

# Markov kernels, convolution semigroups, and projective families of probability measures

Jordan Bell

jordan.bell@gmail.com

Department of Mathematics, University of Toronto

June 9, 2015

## 1 Transition kernels

For a measurable space  $(E, \mathcal{E})$ , we denote by  $\mathcal{E}_+$  the set of functions  $E \rightarrow [0, \infty]$  that are  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$  measurable. It can be proved that if  $I : \mathcal{E}_+ \rightarrow [0, \infty]$  is a function such that (i)  $f = 0$  implies that  $I(f) = 0$ , (ii) if  $f, g \in \mathcal{E}_+$  and  $a, b \geq 0$  then  $I(af + bg) = aI(f) + bI(g)$ , and (iii) if  $f_n$  is a sequence in  $\mathcal{E}_+$  that increases pointwise to an element  $f$  of  $\mathcal{E}_+$  then  $I(f_n)$  increases to  $I(f)$ , then there is a unique measure  $\mu$  on  $\mathcal{E}$  such that  $I(f) = \int \mu f$  for each  $f \in \mathcal{E}_+$ .<sup>1</sup>

Let  $(E, \mathcal{E})$  and  $(F, \mathcal{F})$  be a measurable space. A **transition kernel** is a function

$$K : E \times \mathcal{F} \rightarrow [0, \infty]$$

such that (i) for each  $x \in E$ , the function  $K_x : \mathcal{F} \rightarrow [0, \infty]$  defined by

$$B \mapsto K(x, B)$$

is a measure on  $\mathcal{F}$ , and (ii) for each  $B \in \mathcal{F}$ , the map

$$x \mapsto K(x, B)$$

is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$ .

If  $\mu$  is a measure on  $\mathcal{E}$ , define

$$(K_*\mu)(B) = \int_E K(x, B) d\mu(x), \quad B \in \mathcal{F}.$$

If  $B_n$  are pairwise disjoint elements of  $\mathcal{F}$ , then using that  $B \mapsto K(x, B)$  is a

---

<sup>1</sup>Erhan Çinlar, *Probability and Stochastics*, p. 28, Theorem 4.21.

measure and the monotone convergence theorem,

$$\begin{aligned}
(K_*\mu)\left(\bigcup_n B_n\right) &= \int_E K\left(x, \bigcup_n B_n\right) d\mu(x) \\
&= \int_E \sum_n K(x, B_n) d\mu(x) \\
&= \sum_n \int_E K(x, B_n) d\mu(x) \\
&= \sum_n (K_*\mu)(B_n),
\end{aligned}$$

showing that  $K_*\mu$  is a measure on  $\mathcal{F}$ .

If  $f \in \mathcal{F}_+$ , define  $K^*f : E \rightarrow [0, \infty]$  by

$$(K^*f)(x) = \int_F f(y) dK_x(y), \quad x \in E. \quad (1)$$

For  $\phi = \sum_{j=1}^k b_j 1_{B_j}$  with  $b_j \geq 0$  and  $B_j \in \mathcal{F}$ , because  $x \mapsto K(x, B_j)$  is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$  for each  $j$ ,

$$(K^*\phi)(x) = \int_F \sum_{j=1}^k b_j 1_{B_j}(y) dK_x(y) = \sum_{j=1}^k b_j K_x(B_j) = \sum_{j=1}^k b_j K(x, B_j),$$

is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$ . For  $f \in \mathcal{F}_+$ , there is a sequence of simple functions  $\phi_n$  with  $0 \leq \phi_1 \leq \phi_2 \leq \dots$  that converges pointwise to  $f$ ,<sup>2</sup> and then by the monotone convergence theorem, for each  $x \in E$  we have

$$(K^*\phi_n)(x) = \int_F \phi_n(y) dK_x(y) \rightarrow \int_F f(y) dK_x(y) = (K^*f)(x),$$

showing  $K^*\phi_n$  converges pointwise to  $K^*f$ , and because each  $K^*\phi_n$  is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$ ,  $K^*f$  is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{[0, \infty]}$ .<sup>3</sup> Therefore, if  $f \in \mathcal{F}_+$  then  $K^*f \in \mathcal{E}_+$ . In particular, if  $K$  is a transition kernel from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$ ,

$$(K^*1_B)(x) = \int_F 1_B(y) dK_x(y) = K_x(B) = K(x, B), \quad x \in E, \quad B \in \mathcal{F}. \quad (2)$$

The following gives conditions under which (2) defines a transition kernel.<sup>4</sup>

**Lemma 1.** *Suppose that  $N : \mathcal{F}_+ \rightarrow \mathcal{E}_+$  satisfies the following properties:*

1.  $N(0) = 0$ .

