Stochastic Processes

February 9, 2016

Contents

1 Basic Principles

Definition 1.0.1. Let $T \subset \mathbb{R}$, and $(\Omega, \mathcal{F}, P)$ a probability space. We call a family $\mathcal{X} = (X_t)_{t \in T}$ of S-valued random variables a *stochastic process*. $\pi_t(\mathcal{X}) = X_t$ is the *projection on the t-th coordinate*.

$\mathcal{X}$ is a random variable on $(S^T, \mathcal{S}^T)$.

Lemma 1.0.1. *$\mathcal{X}$ is $\mathcal{F} - \mathcal{S}^T$-measurable iff for all $t \in T$: X_t is $\mathcal{F} - \mathcal{S}$-measurable.*

Definition 1.0.2. The law of the stochastic process $\mathcal{X}$ is the law of $\mathcal{X}$ on $(S^T, \mathcal{S}^T)$.

For a finite vector $t = (t_1, ..., t_n) \in T^n$ let $\pi^t = (\pi_{t_1}, ..., \pi_{t_n})$. This induces a probability measure $P^t = P \circ (\pi^t)^{-1}$ on $(S^n, \mathcal{S}^n)$. We call $\{P^t : t \in T^n, n \in \mathbb{N}\}$ the set of finite dimensional distributions (fdds) associated with P.

Questions:

1. Do the fdds uniquely determine P?
2. Given a family of fdds, $(P^t)_t$, is there a measure P on $(S^T, \mathcal{S}^T)$ such that $(P^t)_t$ is associated with P?

Answers:

1. Yes: Let $\mathfrak{a} = \bigcup_{f \subset_{\text{fin}} T} (\pi^t)^{-1}(\mathcal{S}^{|t|})$. It's an $\cap$-stable algebra of *cylinder events*. If two measures agree on a $\cap$-stable generator $\mathfrak{a}$ then they agree on $\sigma(\mathfrak{a})$. Hence the fdds uniquely determine P, since $\mathcal{S}^T$ is in fact generated by $\mathfrak{a}$.
2. One necessary condition is *compatibility* of $(P^t)_{|t|<\infty}$: $\forall t_1 \subset t_2 \subset T,\ A \in \mathcal{S}^{|t_1|}$: $P^{|t_2|}(A \times S^{|t_2 \setminus t_1|}) = P^{|t_1|}(A)$. For reasonable spaces, this is already sufficient.

Theorem 1.0.2 (Kolmogorov's Extension Theorem)**.** *If S is a Polish space, then for any compatible family of fdds there exists a measure P on $(S^T, \mathcal{S}^T)$ associated with the fdds.*

Proof. Durrett, Theorems 2.1.14 and 2.1.15
See also: Dudley, *Real analysis and probability*, Theorem 12.1.2 $\square$

1.1 Preview

Brownian Motion: A stochastic process $(X_t)_{t \in [0,\infty)}$ such that

1. $\forall t_1 < t_2$: $X_{t_2} - X_{t_1} \sim \mathcal{N}(0, t_2 - t_1)$,
2. $\forall t_1 < t_2 < ... < t_n$: $X_{t_n} - X_{t_{n-1}}, ..., X_{t_2} - X_{t_1}$ are independent,
3. the map $t \mapsto X_t$ is continuous almost surely,

is called *Brownian Motion*.

We will show *Donsker's invariance principle*: Let $(\xi_n)_n$ be iid with $E(\xi_1) = 0, \mathrm{Var}(\xi_n) = 1$ and $S_n = \sum_{k=1}^n \xi_i$. Then $\frac{S_{\lfloor nt \rfloor}}{\sqrt{n}} \xrightarrow{d} (X_t)_{t \in [0,1]}$ where X_t is Brownian motion.

Markov Chains: A family $(X_t)_t$ such that for $s < t$, $E(X_t \mid \mathcal{F}_s) = E(X_t \mid X_s)$ is called *Markov Chain*, where $\mathcal{F}_s = \sigma(X_u : u \leq s)$

1.2 Continuous modifications of stochastic processes

So far we have found a probability measure P on $(S^T, \mathcal{S}^T)$. However, the event $\{\sup_{t\in T} X_t < c\}$ is not an element of $\mathcal{S}^T$. Neither is, in general, $\{\omega : t \mapsto X_t(\omega) \text{ is continuous}\}$.

