset of $\mathcal{P}$ consisting of those p's with rational coefficients, then $\mathcal{P}_0$ is coutable, and $\mathcal{P}_0$ is dense in $\mathcal{P}$. Thus, it suffices to show that $\{p \restriction K : p \in \mathcal{P}\}$ is dense in $C(K;\mathbb{R})$. But $\mathcal{P}$ is obviously an algebra. In addition, if $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are distinct points in $[0,1]^{\mathbb{Z}^+}$, it is an easy (in fact, a one dimensional) matter to see that there is a $p \in \mathcal{P}$ for which $p(\boldsymbol{\xi}) \ne p(\boldsymbol{\eta})$. Hence, the desired density follows from the Stone–Weierstrass Approximation Theorem. Finally, for an arbitrary compact metric space E, define $\mathbf{h} : E \longrightarrow [0,1]^{\mathbb{Z}^+}$ as above, note that $K \equiv \mathbf{h}(E)$ is compact, and conclude that the map $\varphi \in C(K;\mathbb{R}) \longmapsto \varphi\circ\mathbf{h} \in C(E;\mathbb{R})$ is a homeomorphism between the uniform topologies on these spaces. Since we already know that $C(K;\mathbb{R})$ is separable, this completes (**i**).

The proof of (**ii**) is easy. Namely, define

$$D(\mathbf{x},\boldsymbol{\eta}) = \sum_{n=1}^\infty \frac{|\xi_n - \eta_n|}{2^n} \quad \text{for } \mathbf{x},\ \boldsymbol{\eta} \in [0,1]^{\mathbb{Z}^+}.$$

Clearly, D is a metric for $[0,1]^{\mathbb{Z}^+}$, and therefore

$$(x,y) \in E^2 \longmapsto \hat\rho(x,y) \equiv D\big(\mathbf{h}(x),\mathbf{h}(y)\big)$$

is a metric for E. At the same time, since $[0,1]^{\mathbb{Z}^+}$ is compact, and therefore the restriction of D to any subset is totally bounded, it is clear that $\hat{\rho}$ is totally bounded on E.

To prove (**iii**), let $\hat{E}$ denote the completion of E with respect to the totally bounded metric $\hat{\rho}$. Then, because E is dense in $\hat{E}$, $\hat{E}$ is both complete and totally bounded and therefore compact. In addition, $\hat{\varphi} \in C(\hat{E};\mathbb{R}) \longmapsto \hat{\varphi} \restriction E \in U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$ is a surjective homeomorphism; and so (**iii**) now follows from (**i**). $\square$

One of the main reasons why Lemma 9.1.4 will be important to us is that it will enable us to show that, for separable metric spaces E, the weak topology on $\mathbf{M}_1(E)$ is also a separable metric topology. However, thus far we do not even know that the neighborhood bases are countably generated, and so, for a moment longer, I must continue to consider nets when discussing convergence. In order to indicate that a net $\{\mu_\sigma : \alpha \in A\} \subseteq \mathbf{M}_1(E)$ is converging weakly (i.e., in the weak topoology) to μ, I will write $\mu_\alpha \Longrightarrow \mu$.

THEOREM 9.1.5. *Let E be any metric space and $\{\mu_\alpha : \alpha \in A\}$ a net in $\mathbf{M}_1(E)$. Given any $\mu \in \mathbf{M}_1(E)$, the following statements are equivalent:*

(**i**) $\mu_\alpha \Longrightarrow \mu$.

(**ii**) *If ρ is any metric for E, then $\langle \varphi, \mu_\alpha \rangle \longrightarrow \langle \varphi, \mu \rangle$ for every $\varphi \in U_{\mathrm{b}}^{\rho}(E;\mathbb{R})$.*

(**iii**) *For every closed set $F \subseteq E$, $\overline{\lim_\alpha}\, \mu_\alpha(F) \leq \mu(F)$.*

(**iv**) *For every open set $G \subseteq E$, $\underline{\lim_\alpha}\, \mu_\alpha(G) \geq \mu(G)$.*

(**v**) *For every upper semicontinuous function $f : E \longrightarrow \mathbb{R}$ which is bounded above, $\overline{\lim_\alpha} \langle f, \mu_\alpha \rangle \leq \langle f, \mu \rangle$.*

(**vi**) *For every lower semicontinuous function $f : E \longrightarrow \mathbb{R}$ which is bounded below, $\underline{\lim_\alpha} \langle f, \mu_\alpha \rangle \geq \langle f, \mu \rangle$.*

(**vii**) *For every $f \in B(E;\mathbb{R})$ which is continuous at μ-almost every $x \in E$, $\langle f, \mu_\alpha \rangle \longrightarrow \langle f, \mu \rangle$.*

Finally, assume that E is separable, and let $\hat{\rho}$ be a totally bounded metric for E. Then there exists a countable subset $\{\varphi_n : n \geq 1\} \subseteq U_{\mathrm{b}}^{\hat{\rho}}(E;[0,1])$ which is dense in $U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$, and therefore the mapping $\mathbf{H} : \mathbf{M}_1(E) \longrightarrow [0,1]^{\mathbb{Z}^+}$ given by $\mathbf{H}(\mu) = \big(\langle \varphi_1, \mu \rangle, \dots, \langle \varphi_n, \mu \rangle, \dots\big)$ is a homeomorphism from the weak topology on $\mathbf{M}_1(E)$ into $[0,1]^{\mathbb{Z}^+}$. In particular, when E is separable, $\mathbf{M}_1(E)$ with the weak topology is itself a separable metric space and, in fact, one can take

$$(\mu, \nu) \in \mathbf{M}_1(E)^2 \longmapsto R(\mu, \nu) \equiv \sum_{n=1}^{\infty} \frac{\big|\langle \varphi_n, \mu \rangle - \langle \varphi_n, \nu \rangle\big|}{2^n}$$

to be a metric for $\mathbf{M}_1(E)$.

Proof: The implications

$$(\mathbf{vii}) \implies (\mathbf{i}) \implies (\mathbf{ii}), \qquad (\mathbf{iii}) \iff (\mathbf{iv}), \quad \text{and } (\mathbf{v}) \iff (\mathbf{vi})$$

are all trivial. Thus, the first part will be complete once I check that (**ii**) $\implies$ (**iii**), (**iv**) $\implies$ (**vi**), and that (**v**) together with (**vi**) imply (**vii**). To see the first of these, let F be a closed subset of E, and set

$$\psi_n(x) = 1 - \left(\frac{\rho(x,F)}{1+\rho(x,F)}\right)^{\frac{1}{n}} \quad \text{for } n \in \mathbb{Z}^+ \text{ and } x \in E.$$

It is then clear that $\psi_n \in U^\rho_{\mathrm{b}}(E;\mathbb{R})$ for each $n \in \mathbb{Z}^+$ and that $1 \geq \psi_n(x) \searrow \mathbf{1}_F(x)$ as $n \to \infty$ for each $x \in E$. Thus, The Monotone Convergence Theorem followed by (**ii**) imply that

$$\mu(F) = \lim_{n\to\infty} \langle \psi_n, \mu\rangle = \lim_{n\to\infty} \lim_{\alpha} \langle \psi_n, \mu_\alpha\rangle \geq \overline{\lim_{\alpha}}\, \mu_\alpha(F).$$

In proving that (**iv**) $\implies$ (**vi**), I may and will assume that f is a non-negative, lower semicontinuous function. For $n \in \mathbb{N}$, define

$$f_n = \sum_{\ell=0}^{\infty} \frac{\ell \wedge 4^n}{2^n} \mathbf{1}_{I_{\ell,n}} \circ f = \frac{1}{2^n} \sum_{\ell=0}^{4^n} \mathbf{1}_{J_{\ell,n}} \circ f,$$

where

$$I_{\ell,n} = \left(\frac{\ell}{2^n}, \frac{\ell+1}{2^n}\right] \quad \text{and} \quad J_{\ell,n} = \left(\frac{\ell}{2^n}, \infty\right).$$

It is then clear that $0 \leq f_n \nearrow f$ and therefore that $\langle f_n, \mu\rangle \longrightarrow \langle f, \mu\rangle$ as $n \to \infty$. At the same time, by lower semicontinuity, the sets $\{f \in J_{\ell,n}\}$ are open, and so (**iv**) implies

$$\langle f_n, \mu\rangle \leq \underline{\lim_{\alpha}} \langle f_n, \mu_\alpha\rangle \leq \underline{\lim_{\alpha}} \langle f, \mu_\alpha\rangle$$

for each $n \in \mathbb{Z}^+$. After letting $n \to \infty$, one sees that (**iv**) $\implies$ (**vi**).

Turning to the proof that (**v**) & (**vi**) $\implies$ (**vii**), suppose that $f \in B(E;\mathbb{R})$ is continuous at μ-almost every $x \in E$, and define

$$\underline{f}(x) = \underline{\lim_{y\to x}}\, f(y) \quad \text{and} \quad \overline{f}(x) = \overline{\lim_{y\to x}}\, f(y) \quad \text{for } x \in E.$$

It is then an easy matter to check that $\underline{f} \leq f \leq \overline{f}$ everywhere and that equality holds μ-almost surely. Furthermore, $\underline{f}$ is lower semicontinuous, $\overline{f}$ is upper semicontinuous, and both are bounded. Hence, by (**v**) and (**vi**),

$$\overline{\lim_{\alpha}} \langle f, \mu_\alpha\rangle \leq \overline{\lim_{\alpha}} \langle \overline{f}, \mu_\alpha\rangle \leq \langle \overline{f}, \mu\rangle = \langle \underline{f}, \mu\rangle \leq \underline{\lim_{\alpha}} \langle \underline{f}, \mu_\alpha\rangle \leq \underline{\lim_{\alpha}} \langle f, \mu_\alpha\rangle;$$

and so we have now completed the proof that conditions (**i**) through (**vii**) are equivalent.

Now assume that E is separable, and let $\hat{\rho}$ be a totally bounded metric for E. By (**iii**) of Lemma 9.1.4, $U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$ is separable. Hence, we can find a countable set $\{\varphi_n : n \geq 1\}$ which is dense in $U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$. In particular, by the equivalence of (**i**) and (**ii**) above, we see that $\langle \varphi_n, \mu_\alpha \rangle \longrightarrow \langle \varphi_n, \mu \rangle$ for all $n \in \mathbb{Z}^+$ if and only if $\mu_\alpha \Longrightarrow \mu$; which is to say that the corresponding map $\mathbf{H} : \mathbf{M}_1(E) \longrightarrow [0,1]^{\mathbb{Z}^+}$ is a homeomorphism. Since $[0,1]^{\mathbb{Z}^+}$ is a compact metric space and D (cf. the proof of (**ii**) in Lemma 9.1.4) is a metric for it, we also see that the R described is a totally bounded metric for $\mathbf{M}_1(E)$. In particular, $\mathbf{M}_1(E)$ is separable. Finally, since, by (**ii**) in Lemma 9.1.4, it is always possible to find a totally bounded metric for E, the last assertion needs no further comment. $\square$

The reader would do well to pay close attention to what (**iii**) and (**iv**) say about the nature of weak convergence. Namely, even though $\mu_\alpha \Longrightarrow \mu$, it is possible that some or all of the mass which the μ_α's assign to the interior of a set may gravitate to the boundary in the limit. This phenomenon is most easily understood by taking $E = \mathbb{R}$, μ_α to be the unit point mass δ_α at $\alpha \in [0,1)$, checking that $\delta_\alpha \Longrightarrow \delta_1$, and noting that $\delta_1\big((0,1)\big) = 0 < 1 = \delta_\alpha\big((0,1)\big)$ for each $\alpha \in [0,1)$.

Remark 9.1.6. Those who find nets distasteful will be pleased to learn that, from now on, I will be restricting my attention to separable metric spaces E and therefore need only discuss sequential convergence when working with the weak topology on $\mathbf{M}_1(E)$. Furthermore, unless the contrary is explicitly stated, *I will always be thinking of the weak topology when working with* $\mathbf{M}_1(E)$.

Given a separable metric space E, I next want to find conditions which guarantee that a subset of $\mathbf{M}_1(E)$ is compact; and at this point it will be convenient to have introduced the notation $K \subset\subset E$ to indicate that K is a compact subset of E. The key to my analysis is the following extension of the sort of Riesz Representation result in Theorem 3.1.1 combined with a crucial observation made by S. Ulam.*

Lemma 9.1.7. *Let E be a separable metric space, ρ a metric for E, and Λ a non-negative linear functional on $U_{\mathrm{b}}^{\rho}(E;\mathbb{R})$ (i.e., Λ is a linear map which assigns any non-negative value to a non-negative $\varphi \in U_{\mathrm{b}}^{\rho}(E;\mathbb{R})$) with $\Lambda(\mathbf{1}) = 1$. Then in order for there to be a (necessarily unique) $\mu \in \mathbf{M}_1(E)$ satisfying $\Lambda(\varphi) = \langle \varphi, \mu \rangle$ for all $\varphi \in U_{\mathrm{b}}^{\rho}(E;\mathbb{R})$, it is sufficient that, for every $\epsilon > 0$, there exist a $K \subset\subset E$*

* It is no accident that Ulam was the first to make this observation. Indeed, the term *Polish space* was coined by Bourbaki in recognition of the contribution made to this subject by the Polish school in general and C. Kuratowski in particular (cf. Kuratowski's *Topologie, Vol. I*, Warszawa–Lwow, (1933)). Ulam had studied with Kuratowski.

such that

$$\big|\Lambda(\varphi)\big| \le \sup_{x\in K} |\varphi(x)| + \epsilon\|\varphi\|_{\mathrm{u}}, \quad \varphi \in U^\rho_{\mathrm{b}}(E;\mathbb{R}). \tag{9.1.8}$$

Conversely, if E is a Polish space and $\mu \in \mathbf{M}_1(E)$, then for every $\epsilon > 0$ there is a $K \subset\subset E$ such that $\mu(K) \ge 1-\epsilon$. In particular, if $\mu \in \mathbf{M}_1(E)$ and $\Lambda(\varphi) = \langle\varphi,\mu\rangle$ for $\varphi \in C_{\mathrm{b}}(E;\mathbb{R})$, then, for each $\epsilon > 0$, (9.1.8) holds for some $K \subset\subset E$.