<sup>2</sup>Gerald B. Folland, *Real Analysis: Modern Techniques and Their Applications*, second ed., p. 47, Theorem 2.10.

<sup>3</sup>Gerald B. Folland, *Real Analysis: Modern Techniques and Their Applications*, second ed., p. 45, Proposition 2.7.

<sup>4</sup>Heinz Bauer, *Probability Theory*, p. 308, Lemma 36.2.

2.  $N(af + bg) = aN(f) + bN(g)$  for  $f, g \in \mathcal{F}_+$  and  $a, b \geq 0$ .

3. If  $f_n$  is a sequence in  $\mathcal{F}_+$  increasing to  $f \in \mathcal{F}_+$ , then  $N(f_n) \uparrow N(f)$ .

Then

$$K(x, B) = (N(1_B))(x), \quad x \in E, \quad B \in \mathcal{F},$$

is a transition kernel from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$ .  $K$  is the unique transition kernel satisfying

$$K^*f = N(f), \quad f \in \mathcal{F}_+.$$

If  $K$  is a transition kernel from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$  and  $L$  is a transition kernel from  $(F, \mathcal{F})$  to  $(G, \mathcal{G})$ , the function  $K^* \circ L^* : \mathcal{G}_+ \rightarrow \mathcal{E}_+$  satisfies (i)  $(K^* \circ L^*)(0) = K^*(0) = 0$ , (ii) if  $f, g \in \mathcal{G}_+$  and  $a, b \geq 0$ ,

$$\begin{aligned} (K^* \circ L^*)(af + bg) &= K^*(aL^*(f) + bL^*(g)) \\ &= aK^*(L^*(f)) + K^*(L^*(g)) \\ &= a(K^* \circ L^*)(f) + b(K^* \circ L^*)(g), \end{aligned}$$

and (iii) if  $f_n \uparrow f$  in  $\mathcal{G}_+$ , then by the monotone convergence theorem,  $L^*(f_n) \uparrow L^*(f)$ , and then again applying the monotone convergence theorem,  $K^*(L^*(f_n)) \uparrow K^*(L^*(f))$ , i.e.

$$(K^* \circ L^*)(f_n) \uparrow (K^* \circ L^*)(f).$$

Therefore, from Lemma 1 we get that there is a unique transition kernel from  $(E, \mathcal{E})$  to  $(G, \mathcal{G})$ , denoted  $KL$  and called the **product of  $K$  and  $L$** , such that

$$(KL)^*f = (K^* \circ L^*)(f), \quad f \in \mathcal{G}_+.$$

For  $f \in \mathcal{G}_+$  and  $x \in E$ ,

$$\begin{aligned} (KL)^*(f)(x) &= (K^*(L^*f))(x) \\ &= \int_F (L^*f)(y) dK_x(y) \\ &= \int_F \left( \int_G f(z) dL_y(z) \right) dK_x(y). \end{aligned}$$

In particular, for  $C \in \mathcal{G}$ ,

$$(KL)^*(1_C)(x) = \int_F L_y(C) dK_x(y) = \int_F L(y, C) dK_x(y). \quad (3)$$

## 2 Markov kernels

A **Markov kernel** from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$  is a transition kernel  $K$  such that for each  $x \in E$ ,  $K_x$  is a probability measure on  $\mathcal{F}$ . The **unit kernel** from  $(E, \mathcal{E})$  to  $(E, \mathcal{E})$  is

$$I(x, A) = \delta_x(A). \quad (4)$$

It is apparent that the unit kernel is a Markov kernel.

If  $K$  is a Markov kernel from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$  and  $L$  is a Markov kernel from  $(F, \mathcal{F})$  to  $(G, \mathcal{G})$ , then for  $x \in E$ , by (3) we have

$$(KL)^*(1_G)(x) = \int_F dK_x(y) = K_x(F) = K(x, F) = 1,$$

and thus by (2),

$$(KL)_x(G) = (KL)(x, G) = 1,$$

showing that for each  $x \in E$ ,  $(KL)_x$  is a probability measure. Therefore, the product of two Markov kernels is a Markov kernel.