Definition 1.2.1. Stochastic processes X, Y are *stochastic equivalent* if $\forall t \in T : P(X_t = Y_t) = 1$. We then call Y a *modification* of X.

Remark. If X, Y are stochastic equivalent then they have the same fdds.

Example 1.1. Let $S = [0,1]$, $T = [0,1]$, $\omega \sim \mathcal{U}([0,1])$ and $X_t(\omega) = 1_{\{\omega=t\}}$, $Y_t(\omega) = 0$. X, Y are stochastic equivalent, but $\sup_t X_t = 1 > 0 = \sup_t Y_t$ and $P(X_t \in \mathcal{C}(T)) = 0 \neq 1 = P(Y_t \in \mathcal{C}(T))$.

Suppose for some continuous time process X we have a continuous modification Y, (i.e. Y has almost surely continuous paths). Then Y lives on a bigger probability space $(S^T, \widetilde{\mathcal{S}}^T, \widetilde{P})$ where $\mathcal{C}(T) \in \widetilde{\mathcal{S}}^T$ and $\widetilde{P}(Y \in \mathcal{C}(T)) = 1$. $\widetilde{P}$ is not very handy, so we would rather work on $(\mathcal{C}(T), \mathcal{F}_c)$ where $\mathcal{F}_c = \sigma(\text{cylinder sets of } \mathcal{C}(T))$ and $\widehat{P}(A) := \widetilde{P}(A \cap \mathcal{C}(T))$ for $A \in \mathcal{F}_c$. If T is a bounded interval, then $\rho(x,y) := \sup_{t\in T} |x(t) - y(t)|$ is a metric on $\mathcal{C}(T)$ and it turns out that $\mathcal{B}(\mathcal{C}(T)) = \mathcal{F}_c$.

Question: For which processes is there a continuous modification?

Theorem 1.2.1 (Kolmogorov's continuity theorem)**.** *Let X be a stochastic process on $(\mathbb{R}^{[0,1]}, \mathcal{B}^{[0,1]})$. If*

$$\exists a, b > 0\ \exists c < \infty\ \forall t, t+h \in [0,1] : E(|X_{t+h} - X_t|^a) \leq c\,|h|^{1+b}$$

then there exists a continuous modification Y of X.

Proof. Exercise. □

Theorem 1.2.2. *Let $\epsilon(h), q(h)$ be increasing such that $\sum_{n=1}^\infty \epsilon(2^{-n}) < \infty > \sum_{n=1}^\infty 2^n q(2^{-n})$. If $\forall t, t+h \in [0,1]$: $P(|X_{t+h} - X_t| > \epsilon(h)) \leq q(h)$ then a continuous modification exists.*

Proof. Define $t_{n,r} = \frac{r}{2^n}$ for $r = 0 \ldots 2^n, n \geq 1$ and $X_t^n = X_{t_{n,r}} + (t - t_{n,r})(X_{t_{n,r+1}} - X_{t_{n,r}})$ for $t \in [t_{n,r}, t_{n,r+1}]$. Now

$$\begin{aligned} Z_{n,r} := \max_{t\in[t_{n,r}, t_{n,r+1}]} \left|X_t^{n+1} - X_t^n\right| &= \left|X_{t_{n+1,2r+1}} - \frac{1}{2}(X_{t_{n+1,2r}} - X_{t_{n+1,2r+2}})\right| \\ &\leq \frac{1}{2}\left|X_{t_{n+1,2r+1}} - X_{t_{n+1,2r}}\right| + \frac{1}{2}\left|X_{t_{n+1,2r+2}} - X_{t_{n+1,2r+1}}\right| \end{aligned}$$

Thus $P\left(Z_{n,r} > \epsilon(2^{-n})\right) \leq P\left(\frac{1}{2}\left|X_{t_{n+1,2r+1}} - X_{t_{n+1,2r}}\right| > \epsilon(2^{-n})\right) + P\left(\frac{1}{2}\left|X_{t_{n+1,2r+2}} - X_{t_{n+1,2r+1}}\right| > \epsilon(2^{-n})\right) \leq 2q(2^{-n})$. Now

$$P\left(\sup_{t\in[0,1]} \left|X_t^{n+1} - X_t^n\right| > \epsilon(2^{-n})\right) = P\left(\bigcup_{r=0}^{2^n} \{Z_{n,r} > \epsilon(2^{-n})\}\right) \leq 2^{n+1} q(2^{-n})$$