PROOF: I begin with the trivial observation that, because Λ is non-negative and $\Lambda(\mathbf{1}) = 1$, $\big|\Lambda(\varphi)\big| \le \|\varphi\|_{\mathrm{u}}$. Next, according to the Daniell theory of integration, the first statement will be proved as soon as we know that $\Lambda(\varphi_n) \searrow 0$ whenever $\{\varphi_n : n \ge 1\}$ is a non-increasing sequence of functions from $U^\rho_{\mathrm{b}}\big(E;[0,\infty)\big)$ which tend pointwise to 0 as $n \to \infty$. To this end, let $\epsilon > 0$ be given, and choose $K \subset\subset E$ so that (9.1.8) holds. One then has that

$$\varlimsup_{n\to\infty} \big|\Lambda\big(\varphi_n\big)\big| \le \lim_{n\to\infty} \sup_{x\in K} |\varphi_n(x)| + \epsilon\|\varphi_1\|_{\mathrm{u}} = \epsilon\|\varphi_1\|_{\mathrm{u}},$$

since, by Dini's Lemma, $\varphi_n \searrow 0$ uniformly on compact subsets of E.

Turning to the second part, assume that E is Polish, and use $B(x,r)$ to denote the open ball of radius $r > 0$ around $x \in E$, computed with respect to a complete metric ρ for E. Next, let $\{p_k : k \ge 1\}$ be a countable dense subset of E, and set $B_{k,n} = B\left(p_k, \frac{1}{n}\right)$ for $k,\, n \in \mathbb{Z}^+$. Given $\mu \in \mathbf{M}_1(E)$ and $\epsilon > 0$, we can choose, for each $n \in \mathbb{Z}^+$, an $\ell_n \in \mathbb{Z}^+$ so that

$$\mu\left(\bigcup_{k=1}^{\ell_n} B_{k,n}\right) \ge 1 - \frac{\epsilon}{2^n}.$$

Hence, if

$$C_n \equiv \bigcup_{k=1}^{\ell_n} \overline{B}_{k,n} \quad \text{and} \quad K = \bigcap_{n=1}^{\infty} C_n,$$

then $\mu(K) \ge 1-\epsilon$. At the same time, it is obvious that, on the one hand, K is closed (and therefore ρ-complete) and that, on the other hand, $K \subseteq \bigcup_{k=1}^{\ell_n} B\left(p_k, \frac{2}{n}\right)$ for every $n \in \mathbb{Z}^+$. Hence, K is both complete and totally bounded with respect to ρ and, as such, is compact. $\square$

As Lemma 9.1.7 makes clear, probability measures on a Polish space like to be *nearly concentrated on a compact set*. Following Prohorov and Varadarajan,*

* See Yu. V. Prohorov's article "Convergence of random processes and limit theorems in probability theory," *Theory of Prob. & Appl.*, which appeared in 1956. Independently, V.S. Varadarajan developed essentially the same theory in "Weak convergence of measures on a separable metric spaces," *Sankhyă*, which was published in 1958. Although Prohorov got into print first, subsequent expositions, including this one, rely heavily on Varadarajan.

what we are about to see is that, for a Polish space E, relatively compact subsets of $\mathbf{M}_1(E)$ are those whose elements are *nearly concentrated on the same compact set of* E. More precisely, given a separable metric space E, say that $M \subseteq \mathbf{M}_1(E)$ is **tight** if, for every $\epsilon > 0$, there exists a $K \subset\subset E$ such that $\mu(K) \geq 1 - \epsilon$ for all $\mu \in M$.

THEOREM 9.1.9. *Let E be a separable metric space and $M \subseteq \mathbf{M}_1(E)$. Then $\overline{M}$ is compact if M is tight. Conversely, when E is Polish, M is tight if $\overline{M}$ is compact.*[*]

PROOF: Since it is clear, from (**iii**) in Theorem 9.1.5, that $\overline{M}$ is tight if and only if M is, I will assume throughout that M is closed in $\mathbf{M}_1(E)$.

To prove the first statement, take $\hat{\rho}$ to be a totally bounded metric on E, choose $\{\varphi_n : n \geq 1\} \subseteq U_{\mathrm{b}}^{\hat{\rho}}\big(E;[0,1]\big)$ accordingly, as in the last part of Theorem 9.1.5, and let $\varphi_0 = \mathbf{1}$. Given a sequence $\{\mu_\ell : \ell \geq 1\} \subseteq \mathbf{M}_1(E)$, we can use a standard diagonalization procedure to extract a subsequence $\big\{\mu_{\ell_k} : k \geq 1\big\}$ such that

$$\Lambda(\varphi_n) \equiv \lim_{k\to\infty} \langle \varphi_n, \mu_{\ell_k}\rangle$$

exists for each $n \in \mathbb{N}$. Since $\Lambda(\varphi) \equiv \lim_{k\to\infty}\langle \varphi, \mu_{\ell_k}\rangle$ continues to exist for every φ in the uniform closure of the span of $\{\varphi_n : n \geq 1\}$, we now see that Λ determines a non-negative linear functional on $U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$ and that $\Lambda(\mathbf{1}) = 1$. Moreover, because M is tight, we can find, for any $\epsilon > 0$, a $K \subset\subset E$ such that $\mu(K) \geq 1 - \epsilon$ for every $\mu \in M$; and therefore (9.1.8) holds with this choice of K. Hence, by Lemma 9.1.7, we know that there is a $\mu \in \mathbf{M}_1(E)$ for which $\Lambda(\varphi) = \langle \varphi, \mu\rangle$, $\varphi \in U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$. Because this means that $\langle \varphi, \mu_{\ell_k}\rangle \longrightarrow \langle \varphi, \mu\rangle$ for every $\varphi \in U_{\mathrm{b}}^{\hat{\rho}}(E;\mathbb{R})$, the equivalence of (**i**) and (**ii**) in Theorem 9.1.5 allows us to conclude that $\mu_{\ell_k} \Longrightarrow \mu$.

Finally, suppose that E is Polish and that M is compact in $\mathbf{M}_1(E)$. To see that M must be tight, repeat the argument used to prove the second part of Lemma 9.1.7. Thus, choose $B_{k,n}$, k, $n \in \mathbb{Z}^+$ as in the proof there, and set

$$f_{\ell,n}(\mu) = \mu\left(\bigcup_{k=1}^{\ell} B_{k,n}\right) \quad \text{for } \ell,\, n \in \mathbb{Z}^+.$$

By (**iv**) in Theorem 9.1.5, $\mu \in \mathbf{M}_1(E) \longmapsto f_{\ell,n}(\mu) \in [0,1]$ is lower semicontinuous. Moreover, for each $n \in \mathbb{Z}^+$, $f_{\ell,n} \nearrow 1$ as $\ell \nearrow \infty$. Thus, by Dini's Lemma, we can choose, for each $n \in \mathbb{Z}^+$, one $\ell_n \in \mathbb{Z}^+$ so that $f_{\ell_n,n}(\mu) \geq 1 - \frac{\epsilon}{2^n}$ for all

[*] For the reader who wishes to investigate just how far these results can be pushed before they start of break down, a good place to start is Appendix III in P. Billingsley's *Convergence of Probability Measures*, publ. by J. Wiley, (1968) . In particular, although it is reasonably clear that completeness is more or less essential for the necessity, the havoc which results from dropping separability may come as a surprise.

$\mu \in M$; and at this point the rest of the argument is precisely the same as the one given at the end of the proof of Lemma 9.1.7. $\square$

§ **9.1.3. The Lévy Metric and Completeness of $\mathbf{M}_1(E)$.** We have now seen that $\mathbf{M}_1(E)$ inherits properties from E. To be more specific, if E is a metric space, then $\mathbf{M}_1(E)$ is separable or compact if E itself is. What I want to show next is that completeness also gets transferred. That is, I will show that $\mathbf{M}_1(E)$ is Polish if E is. In order to do this, I will need a lemma which is of considerable importance in its own right.

LEMMA 9.1.10. *Let E be a Polish space and Φ a bounded subset of $C_b(E;\mathbb{R})$ which is equicontinuous at each $x \in E$. (That is, for each $x \in E$, $\sup_{\varphi\in\Phi} |\varphi(y) - \varphi(x)| = 0$ as $y \to x$.) If $\{\mu_n : n \geq 1\} \cup \{\mu\} \subseteq \mathbf{M}_1(E)$ and $\mu_n \Longrightarrow \mu$, then*

$$\lim_{n\to\infty} \sup_{\varphi\in\Phi} \Big|\langle\varphi,\mu_n\rangle - \langle\varphi,\mu\rangle\Big| = 0.$$

PROOF: Let $\epsilon > 0$ be given, and use the second part of Theorem 9.1.9 to choose $K \subset\subset E$ so that

$$\left(\sup_{\varphi\in\Phi} \|\varphi\|_{\mathrm{u}}\right)\left(\sup_{n\in\mathbb{Z}^+} \mu_n(K\complement)\right) < \frac{\epsilon}{4}.$$

By (**iv**) of Theorem 9.1.5, $\mu\big(K\complement\big)$ satisfies the same estimate. Next, choose a metric ρ for E and a countable dense set $\{p_k : k \geq 1\}$ in K. Using equicontinuity together with compactness, find $\ell \in \mathbb{Z}^+$ and $\delta_1,\dots,\delta_\ell > 0$ so that $K \subseteq \big\{x : \rho(x,p_k) < \delta_k \text{ for some } 1 \leq k \leq \ell\big\}$ and

$$\sup_{\varphi\in\Phi} \big|\varphi(x) - \varphi(p_k)\big| < \frac{\epsilon}{4} \quad \text{for } 1 \leq k \leq \ell \text{ and } x \in K \text{ with } \rho(x,p_k) < 2\delta_k.$$

Because $r \in (0,\infty) \longmapsto \mu\big(\{y \in K : \rho(y,x) \leq r\}\big) \in [0,1]$ is non-decreasing for each $x \in K$, we can find, for each $1 \leq k \leq \ell$, an $r_k \in \big(\delta_k, 2\delta_k\big)$ such that $\mu(\partial B_k) = 0$ when $B_k \equiv \big\{x \in K : \rho\big(x,p_k\big) < r_k\big\}$. Finally, set $A_1 = B_1$ and $A_{k+1} = B_{k+1} \setminus \bigcup_{j=1}^k B_j$ for $1 \leq k < \ell$. Then, $K \subseteq \bigcup_{k=1}^{\ell} A_k$, the A_k's are disjoint, and, for each $1 \leq k \leq \ell$,

$$\sup_{\varphi\in\Phi} \sup_{x\in A_k} \big|\varphi(x) - \varphi\big(p_k\big)\big| < \frac{\epsilon}{4} \quad \text{and} \quad \mu\big(\partial A_k\big) = 0.$$

Hence, by (**vii**) in Theorem 9.1.5 applied to the $\mathbf{1}_{A_k}$'s:

$$\varlimsup_{n\to\infty} \sup_{\varphi\in\Phi} \Big|\langle\varphi,\mu_n\rangle - \langle\varphi,\mu\rangle\Big| < \epsilon + \varlimsup_{n\to\infty} \sum_{k=1}^{\ell} \sup_{\varphi\in\Phi} \big|\varphi\big(p_k\big)\big|\,\big|\mu_n\big(A_k\big) - \mu\big(A_k\big)\big| = \epsilon. \quad \square$$

THEOREM 9.1.11. *Let E be a Polish space and ρ a complete metric for E. Given $(\mu,\nu)\in\mathbf{M}_1(E)^2$, define*

$$L(\mu,\nu)=\inf\Big\{\delta:\mu(F)\le\nu\big(F^{(\delta)}\big)+\delta \\ \text{and } \nu(F)\le\mu\big(F^{(\delta)}\big)+\delta \text{ for all closed } F\subseteq E\Big\},$$

where $F^{(\delta)}$ denotes the set of $x\in E$ which lie a ρ-distance less than δ from F. Then L is a complete metric for $\mathbf{M}_1(E)$, and therefore $\mathbf{M}_1(E)$ is Polish.

PROOF: It is clear that L is symmetric and that it satisfies the triangle inequality. Thus, we will know that it is a metric for $\mathbf{M}_1(E)$ as soon as we show that $L\big(\mu_n,\mu\big)\longrightarrow 0$ if and only if $\mu_n\Longrightarrow\mu$. To this end, first suppose that $L\big(\mu_n,\mu\big)\longrightarrow 0$. Then, for every closed F, $\mu\big(F^{(\delta)}\big)+\delta\ge\overline{\lim}_{n\to\infty}\mu_n(F)$ for all $\delta>0$; and therefore, by countable additivity, $\mu(F)\ge\overline{\lim}_{n\to\infty}\mu_n(F)$ for every closed F. Hence, by the equivalence of (**i**) and (**iii**) in Theorem 9.1.5, $\mu_n\Longrightarrow\mu$. Now suppose that $\mu_n\Longrightarrow\mu$, and let $\delta>0$ be given. Given a closed F in E, define

$$\psi_F(x)=\frac{\rho\big(x,F^{(\delta)}\complement\big)}{\rho\big(x,F^{(\delta)}\complement\big)+\rho(x,F)}\quad\text{for}\quad x\in E.$$

It is then an easy matter to check that both

$$\mathbf{1}_F\le\psi_F\le\mathbf{1}_{F^{(\delta)}}\quad\text{and}\quad\big|\psi_F(x)-\psi_F(y)\big|\le\frac{\rho(x,y)}{\delta}.$$

In particular, by Lemma 9.1.10, we can choose $m\in\mathbb{Z}^+$ so that

$$\sup_{n\ge m}\,\sup\Big\{\Big|\langle\psi_F,\mu_n\rangle-\langle\psi_F,\mu\rangle\Big|:\,F\text{ closed in }E\Big\}<\delta;$$

from which it is an easy matter to see that, for all $n\ge m$,

$$\mu(F)\le\mu_n\big(F^{(\delta)}\big)+\delta\quad\text{and}\quad\mu_n(F)\le\mu\big(F^{(\delta)}\big)+\delta.$$

In other words, $\sup_{n\ge m}L\big(\mu_n,\mu\big)\le\delta$; and, since $\delta>0$ was arbitrary, we have shown that $L\big(\mu_n,\mu\big)\longrightarrow 0$.

In order to finish the proof, I must show that if $\{\mu_n:\,n\ge 1\}\subseteq\mathbf{M}_1(E)$ is L-Cauchy convergent, then it is tight. Thus, let $\epsilon>0$ be given, and choose, for each $\ell\in\mathbb{Z}^+$, an $m_\ell\in\mathbb{Z}^+$ and a $K_\ell\subset\subset E$ so that

$$\sup_{n\ge m_\ell}L\big(\mu_n,\mu_{m_\ell}\big)\le\frac{\epsilon}{2^{\ell+1}}\quad\text{and}\quad\max_{1\le n\le m_\ell}\mu_n\big(K_\ell\complement\big)\le\frac{\epsilon}{2^{\ell+1}}.$$

Setting $\epsilon_\ell = \frac{\epsilon}{2^\ell}$, one then has that:

$$\sup_{n\in\mathbb{Z}^+} \mu_n\big(K_\ell^{(\epsilon_\ell)}\complement\big) \le \epsilon_\ell \quad \text{for each} \quad \ell \in \mathbb{Z}^+.$$

In particular, if

$$K \equiv \bigcap_{\ell=1}^{\infty} \overline{K_\ell^{(\epsilon_\ell)}},$$

then $\mu_n(K) \ge 1 - \epsilon$ for all $n \in \mathbb{Z}^+$. Finally, because each K_ℓ is compact, it is easy to see that K is both ρ-complete and totally bounded and therefore also compact. $\square$

When $E = \mathbb{R}$, P. Lévy was the first one to construct a complete metric on $\mathbf{M}_1(E)$, and it is for this reason that I will call the metric L described in Theorem 9.1.11 the **Lévy metric** determined by ρ. Using an abstract argument, Varadarajan showed that $\mathbf{M}_1(E)$ must be Polish whenever E is, and the explicit construction which I have used is essentially the one first produced by Prohorov.