Let  $(E, \mathcal{E})$  be a measurable space and let

$$B_b(\mathcal{E})$$

be the set of bounded functions  $E \rightarrow \mathbb{R}$  that are measurable  $\mathcal{E} \rightarrow \mathcal{B}_{\mathbb{R}}$ .  $B_b(\mathcal{E})$  is a Banach space with the **uniform norm**

$$\|f\|_u = \sup_{x \in E} |f(x)|.$$

For  $K$  a Markov kernel from  $(E, \mathcal{E})$  to  $(F, \mathcal{F})$  and for  $f \in B_b(\mathcal{F})$ , define  $K^*f : E \rightarrow \mathbb{R}$  by

$$(K^*f)(x) = \int_F f(y) dK_x(y), \quad x \in E,$$

for which

$$|(K^*f)(x)| \leq \int_F |f(y)| dK_x(y) \leq \|f\|_u K_x(F) = \|f\|_u,$$

showing that  $\|K^*f\|_u \leq \|f\|_u$ . Furthermore, there is a sequence of simple functions  $\phi_n \in B_b(\mathcal{F})$  that converges to  $f$  in the norm  $\|\cdot\|_u$ .<sup>5</sup> For  $x \in E$ , by the dominated convergence theorem we get that

$$(K^*\phi_n)(x) = \int_F \phi_n(y) dK_x(y) \rightarrow \int_F f(y) dK_x(y) = (K^*f)(x).$$

Each  $K^*\phi_n$  is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{\mathbb{R}}$ , hence  $K^*f$  is measurable  $\mathcal{E} \rightarrow \mathcal{B}_{\mathbb{R}}$  and so belongs to  $B_b(\mathcal{E})$ .

### 3 Markov semigroups

Let  $(E, \mathcal{E})$  be a measurable space and for each  $t \geq 0$ , let  $P_t$  be a Markov kernel from  $(E, \mathcal{E})$  to  $(E, \mathcal{E})$ . We say that the family  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  is a **Markov semigroup** if

$$P_{s+t} = P_s P_t, \quad s, t \in \mathbb{R}_{\geq 0}.$$

<sup>5</sup>V. I. Bogachev, *Measure Theory*, p. 108, Lemma 2.1.8.

For  $x \in E$  and  $A \in \mathcal{E}$  and for  $s, t \geq 0$ , by (2) and (3),

$$(P_s P_t)(x, A) = ((P_s P_t)^* 1_A)(x) = \int_E P_t(y, A) d(P_s)_x(y)$$

Thus

$$P_{s+t}(x, A) = \int_E P_t(y, A) d(P_s)_x(y), \quad (5)$$

called the **Chapman-Kolmogorov equation**.

## 4 Infinitely divisible distributions

Let  $\mathcal{P}(\mathbb{R}^d)$  be the collection of Borel probability measures on  $\mathbb{R}^d$ . For  $\mu \in \mathcal{P}(\mathbb{R}^d)$ , its **characteristic function**  $\tilde{\mu} : \mathbb{R}^d \rightarrow \mathbb{C}$  is defined by

$$\tilde{\mu}(x) = \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} d\mu(y).$$

$\tilde{\mu}$  is uniformly continuous on  $\mathbb{R}^d$  and  $|\tilde{\mu}(x)| \leq \tilde{\mu}(0) = 1$  for all  $x \in \mathbb{R}^d$ .<sup>6</sup> For  $\mu_1, \dots, \mu_n \in \mathcal{P}(\mathbb{R}^d)$ , let  $\mu$  be their **convolution**:

$$\mu = \mu_1 * \dots * \mu_n,$$

which for  $A$  a Borel set in  $\mathbb{R}^d$  is defined by

$$\mu(A) = \int_{(\mathbb{R}^d)^n} 1_A(x_1 + \dots + x_n) d(\mu_1 \times \dots \times \mu_n)(x_1, \dots, x_n).$$

One computes that<sup>7</sup>

$$\tilde{\mu} = \tilde{\mu}_1 \dots \tilde{\mu}_n.$$

An element  $\mu$  of  $\mathcal{P}(\mathbb{R}^d)$  is called **infinitely divisible** if for each  $n \geq 1$ , there is some  $\mu_n \in \mathcal{P}(\mathbb{R}^d)$  such that

$$\mu = \underbrace{\mu_n * \dots * \mu_n}_n. \quad (6)$$

Thus,

$$\tilde{\mu} = (\tilde{\mu}_n)^n. \quad (7)$$

On the other hand, if  $\mu_n \in \mathcal{P}(\mathbb{R}^d)$  is such that (7) is true, then because the characteristic function of  $\mu_n * \dots * \mu_n$  is  $(\tilde{\mu}_n)^n$  and the characteristic function of  $\mu$  is  $\tilde{\mu}$  and these are equal, it follows that  $\mu_n * \dots * \mu_n$  and  $\mu$  are equal.

The following theorem is useful for doing calculations with the characteristic function of an infinitely divisible distribution.<sup>8</sup>

<sup>6</sup>Heinz Bauer, *Probability Theory*, p. 183, Theorem 22.3.

<sup>7</sup>Heinz Bauer, *Probability Theory*, p. 184, Theorem 22.4.

<sup>8</sup>Heinz Bauer, *Probability Theory*, p. 246, Theorem 29.2.

