Lecture 22. Semigroup approach.

Proof of Theorem 21.4. Let us introduce a measure

$$\nu(C) = \int_X P(s,\, x,\, dy)\; P(t,\, y,\, C) \tag{22.1}$$

(this measure depends on x, t, and s). We have for every $f \in \mathbf{B}$:

$$P^s(P^t f)(x) = \int_X f(z)\; \nu(dz) \tag{22.2}$$

(proof: for simple measurable functions, which are linear combinations of indicators; and then, by the definition of the Lebesgue integral, we approximate every function $f \in \mathbf{B}$ by simple ones, and take the limit). For every $f \in \mathbf{C}$ we have:

$$P^s(P^t f)(x) = \int_X f(z)\; \nu(dz) = P^{s+t} f(x) = \int_X f(z)\; P(s+t,\, x,\, dz), \tag{22.3}$$

and from this we get $\nu = P(s+t,\, x,\, \bullet)$: the Chapman – Kolmogorov equation.

Let us return to diffusion processes.

The equation (21.7) (with $c(\boldsymbol{x}) \equiv 0$) for the transition density of a diffusion process, as well as the equation (21.3) for $u(t,\, \boldsymbol{x}) = E_{\boldsymbol{x}} f(\boldsymbol{\xi}_t)$, are quite similar to the equations (12.1) and $\dfrac{d}{dt} P^t f(x) = A P^t f(x)$ for continuous-time Markov chains; and they are called, as those for continuous-time Markov chains, the *backward Kolmogorov equation*.

One may wonder what was so interesting in writing these equations for diffusion processes: when we introduced diffusion processes, we *started* with equation (21.3). The fact is that Kolmogorov introduced in the 1930's diffusion processes differently, and deducing from his definition the equations (21.3), (21.7) was not trivial. We don't need to follow Kolmogorov's way, because since the 1930's there was much progress in the theory of parabolic partial differential equations (when Kolmogorov deduced these equations, the existence theorems that we cited did not exist yet). What about the *forward* Kolmogorov equations?

For the differential operator L defined by

$$L f(\boldsymbol{x}) = \tfrac{1}{2} \sum_{i,\, j=1}^{d} a_{ij}(\boldsymbol{x})\, \frac{\partial^2 f}{\partial x_i \partial x_j}(\boldsymbol{x}) + \sum_{i=1}^{d} b_i(\boldsymbol{x})\, \frac{\partial f}{\partial x_i}(\boldsymbol{x}) + c(\boldsymbol{x})\, f(\boldsymbol{x}) \tag{22.4}$$

(I did not write this formula in the previous lecture note: the operator L was introduced as applied to the function $u(t,\, \boldsymbol{x})$ in its second argument, in the differential equation (21.3)) we define the *adjoint* differential operator L^* by

$$L^* g(\boldsymbol{x}) = \tfrac{1}{2} \sum_{i,\, j=1}^{d} \frac{\partial^2}{\partial x_i \partial x_j} \big(a_{ij}(\boldsymbol{x})\, g(\boldsymbol{x}) \big) - \sum_{i=1}^{d} \frac{\partial}{\partial x_i} \big(b_i(\boldsymbol{x})\, g(\boldsymbol{x}) \big) + c(\boldsymbol{x})\, g(\boldsymbol{x}). \tag{22.5}$$

What is it done for? If in the integral $\int_{\mathbb{R}^d} g(\boldsymbol{x}) \cdot Lf(\boldsymbol{x})\, d\boldsymbol{x}$ we take every integral containing the second derivative by part twice, and the one containing the first derivative once, we get $\int_{\mathbb{R}^d} f(\boldsymbol{x}) \cdot L^* g(\boldsymbol{x})\, d\boldsymbol{x}$, plus some terms outside the integral (at infinity); under some not very restrictive conditions these extra terms vanish.