Since $\sum_n 2^n q(2^{-n}) < \infty$ Borel-Cantelli implies that there is an A such that $P(A) = 1$ and $\forall \omega \in A\ \exists n_0(\omega)\ \forall n \geq n_0(\omega) : \rho(X^n, X^{n+1}) < \epsilon(2^{-n})$. In particular for $m > n \geq n_0 : \rho(X^n, X^m) \leq \sum_{k=n}^\infty \epsilon(2^{-k}) \to 0$.
Thus for $\omega \in A$, $X^n(\omega)$ is a Cauchy sequence and a limiting function $Y(\omega) = \lim_{n\to\infty} X^n(\omega)$ exists.
It remains to show that Y is a modification of X. If $t = t_{n,r}$ we are done. If $t \neq t_{n,r}$ for all n, r then there exists a sequence r_n such that $t_{n,r_n} \to t$ and $0 < t - t_{n,r_n} < 2^{-n}$. Then $P\left(\left|X_{t_{n,r_n}} - X_t\right| > \epsilon(t - t_{n,r_n})\right) \leq q(t - t_{n,r_n}) \leq q(2^{-n})$. Borel-Cantelli tells us that $X_{t_{n,r_n}} \to X_t$ almost surely, and by continuity $Y_{t_{n,r_n}} \to Y_t$ almost surely. Therefore, since $X_{t_{n,r_n}} = Y_{t_{n,r_n}}$, the limiting points are the same: $P(X_t = Y_t) = 1$. □

Example 1.2. Let $\omega \sim \mathcal{U}([0,1])$ and $X_t = 1_{\{t\geq\omega\}}$. Then $E(|X_{t+h} - X_t|^a) = P(|X_{t+h} - X_t| > 0) = h$.

There are other criteria with (weaker) conditions which give weaker regularity properties

Definition 1.2.2. A process X is called *stochastic continuous* if $\forall t \in T : X_{t+h} \overset{h\to 0}{\longrightarrow} X_t$ in probability. It is L^p-*continuous* if $\forall t \in T : X_{t+h} \overset{h\to 0}{\longrightarrow} X_t$ in L^p.

1.3 Processes with stationary independent increments

Definition 1.3.1. A process $(X_t)_{t\in T}$ has stationary independent increments if $\mathcal{L}(X_t - X_s)$ depends only on $t - s$ and $\forall 0 = t_0 < t_1 < \ldots < t_n :\ (X_{t_i} - X_{t_{i-1}})_{i=1,\ldots,n}$ are independent.

Example 1.3 (Poisson process on $[0,\infty)$)**.** There are three different constructions:

- $N_t(\omega)$ as an increasing right-continuous step functions with jumps of size 1. Then $N_t - N_s \sim \mathrm{Poi}(\lambda(t-s))$ and $N_0 = 0$.
- $(\tau_i)_{i\in\mathbb{N}} \sim \mathrm{Exp}(\lambda)$ iid and $N_t = |\{k \geq 1 : \tau_1 + \ldots + \tau_k \leq t\}|$. The stationarity follows from the Markov property of the exponential distribution.
- On any interval $[i, i+1]$ place $\mathrm{Poi}(\lambda)$ number of jump points uniformly distributed over the interval.

2 Brownian Motion

2.1 Multivariate Gaussian distributions

Definition 2.1.1. A vector $X = (X_1, ..., X_n)$ of $\mathbb{R}$-valued random variables has a multivariate Gaussian distribution if $a \cdot X$ is a univariate Gaussian for any $a \in \mathbb{R}^n$.

Remark.

- If $X_1, ..., X_n$ are independent Gaussians then $(X_1, ..., X_n)$ is a multivariate Gaussian.
- $(X_1, ..., X_n)$ being Gaussian is much stronger than all X_j being Gaussian.
- It does not require a density, e.g. (Z, Z) is a multivariate Gaussian, but it lives on $\Delta\mathbb{R} \subset \mathbb{R}^2$ which is a nullset.
- If X is Gaussian then so is XA for any $A \in \mathbb{R}^{n\times m}$
- If X is Gaussian then its distribution is characterized by its mean and covariance matrix $E(X), \Sigma = (\mathrm{Cov}(X_i, X_j))_{i,j}$. This follows from the represantation of its characteristic function.