**Theorem 2.** Suppose that  $\mu$  is an infinitely divisible distribution on  $\mathbb{R}^d$ . First,

$$\tilde{\mu}(x) \neq 0, \quad x \in \mathbb{R}^d.$$

Second, there is a unique continuous function  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  satisfying  $\phi(0) = 0$  and

$$\tilde{\mu} = |\tilde{\mu}|e^{i\phi}.$$

Third, for each  $n \geq 1$ , there is a unique  $\mu_n \in \mathcal{P}(\mathbb{R}^d)$  for which  $\mu = \mu_n * \cdots * \mu_n$ . The characteristic function of this unique  $\mu_n$  is

$$\tilde{\mu}_n = |\tilde{\mu}|^{\frac{1}{n}} e^{i\frac{\phi}{n}}.$$

A **convolution semigroup** is a family  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  of elements of  $\mathcal{P}(\mathbb{R}^d)$  such that for  $s, t \in \mathbb{R}_{\geq 0}$ ,

$$\mu_{s+t} = \mu_s * \mu_t.$$

The convolution semigroup is called **continuous** when  $t \mapsto \mu_t$  is continuous  $\mathbb{R}_{\geq 0} \rightarrow \mathcal{P}(\mathbb{R}^d)$ , where  $\mathcal{P}(\mathbb{R}^d)$  has the **narrow topology**.<sup>9</sup>

The following theorem connects convolution semigroups and infinitely divisible distributions.<sup>10</sup>

**Theorem 3.** If  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a convolution semigroup on  $\mathcal{B}_{\mathbb{R}^d}$ , then for each  $t$ , the measure  $\mu_t$  is infinitely divisible.

If  $\mu \in \mathcal{P}(\mathbb{R}^d)$  is infinitely divisible and  $t_0 > 0$ , then there is a unique continuous convolution semigroup  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  such that  $\mu_{t_0} = \mu$ .

It follows from the above theorem that for a convolution semigroup  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  on  $\mathcal{B}_{\mathbb{R}^d}$ ,  $\mu_1$  is infinitely divisible and therefore by Theorem 2,  $\tilde{\mu}_1(x) \neq 0$  for all  $x$ . But  $\mu_0 * \mu_1 = \mu_1$ , so  $\tilde{\mu}_0 \tilde{\mu}_1 = \tilde{\mu}_1$ , and  $\tilde{\mu}_0(x) = 1$  for each  $x$ . But  $\delta_0(x) = 1$  for all  $x$ , so

$$\mu_0 = \delta_0. \tag{8}$$

## 5 Translation-invariant semigroups

Let  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  be a Markov semigroup on  $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ . We say that  $(P_t)_{t \in \mathbb{R}}$  is **translation-invariant** if for all  $x, y \in \mathbb{R}^d$ ,  $A \in \mathcal{B}_{\mathbb{R}^d}$ , and  $t \in \mathbb{R}_{\geq 0}$ ,

$$P_t(x, A) = P_t(x + y, A + y).$$

In this case, for  $t \geq 0$  and for  $A \in \mathcal{B}_{\mathbb{R}^d}$ , define

$$\mu_t(A) = P_t(0, A).$$

<sup>9</sup><http://individual.utoronto.ca/jordanbell/notes/narrow.pdf>

<sup>10</sup>Heinz Bauer, *Probability Theory*, p. 248, Theorem 29.6.

Each  $\mu_t$  is a probability measure on  $\mathcal{B}_{\mathbb{R}^d}$ , and

$$\mu_t(A - x) = P_t(0, A - x) = P_t(x, (A - x) + x) = P_t(x, A).$$

Using that the Chapman-Kolmogorov equation (5) and as  $(P_s)_0(B) = P_s(0, B) = \mu_s(B)$ ,

$$\begin{aligned} \mu_{s+t}(A) &= P_{s+t}(0, A) \\ &= \int_{\mathbb{R}^d} P_t(y, A) d(P_s)_0(y) \\ &= \int_{\mathbb{R}^d} \mu_t(A - y) d\mu_s(y) \\ &= (\mu_t * \mu_s)(A), \end{aligned}$$

showing that  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a convolution semigroup on  $\mathcal{B}_{\mathbb{R}^d}$ .