Before closing this subsection, it seems appropriate to introduce and explain some of the more classical terminology connected with applications of weak convergence to probability theory. For this purpose, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and E a metric space. Given E-valued random variables $\{X_n : n \ge 1\} \cup \{X\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$, one says that the sequence X_n **tends to** X **in law** (or **distribution**) and writes $X_n \xrightarrow{\mathcal{L}} X$ if (cf. Exercise 1.1.16) $(X_n)_*\mathbb{P} \Longrightarrow X_*\mathbb{P}$. The idea here is that, when the measures under consideration are the distributions of random variables, one wants to think of weak convergence of the distributions as determining a kind of convergence of the corresponding random variables. Thus, one can add *convergence in law* to the list of possible ways in which random variables might converge. In order to elucidate the relationship between convergence in law, $\mathbb{P}$-almost sure convergence, and convergence in $\mathbb{P}$-measure, it will be convenient to have the following lemma.

LEMMA 9.1.12. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and E a metric space. Given any E-valued random variables $\{X_n : n \ge 1\} \cup \{X\}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ and any pair of topologically equivalent metrics ρ and σ for E, $\rho\big(X_n, X\big) \longrightarrow 0$ in $\mathbb{P}$-measure if and only if $\sigma\big(X_n, X\big) \longrightarrow 0$ in $\mathbb{P}$-measure. In particular, convergence in $\mathbb{P}$-measure does not depend on the choice of metric, and so one can write $X_n \longrightarrow X$ in $\mathbb{P}$-measure without specifying a metric. Moreover, if $X_n \longrightarrow X$ in $\mathbb{P}$-measure, then $X_n \xrightarrow{\mathcal{L}} X$. In fact, if E is a Polish space and L is the Lévy metric on $\mathbf{M}_1(E)$ associated with a complete metric ρ for E, then*

$$L\big(X_*\mathbb{P}, Y_*\mathbb{P}\big) \le\le \delta \vee \mathbb{P}\big(\rho(X, Y) \ge \delta\big)$$

and all $\delta > 0$ and E-valued random variables X and Y.

PROOF: To prove the first assertion, suppose that

$$\rho(X_n, X) \longrightarrow 0 \text{ in } \mathbb{P}\text{-measure but that } \sigma(X_n, X) \not\longrightarrow 0 \text{ in } \mathbb{P}\text{-measure.}$$

After passing to a subsequence if necessary, we could then arrange that $\rho(X_n, X) \longrightarrow 0$ (a.s., $\mathbb{P}$) but $\mathbb{P}\big(\sigma(X_n, X) \geq \epsilon\big) \geq \epsilon$ for all $n \in \mathbb{Z}^+$ and some $\epsilon > 0$. But this is impossible, since then we would have that $\sigma(X_n, X) \longrightarrow 0$ $\mathbb{P}$-almost surely but not in $\mathbb{P}$-measure. Hence, we now know that convergence in $\mathbb{P}$-measure does not depend on the choice of metric. To complete the first part, suppose that $\rho(X_n, X) \longrightarrow 0$ in $\mathbb{P}$-measure. Then, for every $\varphi \in U^\rho_{\mathrm{b}}(E; \mathbb{R})$ and $\delta > 0$,

$$\begin{aligned}\varlimsup_{n\to\infty} \big|\mathbb{E}^{\mathbb{P}}\big[\varphi(X_n)\big] - \mathbb{E}^{\mathbb{P}}\big[\varphi(X)\big]\big| &\leq \varlimsup_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\big|\varphi(X_n) - \varphi(X)\big|\big] \\ &\leq \epsilon(\delta) + \|\varphi\|_{\mathrm{u}} \varlimsup_{n\to\infty} \mathbb{P}\Big(\rho\big(X_n, X\big) \geq \delta\Big) = \epsilon(\delta),\end{aligned}$$

where

$$\epsilon(\delta) \equiv \sup\big\{|\varphi(y) - \varphi(x)| : \rho(x, y) \leq \delta\big\} \longrightarrow 0 \quad \text{as} \quad \delta \searrow 0.$$

Thus, by (**ii**) in Theorem 9.1.5, $(X_n)_*\mathbb{P} \Longrightarrow X_*\mathbb{P}$.

Now assume that E is Polish, and take ρ and L accordingly. Then, for any closed set F and $\delta > 0$,

$$\begin{aligned}X_*\mathbb{P}(F) = \mathbb{P}(X \in F) &\leq \mathbb{P}\big(\rho(Y, F) < \delta\big) + \mathbb{P}\big(\rho(X, Y) \geq \delta\big) \\ &= Y_*\big(F^{(\delta)}\big) + \mathbb{P}\big(\rho(X, Y) \geq \delta\big).\end{aligned}$$

Hence, since the same is true when the roles of X and Y are reversed, the asserted estimate on $L\big(X_*\mathbb{P}, Y_*\mathbb{P}\big)$ holds. □

As a demonstration of the sort of use to which one can put these ideas, I present the following version of the **Principle of Accompanying Laws**.

THEOREM 9.1.13. *Let E be a Polish space and, for each $k \in \mathbb{Z}^+$, let $\{Y_{k,n} : n \geq 1\}$ be a sequence of E-valued random variables on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Further, assume that, for each $k \in \mathbb{Z}^+$, there is a $\mu_k \in \mathbf{M}_1(E)$ such that $Y^*_{k,n}\mathbb{P} \Longrightarrow \mu_k$ as $n \to \infty$. Finally, let ρ be a complete metric for E, and suppose that $\{X_n : n \geq 1\}$ is a sequence of E-valued random variables on $(\Omega, \mathcal{F}, \mathbb{P})$ with the property that*

$$\lim_{k\to\infty} \varlimsup_{n\to\infty} \mathbb{P}\Big(\rho\big(X_n, Y_{k,n}\big) \geq \delta\Big) = 0 \quad \textit{for every } \delta > 0. \tag{9.1.14}$$

Then there is a $\mu \in \mathbf{M}_1(E)$ such that $\mu_k \Longrightarrow \mu$ as $k \to \infty$ and $(X_n)_*\mathbb{P} \Longrightarrow \mu$ as $n \to \infty$. In particular, if, as $n \to \infty$, $Y_n \xrightarrow{\mathcal{L}} X$ and $\mathbb{P}\big(\rho(X_n, Y_n) \geq \delta\big) \longrightarrow 0$ for each $\delta > 0$, then $X_n \xrightarrow{\mathcal{L}} X$.

PROOF: Let L be the Lévy metric associated with a complete metric ρ for E. By the second part of Lemma 9.1.12,

$$\sup_{\ell \geq k} L\big((Y_{\ell,n})_*\mathbb{P}, (X_n)_*\mathbb{P}\big) \leq \delta \vee \left(\sup_{\ell \geq k} \overline{\lim_{n\to\infty}}\, \mathbb{P}\big(\rho(Y_{\ell,n}, X_n) \geq \delta\big)\right),$$

and therefore, by (9.1.14),

$$(*) \qquad \overline{\lim_{k\to\infty}}\, \overline{\lim_{n\to\infty}}\, L\big((Y_{\ell,n})_*\mathbb{P}, (X_n)_*\mathbb{P}\big) = 0.$$

Thus, since for any $k \in \mathbb{Z}^+$,

$$\sup_{\ell \geq k} L\big(\mu_\ell, \mu_k\big) = \sup_{\ell \geq k} \lim_{n\to\infty} L\big((Y_{\ell,n})_*\mathbb{P}, (Y_{k,n})_*\mathbb{P}\big),$$

$\{\mu_k : k \geq 1\}$ is an L-Cauchy sequence and, as such, converges to some μ. Finally, for every $k \in \mathbb{Z}^+$,

$$L\big(\mu, (X_n)_*\mathbb{P}\big) \leq L(\mu, \mu_k) + L\big(\mu_k, (Y_{k,n})_*\big) + L\big((Y_{k,n})_*\mathbb{P}, (X_n)_*\mathbb{P}\big),$$

and so

$$\overline{\lim_{n\to\infty}}\, L\big(\mu, (X_n)_*\mathbb{P}\big) \leq L(\mu, \mu_k) + \overline{\lim_{n\to\infty}}\, L\big((Y_{k,n})_*\mathbb{P}, (X_n)_*\mathbb{P}\big).$$

Thus, after letting $k \to \infty$ and applying (*), one concludes that $(X_n)_*\mathbb{P} \implies \mu$. $\square$

Exercises for § 9.1

EXERCISE 9.1.15. Let $(E, \mathcal{B})$ be a measurable space with the property that $\{x\} \in \mathcal{B}$ for all $x \in E$. In this exercise, we will investigate the strong topology in a little more detail. In particular, in part (**iv**) below, we will show that when $\mu \in \mathbf{M}_1(E)$ is **non-atomic** (i.e., $\mu\big(\{x\}\big) = 0$ for every $x \in E$), then there is no countable neighborhood basis of μ in the strong topology. Obviously, this means that *the strong topology for $\mathbf{M}_1(E)$ admits no metric whenever $\mathbf{M}_1(E)$ contains a non-atomic element.*

(**i**) Show that, in general,

$$\|\nu - \mu\|_{\text{var}} = 2\max\big\{\nu(A) - \mu(A) : A \in \mathcal{B}\big\}$$

and that in the case when E is a metric space, $\mathcal{B}$ its Borel field, and ρ a metric for E,

$$\|\nu - \mu\|_{\text{var}} = \sup\big\{\langle\varphi, \nu\rangle - \langle\varphi, \mu\rangle : \varphi \in U^\rho_{\text{b}}(E;\mathbb{R}) \text{ and } \|\varphi\|_{\text{u}} \leq 1\big\}.$$

(**ii**) Show that if $\{\mu_n : n \geq 1\}$ is a sequence in $\mathbf{M}_1(E)$ which tends in the strong topology to $\mu \in \mathbf{M}_1(E)$, then $\mu \ll \sum_{n=1}^\infty 2^{-n}\mu_n$.

(**iii**) Given $\mu \in \mathbf{M}_1(E)$, show that μ admits a countable neighborhood basis in the strong topology if and only if there exists a countable $\{\varphi_k : k \geq 1\} \subseteq B(E;\mathbb{R})$ such that, for any net $\{\mu_\alpha : \alpha \in A\} \subseteq \mathbf{M}_1(E)$, $\mu_\alpha \longrightarrow \mu$ in the strong topology as soon as $\lim_\alpha \langle \varphi_k, \mu_\alpha \rangle = \langle \varphi_k, \mu \rangle$ for every $k \in \mathbb{Z}^+$.

(**iv**) Referring to Exercises 1.1.14 and 1.1.16, set $\Omega = E^{\mathbb{Z}^+}$ and $\mathcal{F} = \mathcal{B}^{\mathbb{Z}^+}$. Next, let $\mu \in \mathbf{M}_1(E)$ be given, and define $\mathbb{P} = \mu^{\mathbb{Z}^+}$ on $(\Omega, \mathcal{F})$. Show that, for any $\varphi \in B(E;\mathbb{R})$, the random variables $\mathbf{x} \in \Omega \longmapsto X_n^\varphi(\mathbf{x}) \equiv \varphi(x_n)$, $n \in \mathbb{Z}^+$, are mutually $\mathbb{P}$-independent and all have distribution $\varphi_*\mu$. In particular, use the Strong Law of Large Numbers to conclude that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{m=1}^{n} X_m^\varphi(\mathbf{x}) = \langle \varphi, \mu \rangle$$

for each $\mathbf{x}$ outside of a $\mathbb{P}$-null set.

Now assume that μ is non-atomic, and suppose that μ admitted a countable neighborhood basis in the strong topology. Choose $\{\varphi_k : k \geq 1\} \subseteq B(E;\mathbb{R})$ accordingly, as in (**iii**), and (using the preceding) conclude that there exists at least one $\mathbf{x} \in \Omega$ for which the measures μ_n given by $\mu_n \equiv \frac{1}{n}\sum_{m=1}^n \delta_{x_m}$, $n \in \mathbb{Z}^+$, converge in the strong topology to μ. Finally, apply (**ii**) to see that this is impossible.

Exercise 9.1.16. Throughout this exercise, E is a separable metric space.

(**i**) We already know that $\mathbf{M}_1(E)$ is separable; however our proof was non-constructive. Show that if $\{p_k : k \geq 1\}$ is a dense subset of E, then the set of all convex combinations $\sum_{k=1}^n \alpha_k \delta_{p_k}$, where $n \in \mathbb{Z}^+$ and $\{\alpha_k : 1 \leq k \leq n\} \subset [0,1] \cap \mathbb{Q}$ with $\sum_1^n \alpha_k = 1$, is a countable dense set in $\mathbf{M}_1(E)$.

(**ii**) We have seen that $\mathbf{M}_1(E)$ is compact if E is. To see that the converse is also true, show that $x \in E \longmapsto \delta_x \in \mathbf{M}_1(E)$ is a homeomorphism whose image is closed.

(**iii**) Although it is a little off our track, it is amusing to show that E being compact is equivalent to $C_b(E;\mathbb{R})$ being separable; and, in view of (**i**) in Lemma 9.1.4, this comes down to checking that E is compact if $C_b(E;\mathbb{R})$ is separable.

Hint: Let $\hat{\rho}$ be a totally bounded metric on E, and use $\hat{E}$ to denote the $\hat{\rho}$-completion of E. Show that if $\{x_n : n \geq 1\} \subseteq E$ has the properties that $x_n \longrightarrow \hat{x} \in \hat{E}$ and $\lim_{n\to\infty} \varphi(x_n)$ exists for every $\varphi \in C_b(E;\mathbb{R})$, then $\hat{x} \in E$. (Suppose not, set $\psi(x) = \frac{1}{\hat{\rho}(x,\hat{x})}$, and consider functions of the form $f \circ \psi$ for $f \in C_b(\mathbb{R};\mathbb{R})$.) Finally, assuming that $C_b(E;\mathbb{R})$ is separable, and using a diagonalization procedure, show that every sequence $\{x_n : n \geq 1\} \subseteq E$ admits a subsequence $\{x_{n_m} : m \geq 1\}$ which converges to some $\hat{x} \in \hat{E}$ and $\lim_{m\to\infty} \varphi(x_{n_m})$ exists for every $\varphi \in C_b(E;\mathbb{R})$.