Lemma 2.1.1. *If X is Gaussian then $X_1, ..., X_n$ are independent iff they are uncorrelated.*

Proof. If $\mathrm{Cov}(X_i, X_j) = \sigma_i\delta_{ij}$ then we can take $Y = (Y_1, ..., Y_n)$ independent Gaussians where $\mathcal{L}(Y) = \mathcal{L}(X)$. Now use the last point in the remark. $\square$

Definition 2.1.2. A *Gaussian process* is a process such that all fdds are multivariate Gaussian.

2.2 Definition of Brownian Motion

Proposition 2.2.1. *For (X_t) the following are equivalent:*

1. *(X_t) has stationary independent increments such that $X_t \sim \mathcal{N}(0, t)$*
2. *(X_t) is a Gaussian process and $E(X_t) = 0$, $Cov(X_s, X_t) = s \wedge t$.*

Proof.

1. $\Rightarrow$ 2.: $\sum a_k X_{t_k} = \sum b_k(X_{t_k} - X_{t_{k-1}})$ for suitable b_k.
2. $\Rightarrow$ 1.: For $s < t : X_t - X_s$ is Gaussian with zero mean and $\mathrm{Var}(X_s, X_t) = EX_t^2 - 2EX_sX_t - EX_s^2 = t - s$. Furthermore for $u < v \leq s < t$: $\mathrm{Cov}(X_v - X_u, X_t - X_s) = v - u - v + u = 0$

$\square$

Definition 2.2.1. *Standard Brownian motion* is a stochastic process satisfying the conditions of the Proposition with almost surely continuous paths.

Theorem 2.2.2. *Standard Brownian motion exists on $(\mathcal{C}(T), \mathcal{B}^T)$ and is unique.*

Proof. There exists a Gaussian process with $\mathrm{Cov}(B_s, B_t) = s \wedge t$ because multivariate Gaussians form a compatible family of fdds and they determine the process uniquely by Kolmogorov's consistency theorem. Now

$$E((X_{t+h} - X_t)^{2k}) = h^k E(Z^{2k}) \leq ch^k$$

For $k = 2$ we satisfy Kolmogorov's continuity theorem with $a = 4, b = 1$. $\square$

Lemma 2.2.3. *If $(B_t)_t$ is standard Brownian motion then so are*

1. *$B_{t+s} - B_s$*
2. *cB_{t/c^2}*
3. *$X_t = tB_{1/t}$*

Proof. $\mathrm{Cov}(X_t, X_s) = st(1/s \wedge 1/t) = s \wedge t$ with continuous paths on $(0, \infty)$. To check continuity in 0, write $\{\omega : \lim_{t\downarrow 0} X_t = 0\} = \bigcap_{m\geq 1} \bigcup_{n\geq 1} \{\omega : |B_t| \leq 1/m \;\forall t \in \mathbb{Q} \cap (0, 1/n)\}$. Now the right-hand side has the same probability as for B_t, and thus so does the left-hand side. $\square$

Some further properties of Brownian motion

- Almost surely the paths are nowhere differentiable
- Quadratic variation: Letting (π_n) be a sequence of partitions of $[0, t]$ with $|\pi_n| \to 0$, we can let $\langle B_t\rangle := \lim_{n\to\infty} \sum_{(s,u)\in\pi_n} (B_u - B_s)^2$ which exists in L^p (and a.s. if $\sum_n |\pi_n| < \infty$) and $\langle B_t\rangle = t$ a.s.
- Almost surely the paths are monotone in no interval.
- The set of local maxima is dense and countable.
- Every local maximum is strict.

2.3 The Markov property

Let B be standard Brownian motion defined on $(\mathcal{C}([0,\infty)), \mathcal{B}^{[0,\infty)})$ where $\mathcal{B}^{[0,\infty)}$ is the smallest σ-algebra such that the projections $\omega \mapsto \omega(t)$ are measurable. Consider the family of measures $\{P^x\}_{x\in\mathbb{R}}$ where P^x is the measure of $x+B$. Write $(X_t)_t$ for Brownian motion starting in x. In particular $P^x(X_0 = x) = 1$.