On the other hand, if  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a convolution semigroup of probability measures on  $\mathcal{B}_{\mathbb{R}^d}$ , for  $t \geq 0$ ,  $x \in \mathbb{R}^d$ , and  $A \in \mathcal{B}_{\mathbb{R}^d}$  define

$$P_t(x, A) = \mu_t(A - x).$$

Let  $t \geq 0$ . For  $x \in \mathbb{R}^d$ , the map  $A \mapsto P_t(x, A) = \mu_t(A - x)$  is a probability measure on  $\mathcal{B}_{\mathbb{R}^d}$ . The map  $(x, y) \mapsto x + y$  is continuous  $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ , and for  $A \in \mathcal{B}_{\mathbb{R}^d}$ , the map  $1_A : \mathbb{R}^d \rightarrow \mathbb{R}$  is measurable  $\mathcal{B}_{\mathbb{R}^d} \rightarrow \mathcal{B}_{\mathbb{R}}$ . Hence, as  $\mathcal{B}_{\mathbb{R}^d \times \mathbb{R}^d} = \mathcal{B}_{\mathbb{R}^d} \otimes \mathcal{B}_{\mathbb{R}^d}$ , the map  $(x, y) \mapsto 1_A(x + y)$  is measurable  $\mathcal{B}_{\mathbb{R}^d} \otimes \mathcal{B}_{\mathbb{R}^d} \rightarrow \mathcal{B}_{\mathbb{R}}$ . Thus by Fubini's theorem,

$$x \mapsto \int_{\mathbb{R}^d} 1_A(x + y) d\mu_t(y) = \int_{\mathbb{R}^d} 1_{A-x}(y) d\mu_t(y) = \mu_t(A - x)$$

is measurable  $\mathcal{B}_{\mathbb{R}^d} \rightarrow \mathcal{B}_{\mathbb{R}}$ . Hence  $P_t$  is a Markov kernel, and thus  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  is a translation-invariant Markov semigroup.

Define  $S : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by  $S(x) = -x$ . For  $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ ,

$$\begin{aligned} S_*(\mu * \nu)(A) &= (\mu * \nu)(-A) \\ &= \int_{\mathbb{R}^d} \mu(-A - y) d\nu(y) \\ &= \int_{\mathbb{R}^d} \mu(-A + y) d\bar{\nu}(y) \\ &= \int_{\mathbb{R}^d} \bar{\mu}(A - y) d\bar{\nu}(y) \\ &= (\bar{\mu} * \bar{\nu})(A), \end{aligned}$$

thus

$$S_*(\mu * \nu) = (S_*\mu) * (S_*\nu). \quad (9)$$

For  $\mu \in \mathcal{P}(\mathbb{R}^d)$ , write

$$\bar{\mu} = S_*\mu \in \mathcal{P}(\mathbb{R}^d),$$

i.e.,

$$\bar{\mu}(A) = \mu(S^{-1}(A)) = \mu(S(A)) = \mu(-A).$$

We calculate

$$(P_t^* 1_A)(x) = P_t(x, A) = \mu_t(A - x) = \int_{\mathbb{R}^d} 1_A(x + y) d\mu_t(y).$$

Then if  $f$  is a simple function,  $f = \sum_k a_k 1_{A_k}$ ,

$$(P_t^* f)(x) = \sum_k a_k \int_{\mathbb{R}^d} 1_{A_k}(x + y) d\mu_t(y) = \int_{\mathbb{R}^d} f(x + y) d\mu_t(y).$$

For  $f \in B_b(\mathcal{B}_{\mathbb{R}^d})$ , there is a sequence of simple functions  $f_n$  that converge to  $f$  in the uniform norm, and then by the dominated convergence theorem we get

$$(P_t^* f)(x) = \int_{\mathbb{R}^d} f(x + y) d\mu_t(y).$$

But

$$\begin{aligned} \int_{\mathbb{R}^d} f(x + y) d\mu_t(y) &= \int_{\mathbb{R}^d} f(x + S(S(y))) d\mu_t(y) \\ &= \int_{\mathbb{R}^d} f(x + S(y)) d(S_* \mu_t)(y) \\ &= \int_{\mathbb{R}^d} f(x - y) d\bar{\mu}_t(y) \\ &= (f * \bar{\mu}_t)(x). \end{aligned}$$

Therefore for  $t \geq 0$  and  $f \in B_b(\mathcal{B}_{\mathbb{R}^d})$ ,

$$P_t^* f = f * \bar{\mu}_t. \quad (10)$$

For  $s, t \geq 0$  and  $f \in B_b(\mathcal{B}_{\mathbb{R}^d})$ , by (10), the fact that  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a convolution semigroup, and (9), we get

$$\begin{aligned} P_{s+t}^* f &= f * (S_* \mu_{s+t}) \\ &= f * (S_* (\mu_s * \mu_t)) \\ &= f * ((S_* \mu_s) * (S_* \mu_t)) \\ &= (f * (S_* \mu_s)) * (S_* \mu_t) \\ &= (P_s^* f) * (S_* \mu_t) \\ &= P_t^* (P_s^* f). \end{aligned}$$

This shows that  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  is a Markov semigroup. Moreover, by (8) it holds that  $\mu_0 = \delta_0$ , and hence

$$P_0(x, A) = \mu_0(A - x) = \delta_0(A - x) = \delta_x(A).$$

Namely,  $P_0$  is the unit kernel (4).