(**iv**) Let $\{M_n : n \geq 1\}$ be a sequence of finite, non-negative measures on $(E, \mathcal{B})$. Assuming that $\{M_n : n \geq 1\}$ is **tight** in the sense that $\{M_n(E) : n \geq 1\}$ is bounded and that, for each $\epsilon > 0$, there is a $K \subset\subset E$ such that $\sup_n M_n\big(K\complement\big) \leq \epsilon$, show that there is a subsequence $\{M_{n_k} : k \geq 1\}$ and a finite measure M such that

$$\int_E \varphi \, dM = \lim_{k\to\infty} \int_E \varphi \, dM_{n_k}, \quad \text{for all } \varphi \in C_b(E;\mathbb{R}).$$

Conversely, if E is Polish and there is a finite measure M such that $\int_E \varphi \, dM_n \longrightarrow \int_E \varphi \, dM$ for every $\varphi \in C_b(E;\mathbb{R})$, show that $\{M_n : n \geq 1\}$ is tight.

EXERCISE 9.1.17. Let $\{E_\ell : \ell \geq 1\}$ be a sequence of Polish spaces, set $\mathbf{E} = \prod_1^\infty E_\ell$, and give $\mathbf{E}$ the product topology.

(**i**) For each $\ell \in \mathbb{Z}^+$, let ρ_ℓ be a complete metric for E_ℓ, and define

$$\mathbf{R}(\mathbf{x}, \mathbf{y}) = \sum_{\ell=1}^\infty \frac{1}{2^\ell} \frac{\rho_\ell(x_\ell, y_\ell)}{1 + \rho_\ell(x_\ell, y_\ell)} \quad \text{for} \quad \mathbf{x},\, \mathbf{y} \in \mathbf{E}.$$

Show that $\mathbf{R}$ is a complete metric for $\mathbf{E}$, and conclude that $\mathbf{E}$ is a Polish space. In addition, check that $\mathcal{B}_{\mathbf{E}} = \prod_1^\infty \mathcal{B}_{E_\ell}$.

(**ii**) For $\ell \in \mathbb{Z}^+$, let π_ℓ be the natural projection map from $\mathbf{E}$ onto E_ℓ, and show that $\mathbf{K} \subset\subset \mathbf{E}$ if and only if

$$\mathbf{K} = \bigcap_{\ell \in \mathbb{Z}^+} \pi_\ell^{-1}(K_\ell) \quad \text{where} \quad K_\ell \subset\subset E_\ell \text{ for each } \ell \in \mathbb{Z}^+.$$

Also, show that the span of the functions

$$\prod_{k=1}^\ell \varphi_k \circ \pi_k \quad \text{where } \ell \in \mathbb{Z}^+ \text{ and } \varphi_k \in U_b^{\rho_k}(E_k;\mathbb{R}),\ 1 \leq k \leq \ell,$$

is dense in $U_b^{\mathbf{R}}(\mathbf{E};\mathbb{R})$. In particular, conclude from these that $\mathbf{A} \subseteq \mathbf{M}_1(\mathbf{E})$ is tight if and only if $\big\{(\pi_\ell)_*\mu : \mu \in \mathbf{A}\big\} \subseteq \mathbf{M}_1(E_\ell)$ is tight for every $\ell \in \mathbb{Z}^+$ and that $\mu_n \Longrightarrow \mu$ in $\mathbf{M}_1(\mathbf{E})$ if and only if

$$\left\langle \prod_{k=1}^\ell \varphi_k \circ \pi_k, \mu_n \right\rangle \longrightarrow \left\langle \prod_{k=1}^\ell \varphi_k \circ \pi_k, \mu \right\rangle$$

for every $\ell \in \mathbb{Z}^+$ and choice of $\varphi_k \in U_b^{\rho_k}(E_k;\mathbb{R}),\ 1 \leq k \leq \ell$.

(**iii**) For each $\ell \in \mathbb{Z}^+$, set $\mathbf{E}_\ell = \prod_{k=1}^\ell E_k$, and let $\boldsymbol{\pi}_\ell$ denote the natural projection map from $\mathbf{E}$ onto $\mathbf{E}_\ell$. Next, let $\mu_{[1,\ell]}$ be an element of $\mathbf{M}_1\big(\mathbf{E}_\ell\big)$, and assume that the $\mu_{[1,\ell]}$'s are **consistent** in the sense that, for every $\ell \in \mathbb{Z}^+$,

$$\mu_{[1,\ell+1]}\big(\Gamma \times E_{\ell+1}\big) = \mu_{[1,\ell]}(\Gamma) \quad \text{for all } \Gamma \in \mathcal{B}_{\mathbf{E}_\ell}.$$

Show that there is a unique $\mu \in \mathbf{M}_1(\mathbf{E})$ such that $\mu_{[1,\ell]} = (\boldsymbol{\pi}_\ell)_*\mu$ for every $\ell \in \mathbb{Z}^+$.

Hint: Choose and fix an $\mathbf{e} \in \mathbf{E}$, and define $\Phi_\ell : \mathbf{E}_\ell \longrightarrow \mathbf{E}$ so that

$$\Big(\Phi_\ell\big(x_1, \dots, x_\ell\big)\Big)_n = \begin{cases} x_n & \text{if} \quad n \le \ell \\ e_n & \text{otherwise.} \end{cases}$$

Show that $\big\{(\Phi_\ell)_*\mu_{[1,\ell]} : \ell \in \mathbb{Z}^+\big\} \in \mathbf{M}_1(\mathbf{E})$ is tight and that any limit must be the desired μ.

The conclusion drawn in (**iii**) is the renowned **Kolmogorov Extension** (or **Consistency**) **Theorem**. Notice that, at least for Polish spaces, it represents a vast generalization of the result obtained in Exercise 1.1.14.

EXERCISE 9.1.18. In this exercise we will use the theory of weak convergence to develop variations on The Strong Law of Large Numbers (cf. Theorem 1.4.9). Thus, let E be a Polish space, $(\Omega, \mathcal{F}, P)$ a probability space, and $\{X_n : n \ge 1\}$ a sequence of mutually independent E-valued random variables on $(\Omega, \mathcal{F}, P)$ with common distribution $\mu \in \mathbf{M}_1(E)$. Next, define the *empirical distribution functional*

$$\omega \in \Omega \longmapsto \mathbf{L}_n(\omega) \equiv \frac{1}{n} \sum_{m=1}^{n} \delta_{X_m(\omega)} \in \mathbf{M}_1(E);$$

and observe that, for any $\varphi \in B(E;\mathbb{R})$,

$$\big\langle \varphi, \mathbf{L}_n(\omega) \big\rangle = \frac{1}{n} \sum_{m=1}^{n} \varphi\big(X_m(\omega)\big), \quad n \in \mathbb{Z}^+ \text{ and } \omega \in \Omega.$$

As a consequence of The Strong Law, show that

$$\mathbf{L}_n(\omega) \Longrightarrow \mu \quad \text{for} \quad P\text{-almost every } \omega \in \Omega, \tag{9.1.19}$$

which is the **Strong Law of Large Numbers for the empirical distribtution**.

Next show that (9.1.19) provides another (cf. Exercises 6.1.16 and 6.2.19) proof of the Strong Law of Large Numbers for Banach space valued random variables. Thus, let E be a real, separable, Banach space with dual space E^*, and set $\overline{S}_n(\omega) = \frac{1}{n} \sum_1^n X_m(\omega)$ for $n \in \mathbb{Z}^+$ and $\omega \in \Omega$.

(**i**) As a preliminary step, begin with the case when

$$\mu\Big(B_E(0, R)\complement\Big) = 0 \quad \text{for some} \quad R \in (0, \infty). \tag{*}$$

Choose $\eta \in C_b\big(\mathbb{R};\mathbb{R}\big)$ so that $\eta(t) = t$ for $t \in [-R, R]$ and $\eta(t) = 0$ when $|t| \ge R + 1$, and define $\psi_{x^*} \in C_b(E;\mathbb{R})$ for $x^* \in E^*$ by $\psi_{x^*}(x) = \eta\big(\langle x, x^* \rangle\big)$, $x \in E$, where $\langle x, x^* \rangle$ is used here to denote the action of $x^* \in E^*$ on $x \in E$. Taking (*) into account and applying (9.1.19) and Lemma 9.1.10, show that

$$\lim_{n\to\infty} \sup_{\|x^*\|_{E^*}\le 1} \left| \langle \psi_{x^*}, \mathbf{L}_n(\omega) \rangle - \int_E \langle x, x^* \rangle \, \mu(dx) \right| = 0$$

for $\mathbb{P}$-almost every $\omega \in \Omega$; and conclude from this that

$$\lim_{n\to\infty} \left\| \overline{S}_n(\omega) - \mathbf{m} \right\|_E = 0 \quad \text{for } \mathbb{P}\text{-almost every } \omega \in \Omega,$$

where (cf. Lemma 5.1.10) $\mathbf{m} = \mathbb{E}^\mu[x]$.

(**ii**) The next step is to replace the boundedness assumption in (*) by the hypothesis

$$\int_E \|x\|_E \, \mu(dx) < \infty.$$

Assuming that this holds, define, for $R \in (0, \infty)$, $n \in \mathbb{Z}^+$, and $\omega \in \Omega$:

$$X_n^{(R)}(\omega) = \begin{cases} X_n(\omega) & \text{if } \left\| X_n(\omega) \right\|_E < R \\ 0 & \text{otherwise} \end{cases}$$

and $Y_n^{(R)}(\omega) = X_n(\omega) - X_n^{(R)}(\omega)$. Next, set $\overline{S}_n^{(R)} = \frac{1}{n} \sum_1^n X_m^{(R)}$, $n \in \mathbb{Z}^+$; and, from (**i**), note that $\left\{ \overline{S}_n^{(R)}(\omega) : n \ge 1 \right\}$ converges in E for $\mathbb{P}$-almost every $\omega \in \Omega$. In particular, if $\epsilon > 0$ is given and $R \in (0, \infty)$ is chosen so that

$$\int\limits_{\{\|x\|_E \ge R\}} \|x\|_E \, \mu(dx) < \frac{\epsilon}{8},$$

use the preceding and Theorem 1.4.9 to verify the computation

$$\begin{aligned}
&\varlimsup_{m\to\infty} P\left(\sup_{n\ge m} \left\| \overline{S}_n - \overline{S}_m \right\|_E \ge \epsilon \right) \\
&\quad \le \lim_{m\to\infty} P\left(\sup_{n\ge m} \left\| \overline{S}_n^{(R)} - \overline{S}_m^{(R)} \right\| \ge \frac{\epsilon}{2} \right) \\
&\qquad\qquad + 2 \varlimsup_{m\to\infty} P\left(\sup_{n\ge m} \left\| \frac{1}{n} \sum_1^n Y_k^{(R)} \right\|_E \ge \frac{\epsilon}{4} \right) \\
&\quad \le 2 \varlimsup_{m\to\infty} P\left(\sup_{n\ge m} \frac{1}{n} \sum_1^n \left\| Y_k^{(R)} \right\|_E \ge \frac{\epsilon}{4} \right) = 0;
\end{aligned}$$

and from this, conclude that $\overline{S}_n \longrightarrow \mathbb{E}^\mu[x]$ $\mathbb{P}$-almost surely.

(**iii**) Finally, repeat the argument given in the proof of Theorem 1.4.9 to show that $\|\mathbf{x}\|$ is μ-integrable if $\{\overline{S}_n : n \geq 1\}$ converges in E on a set of positive $\mathbb{P}$-measure.††

§ 9.2 Regular Conditional Probability Distributions

As I mentioned in the discussion following Theorem 5.1.4, there are quite general situations in which conditional expectation values can be computed as expectation values. The following is a basic result in that direction.

THEOREM 9.2.1. *Suppose that Ω is a Polish space and that $\mathcal{F} = \mathcal{B}_\Omega$. Then, for every sub-$\sigma$-algebra Σ of $\mathcal{F}$, there is a $\mathbb{P}$-almost surely unique Σ-measurable map $\omega \in \Omega \longmapsto \mathbb{P}^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ with the property that*

$$\mathbb{P}\big(A \cap B\big) = \int_A \mathbb{P}^\Sigma_\omega(B)\,\mathbb{P}(d\omega) \quad \text{for all } A \in \Sigma \text{ and } B \in \mathcal{F}.$$

In particular, for each $(-\infty, \infty]$-valued random variable X which is bounded below, $\omega \in \Omega \longmapsto \mathbb{E}^{\mathbb{P}^\Sigma_\omega}[X]$ is a conditional expectation value of X given Σ. Finally, if Σ is countably generated, then there is a $\mathbb{P}$-null set $\mathcal{N} \in \Sigma$ with the property that $\mathbb{P}^\Sigma_\omega(A) = \mathbf{1}_A(\omega)$ for all $\omega \notin \mathcal{N}$ and $A \in \Sigma$.

PROOF: To prove the uniqueness, suppose $\omega \in \Omega \longmapsto \mathbb{Q}^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ were a second such mapping. We would then know that, for each $B \in \mathcal{F}$, $\mathbb{Q}^\Sigma_\omega(B) = \mathbb{P}^\Sigma_\omega(B)$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Hence, since $\mathcal{F}$ (as the Borel field over a second countable topological space) is countably generated, we could find one Σ-measurable $\mathbb{P}$-null set off of which $\mathbb{Q}^\Sigma_\omega = \mathbb{P}^\Sigma_\omega$. Similarly, to prove the final assertion when Σ is countably generated, note (cf. (5.1.7)) that, for each $A \in \Sigma$, $\mathbb{P}^\Sigma_\omega(A) = \mathbf{1}_A(\omega) = \delta_\omega(A)$ for $\mathbb{P}$-almost every $\omega \in \Omega$. Thus, once again countability allows us to choose one Σ-measurable $\mathbb{P}$-null set $\mathcal{N}$ such that $\mathbb{P}^\Sigma_\omega \restriction \Sigma = \delta_\omega \restriction \Sigma$ if $\omega \notin \mathcal{N}$.