If not only the conditions imposed on the coefficients $a_{ij}(\boldsymbol{x})$, $b_i(\boldsymbol{x})$, $c(\boldsymbol{x})$ in Theorems 21.2, 21.3 are satisfied, but also the first and the second derivatives of the coefficients $a_{ij}(\boldsymbol{x})$ and the first derivatives of $b_i(\boldsymbol{x})$ are bounded and Hölder-continuous, the operator L^* can be written as

$$L^* g(\boldsymbol{x}) = \tfrac{1}{2} \sum_{i,\, j=1}^{d} a_{ij}(\boldsymbol{x})\, \frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{x}) + \sum_{i=1}^{d} \tilde{b}_i(\boldsymbol{x})\, \frac{\partial g}{\partial x_i}(\boldsymbol{x}) + \tilde{c}(\boldsymbol{x})\, g(\boldsymbol{x}) \tag{22.6}$$

with the coefficients a_{ij}, $\tilde{b}_i$, and $\tilde{c}$ satisfying the conditions of Theorems 21.2, 21.3 (this is why I did not consider from the very beginning only equations with the coefficient $c(\boldsymbol{x}) \equiv 0$, i. e., equations corresponding to diffusion processes: in order to be able to consider the equation with the adjoint operator L^*); and the fundamental solution $q(t,\, \boldsymbol{x},\, \boldsymbol{y})$ of the equation $\dfrac{\partial v}{\partial t}(t,\, \boldsymbol{x}) = L^* v(t,\, \bullet)(\boldsymbol{x})$ exists, satisfying the equation $\dfrac{\partial q}{\partial t}(t,\, \boldsymbol{x},\, \boldsymbol{y}) = L^* q(t,\, \bullet,\, \boldsymbol{y})(\boldsymbol{x})$.

It turns out (a fact from the theory of parabolic equations) that $q(t,\, \boldsymbol{x},\, \boldsymbol{y}) = p(t,\, \boldsymbol{y},\, \boldsymbol{x})$: the fundamental solutions of equations $\partial u/\partial t = Lu$ and $\partial v/\partial t = L^* v$ are obtained from one another by interchanging their spatial arguments. The proof is not difficult, involving, of course, integration by parts; but it would be out of place here.

It follows from this that the transition density $p(t,\, \boldsymbol{y},\, \boldsymbol{x})$ of a diffusion process $\boldsymbol{\xi}_t$ satisfies, for every fixed x, the equation

$$\frac{\partial p}{\partial t}(t,\, \boldsymbol{x},\, \boldsymbol{y}) = L^* p(t,\, \boldsymbol{x},\, \bullet); \tag{22.7}$$

also the distribution density of the random variable $\boldsymbol{\xi}_t$ evaluated under the assumption that the process starts from a random point $\boldsymbol{\xi}_0$ with distribution μ:

$$v(t,\, \boldsymbol{y}) = \int_{\mathbb{R}^d} \mu(d\boldsymbol{x})\, p(t,\, \boldsymbol{x},\, \boldsymbol{y}) \tag{22.8}$$

satisfies the same equation:

$$\frac{\partial v}{\partial t}(t,\, \boldsymbol{y}) = L^* v(t,\, \bullet)(\boldsymbol{y}). \tag{22.9}$$

Equations (22.8), (22.9) (in fact, it is the same equation, but solved with different initial conditions) are called *the forward Kolmogorov equation.*

As we see, an approach to Markov processes is possible based on linear operators in Banach spaces (a Banach space is, as you know, a complete normed space). There is a whole theory about one-parametric semigroups of linear operators in a Banach space and how they are characterized by their *infinitesimal operators*, created by Kôsaku Yosida in Japan and Einar Hille. You can look up this theory (which can be formulated as one big *theorem*) in the book *Markov Processes* by E.B.Dynkin, 1965, Chapter 1, Section 4.

I am not going to formulate the whole Hille – Yosida theory as one theorem, but rather as a number of small ones. We'll start with the easier part of the theory having to do with the properties of the infinitesimal operator and uniqueness of a semigroup corresponding to a given infinitesimal operator; then we consider examples (in the next lecture) and the more complicated part about existence of a semigroup corresponding to a given operator as its infinitesimal operator.

Let $\mathbf{E}$ be a Banach space (some results of the first part hold for normed spaces without assuming their completeness; but what should we consider a non-complete normed space for?). Let P^t, $t \geq 0$, be a family of bounded linear operators on $\mathbf{E}$ with the norms $\|P^t\| \leq 1$ (the norm of a linear operator P^t is, by definition, $\|P^t\| = \sup\{\|P^t f\|\colon\ f \in \mathbf{E},\ \|f\| \leq 1\}$, so that $\|P^t\| \leq 1$ means no more, no less than $\|P^t f\| \leq \|f\|$ for all $f \in \mathbf{E}$: the operators P^t don't increase the norm); we assume that $P^0 = I$ (the identity operator) and that $P^{s+t} = P^t P^s$ for all $s,\ t \geq 0$.