If  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a convolution semigroup and some  $\mu_t$  has density  $q_t$  with respect to Lebesgue measure  $\lambda_d$  on  $\mathbb{R}^d$ ,

$$\mu_t = q_t \lambda_d,$$

then writing  $\bar{q}_t(x) = q_t(-x)$ , for  $f \in B_b(\mathcal{B}_{\mathbb{R}^d})$  by (10) we have

$$(P_t^* f)(x) = (f * \bar{\mu}_t)(x) = \int_{\mathbb{R}^d} f(x-y) d\bar{\mu}_t(y) = \int_{\mathbb{R}^d} f(x+y) q_t(y) d\lambda_d(y)$$

so

$$P_t * f = f * \bar{q}_t. \quad (11)$$

## 6 The Brownian semigroup

For  $a \in \mathbb{R}$  and  $\sigma > 0$ , let  $\gamma_{a,\sigma^2}$  be the Gaussian measure on  $\mathbb{R}$ , the probability measure on  $\mathbb{R}$  whose density with respect to Lebesgue measure is

$$p(x, a, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

For  $\sigma = 0$ , let

$$\gamma_{a,0} = \delta_a.$$

Define for  $t \in \mathbb{R}_{\geq 0}$ ,

$$\mu_t = \prod_{k=1}^d \gamma_{0,t},$$

which is an element of  $\mathcal{P}(\mathbb{R}^d)$ . For  $s, t \in \mathbb{R}_{\geq 0}$ , we calculate

$$\mu_s * \mu_t = \left( \prod_{k=1}^d \gamma_{0,s} \right) * \left( \prod_{k=1}^d \gamma_{0,t} \right) = \prod_{k=1}^d (\gamma_{0,s} * \gamma_{0,t}) = \prod_{k=1}^d \gamma_{0,s+t} = \mu_{s+t}.$$

**Lévy's continuity theorem** states that if  $\nu_n$  is a sequence in  $\mathcal{P}(\mathbb{R}^d)$  and there is some  $\phi : \mathbb{R}^d \rightarrow \mathbb{C}$  that is continuous at 0 and to which  $\tilde{\nu}_n$  converges pointwise, then there is some  $\nu \in \mathcal{P}(\mathbb{R}^d)$  such that  $\phi = \tilde{\nu}$  and such that  $\nu_n \rightarrow \nu$  narrowly.<sup>11</sup> But for  $t \in \mathbb{R}_{\geq 0}$  and  $x \in \mathbb{R}^d$ , we calculate

$$\tilde{\mu}_t(x) = \int_{\mathbb{R}^d} e^{i\langle x, y \rangle} d\mu_t(y) = \exp\left(-\frac{t|x|^2}{2}\right). \quad (12)$$

Let  $\phi(x) = 1$  for all  $x$ , for which  $\tilde{\delta}_0 = \phi$ . For  $t_n \in \mathbb{R}_{\geq 0}$  tending to 0, let  $\nu_n = \mu_{t_n}$ . Then by (12),  $\tilde{\nu}_n$  converges pointwise to  $\phi$ , so by Lévy's continuity theorem,  $\nu_n$  converges narrowly to  $\delta_0$ . Moreover, because  $\mathbb{R}^d$  is a Polish space,

<sup>11</sup><http://individual.utoronto.ca/jordanbell/notes/martingaleCLT.pdf>, p. 19, Theorem 15.

$\mathcal{P}(\mathbb{R}^d)$  is a Polish space, and in particular is metrizable. It thus follows that  $\mu_t$  converges narrowly to  $\delta_0$  as  $t \rightarrow 0$ . It then follows that  $t \mapsto \mu_t$  is continuous  $\mathbb{R}_{\geq 0} \rightarrow \mathcal{P}(\mathbb{R}^d)$ . Summarizing,  $(\mu_t)_{t \in \mathbb{R}_{\geq 0}}$  is a continuous convolution semigroup.