I turn next to the question of existence. For this purpose, first choose (cf. (**ii**) of Lemma 9.1.4) ρ to be a totally bounded metric for Ω, and let $\mathcal{U} = U^\rho_{\mathrm{b}}(\Omega;\mathbb{R})$ be the space of bounded, ρ-uniformly continuous, $\mathbb{R}$-valued functions on Ω. Then (cf. (**iii**) of Lemma 9.1.4) $\mathcal{U}$ is a separable Banach space with respect to the uniform norm. In particular, we can choose a sequence $\{f_n : n \geq 0\} \subseteq \mathcal{U}$ so that $f_0 = \mathbf{1}$, the functions $f_0, \dots, f_n$ are linearly independent for each $n \in \mathbb{Z}^+$, and the linear span $\mathcal{S}$ of $\{f_n : n \geq 0\}$ is dense in $\mathcal{U}$. Set $g_0 = \mathbf{1}$, and, for each $n \in \mathbb{Z}^+$, let g_n be some fixed representative of $\mathbb{E}^{\mathbb{P}}[f_n \,|\, \Sigma]$. Next, set

$$\mathfrak{R} = \Big\{\boldsymbol{\alpha} \in \mathbb{R}^{\mathbb{N}} : \exists m \in \mathbb{N}\ \alpha_n = 0 \text{ for all } n \geq m\Big\}$$

†† The beautiful argument which I have just given is due to Ranga Rao. See his 1963 article "The law of large numbers for $D[0,1]$-valued random variables," *Theory of Prob. & Appl.* **VIII, 1**, where he shows that this method applies even outside the separable context.

and define

$$f_{\boldsymbol{\alpha}} = \sum_{n=0}^{\infty} \alpha_n \, f_n \quad \text{and} \quad g_{\boldsymbol{\alpha}} = \sum_{n=0}^{\infty} \alpha_n \, g_n$$

for $\boldsymbol{\alpha} \in \mathfrak{R}$. Because of the linear independence of the f_n's, we know that $f_{\boldsymbol{\alpha}} = f_{\boldsymbol{\beta}}$ if and only if $\boldsymbol{\alpha} = \boldsymbol{\beta}$. Hence, for each $\omega \in \Omega$, we can define the (not necessarily continuous) linear functional $\Lambda_\omega : \mathcal{S} \longrightarrow \mathbb{R}$ so that

$$\Lambda_\omega\bigl(f_{\boldsymbol{\alpha}}\bigr) = g_{\boldsymbol{\alpha}}(\omega), \qquad \boldsymbol{\alpha} \in \mathfrak{R}.$$

Clearly, $\Lambda_\omega(\mathbf{1}) = \mathbf{1}$ for all $\omega \in \Omega$. On the other hand, we cannot say that Λ_ω is always non-negative as a linear functional on $\mathcal{S}$. In fact, the best we can do is extract a Σ-measurable $\mathbb{P}$-null set $\mathcal{N}$ so that Λ_ω is a non-negative linear functional on $\mathcal{S}$ whenever $\omega \notin \mathcal{N}$. To this end, let $\mathbb{Q}$ denote the rational reals and set

$$\mathfrak{Q}^+ = \Bigl\{\boldsymbol{\alpha} \in \mathfrak{R} \cap \mathbb{Q}^{\mathbb{N}} : \, f_{\boldsymbol{\alpha}} \geq 0\Bigr\}.$$

Since $g_{\boldsymbol{\alpha}} \geq 0$ (a.s., $\mathbb{P}$) for every $\boldsymbol{\alpha} \in \mathfrak{Q}^+$ and $\mathfrak{Q}^+$ is countable,

$$\mathcal{N} \equiv \Bigl\{\omega \in \Omega : \, \exists \boldsymbol{\alpha} \in \mathfrak{Q}^+ \quad g_{\boldsymbol{\alpha}}(\omega) < 0\Bigr\}$$

is a Σ-measurable, $\mathbb{P}$-null set. In addition, it is obvious that, for every $\omega \notin \mathcal{N}$, $\Lambda_\omega(f) \geq 0$ whenever f is a non-negative element of $\mathcal{S}$. In particular, for $\omega \notin \mathcal{N}$,

$$\|f\|_{\mathrm{u}} \pm \Lambda_\omega(f) = \Lambda_\omega\bigl(\|f\|_{\mathrm{u}} \, \mathbf{1} \pm f\bigr) \geq 0, \quad f \in \mathcal{S},$$

and therefore Λ_ω admits a unique extension as a non-negative, continuous linear functional on $\mathcal{U}$ which takes $\mathbf{1}$ to 1. Furthermore, it is an easy matter to check that, for every $f \in \mathcal{U}$, the function

$$g(\omega) = \begin{cases} \Lambda_\omega(f) & \text{for} \quad \omega \notin \mathcal{N} \\ \mathbb{E}^{\mathbb{P}}[f] & \text{for} \quad \omega \in \mathcal{N} \end{cases}$$

is a conditional expectation value of f given Σ.

At this point, all that remains is to show that, for $\mathbb{P}$-almost every $\omega \notin \mathcal{N}$, Λ_ω is given by integration with respect to a $\mathbb{P}_\omega \in \mathbf{M}_1(\Omega)$. In particular, by the Riesz Representation Theorem, there is nothing more to do in the case when Ω is compact. To treat the case when Ω is not compact, I will use Lemma 9.1.7. Namely, first choose (cf. the last part of Lemma 9.1.7) a non-decreasing sequence of sets $K_n \subset\subset \Omega$, $n \in \mathbb{Z}^+$, with the property that $\mathbb{P}\bigl(K_n\complement\bigr) \leq \frac{1}{2^n}$. Next, define

$$\eta_{m,n}(\omega) = \frac{m \, \rho(\omega, K_n)}{1 + m \, \rho(\omega, K_n)} \quad \text{for } m, \, n \in \mathbb{Z}^+.$$

Clearly, $\eta_{m,n} \in \mathcal{U}$ for each pair (m,n) and $0 \le \eta_{m,n} \nearrow \mathbf{1}_{K_n\complement}$ as $m \to \infty$ for each $n \in \mathbb{Z}^+$. Thus, by the Monotone Convergence Theorem, for each $n \in \mathbb{Z}^+$,

$$\begin{aligned}\int_{\mathcal{N}\complement} \sup_{m\in\mathbb{Z}^+} \Lambda_\omega\big(\eta_{m,n}\big)\,\mathbb{P}(d\omega) &= \lim_{m\to\infty} \int_{\mathcal{N}\complement} \Lambda_\omega\big(\eta_{m,n}\big)\,\mathbb{P}(d\omega) \\ &= \lim_{m\to\infty} \mathbb{E}^{\mathbb{P}}\big[\eta_{m,n}\big] \le \frac{1}{2^n};\end{aligned}$$

and so, by the Borel–Cantelli Lemma, we can find a Σ-measurable $\mathbb{P}$-null set $\mathcal{N}' \supseteq \mathcal{N}$ such that

$$M(\omega) \equiv \sup_{n\in\mathbb{Z}^+} n\left(\sup_{m\in\mathbb{Z}^+} \Lambda_\omega\big(\eta_{m,n}\big)\right) < \infty \quad \text{for every } \omega \notin \mathcal{N}'.$$

Hence, if $\omega \notin \mathcal{N}'$, then, for every $f \in \mathcal{U}$ and $n \in \mathbb{Z}^+$,

$$\begin{aligned}\big|\Lambda_\omega(f)\big| &\le \big|\Lambda_\omega\big((1-\eta_{m,n})\,f\big)\big| + \big|\Lambda_\omega\big(\eta_{m,n}\,f\big)\big| \\ &\le \big\|(1-\eta_{m,n})\,f\big\|_{\mathrm{u}} + \frac{M(\omega)}{n}\,\|f\|_{\mathrm{u}}\end{aligned}$$

for all $m \in \mathbb{Z}^+$. But $\big\|(1-\eta_{m,n})\,f\big\|_{\mathrm{u}} \longrightarrow \|f\|_{\mathrm{u},K_n}$ as $m \to \infty$, and so we now see that the condition in (9.1.8) is satisfied by Λ_ω for every $\omega \notin \mathcal{N}'$. In other words, we have shown that, for each $\omega \notin \mathcal{N}'$, there is a unique $\mathbb{P}^{\Sigma}_{\omega} \in \mathbf{M}_1(\Omega)$ such that $\Lambda_\omega(f) = \mathbb{E}^{\mathbb{P}^{\Sigma}_{\omega}}[f]$ for all $f \in \mathcal{U}$. Finally, if we complete the definition of the map $\omega \in \Omega \longmapsto \mathbb{P}^{\Sigma}_{\omega}$ by taking $\mathbb{P}^{\Sigma}_{\omega} = \mathbb{P}$ for $\omega \in \mathcal{N}'$, then this map is Σ-measurable and

$$\mathbb{E}^{\mathbb{P}}\big[f,\,A\big] = \int_{\Omega} \mathbb{E}^{\mathbb{P}^{\Sigma}_{\omega}}[f]\,\mathbb{P}(d\omega), \quad A \in \Sigma,$$

first for all $f \in \mathcal{U}$ and thence for all $\mathcal{F}$-measurable f's which are bounded below. □

If $\mathbb{P}$ is a probability measure on $(\Omega,\mathcal{F})$ and Σ is a sub σ-algebra of $\mathcal{F}$, then a **conditional probability distribution of $\mathbb{P}$ given Σ** is a map $(\omega,B) \longmapsto \mathbb{P}^{\Sigma}_{\omega}(B)$ such that $\mathbb{P}^{\Sigma}_{\omega}$ is a probability measure on $(\Omega,\mathcal{F})$ for each $\omega \in \Omega$ and $\omega \rightsquigarrow \mathbb{P}^{\Sigma}_{\omega}(B)$ a conditional probability of B given Σ for all $B \in \mathcal{F}$. If, in addition, for ω outside a Σ-measurable, $\mathbb{P}$-null set and all $A \in \Sigma$, $\mathbb{P}^{\Sigma}(A) = \mathbf{1}_A(\omega)$, then the conditional probability distribution is said to be **regular**. Notice that, although they may not always exist, conditional probability distributions are always unique up to a Σ-measurable $\mathbb{P}$-null set so long as $\mathcal{F}$ is countably generated. Moreover, Theorem 9.2.1 says that they will always exist if Ω is Polish and $\mathcal{F} = \mathcal{B}_\Omega$. Finally, whenever a conditional probability distribution of $\mathbb{P}$ given Σ exists, the argument leading to the last part of Theorem 9.2.1 when Σ is countably generated is completely general and shows that a regular version can be found.

§ **9.2.1. Fibering a Measure.** When Ω is a product space $E_1 \times E_2$ of two Polish spaces and Σ is the σ-algebra generated by the second coordinate, then the conclusion of Theorem 9.2.1 takes a particularly pleasing form.

THEOREM 9.2.2. *Let E_1 and E_2 be a pair of Polish spaces, and take Ω to be the Polish space $E_1 \times E_2$. Given $\mu \in \mathbf{M}_1(\Omega)$, use μ_2 to denote the marginal distribution of μ on E_2: $\mu_2(\Gamma) = \mu(E_1 \times \Gamma)$ for $\Gamma \in \mathcal{B}_{E_2}$. Then there is a measurable map $x_2 \in E_2 \longmapsto \mu(x_2, \cdot\,) \in \mathbf{M}_1(E_1)$ such that $\mu(dx_1 \times dx_2) = \mu(x_2, dx_1)\,\mu_2(dx_2)$.*

PROOF: Referring to Theorem 9.2.1, take $\mathbb{P} = \mu$, $\Sigma = \{E_1 \times \Gamma : \Gamma \in \mathcal{B}_{E_2}\}$, and let $\omega \in \Omega \longmapsto \mathbb{P}^\Sigma_\omega \in \mathbf{M}_1(\Omega)$ be the map guaranteed by the result there. Next, choose and fix a point $x_1^0 \in E_1$. Then, because $\omega \rightsquigarrow \mathbb{P}^\Sigma_\omega$ is Σ-measurable, we know that $\mathbb{P}^\Sigma_{(x_1,x_2)} = \mathbb{P}^\Sigma_{(x_1^0,x_2)}$. In addition, because Σ is countably generated, the final part of Theorem 9.2.1 guarantees that there exists a μ_2-null set $B \in \mathcal{B}_{E_2}$ such that $\mathbb{P}^\Sigma_{(x_1^0,x_2)}\big(E_1 \times \{x_2\}\big) = 1$ for all $x_2 \notin B$. Hence, if we define $x_2 \rightsquigarrow \mu(x_2, \cdot\,)$ by $\mu(x_2, \Gamma) = \mathbb{P}^\Sigma_{(x_1^0,x_2)}(\Gamma \times E_2)$, then, for any Borel measurable $\varphi : E_1 \times E_2 \longrightarrow [0,\infty)$, $\langle \varphi, \mu \rangle$ equals

$$\int \left(\int \varphi(\omega')\mathbb{P}^\Sigma_\omega(d\omega') \right) \mathbb{P}(d\omega) = \int_{E_2} \left(\int_{E_1} \varphi(x_1, x_2)\,\mu(x_2, dx_1) \right) \mu_2(dx_2). \quad \square$$

In the older literature, the result in Theorem 9.2.2 would be called a **fibering** of μ. The name derives from the idea that μ on $E_1 \times E_2$ can be decomposed into its "vertical component" μ_2 and its "restrictions" $\mu(x_2, \cdot\,)$ to "horizontal fibers" $E_1 \times \{x_2\}$. Alternatively, Theorem 9.2.2 can be interpreted as saying that any $\mu \in \mathbf{M}_1(E_1 \times E_2)$ can be decomposed into its marginal distribution on E_2 and a transition probability $x_2 \in E_2 \longmapsto \mu(x_2, \cdot\,) \in \mathbf{M}_1(E_1)$. The two extreme cases are when the coordinates are independent, in which case $\mu(x_2, \cdot\,)$ is independent of x_2, and the case when the coordinates are equal, in which case $\mu(x_2, \cdot\,) = \delta_{x_2}$.

As an application of Theorem 9.2.2, I present the following important special case of a more general result which indicates just how remarkably fungible non-atomic measures are.