We define the subspace $\mathbf{E}_0 \subseteq \mathbf{E}$ as the set on which the semigroup P^t is right-continuous at $t = 0$:

$$\mathbf{E}_0 = \{f \in \mathbf{E}\colon\ P^h f \to f\ (h \to 0^+)\} \tag{22.10}$$

(let me remind you that, in a normed space, $P^h f \to f$ means that $\lim_{h\to 0^+} \|P^h f - f\| = 0$). Clearly $\mathbf{E}_0$ is a *linear* subspace of $\mathbf{E}$.

Theorem 22.1. *The space* $\mathbf{E}_0$ *is a closed subset of* $\mathbf{E}$.

Proof (not given in the lecture). We need to prove that if $f_n \in \mathbf{E}_0$, $\lim_{n\to\infty} f_n = f$, then $f \in \mathbf{E}_0$. And this means that for every positive ε there exists a positive h_0 such that $\|P^h f - f\| < \varepsilon$ for $h \leq h_0$. Let n be such that $\|f_n - f\| < \varepsilon/3$. We have $f_n \in \mathbf{E}_0$, so there exists an $h_0 > 0$ such that $\|P^h f_n - f_n\| < \varepsilon/3$ for $h \leq h_0$. Then we have: $\|P^h f - f\| \leq \|P^h(f - f_n)\| + \|P^h f_n - f_n\| + \|f_n - f\| \leq \|f - f_n\| + \|P^h f_n - f_n\| + \|f_n - f\| < \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon$.

The space $\mathbf{E}_0$, being a closed subspace of a complete space $\mathbf{E}$, is certainly also complete (i. e. a Banach space).

Theorem 22.2. *The operators* P^t *take the space* $\mathbf{E}_0$ *into itself*; *and for* $f \in \mathbf{E}_0$ *the function* $P^t f$, $t \geq 0$, *is uniformly continuous.*

Proof. Let $f \in \mathbf{E}_0$. Let us take $s,\ t \in [0,\ \infty)$ and estimate the norm $\|P^s f - P^t f\|$. Without restriction of generality we can assume that $t < s$. We have:

$$\|P^s f - P^t f\| = \|P^t(P^{s-t} f - f)\| \leq \|P^{s-t} f - f\| \qquad (t - s \to 0). \tag{22.11}$$

Here is the uniform continuity, and also that $P^t f \in \mathbf{E}_0$.

Now let us define the infinitesimal operator A of the semigroup P^t (sometimes other terms are used here: *infinitesimal generator*, or just *generator*; but some authors use the words *infinitesimal operator* and *generator* with (slightly) different meaning, so I'll stick to the term *infinitesimal operator*). The operator A will be a linear operator, but not necessarily a bounded one, and not necessarily defined on the whole space: we'll denote its domain of definition as D_A.

By definition,

$$D_A = \{f \in \mathbf{E}\colon \text{ there exists } \lim_{h\to 0^+} h^{-1}(P^h f - f)\}; \tag{22.12}$$

and for $f \in D_A$

$$Af = \lim_{h \to 0^+} h^{-1}(P^h f - f). \tag{22.13}$$

In other words, Af is the right-hand derivative (in the sense of norm convergence) of the function $P^t f$ at $t = 0$.

Clearly D_A is a linear subspace of $\mathbf{E}$ (not necessarily a closed one), and A is a linear operator: if f_1, $f_2 \in D_A$, c_1 and c_2 are two scalars (numbers), then $c_1 f_1 + c_2 f_2 \in D_A$, and $A(c_1 f_1 + c_2 f_2) = c_1 A f_1 + c_2 f_2$.

Theorem 22.3. $D_A \subseteq \mathbf{E}_0$.

Proof. Nothing to prove, in fact: from (right) differentiability (right) continuity follows.

Theorem 22.4. *For* $f \in D_A$ *we have*: $Af \in \mathbf{E}_0$ (in other words, or rather formulas: $AD_A \subseteq \mathbf{E}_0$).