For  $t > 0$ ,  $\mu_t$  has density

$$g_t(x) = \prod_{j=1}^d (2\pi t)^{-1/2} e^{-\frac{x_j^2}{2t}} = (2\pi t)^{-d/2} e^{-\frac{|x|^2}{2t}}$$

with respect to Lebesgue measure  $\lambda_d$  on  $\mathbb{R}^d$ . For  $t \geq 0$ , let

$$P_t(x, A) = \mu_t(A - x).$$

We have established that  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  is a translation-invariant Markov semigroup for which  $P_0(x, A) = \delta_x(A)$ . We call  $(P_t)_{t \in \mathbb{R}_{\geq 0}}$  the **Brownian semigroup**. For  $t > 0$  and  $f \in B_b(\mathcal{B}_{\mathbb{R}^d})$ , because  $\bar{g}_t = g_t$  we have by (11),

$$(P_t f)(x) = (f * g_t)(x) = (2\pi t)^{-d/2} \int_{\mathbb{R}^d} f(x - y) e^{-\frac{|y|^2}{2t}} d\lambda_d(y).$$

## 7 Projective families

For a nonempty set  $I$ , let  $\mathcal{K}(I)$  denote the family of finite nonempty subsets of  $I$ . We speak in this section about **projective families** of probability measures.<sup>12</sup>

The following theorem shows how to construct a projective family from a Markov semigroup on a measurable space and a probability measure on this measurable space.<sup>13</sup>

**Theorem 4.** *Let  $I = \mathbb{R}_{\geq 0}$ , let  $(E, \mathcal{E})$  be a measurable space, let  $(P_t)_{t \in I}$  be a Markov semigroup on  $\mathcal{E}$ , and let  $\mu$  be a probability measure on  $\mathcal{E}$ . For  $J \in \mathcal{K}(I)$ , with elements  $t_1 < \dots < t_n$ , and for  $A \in \mathcal{E}^J$ , let*

$$P_J(A) = \underbrace{\int_E \int_E \cdots \int_E}_{n+1} 1_A(x_1, \dots, x_n) d(P_{t_n - t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1})_{x_0}(x_1) d\mu(x_0).$$

*Then  $(P_J)_{J \in \mathcal{K}(I)}$  is a projective family of probability measures.*

*Proof.* Let  $A_k$  be pairwise disjoint elements of  $\mathcal{E}^J$ , and call their union  $A$ . Then  $1_A = \sum_k 1_{A_k}$ , and applying the monotone convergence theorem  $n + 1$  times,

$$\begin{aligned} & \underbrace{\int_E \int_E \cdots \int_E}_{n+1} 1_A(x_1, \dots, x_n) d(P_{t_n - t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1})_{x_0}(x_1) d\mu(x_0) \\ &= \sum_k \underbrace{\int_E \int_E \cdots \int_E}_{n+1} 1_{A_k}(x_1, \dots, x_n) d(P_{t_n - t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1})_{x_0}(x_1) d\mu(x_0), \end{aligned}$$

<sup>12</sup>See <http://individual.utoronto.ca/jordanbell/notes/finitedimdistributions.pdf>

<sup>13</sup>Heinz Bauer, *Probability Theory*, p. 314, Theorem 36.4.

i.e.

$$P_J(A) = \sum_k P_J(A_k).$$

Furthermore, because  $(P_t)_x$  is a probability measure for each  $t$  and for each  $x$  and  $\mu$  is a probability measure, we calculate that

$$P_J(E^J) = 1.$$

Thus,  $P_J$  is a probability measure on  $\mathcal{E}^J$ .

To prove that  $(P_J)_{J \in \mathcal{K}(I)}$  is a projective family, it suffices to prove that when  $J, K \in \mathcal{K}(I)$ ,  $J \subset K$ , and  $K \setminus J$  is a singleton, then  $(\pi_{K,J})_* P_K = P_J$ .<sup>14</sup> Moreover, because (i) the product  $\sigma$ -algebra  $\mathcal{E}^J$  is generated by the collection of **cylinder sets**, i.e. sets of the form  $\prod_{t \in J} A_t$  for  $A_t \in \mathcal{E}$ , and (ii) the intersection of finitely many cylinder sets is a cylinder set, it is proved using the monotone class theorem that if two probability measures on  $\mathcal{E}^J$  coincide on the cylinder sets, then they are equal.<sup>15</sup> Let  $t_1 < \dots < t_n$  be the elements of  $J$ . To prove that  $(\pi_{K,J})_* P_K$  and  $P_J$  are equal, it suffices to prove that for any  $A_1, \dots, A_n \in \mathcal{E}$ ,

$$(\pi_{K,J})_* P_K \left( \prod_{j=1}^n A_j \right) = P_J \left( \prod_{j=1}^n A_j \right).$$

Moreover, for  $A = \prod_{j=1}^n A_j$ ,

$$1_A = 1_{A_1} \otimes \dots \otimes 1_{A_n},$$

thus

$$\begin{aligned} & P_J \left( \prod_{j=1}^n A_j \right) \\ &= \underbrace{\int_E \int_E \dots \int_E}_{n+1} 1_{A_1}(x_1) \cdots 1_{A_n}(x_n) d(P_{t_n - t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1})_{x_0}(x_1) d\mu(x_0) \\ &= \int_E \int_{A_1} \cdots \int_{A_n} d(P_{t_n - t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1})_{x_0}(x_1) d\mu(x_0). \end{aligned}$$

Let  $K \setminus J = \{t'\}$ . Either  $t' < t_1$ , or  $t' > t_n$ , or there is some  $1 \leq j \leq n-1$  for which  $t_j < t' < t_{j+1}$ . Take the case  $t' < t_1$ . Then

$$\pi_{K,J}^{-1} \left( \prod_{j=1}^n A_j \right) = \prod_{k=0}^n B_k,$$

<sup>14</sup><http://individual.utoronto.ca/jordanbell/notes/finitedimdistributions.pdf>, p. 4, Lemma 2.