Proposition 2.3.1. *If Y is a bounded random variable then $x \mapsto E^x Y$ is measurable.*

Proposition 2.3.2 (Monotone class theorem)**.** *Let $\Omega \in \mathcal{P}$ be a π-system and $\mathcal{H}$ a linear space with*

1. *$A \in \mathcal{P} \Rightarrow 1_A \in \mathcal{H}$*
2. *$X_n \in \mathcal{H}$ bounded, $X_n \uparrow X$, then $X \in \mathcal{H}$*

Then $\{X : X$ is bounded and $\sigma(P)$-measurable$\} \subset \mathcal{H}$.

Proof of 2.3.2. From $\mathcal{H}$ linear and 2. it follows that $\mathcal{G} = \{A : 1_A \in \mathcal{H}\} \supset \mathcal{P}$ is a Dynkin system. Since $\mathcal{H}$ contains all simple functions we win. □

Lemma 2.3.3. *Let $f(x,y)$ be a bounded measurable function, X $\mathcal{G}$-measurable and $Y \amalg \mathcal{G}$. Then $E(f(X,Y) \mid \mathcal{G}) = g(X)$ almost surely where $g(x) = Ef(x,Y)$.*

Proof of 2.3.3. Exercise. □

Proof of 2.3.1. Call Y *special* if $Y(\omega) = \prod_{m=1}^n f_m(\omega(t_m))$ for $0 < t_1 < ... < t_n$ and $f_m \in \mathcal{C}_0(\mathbb{R})$. Let $p_t(x,y)$ the density of a $\mathcal{N}(x,t)$ random variable. If Y is special then $E^x Y = E^x \prod_{m=1}^n f_m(X_{t_m}) = E^x \prod_{m=1}^n f_m(x + B_{t_m})$. We show continuity in x by induction on n. For $n = 1$:

$$Ef_1(x + B_{t_1}) = \int f_1(x+y) p_{t_1}(0,y)dy = \int f_1(z) p_{t_1}(x,z)dz$$

which is continuous in x. For $n \geq 2$, using Lemma 2.3.3 and independent increments yields

$$E\left(E\left(\prod_{m=1}^n f_m(x+B_{t_m}) \mid B_{t_i} : 1 \leq i \leq n-1\right)\right) = E\left(\prod_{m=1}^{n-1} f_m(x+B_{t_m}) h(x+B_{t_{n-1}})\right)$$

where $h(u) = Ef_n(u + B_{t_n - t_{n-1}})$. Now extend to $Y \in \mathcal{C}_0(\mathbb{R})$ using the monotone class theorem. □

Definition 2.3.1. $(\mathcal{F}_t)_{t\in T}$ is called a *filtration* if $\mathcal{F}_t$ is a σ-field such that $\mathcal{F}_s \subset \mathcal{F}_t$ for any $s \leq t$. It is called *right-continuous* if $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.

For Brownian motion a natural candidate for a filtration is
$\mathcal{F}_t^0 = \{\text{smallest } \sigma-\text{algebra such that } \omega \mapsto \omega(s) \text{ are measurable for } s \in [0,t]\}$. However, this is not right-continuous. Therefore, define $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s^0$.
Now let $(\theta_s)_{s\in[0,\infty)}$ be the time shift defined by $\theta_s\omega(t) = \omega(s+t)$. Note that $X_t(\omega_s) = X_{t+s}(\omega)$.

Theorem 2.3.4 (Markov property for Brownian motion)**.** *Let Y be a bounded random variable. Then for all $x \in \mathbb{R}$, $s \geq 0$:*

$$E^x(Y \circ \theta_s \mid \mathcal{F}_s) = E^{X_s} Y := E^y Y|_{y=X_s}$$

Proof. It suffices to show that $E(Y \circ \theta_s 1_A) = E^x(E^{X_s} Y 1_A)$ for $s \geq 0$, Y bounded and $A \in \mathcal{F}_s$.