Proof. By Theorem 22.3 $f \in \mathbf{E}_0$; by Theorem 22.2 $P^h f \in \mathbf{E}_0$, also $h^{-1}(P^h f - f) \in \mathbf{E}_0$; and by Theorem 22.1 $Af = \lim_{h \to 0^+} h^{-1}(P^h f - f) \in \mathbf{E}_0$.

Theorem 22.5. *If* $f \in D_A$, *then* $P^t f \in D_A$ *for every* $t \geq 0$, *and*

$$\frac{d}{dt} P^t f = P^t A f = A P^t f. \tag{22.14}$$

Proof. We have:

$$\|h^{-1}(P^h f - f) - Af\| \to 0 \qquad (h \to 0^+), \tag{22.15}$$

$$\begin{aligned} \|h^{-1}(P^h P^t f - P^t f) - P^t A f\| &= \|h^{-1}(P^{t+h} f - P^t f) - P^t A f\| \\ &= \|P^t[h^{-1}(P^h f - f) - Af]\| \leq \|h^{-1}(P^h f - f) - Af\| \to 0 \quad (h \to 0^+), \end{aligned} \tag{22.16}$$

which means two things: that $P^t f \in D_A$ with $AP^t f = P^t A f$; and that the *right-hand derivative* $\dfrac{d^+}{dt} P^t f = P^t A f$. Since $Af \in \mathbf{E}_0$ (Theorem 22.4), the function $P^t A f$ is (uniformly) continuous in t. And now we use the following theorem, which I am giving without proof (it is pure analysis, and has even nothing to do with the theory of semigroups):

Theorem 22.6 (here the numbering becomes different from that in the lecture). *Let* $f(t)$ *be a continuous function with values in a Banach space* $\mathbf{E}$; *suppose its right-hand derivative* $\dfrac{d^+}{dt} f(t)$ *exists everywhere and is continuous in* t. *Then there exists the two-sided derivative* (equal, of course, to the right-hand on).

Possible **proof**: First of all you prove this theorem for real-valued functions. Then either you carry the proof over to a Banach space; or you apply this real-variable theorem to the function $l\big(f(t)\big)$, where l is an arbitrary bounded linear functional on the space $\mathbf{E}$.

Now, for $f \in \mathbf{E}_0$ we can define the following family of bounded linear operators $R_\lambda \colon \mathbf{E}_0 \mapsto \mathbf{E}_0$ (forming the Laplace transform of the semigroup P^t):

$$R_\lambda f = \int_0^\infty e^{-\lambda t} P^t f \, dt, \tag{22.17}$$

where $\lambda > 0$ (we can also define the operators R_λ for complex λ with $\mathrm{Re}\lambda > 0$, but we don't need it now).

The integral in (22.17) is understood as an improper Riemann integral in the Banach space. A *proper* Riemann integral is defined in the same way as the integral for real-valued functions; and in the same way it is proved that every continuous function is Riemann integrable (the standard proof: the difference of the Riemann sums corresponding to two different partitions of the interval of integration converges to 0 as both partitions become infinitely small, and the existence of the limit follows from the Cauchy principle; the same works for a *complete* space $\mathbf{E}$). The improper integral $\int_0^\infty f(t)\,dt$ converges if (but not *only* if) $\int_0^\infty \|f(t)\|\,dt < \infty$.

It is clear that $\|R_\lambda\| \le 1/\lambda$.

Theorem 22.7. *Let* $f \in \mathbf{E}_0$. *Then* $R_\lambda f \in D_A$, *and* $AR_\lambda f = \lambda R_\lambda f - f$.