COROLLARY 9.2.3. *Let $\lambda_{[0,1)}$ denote Lebesgue measure on $[0,1)$. For each $N \in \mathbb{Z}^+$ and $\mu \in \mathbf{M}_1(\mathbb{R}^N)$, there is a Borel measurable map $f : [0,1) \longrightarrow \mathbb{R}^N$ such that $\mu = f_*\lambda_{[0,1)}$.*

PROOF: I will work by induction on $N \in \mathbb{Z}^+$. When $N = 1$, take

$$f(u) = \inf\big\{t \in \mathbb{R} : \mu\big((-\infty, t]\big) \geq u\big\}, \quad u \in [0,1).$$

Next, assume the result is true for N, take $E_1 = \mathbb{R}$ and $E_2 = \mathbb{R}^N$ in Theorem 9.2.2, and, given $\mu \in \mathbf{M}_1(\mathbb{R}^N)$, define $\mu_2 \in \mathbf{M}_1(\mathbb{R}^N)$ and $\mathbf{y} \in \mathbb{R}^N \longmapsto \mu(\mathbf{y}, \cdot\,) \in \mathbf{M}_1(\mathbb{R})$ accordingly. By the induction hypothesis, $\mu_2 = f_2(\,\cdot\,)_*\lambda_{[0,1)}$ for some $f_2 : [0,1) \longrightarrow \mathbb{R}^N$. Thus, if $g : [0,1)^2 \longrightarrow \mathbb{R} \times \mathbb{R}^N$ is given by

$$g(u_1, u_2) = \Big(\inf\big\{t \in \mathbb{R} : \mu\big(f_2(u_2), (-\infty, t]\big) \geq u_1\big\}, f_2(u_2)\Big)$$

for $(u_1, u_2) \in [0,1)^2$, then g is Borel measurable on $[0,1)^2$ and $\mu = g_*\lambda^2_{[0,1)}$. Finally, by Lemma 1.1.6 or part (**ii**) of Exercise 1.1.11, we know that there is a Borel measurable map $u \in [0,1) \longmapsto \mathbf{U}(u) = \big(U_1(u), U_2(u)\big) \in [0,1)^2$ such that $\mathbf{U}_*\lambda_{[0,1)} = \lambda^2_{[0,1)}$, and so we can take $f(u) = g \circ \mathbf{U}$. $\square$

§ **9.2.2. Representing Lévy Measures via the Itô Map.** There is another way of thinking about the construction of the Poisson jump processes, one which is based on Corollary 9.2.3 and the transformation property described in Lemma 4.2.12. The advantage of this approach is that it provides a method of coupling Lévy processes corresponding to different Lévy measures. Indeed, it is this coupling procedure which underlies K. Itô's construction of Markov processes modeled on Lévy processes.*

Let $M_0(d\mathbf{y}) = |\mathbf{y}|^{-N-1}\,d\mathbf{y}$, which is the Lévy measure for a (cf. Corollary 3.3.9) symmetric 1-stable law. My first goal is to show that every $M \in \mathfrak{M}_\infty(\mathbb{R}^N)$ can be realized as (cf. the notation in Lemma 4.2.6) M_0^F for some Borel measurable $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ satisfying $F(\mathbf{0}) = \mathbf{0}$.*

Theorem 9.2.4. *For each $M \in \mathfrak{M}_\infty(\mathbb{R}^N)$ there exists a Borel measurable map $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ such that $F(\mathbf{0}) = \mathbf{0}$ and*

$$M(\Gamma) = M_0^F \equiv M_0\big(F^{-1}(\Gamma \setminus \{\mathbf{0}\})\big), \quad \Gamma \in \mathcal{B}_{\mathbb{R}^N}.$$

Proof: I begin with the case when $N = 1$. Given $M \in \mathfrak{M}_\infty(\mathbb{R})$, define $\rho(r, \pm 1)$ for $r > 0$ by

$$\begin{aligned} \rho(r,1) &= \sup\big\{\rho \in [0,\infty) :\ M\big([\rho,\infty)\big) \ge r^{-1}\big\} \\ \rho(r,-1) &= \sup\big\{\rho \in [0,\infty) :\ M\big((-\infty,-\rho]\big) \ge r^{-1}\big\}, \end{aligned}$$

where I have taken the supremum over the empty set to be 0. Applying Exercise 9.2.6 with $\nu(dr) = r^{-2}\lambda_{(0,\infty)}(dr)$, one sees that $M = M_0^F$ when $F(0) = 0$ and $F(y) = \rho\big(|y|, \frac{y}{|y|}\big)$ for $y \in \mathbb{R} \setminus \{0\}$.

Now assume that $N \ge 2$, and let $M \in \mathfrak{M}_\infty(\mathbb{R}^N)$. If $M = 0$, simply take $F \equiv 0$. If $M \neq 0$, choose a non-decreasing function $h : (0,\infty) \longrightarrow (0,\infty)$ so that

$$\int h\big(|\mathbf{y}|\big)\, M(d\mathbf{y}) = 1,$$

and define $\mu \in \mathbf{M}_1\big((0,\infty) \times \mathbb{S}^{N-1}\big)$ so that

$$\langle \varphi, \mu \rangle = \int_{\mathbb{R}^N} h\big(|\mathbf{y}|\big)\varphi(\mathbf{y}) M(d\mathbf{y}).$$

---

* See K. Itô's *On stochastic differential equations*, Memoirs of the A.M.S. #4 (1951) or my *Markov Processes from K. Itô's Perspective*, Annals of Math. Studies #155 (2003).

* There is nothing sacrosanct about the choice of M_0 as our reference measure. For instance, it should obvious that one can choose any Lévy measure M with the property that $M_0 = M^F$ for some Borel measurable $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ which takes $\mathbf{0}$ to $\mathbf{0}$.

Using μ_2 to denote the marginal distribution of μ on $\mathbb{S}^{N-1}$, apply Corollary 9.2.3 to find a Borel measurable $\mathbf{f} : [0,1) \longrightarrow \mathbb{R}^N$ so that $\mu_2 = \mathbf{f}_*\lambda_{[0,1)}$. Since μ_2 lives on $\mathbb{S}^{N-1}$, I may and will assume that $\mathbf{f}(u) \in \mathbb{S}^{N-1}$ for all $u \in [0,1)$. Next, use Theorem 9.2.2 to find a measurable map $\boldsymbol{\eta} \in \mathbb{S}^{N-1} \longmapsto \mu(\boldsymbol{\eta}, \cdot\,) \in \mathbf{M}_1\big((0,\infty)\big)$ so that $\mu(dr \times d\boldsymbol{\eta}) = \mu(\boldsymbol{\eta}, dr)\,\mu_2(d\boldsymbol{\eta})$, and define $\rho : (0,\infty) \times \mathbb{S}^{N-1} \longrightarrow [0,\infty)$ by

$$\rho(r,\boldsymbol{\eta}) = \sup\left\{\rho \in [0,\infty) : \int_{[\rho,\infty)} \frac{1}{h(r)}\,\mu(\boldsymbol{\eta}, dr) \geq \frac{\omega_{N-1}}{r}\right\}.$$

Then, again by Exercise 9.2.6, but this time with $\nu(dr) = \omega_{N-1}r^{-2}\lambda_{(0,\infty)}(dr)$, for any continuous $\varphi : \mathbb{R}^N \longrightarrow [0,\infty)$ which vanishes in a neighborhood of $\mathbf{0}$,

$$\int_{(0,\infty)} \frac{\varphi(r\boldsymbol{\eta})}{h(r)}\,\mu(\boldsymbol{\eta}, dr) = \omega_{N-1}\int_{(0,\infty)} \varphi\big(\rho(r,\boldsymbol{\eta})\boldsymbol{\eta}\big)r^{-2}\,dr, \quad \boldsymbol{\eta} \in \mathbb{S}^{N-1},$$

and so

$$\begin{aligned}\int_{\mathbb{R}^N} \varphi(\mathbf{y})\,M(dy) &= \omega_{N-1}\int_{\mathbb{S}^{N-1}}\left(\int_{(0,\infty)} \varphi\big(\rho(r,\boldsymbol{\eta})\boldsymbol{\eta}\big)r^{-2}\,dr\right)\mu_2(d\boldsymbol{\eta})\\ &= \omega_{N-1}\int_{[0,1)}\left(\int_{(0,\infty)} \varphi\big(\rho(r,\boldsymbol{\eta})\mathbf{f}(t)\big)r^{-2}\,dr\right)\lambda_{[0,1)}(dt).\end{aligned}$$

Finally, define $g : \mathbb{S}^{N-1} \longrightarrow [0,\omega_{N-1})$ by $g(\boldsymbol{\eta}) = \lambda_{\mathbb{S}^{N-1}}\big(\{\boldsymbol{\eta}' \in \mathbb{S}^{N-1} : \eta_1' \leq \eta_1\}\big)$, note that $\omega_{N-1}\lambda_{[0,1)} = g_*\lambda_{\mathbb{S}^{N-1}}$, and conclude that $M = M_0^F$ when

$$F(\mathbf{0}) = 0 \quad \text{and} \quad F(\mathbf{y}) = \rho\big(|\mathbf{y}|, \tfrac{\mathbf{y}}{|\mathbf{y}|}\big)\mathbf{f} \circ g\big(\tfrac{\mathbf{y}}{|\mathbf{y}|}\big) \text{ for } \mathbf{y} \in \mathbb{R}^N \setminus \{\mathbf{0}\}. \quad \square$$

We can now prove the following theorem, which is the simplest example of Itô's procedure.

Theorem 9.2.5. *Let $\{j_0(t,\cdot\,) : t \geq 0\}$ be a Poisson jump process associated with M_0. Then for each $M \in \mathfrak{M}_\infty(\mathbb{R}^N)$, there is a Borel measurable map $F : \mathbb{R}^N \longrightarrow \mathbb{R}^N$ with $F(\mathbf{0}) = \mathbf{0}$ and a Poisson jump process $\{j(t,\cdot\,) : t \geq 0\}$ associated with M such that $j(t,\cdot\,) = j_0^F(t,\cdot\,)$, $t \geq 0$, $\mathbb{P}$-almost surely.*

Proof: Choose F as in Theorem 9.2.4 so that $M = M_0^F$. For $R > 0$, set $F_R(\mathbf{y}) = \mathbf{1}_{[R,\infty)}(\mathbf{y})F(\mathbf{y})$. By Lemma 4.2.12, we know that $\{j_0^{F_R}(t,\cdot\,) : t \geq 0\}$ is a Poisson jump process associated with M^{F_R}. In particular, for each $r > 0$,

$$\mathbb{E}^{\mathbb{P}}\big[j_0^F\big(t, \mathbb{R}^N \setminus B(\mathbf{0},r)\big)\big] = \lim_{R \searrow 0}\mathbb{E}^{\mathbb{P}}\big[j_0^{F_R}\big(t, \mathbb{R}^N \setminus B(\mathbf{0},r)\big)\big] = M\big(\mathbb{R}^N \setminus B(0,r)\big) < \infty.$$

Hence, there exists a $\mathbb{P}$-null set $\mathcal{N}$ such that $t \rightsquigarrow j_0^F(t,\cdot\,,\omega)$ is a jump function for all $\omega \notin \mathcal{N}$. Finally, if $j(t,\cdot\,,\omega) = j_0^F(t,\cdot\,,\omega)$ when $\omega \notin \mathcal{N}$ and $j(t,\cdot\,,\omega) = 0$ for $\omega \in \mathcal{N}$, then $\{j(t,\cdot\,) : t \geq 0\}$ is a jump process associated with M and $j(t,\cdot\,) = j_0^F(t,\cdot\,)$, $t \geq 0$, for $\mathbb{P}$-almost every $\omega \in \Omega$. $\square$

Exercises for § 9.2

EXERCISE 9.2.6. Let ν be an infinite non-negative, non-atomic, Borel measure on $[0,\infty)$ with the property that $\nu\big([r_2,\infty)\big) < \nu\big([r_1,\infty)\big) < \infty$ for all $0 < r_1 < r_2 < \infty$. Given any other non-negative, Borel measure on $[0,\infty)$ with the properties that $\mu(\{0\}) = 0$ and $\mu\big([r,\infty)\big) < \infty$ for all $r > 0$, define

$$\rho(r) = \sup\big\{\rho \in (0,\infty) : \mu\big([\rho,\infty)\big) \geq \nu\big([r,\infty)\big)\big\}, \quad r \geq 0,$$

where the supremum over the empty set is taken to be 0. Show that $\mu\big([t,\infty)\big) = \nu\big(\{r : \rho(r) \geq t\}\big)$ for all $t > 0$, and therefore that $\langle \varphi, \mu\rangle = \langle \varphi \circ \rho, \nu\rangle$ for all Borel measurable $\varphi : [0,\infty) \longrightarrow [0,\infty)$ which vanish at 0.

Hint: Determine $g : (0,\infty) \longrightarrow (0,\infty)$ so that $\nu\big(\big[g(r),\infty\big)\big) = r$, and check that $\{r : \rho(r) \geq t\} = \big[g\big(\mu([t,\infty))\big),\infty\big)$ for all $t > 0$.

§9.3 Donsker's Invariance Principle

The content of this section is my main justification for presenting the material in § 9.1. Namely, as we saw in Chapter VIII, there is good reason to think that Wiener measure is the infinite dimensional version of the standard Gauss measure in $\mathbb{R}^N$, and as such one might suspect that there is a version of the Central Limit Theorem which applies to it. In this section I will prove such a Central Limit Theorem for Wiener measure. The result is due to M. Donsker and is known as **Donsker's Invariance Principle** (cf. Theorem 9.3.1).

Before getting started, I need to make a couple of simple preparatory remarks. In the first place, I will be thinking of Wiener measure $\mathcal{W}^{(N)}$ as a Borel probability measure on $C(\mathbb{R}^N) = C\big([0,\infty);\mathbb{R}^N\big)$ with the topology of uniform convergence on compact intervals. Equivalently, $C(\mathbb{R}^N)$ is given the topology for which

$$\rho(\boldsymbol{\psi},\boldsymbol{\psi}') = \sum_{n=1}^{\infty} \frac{1}{2^n} \frac{\|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_{[0,n]}}{1+\|\boldsymbol{\psi}-\boldsymbol{\psi}'\|_{[0,n]}}$$

is a metric, which, just as in the case of $D(\mathbb{R}^N)$ (cf. 4.1.1), is complete on $C(\mathbb{R}^N)$ and, as distinguished from $D(\mathbb{R}^N)$, is separable there. One way to check separability is to note that the set of paths $\boldsymbol{\psi}$ which, for some $n \in \mathbb{N}$, are linear on $[(m-1)2^{-n}, m2^{-n}]$ and satisfy $\boldsymbol{\psi}(m2^{-n}) \in \mathbb{Q}^N$ for all $m \in \mathbb{Z}^+$ is a countable, dense subset. In particular, this means that $C(\mathbb{R}^N)$ is a Polish space, and so the theory developed in § 9.1 applies to it. In addition, the Borel field $\mathcal{B}_{C(\mathbb{R}^N)}$ coincides with $\sigma\big(\{\boldsymbol{\psi}(t) : t \geq 0\}\big)$, the σ-algebra which $C(\mathbb{R}^N)$ inherits as a subset of $(\mathbb{R}^N)^{[0,\infty)}$ (cf. § 4.1). Indeed, since $\boldsymbol{\psi} \rightsquigarrow \boldsymbol{\psi}(t)$ is continuous for every $t \geq 0$, it is obvious that $\sigma\big(\{\boldsymbol{\psi}(t) : t \geq 0\}\big) \subseteq \mathcal{B}_{C(\mathbb{R}^N)}$. At the same time, since $\|\boldsymbol{\psi}\|_{[0,t]} = \sup\{|\boldsymbol{\psi}(\tau) : \tau \in [0,t] \cap \mathbb{Q}\}$, it is easy to check that open balls are $\sigma\big(\{\boldsymbol{\psi}(t) : t \geq 0\}\big)$-measurable. Hence, since every open set is the countable union of open balls, $\mathcal{B}_{C(\mathbb{R}^N)}$ is contained in $\sigma\big(\{\boldsymbol{\psi}(t) : t \geq$

$0\}\big)$. Knowing that these σ-algebras coincide, we know that two probability measures μ, $\nu \in \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ are equal if they determine the same distribution on $(\mathbb{R}^N)^{[0,\infty)}$. That is, if, for each $n \in \mathbb{Z}^+$ and $0 = t_0 < t_1 < t_n$, the distribution of $\boldsymbol{\psi} \in C(\mathbb{R}^N) \longmapsto \big(\boldsymbol{\psi}(t_0, \dots, \boldsymbol{\psi}(t_n)\big) \in (\mathbb{R}^N)^n$ is the same under μ and ν.