1. Choose first Y special and A finite dimensional. Let $0 < r_1 < ... < r_k < s + h < s + t_1 < s + t_2 < ... < s + t_n$. Let $\phi(y,h) := E^y f_1(X_{t_1-h}) \cdots f_n(X_{t_n-h})$.
2. ϕ is jointly continuous in $(y,h) \in \mathbb{R} \times [0,t_1)$ and $E^x(Y \circ \theta_s 1_A) = E^x(\phi(X_{s+h}, h)1_A)$.
For $k = n = 1$ we have $\phi(y,h) = \int_{\mathbb{R}} dz p_{t_1-h}(y,z) f_1(z)$ and
$$E^x(Y \circ \theta_s 1_A) = E^x(f_1(\omega(t+s))1_{\{\omega(r_1)\in A_1\}}) = \int_{A_1} du p_{r_1}(x,u) \int_{\mathbb{R}} dv p_{s+h-r_1}(u,v) \int_{\mathbb{R}} dz f_1(z) p_{t_1-h}(v,z)$$
$$= E^x(\phi(X_{s+h}, h)1_A)$$
which is what we wanted to show for $n = k = 1$. Now do induction on k, n.
3. Apply the Dynkin lemma to $\mathcal{P} = \left\{\text{finite dimensional subsets of } \mathcal{F}^0_{s+h/2}\right\}$ and $\mathcal{L} = \left\{A \in \mathcal{F}^0_{s+h/2} : 2. \text{ holds}\right\}$
4. Letting $h \downarrow 0$ in 2. we use the following properties
    - The paths of X are right-continuous

- ϕ is jointly continuous
- $\phi(y, 0) = E^x Y$

to get 1. for the special Y's.

5. To go from special Y to general bounded Y we apply the monotone class theorem to $\mathcal{P} = \{\text{finite dimensional sets}\}$, $\mathcal{H} = \{\text{bounded random variables for which 1. holds}\}$.

□

Remark.

- We have used only the *right*-continuity of the paths.
- $p_{t+s}(x, y) = \int dz p_t(x, z) p_s(z, y)$, which is known as the semi-group property.

Proposition 2.3.5. *If Y is a bounded random variable and $x \in \mathbb{R}$ then $E^x(Y \mid \mathcal{F}_s) = E^x(Y \mid \mathcal{F}_s^0)$.*

Proof. First let Y be special. Write $Y(\omega) = Y_1(\omega)(Y_2 \circ \theta_s)(\omega)$ with $Y_1(\omega) = \prod_{m:t_m \leq s} f_m(\omega(t_m))$ and $Y_2(\omega) = \prod_{m:t_m > s} f(\omega(t_m - s))$. Then using Markov property:

$$E^x(Y \mid \mathcal{F}_s) = Y_1 E^x(Y_2 \circ \theta_s \mid \mathcal{F}_s) = Y_1 E^{X_s} Y_2$$

So $E^x(Y \mid \mathcal{F}_s)$ is $\mathcal{F}_s^0$-measurable. Dynkin lemma and Monotone class theorem to win. □

Corollary 2.3.6 (Blumenthal 0-1 law)**.** *If $A \in \mathcal{F}_0$ then $P^x(A) \in \{0, 1\}$ for each $x \in \mathbb{R}$.*

Proof. Let $A \in \mathcal{F}_0$. Then $1_A = E^x(1_A \mid \mathcal{F}_0) = E^x(1_A \mid \mathcal{F}_0^0)$ is constant almost surely. □

Let $\tau_{>0} := \inf\{t : X_t > 0\}$ and $\tau_0 = \inf\{t > 0 : X_t = 0\}$. Then $P^0(\tau_{>0} = 0) = P^0(\tau_0 = 0) = 1$. We know that $P^0(\tau_{>0} \leq t) \geq P^0(X_t > 0) = 1/2$. Now use the Blumenthal 0-1 law.
Furthermore we also get $\tau_{<0} = 0$ P^0-almost surely. Now use continuous paths to conclude that $P^0(\tau_0 = 0) = 1$.

2.4 The Strong Markov Property

Let $(\Omega, \mathcal{F})$ be a measurable space and let $(\mathcal{F}_t)_t$ be a right-continuous filtration.

Definition 2.4.1. A random variable $\tau : \Omega \to [0, \infty]$ is called a stopping time with respect to $\mathcal{F}_t$ if $\{\tau \leq t\} \in \mathcal{F}_t$ for all $t < \infty$.