Proof. We have (skipping some details that are pretty easy to work out):

$$\begin{aligned} P^h R_\lambda f &= P^h \int_0^\infty e^{-\lambda t} P^t f\,dt = \int_0^\infty P^h(e^{-\lambda t}P^t f)\,dt \\ &= \int_0^\infty e^{-\lambda t} P^{t+h} f\,dt = \int_h^\infty e^{-\lambda(s-h)} P^s f\,ds, \end{aligned} \tag{22.18}$$

$$h^{-1}(P^h R_\lambda f - R_\lambda f) = h^{-1}(e^{\lambda h} - 1)\int_0^\infty e^{-\lambda t}P^t f\,dt - h^{-1}\int_0^h e^{-\lambda(t-h)}P^t f\,dt, \tag{22.19}$$

$$AR_\lambda f = \lim_{h\to 0+} h^{-1}(P^h R_\lambda f - R_\lambda f) = \lambda R_\lambda f - f. \tag{22.20}$$

The formula $AR_\lambda f = \lambda R_\lambda f - f$ can be rewritten as

$$(\lambda I - A)R_\lambda f = f, \tag{22.21}$$

or

$$(\lambda I - A)R_\lambda = I. \tag{22.22}$$

So R_λ is a one-to-one operator, it is "neutralized" by applying to $R_\lambda f$ the operator $\lambda I - A$.

Theorem 22.8. *Let* $F \in D_A$. *Then* $R_\lambda AF = \lambda R_\lambda F - F$.

This can be rewritten as

$$R_\lambda(\lambda I - A)F = F, \tag{22.23}$$

$$R_\lambda(\lambda I - A) = I, \tag{22.24}$$

if we consider the identity operator on the set D_A only. That is, $\lambda I - A$ is a one-to one operator, and

$$R_\lambda = (\lambda I - A)^{-1}. \tag{22.25}$$

The domain D_A of definition of the operator A, which is the same as the domain of the operator $\lambda I - A$, is the same as the range of the operator R_λ (which is the same for every

$\lambda > 0$), and the domain of definition of the operator R_λ, which is the space $\mathbf{E}_0$, is the same as the range of the operator $\lambda I - A$ for every positive λ.

The family of operators $R\lambda$ is called the *resolvent* of the linear operator A.

Proof of the theorem. We have:

$$R_\lambda AF = \int_0^\infty e^{-\lambda t} P^t AF \, dt. \tag{22.26}$$

By Theorem 22.5, $P^t AF = \dfrac{d}{dt} P^t F$; integrating by parts, we get:

$$\begin{aligned} R_\lambda AF = \int_0^\infty e^{-\lambda t} \frac{d}{dt} P^t F \, dt &= [e^{-\lambda t} \cdot P^t F]_0^\infty - \int_0^\infty P^t F \, d(e^{-\lambda t}) \\ &= - P^0 F + \lambda \int_0^\infty e^{-\lambda t} P^t F \, dt, \end{aligned} \tag{22.27}$$

which is precisely the statement of our theorem.

Theorem 22.9. *The domain D_A is dense in the space $\mathbf{E}_0$.*

Proof. According to the previous theorem (or two previous theorems?), this domain is the same as the common range of the operators R_λ, $\lambda > 0$. So it's enough to prove that

$$\lim_{\lambda \to \infty} \lambda R_\lambda f = f \tag{22.28}$$

for every $f \in \mathbf{E}_0$. We have:

$$\lambda R_\lambda f = \int_0^\infty \lambda e^{-\lambda t} P^t f \, dt, \tag{22.29}$$

$$\|\lambda R_\lambda f - f\| = \left\| \int_0^\infty \lambda e^{-\lambda t} (P^t f - f) \, dt \right\| \le \int_0^{h_0} \lambda e^{-\lambda t} \|P^t f - f\| \, dt + \int_{h_0}^\infty \lambda e^{-\lambda t} \|P^t f - f\| \, dt. \tag{22.30}$$

The first integral is estimated using the continuity of $P^t f$ at $t = 0$, the second is not greater than $e^{-\lambda h_0} \cdot 2\|f\|$, which goes to 0 as $\lambda \to \infty$.

Theorem 22.10. *The operators P^t on the space $\mathbf{E}_0$ are determined uniquely by the infinitesimal operator A.*

This is because the operators R_λ, $\lambda > 0$, are determined uniquely (as $(\lambda I - A)^{-1}$), and the Laplace transform determines a continuous function uniquely.

We know this theorem for real-valued functions (or we don't, but it is definitely not a theorem in probability theory or the theory of stochastic processes); to prove it for functions with values in a Banach space we either repeat the proof that we got in the real case, or we apply the real-valued functions theorem to the real-valued function $l\big(f(t)\big)$, where l is an arbitrary bounded linear functional on the Banach space.

In the next lecture we'll consider examples.