§**9.3.1. Donsker's Theorem.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and uppose that $\{\mathbf{X}_n : n \geq 1\}$ is a sequence of independent, $\mathbb{P}$-uniformly square integrable random variables (i.e., as $R \to \infty$, $\mathbb{E}^{\mathbb{P}}\big[|\mathbf{X}_n|^2,\, |\mathbf{X}_n| \geq R\big] \longrightarrow 0$ uniformly in n) with mean value $\mathbf{0}$ and covariance $\mathbf{I}$. Given $n \geq 1$, define $\omega \in \Omega \longmapsto \mathbf{S}_n(\,\cdot\,, \omega) \in C(\mathbb{R}^N)$ so that $\mathbf{S}_n(0) = \mathbf{0}$, $\mathbf{S}_n\big(\frac{m}{n}\big) = n^{-\frac{1}{2}} \sum_{k=1}^m \mathbf{X}_k$, and $\mathbf{S}_n(\,\cdot\,, \omega)$ is linear on each interval $\big[\frac{m-1}{n}, \frac{m}{n}\big]$ for all $m \in \mathbb{Z}^+$. Donsker's theorem is the following.

Theorem 9.3.1 (**Donsker's Invariance Principle**). *If $\mu_n = (\mathbf{S}_n)_*\mathbb{P} \in \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ is the distribution of $\omega \in \Omega \longmapsto \mathbf{S}_n(\,\cdot\,, \omega) \in C(\mathbb{R}^N)$ under $\mathbb{P}$, then $\mu_n \Longrightarrow \mathcal{W}^{(N)}$. Equivalently, for any bounded, continuous $\Phi : C(\mathbb{R}^N) \longrightarrow \mathbb{C}$,*

$$\lim_{n\to\infty} \mathbb{E}^{\mathbb{P}}\big[\Phi \circ \mathbf{S}_n\big] = \langle \Phi, \mathcal{W}^{(N)} \rangle.$$

Proving this result comes down to showing that $\{\mu_n : n \geq 1\}$ is tight and that every limit point is $\mathcal{W}^{(N)}$. The second of these is a rather elementary application of the Central Limit Theorem, and, at least when the $\mathbf{X}_n$'s have uniformly bounded fourth moments, the first is an application of Kolmogorov's Continuity Criterion. Finally, to remove the fourth moment assumption, I will use the Principle of Accompanying Laws. It should be noticed that, at no point in the proof, do I make use of the *a priori* existence of Wiener measure. Thus, Theorem 9.3.1 provides another derivation of its existence, a derivation which includes an an extremely ubiquitous approximation procedure.

Lemma 9.3.2. *Any limit point of $\{\mu_n : n \geq 1\}$ is $\mathcal{W}^{(N)}$.*

Proof: Since a probability on $C(\mathbb{R}^N)$ is uniquely determined by its finite dimensional time marginals, and because $\boldsymbol{\psi}(0) = \mathbf{0}$ with probability 1 under all the μ_n's as well as $\mathcal{W}^{(N)}$, it suffices to show that, for each $\ell \in \mathbb{Z}^+$ and $0 = t_0 < t_1 < \cdots < t_\ell$,

$$\big(\mathbf{S}_n(t_1), \mathbf{S}_n(t_2) - \mathbf{S}_n(t_1), \dots, \mathbf{S}_n(t_\ell) - \mathbf{S}_n(t_{\ell-1})\big)_*\mathbb{P} \Longrightarrow \gamma_{\mathbf{0},\tau_1\mathbf{I}} \times \cdots \times \gamma_{\mathbf{0},\tau_\ell\mathbf{I}},$$

where $\tau_k = t_k - t_{k-1}$, $1 \leq k \leq \ell$. To this end, for $1 \leq k \leq \ell$ and $n > \frac{1}{\tau_k}$, set

$$\boldsymbol{\Delta}_n(k) = n^{-\frac{1}{2}} \sum_{j=[nt_{k-1}]+1}^{[nt_k]} \mathbf{X}_j,$$

where, as usual, I use the notation $[t]$ to denote the integer part of t. Noting that

$$\begin{aligned}
&\Big|\mathbf{S}_n(t_k) - \mathbf{S}_n(t_{k-1}) - \boldsymbol{\Delta}_n(k)\Big| \\
&\quad \le \left|\mathbf{S}_n(t_k) - \mathbf{S}_n\left(\frac{[nt_k]}{n}\right)\right| + \left|\mathbf{S}_n(t_{k-1}) - \mathbf{S}_n\left(\frac{[nt_{k-1}]}{n}\right)\right| \\
&\quad \le \frac{\big|\mathbf{X}_{[nt_k]+1}\big| + \big|\mathbf{X}_{[nt_{k-1}]+1}\big|}{n^{\frac{1}{2}}},
\end{aligned}$$

one sees that, for any $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\sum_{k=1}^{\ell}\Big|\mathbf{S}_n(t_k) - \mathbf{S}_n(t_{k-1}) - \boldsymbol{\Delta}_n(k)\Big|^2 \ge \epsilon^2\right) \le \mathbb{P}\left(\sum_{k=0}^{\ell}\Big|\mathbf{X}_{[nt_k]+1}\Big|^2 \ge \frac{n\epsilon^2}{4}\right) \\
&\le \frac{4}{n\epsilon^2}\sum_{k=0}^{\ell}\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{X}_{[nt_k]+1}\big|^2\Big] = \frac{4(\ell+1)N}{n\epsilon^2} \longrightarrow 0
\end{aligned}$$

as $n \to \infty$. Hence, by the Principle of Accompanying Laws (cf. Theorem 9.1.13), we need only check that

$$\big(\boldsymbol{\Delta}_n(1), \dots, \boldsymbol{\Delta}_n(\ell)\big)_*\mathbb{P} \Longrightarrow \gamma_{\tau_1}^N \times \cdots \times \gamma_{\tau_\ell}^N.$$

Moreover, since

$$\big(\boldsymbol{\Delta}_n(1), \dots, \boldsymbol{\Delta}_n(\ell)\big)_*\mathbb{P} = \big(\boldsymbol{\Delta}_n(1)\big)_*\mathbb{P} \times \cdots \times \big(\boldsymbol{\Delta}_n(\ell)\big)_*\mathbb{P}$$

for all sufficiently large n's, this reduces to checking $\big(\boldsymbol{\Delta}_n(k)\big)_*\mathbb{P} \Longrightarrow \gamma_{\mathbf{0},\tau_k\mathbf{I}}$ for each $1 \le k \le \ell$. Finally, given $1 \le k \le \ell$, set $M_n(k) = [nt_k] - [nt_{k-1}]$, and use Theorem 2.3.8 to see that, as $n \to \infty$,

$$\mathbb{E}^{\mathbb{P}}\left[\exp\left(\frac{\sqrt{-1}}{M_n(k)^{\frac{1}{2}}}\sum_{j=1}^{M_n(k)}\big(\boldsymbol{\xi}, \mathbf{X}_{[nt_k]+j}\big)_{\mathbb{R}^N}\right)\right] \longrightarrow \exp\left[-\frac{|\boldsymbol{\xi}|^2}{2}\right]$$

uniformly for $\boldsymbol{\xi}$ in compact subsets of $\mathbb{R}^N$. Hence, since $\frac{M_n(k)}{n} \longrightarrow \tau_k$, we now see that, for any fixed $\boldsymbol{\xi} \in \mathbb{R}^N$,

$$\mathbb{E}^{\mathbb{P}}\Big[\exp\Big(\sqrt{-1}\big(\boldsymbol{\xi}, \boldsymbol{\Delta}_n(k)\big)_{\mathbb{R}^N}\Big)\Big] \longrightarrow \exp\left[-\frac{\tau_k|\boldsymbol{\xi}|^2}{2}\right] = \widehat{\gamma_{\mathbf{0},\tau_k\mathbf{I}}}(\boldsymbol{\xi}),$$

and therefore $\big(\boldsymbol{\Delta}_n(k)\big)_*\mathbb{P} \Longrightarrow \gamma_{\mathbf{0},\tau_k\mathbf{I}}$. $\square$

I turn next to the problem of showing that $\{\mu_n : n \geq 1\}$ is tight. By the Ascoli–Arzelá Theorem, any subset $K \subseteq C(\mathbb{R}^N)$ of the form

$$\bigcap_{\ell=1}^{\infty} \left\{ \boldsymbol{\psi} : |\boldsymbol{\psi}(0)| \vee \sup_{0 \leq s < t \leq \ell} \frac{|\boldsymbol{\psi}(t) - \boldsymbol{\psi}(s)|}{(t-s)^{\alpha}} \leq R_\ell \right\}$$

is compact for any $\alpha > 0$ and $\{R_\ell : \ell \geq 1\} \subseteq [0,\infty)$. Thus, since $\mu_n\big(\boldsymbol{\psi}(0) = \mathbf{0}\big) = 1$, all that we have to do is show that, for each $T > 0$

$$\sup_{n \geq 1} \mathbb{E}^{\mathbb{P}} \left[\sup_{1 \leq s < t \leq T} \frac{|\mathbf{S}_n(t) - \mathbf{S}_n(s)|}{(t-s)^{\frac{1}{8}}} \right] < \infty,$$

and, by Theorem 4.3.2, this would follow if we knew that

$$(*) \qquad \sup_{n \geq 1} \mathbb{E}^{\mathbb{P}}\big[|\mathbf{S}_n(t) - \mathbf{S}_n(s)|^4\big] \leq C(t-s)^2, \quad s,\, t \in [0,\infty),$$

for some $C < \infty$.

I will prove (*) under the assumption that, for some $M < \infty$ and all $n \geq 1$, $\mathbb{E}^{\mathbb{P}}\big[|\mathbf{X}_n|^4\big] \leq M$. To do this, note that when $k - 1 \leq ns < nt \leq k$,

$$\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{S}_n(t) - \mathbf{S}_n(s)\big|^4\Big] = n^2(t-s)^4 \mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{X}_k\big|^4\Big] \leq M(t-s)^2.$$

On the other hand, when $k - 1 \leq ns \leq k \leq \ell \leq nt \leq \ell + 1$,

$$\begin{aligned}
&\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{S}_n(t) - \mathbf{S}_n(s)\big|^4\Big] \\
&\leq 27\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{S}_n(t) - \mathbf{S}_n\big(\tfrac{\ell}{n}\big)\big|^4\Big] + 27\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{S}_n\big(\tfrac{\ell}{n}\big) - \mathbf{S}_n\big(\tfrac{k}{n}\big)\big|^4\Big] \\
&\qquad + 27\mathbb{E}^{\mathbb{P}}\Big[\big|\mathbf{S}_n\big(\tfrac{k}{n}\big) - \mathbf{S}_n(s)\big|^4\Big] \\
&\leq 27Mn^2\left(t - \frac{\ell}{n}\right)^4 + \frac{27}{n^2}\mathbb{E}^{\mathbb{P}}\left[\left|\sum_{j=1}^{\ell-k} \mathbf{X}_{k+j}\right|^4\right] + 27Mn^2\left(\frac{k}{n} - s\right)^4 \\
&\leq 54M(t-s)^2 + \frac{81N^2M(\ell-k)^2}{n^2} \leq 135N^2M(t-s)^2,
\end{aligned}$$

where, in the passage to the final line, I have chosen an orthonormal basis $\{\mathbf{e}_i : 1 \leq i \leq N\}$ for $\mathbb{R}^N$ and used the estimate

$$\begin{aligned}
\mathbb{E}^{\mathbb{P}}\left[\left|\sum_{j=1}^{\ell-k} \mathbf{X}_{k+j}\right|^4\right] &= \mathbb{E}^{\mathbb{P}}\left[\left(\sum_{i=1}^{N}\left(\sum_{j=1}^{\ell-k}\big(\mathbf{e}_i, \mathbf{X}_{k+j}\big)_{\mathbb{R}^N}\right)^2\right)^2\right] \\
&\leq N\sum_{i=1}^{N}\mathbb{E}^{\mathbb{P}}\left[\left(\sum_{j=1}^{\ell-k}\big(\mathbf{e}_i, \mathbf{X}_{k+j}\big)_{\mathbb{R}^N}\right)^4\right] \leq 3N^2M(\ell-k)^2
\end{aligned}$$

coming from the second inequality in (1.3.2).