Lemma 2.4.1. *τ is a stopping time iff $\{\tau < t\} \in \mathcal{F}_t$ for all $t \geq 0$. Note that right-continuity is necessary for this.*

Proposition 2.4.2. *Let G be an open set. Then $\tau_G = \inf\{t : X_t \in G\}$ is a stopping time.*

Proof. $\{\tau < t\} = \bigcup_{\mathbb{Q} \ni s < t} \{X_s \in G\} \in \mathcal{F}_t$. □

Lemma 2.4.3. *If $(\tau_n)_n$ is a sequence of stopping times, then so are $\inf \tau_n$, $\sup \tau_n$, $\liminf \tau_n$, $\limsup \tau_n$.*

Proposition 2.4.4. *If G is a closed set then τ_G is also a stopping time for Brownian motion.*

Proof. For every n let $G_n = \{x : d(x, G) < 1/n\}$ and $\tau_n = \tau_{G_n}$. Clearly $\sup \tau_n \leq \tau_G$. The converse remains as a maybe-not-so-easy exercise. □

Definition 2.4.2. Let τ be a stopping time. Define $\mathcal{F}_\tau := \{A : \forall t : A \cap \{\tau \leq t\} \in \mathcal{F}_t\}$.

- $\mathcal{F}_\tau$ is a σ-algebra.
- τ is $\mathcal{F}_t$-measurable
- If $\tau_n \downarrow \tau$ then $\mathcal{F}_\tau = \bigcap_n \mathcal{F}_{\tau_n}$.
- If $\tau_1 \leq \tau_2$ then $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

Proposition 2.4.5. *If (Z_t) is adapted to $(\mathcal{F}_t)$ and Z_t has right-continuous paths then $Z_\tau 1_{\tau < \infty}$ is $\mathcal{F}_\tau$-measurable.*

Proof.
- First assume that τ takes on only countably many values $t_1, t_2, \ldots$. Since Z is adapted $\{Z_\tau \leq a\} \cap \{\tau < t\} = \bigcup_{k:t_k < t} \{\tau = t_k, Z_{t_k} \leq a\} \in \mathcal{F}_t$
- Now assume $\tau < \infty$. Now we can approximate τ by $\tau_n := \frac{k+1}{2^n}$ if $\frac{k}{2^n} \leq \tau < \frac{k+1}{2^n}$. Now $\{\tau_n \leq t\} = \{\tau < k/2^n\} \in \mathcal{F}_t$ for $\frac{k}{2^n} \leq t < \frac{k+1}{2^n}$. Moreover, $\tau_n \downarrow \tau$. Every Z_{τ_n} is $\mathcal{F}_{\tau_n}$-measurable and therefore Z_τ is $\mathcal{F}_{\tau_n}$-measurable and therefore also $\mathcal{F}_\tau$-measurable because both $Z, (\mathcal{F}_t)_t$ are right-continuous.

- For arbitrary τ, $Z_{\tau\wedge n}$ is $\mathcal{F}_{\tau\wedge n}$-measurable so that $Z_{\tau\wedge n}1_{\{\tau<\infty\}}$ and $Z_\tau 1_{\{\tau<\infty\}}$ are $\mathcal{F}_\tau$-measurable too.

$\square$

Theorem 2.4.6 (Strong Markov property for Brownian motion)**.** *Suppose $Y_s(\omega)$ is bounded and jointly measurable on $[0,\infty)\times\Omega$ and that τ is a stopping time. Then for all $x\in\mathbb{R}$:*

$$E^x(Y_\tau\circ\theta_\tau \mid \mathcal{F}_\tau) = E^{X_\tau}Y_\tau = E^y Y_t|_{t=\tau,y=X_\tau}$$

P^x-almost surely on $\{\tau<\infty\}$. In particular

$$E^x(Y_\tau\circ\theta_\tau 1_{\tau<\infty}) = E^x(E^{X_\tau}(Y_\tau)1_{\tau<\infty})$$

Proof idea.

- If τ takes on countably many values $t_1, t_2, \ldots$ we condition on $\tau = t_k$ and apply the Markov property.
- General τ we can approximate with a sequence $\tau_n\downarrow\tau$ where each τ_n takes on countably many values.
- Use special Y's and generalize via the monotone class theorem.

$\square$

For a first application we can look at the zeros of Brownian motion: $Z(\omega)=\{t:\omega(t)=0\}$. Then

$$E^x\lambda(Z) = E^x\int_0^\infty 1_Z dt = \int_0^\infty P^x(t\in Z)dt = 0$$

Hence $\lambda(Z)=0$ P^x-almost surely.