<sup>15</sup>V. I. Bogachev, *Measure Theory*, volume I, p. 35, Lemma 1.9.4.

where  $B_0 = E$  and  $B_j = A_j$  for  $1 \leq j \leq n$ . Then

$$\begin{aligned}
& (\pi_{K,J})_* P_K \left( \prod_{j=1}^n A_j \right) \\
&= P_K \left( \prod_{k=0}^n B_k \right) \\
&= \int_E \int_E \int_{A_1} \cdots \int_{A_n} d(P_{t_n-t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_1-t'})_{x'}(x_1) d(P_{t'})_{x_0}(x') d\mu(x_0) \\
&= \int_E \int_E \int_{A_1} f(x_1) d(P_{t_1-t'})_{x'}(x_1) d(P_{t'})_{x_0}(x') d\mu(x_0),
\end{aligned}$$

for

$$f(x_1) = \int_{A_2} \cdots \int_{A_n} d(P_{t_n-t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_2-t_1})_{x_1}(x_2).$$

By (1) and because  $(P_t)_{t \in I}$  is a Markov semigroup,

$$\begin{aligned}
& \int_E \int_{A_1} f(x_1) d(P_{t_1-t'})_{x'}(x_1) d(P_{t'})_{x_0}(x') \\
&= \int_E \int_E f(x_1) 1_{A_1}(x_1) d(P_{t_1-t'})_{x'}(x_1) d(P_{t'})_{x_0}(x') \\
&= \int_E P_{t_1-t'}^*(f 1_{A_1})(x') d(P_{t'})_{x_0}(x') \\
&= P_{t'}^*(P_{t_1-t'}^*(f 1_{A_1}))(x_0) \\
&= P_{t_1}(f 1_{A_1})(x_0) \\
&= \int_E f(x_1) 1_{A_1}(x_1) d(P_{t_1})_{x_0}(x_1) \\
&= \int_{A_1} f(x_1) d(P_{t_1})_{x_0}(x_1) \\
&= \int_{A_1} \int_{A_2} \cdots \int_{A_n} d(P_{t_n-t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_2-t_1})_{x_1}(x_2) d(P_{t_1})_{x_0}(x_1).
\end{aligned}$$

Thus

$$\begin{aligned}
& (\pi_{K,J})_* P_K \left( \prod_{j=1}^n A_j \right) \\
&= \int_E \int_{A_1} \int_{A_2} \cdots \int_{A_n} d(P_{t_n-t_{n-1}})_{x_{n-1}}(x_n) \cdots d(P_{t_2-t_1})_{x_1}(x_2) d(P_{t_1})_{x_0}(x_1) d\mu(x_0) \\
&= P_J \left( \prod_{j=1}^n A_j \right).
\end{aligned}$$

This shows that the claim is true in the case  $t' < t_1$ .  $\square$

Thus, if  $E$  is a Polish space with Borel  $\sigma$ -algebra  $\mathcal{E}$ , let  $I = \mathbb{R}_{\geq 0}$ , let  $(P_t)_{t \in I}$  be a Markov semigroup on  $\mathcal{E}$ , and let  $\mu$  be a probability measure on  $\mathcal{E}$ . The above theorem tells us that  $(P_J)_{\mathcal{K}(I)}$  is a projective family, and then the **Kolmogorov extension theorem**<sup>16</sup> tells us that there is a probability measure<sup>17</sup>  $P^\mu$  on  $\mathcal{E}^I$  such that for any  $J \in \mathcal{K}(I)$ ,  $\pi_{J*} P^\mu = P_J^\mu$ . This implies that there is a stochastic process  $(X_t)_{t \in I}$  whose finite-dimensional distributions are equal to the probability measures  $P_J$  defined in Theorem 4 using the Markov semigroup  $(P_t)_{t \in I}$  and the probability measure  $\mu$ .

---

<sup>16</sup><http://individual.utoronto.ca/jordanbell/notes/finitedimdistributions.pdf>

<sup>17</sup>We write  $P^\mu$  to indicate that this measure involves  $\mu$ ; it also involves the Markov semigroup, which we do not indicate.