In order to complete the proof, I will apply the Principle of Accompanying Laws. Namely, because the $\mathbf{X}_n$'s are uniformly square $\mathbb{P}$-integrable, we can use a truncation procedure to find functions $\{f_{n,\delta} : n \in \mathbb{Z}^+ \text{ and } \delta > 0\} \subseteq C_{\mathrm{b}}(\mathbb{R}^N, \mathbb{R}^N)$ with the properties that, for each $\delta > 0$, $\sup_{n\in\mathbb{Z}^+} \|f_{n,\delta}\|_{\mathrm{u}} < \infty$,

$$\sup_{n\in\mathbb{Z}^+} \mathbb{E}^{\mathbb{P}}\left[\left|\mathbf{X}_n - f_{n,\delta}\circ \mathbf{X}_n\right|^2\right] < \delta,$$

and, for every $n \in \mathbb{Z}^+$, the random variable $\mathbf{X}_{n,\delta} \equiv f_{n,\delta}\circ\mathbf{X}_n$ has mean-value $\mathbf{0}$ and covariance $\mathbf{I}$. Next, for each $\delta > 0$, define the maps $\omega \in \Omega \longmapsto \mathbf{S}_{n,\delta}(\,\cdot\,,\omega) \in C(\mathbb{R}^N)$ relative to $\{\mathbf{X}_{n,\delta} : n \geq 1\}$, and set $\mu_{n,\delta} = (\mathbf{S}_{n,\delta})_*\mathbb{P}$. Then, by the preceding, we know that $\mu_{n,\delta} \Longrightarrow \mathcal{W}^{(N)}$ for each $\delta > 0$. Hence, by Theorem 9.1.13, we will have proved that $\mu_n \Longrightarrow \mathcal{W}^{(N)}$ as soon as we show that

$$\varlimsup_{\delta\searrow 0} \sup_{n\in\mathbb{Z}^+} P\left(\sup_{0\leq t\leq T}\left|\mathbf{S}_n(t) - \mathbf{S}_{n,\delta}(t)\right| \geq \epsilon\right) = 0$$

for every $T \in \mathbb{Z}^+$ and $\epsilon > 0$. To this end, first observe that, because $\mathbf{S}_n(\,\cdot\,)$ and $\mathbf{S}_{n,\delta}(\,\cdot\,)$ are linear on each interval $[(m-1)2^{-n}, m2^{-n}]$,

$$\sup_{t\in[0,T]}\left|\mathbf{S}_n(t) - \mathbf{S}_{n,\delta}(t)\right| = \max_{1\leq m\leq nT} \frac{1}{n^{\frac12}}\left|\sum_{k=1}^m \mathbf{Y}_{k,\delta}\right|,$$

where $\mathbf{Y}_{k,\delta} \equiv \mathbf{X}_k - \mathbf{X}_{k,\delta}$. Next, note that

$$\begin{aligned} &P\left(\max_{1\leq m\leq nT}\frac{1}{n^{\frac12}}\left|\sum_{k=1}^m \mathbf{Y}_{k,\delta}\right| \geq \epsilon\right) \\ &\qquad \leq N \max_{\mathbf{e}\in\mathbb{S}^{N-1}} P\left(\max_{1\leq m\leq nT}\left|\sum_{k=1}^m \left(\mathbf{e}, \mathbf{Y}_{k,\delta}\right)_{\mathbb{R}^N}\right| \geq \frac{n^{\frac12}\epsilon}{N^{\frac12}}\right). \end{aligned}$$

Finally, by Kolmogorov's Inequality,

$$\mathbb{P}\left(\max_{1\leq m\leq nT}\left|\sum_{k=1}^m \left(\mathbf{e}, \mathbf{Y}_{k,\delta}\right)_{\mathbb{R}^N}\right| \geq \frac{n^{\frac12}\epsilon}{N^{\frac12}}\right) \leq \frac{NT\delta}{\epsilon^2}$$

for every $\mathbf{e} \in \mathbb{S}^{N-1}$.

§**9.3.2. Rayleigh's Random Flights Model.** Here is a more picturesque scheme for approximating Brownian motion. Imagine the path $t \rightsquigarrow \mathbf{R}(t)$ of a bird which starts at the origin, flies in a randomly chosen direction at unit speed for a unit exponential random time, then switches to a new randomly chosen direction for a second unit exponential time, etc. Next, given $\epsilon > 0$, rescale time and space so that the path becomes $t \rightsquigarrow \mathbf{R}_\epsilon(t)$, where $\mathbf{R}_\epsilon(t) \equiv \epsilon^{\frac{1}{2}}\mathbf{R}(\epsilon^{-1}t)$. I will show that, as $\epsilon \searrow 0$, the distribution of $\{\mathbf{R}_\epsilon(t) : t \geq 0\}$ becomes Brownian motion. This model was introduced by Rayleigh and is called his random flights model.

In the following, $\{\tau_m : m \geq 1\}$ is a sequence of independent, unit exponential random variables from which their partial sums $\{T_n : n \geq 0\}$ and the associated simple Poisson process $\{N(t) : t \geq 0\}$ are defined as in § 4.2.1. Finally, given $\epsilon > 0$, $N_\epsilon(t) = N(\epsilon^{-1}t)$.

Lemma 9.3.3. *Let $\{\mathbf{X}_n : n \geq 1\}$ a sequence of mutually independent $\mathbb{R}^N$-valued, uniformly square $\mathbb{P}$-integrable random variables with mean-value $\mathbf{0}$ and covariance $\mathbf{I}$, and define $\{\mathbf{S}_n(t) : t \geq 0\}$ accordingly, as in Theorem 9.3.1. (Note that the $\mathbf{X}_n$'s are not assumed to be independent of the τ_n's.) Next, define*

$$\mathbf{X}_\epsilon(t,\omega) = \sqrt{\epsilon}\sum_{m=1}^{N_\epsilon(t,\omega)}\mathbf{X}_m, \quad (t,\omega) \in [0,\infty)\times\Omega,$$

Then, for all $r \in (0,\infty)$ and $T \in [0,\infty)$,

$$\lim_{\epsilon\searrow 0} P\left(\sup_{t\in[0,T]}\left|\mathbf{X}_\epsilon(t) - \mathbf{S}_{n_\epsilon}(t)\right| \geq r\right) = 0 \quad \textit{where } n_\epsilon \equiv [\epsilon^{-1}].$$

Proof: Note that

$$\mathbf{X}_\epsilon(t,\omega) - \mathbf{S}_{n_\epsilon}(t,\omega) = (\sqrt{\epsilon n_\epsilon} - 1)\,\mathbf{S}_{n_\epsilon}\left(\frac{N_\epsilon(t,\omega)}{n_\epsilon},\omega\right) + \left(\mathbf{S}_{n_\epsilon}\left(\frac{N_\epsilon(t,\omega)}{n_\epsilon},\omega\right) - \mathbf{S}_{n_\epsilon}(t,\omega)\right).$$

Hence, for every $\delta \in (0,1]$,

$$\begin{aligned}\mathbb{P}&\left(\sup_{t\in[0,T]}\left|\mathbf{X}_\epsilon(t) - \mathbf{S}_{n_\epsilon}(t)\right| \geq r\right)\\ &\leq \mathbb{P}\left(\sup_{t\in[0,T+\delta]}\left|\mathbf{S}_{n_\epsilon}(t)\right| \geq \frac{r}{2\epsilon}\right) + \mathbb{P}\left(\sup_{t\in[0,T]}\left|\frac{N_\epsilon(t)}{n_\epsilon} - t\right| \geq \delta\right)\\ &\quad + \mathbb{P}\left(\sup_{s\in[0,T]}\ \sup_{|t-s|\leq\delta}\left|\mathbf{S}_{n_\epsilon}(t) - \mathbf{S}_{n_\epsilon}(s)\right| \geq \frac{r}{2}\right).\end{aligned}$$

But, by Theorem 9.3.1 and the converse statement in Theorem 9.1.9, we know that the first term tends to 0 as $\epsilon \searrow 0$ uniformly in $\delta \in (0,1]$ and that the third term tends to 0 as $\delta \searrow 0$ uniformly in $\epsilon \in (0,1]$. Thus, all that remains is to note that, by Exercise 4.2.19,

$$\lim_{\epsilon\searrow 0}\mathbb{P}\left(\sup_{t\in[0,T]}\bigl|\epsilon N_\epsilon(t)-t\bigr|\geq\delta\right)=0. \quad\square \tag{9.3.4}$$

Now suppose that $\{\boldsymbol{\theta}_n : n\geq 1\}$ is a sequence of mutually independent $\mathbb{R}^N$-valued random variables which satisfy the conditions that

$$M\equiv\sup_{n\in\mathbb{Z}^+}\mathbb{E}^{\mathbb{P}}\Bigl[|\tau_n\boldsymbol{\theta}_n|^4\Bigr]<\infty,$$

$$\mathbb{E}^{\mathbb{P}}\bigl[\tau_n\boldsymbol{\theta}_n\bigr]=\mathbf{0},\quad\text{and}\quad\mathbb{E}^{\mathbb{P}}\Bigl[(\tau_n\boldsymbol{\theta}_n)\otimes(\tau_n\boldsymbol{\theta}_n)\Bigr]=\mathbf{I},\quad n\in\mathbb{Z}^+.$$

Finally, define $\omega\in\Omega\longmapsto\mathbf{R}(\,\cdot\,,\omega)\in C(\mathbb{R}^N)$ by

$$\mathbf{R}(t,\omega)=\bigl(t-T_{N(t,\omega)}(\omega)\bigr)\boldsymbol{\theta}_{N(t,\omega)+1}(\omega)+\sum_{m=1}^{N(t,\omega)}\tau_m(\omega)\boldsymbol{\theta}_m(\omega).$$

The process $\{\mathbf{R}(t) : t\geq 0\}$ is our interpretation of Rayleigh's random flight model. A typical choice of the $\boldsymbol{\theta}_n$'s would be to make them independent of the holding times (i.e., the τ_n's) and to choose them to be uniformly distributed over the sphere $\mathbb{S}^{N-1}\left(\sqrt{N}\right)$.

Theorem 9.3.5. *Referring to the preceding, set*

$$\mathbf{R}_\epsilon(t,\omega)=\sqrt{\epsilon}\,\mathbf{R}\left(\tfrac{t}{\epsilon},\omega\right),\quad(t,\omega)\in[0,\infty)\times\Omega.$$

Then $\left(\mathbf{R}_\epsilon\right)_*\mathbb{P}\Longrightarrow\mathcal{W}^{(N)}$ *as* $\epsilon\searrow 0$.

Proof: Set $\mathbf{X}_n=\tau_n\boldsymbol{\theta}_n$, and, using the same notation as in Lemma 9.3.3, observe that

$$\bigl|\mathbf{R}_\epsilon(t)-\mathbf{X}_\epsilon(t)\bigr|\leq\sqrt{\epsilon}\,\bigl|\mathbf{X}_{N_\epsilon(t)+1}\bigr|.$$

Hence, by Lemma 9.3.3 and Theorems 9.3.1 and 9.1.13, all that we have to do is check that

$$\lim_{\epsilon\searrow 0}\mathbb{P}\left(\sup_{t\in[0,T]}\bigl|\sqrt{\epsilon}\,\mathbf{X}_{N_\epsilon(t)+1}\bigr|\geq r\right)=0$$

for every $r\in(0,\infty)$ and $T\in[0,\infty)$. To this end, set $T_\epsilon=\frac{1+T}{\epsilon}$. Then, by (9.3.4), we have that

$$\begin{aligned}\lim_{\epsilon\searrow 0}\mathbb{P}\left(\sup_{t\in[0,T]}\bigl|\sqrt{\epsilon}\,\mathbf{X}_{N_\epsilon(t)+1}\bigr|\geq r\right)&=\lim_{\epsilon\searrow 0}\mathbb{P}\left(\max_{0\leq n\leq T_\epsilon}|\mathbf{X}_{n+1}|\geq\frac{r}{\sqrt{\epsilon}}\right)\\&\leq\lim_{\epsilon\searrow 0}\frac{\sqrt{\epsilon}}{r}\mathbb{E}^{\mathbb{P}}\left[\left(\sum_{0\leq n\leq T_\epsilon}|\mathbf{X}_{n+1}|^4\right)^{\frac14}\right]\leq\lim_{\epsilon\searrow 0}\frac{\bigl(M\epsilon(2+T)\bigr)^{\frac14}}{r}=0.\quad\square\end{aligned}$$

Exercise for § 9.3

EXERCISE 9.3.6. Let $\{\mu_n : n \geq 1\} \subseteq \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$, and, for each $T \in (0,\infty)$, let $\mu_n^T \in \mathbf{M}_1\big(C\big([0,T];E\big)\big)$ denote the distribution of

$$\boldsymbol{\psi} \in C(\mathbb{R}^N) \longmapsto \boldsymbol{\psi} \restriction [0,T] \in C\big([0,T];\mathbb{R}^N\big) \text{ under } \mu_n.$$

Show that there is a $\mu \in \mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ to which $\{\mu_n : n \geq 1\}$ converges in $\mathbf{M}_1\big(C(\mathbb{R}^N)\big)$ if and only if, for each $T \in (0,\infty)$, there is a $\mu^T \in \mathbf{M}_1\big(C([0,T];\mathbb{R}^N)\big)$ with the property that

$$\mu_n^T \Longrightarrow \mu^T \quad \text{in } \mathbf{M}_1\big(C([0,T];\mathbb{R}^N)\big);$$

in which case, μ^T is the distribution of

$$\boldsymbol{\psi} \in C(\mathbb{R}^N) \longmapsto \boldsymbol{\psi} \restriction [0,T] \in C([0,T];\mathbb{R}^N) \text{ under } \mu.$$

In particular, weak convergence of measures on $C(\mathbb{R}^N)$ is really a *local* property.

EXERCISE 9.3.7. *Erdös–Kac Theorem* Donsker's own proof of Theorem 9.3.1 was entirely different from the one given here. Instead it was based on a special case of his result, a case which had been proved already (with a very difficulty argument) by P. Erdös and M. Kac. The result of Erdös and Kac was that if $\{X_n : n \geq 1\}$ is a sequence of independent, uniformly square integrable random variables with mean value 0 and variance 1, then, for all $a \geq 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\max_{1\leq m\leq n} n^{-\frac{1}{2}} \sum_{k=1}^{m} X_k \geq a\right) = \sqrt{\frac{2}{\pi}} \int_a^\infty e^{-\frac{x^2}{2}}\, dx.$$

Prove their result as an application of Donsker's Theorem and part (**iii**) of Exercise 4.3.11. According to Kac, it was G. Uhlenbeck who first suggested that their result might be a consequence of a more general "invariance" principle.

EXERCISE 9.3.8. Here is another version of Rayleigh's random flight model. Again let $\{\tau_k : k \geq 1\}$, $\big\{T_m : m \geq 0\big\}$, and $\big\{N(t) : t \geq 0\big\}$ be as in § 4.2.2; and set

$$R(t) = \int_0^t (-1)^{N(s)}\, ds \quad \text{and} \quad R_\epsilon(t) = \sqrt{\epsilon}\, R\left(\tfrac{t}{\epsilon}\right).$$

Show that $\big(R_\epsilon\big)_* \mathbb{P} \Longrightarrow \mathcal{W}^{(1)}$ as $\epsilon \searrow 0$.

Hint: Set $\beta_k = 0$ or 1 according to whether $k \in \mathbb{N}$ is even or odd, and note that

$$\sum_{k=1}^{n} (-1)^k \tau_k = \sum_{k=1}^{n} \beta_k\big(\tau_{k+1} - \tau_k\big) - \beta_n \tau_n = \sum_{1\leq k\leq \frac{n}{2}} \big(\tau_{2k} - \tau_{2k-1}\big) - \beta_n \tau_{n+1};$$

and now proceed as in the derivations of Lemma 9.3.3 and Theorem 9.3.5.