Proposition 2.4.7. *Z is almost surely perfect, hence uncountable.*

Proof. We show that any point is an accumulation point. Let $a\geq 0, \tau_a = \inf\{t\geq a: X_t=0\}$. Let $Y=1_A$ for $A=\{\omega:\omega(t_n)=0 \text{ for some sequence } t_n\downarrow 0\}$. Then $Y\circ\theta_{\tau_a}=1_{A_a}$ for $A_a=\{\omega:\omega(t_n)=0 \text{ for some sequence } t_n\downarrow\tau_a\}$. Then

$$E^x(Y\circ\theta_{\tau_a}\mid\mathcal{F}_{\tau_a}) = E^{X_{\tau_\alpha}}Y = E^0Y = 1$$

Hence $P^x(A_a)=E^x(E^x(Y\circ\theta_{\tau_a}\mid\mathcal{F}_{\tau_a}))=1$. Hence for all a, τ_a is a limit point in Z from the right. All other elements of Z are accumulation points from the left. $\square$

As another application we can look at

Theorem 2.4.8 (Reflection principle)**.** *Let $M_t := \max_{s\leq t}X_s$, $0<b<a$. Then $P^0(M_t>a, X_t<b) = P^0(X_t>2a-b)$ for all $t\geq 0$.*

Proof. Let $\tau:=\inf\{t: X_t=a\}$. Let $Y_s = 1_{\omega(t-s)>2a-b} - 1_{\omega(t-s)<b}$. We get $E^xY_s = P^x(X_{t-s}>2a-b)-P^x(X_{t-s}<b)$. In particular $E^aY_s=0$. On $\{\tau<t\}$, $0=E^{X_\tau}Y_\tau$ and

$$\begin{aligned}0 = E^aY_s &= E^0\left(E^{X_\tau}Y_\tau 1_{\tau<t}\right) = E^0\left(Y_\tau\circ\theta_\tau 1_{\tau<t}\right) = P^0(X_t>2a-b,\tau<t) - P^0(X_t<b,\tau<t)\\ &= P^0(X_t>2a-b) - P^0(X_t<b, M_t>a)\end{aligned}$$

$\square$

Corollary 2.4.9. *Under P^0, M_t and $|X_t|$ have the same distribution.*

Proof. Using the reflection principle, we get

$$P^0(M_t>a) = P^0(M_t>a, X_t>a) + P^0(M_t>a, X_t<a) = P^0(X_t>a)+P^0(X_t>a) = P^0(|X_t|>a).$$

$\square$

Corollary 2.4.10. *Let $\tau_0 = \inf\{t: X_t=0\}$. Then $P^x(\tau_0<t) = \int\limits_0^t \frac{|x|}{\sqrt{2\pi z^3}}e^{-x^2/2z}dz$*

Proof.

$$\begin{aligned}P^x(\tau_0<t) &= P^{|x|}(\tau_0<t) = P^{|x|}(X_s<0 \text{ for some } s<t) = P^0(X_s>|x| \text{ for some } s<t)\\ &= P^0(M_t>|x|) = 2P^0(X_t>|x|) = 2\int\limits_{|x|}^\infty p_t(0,y)dy\end{aligned}$$

Substituting $y=|x|\sqrt{t/z}$ gives the desired result. $\square$

2.5 The Skorohod Embedding

Definition 2.5.1. $M = (M_t)_t$ is called *submartingale* if for each t:

- M_t is $\mathcal{F}_t$-measurable,
- $M_t \in L^1$ and
- $E(M_t \mid \mathcal{F}_s) \leq M_s$ for all $s < t$

We call M *supermartingale* if $-M$ is a submartingale and martingale if it is a super- as well as a submartingale.

Theorem 2.5.1 (Martingale convergence)**.** *If M is a right-continuous submartingale bounded in L^1 then $M_\infty := \lim_t M_t$ exists and is finite a.s. If M is uniformly integrable, that is $\sup_t E(|M_t|, |M_t| > N) \overset{N\to\infty}{\longrightarrow} 0$, then convergence is also in L^1.*

Remark. If M is bounded in L^2 then M is uniformly integrable.

Theorem 2.5.2 (Stopping time theorem)**.** *Let M be a right-continuous martingale, $\sigma \leq \tau$ a stopping time. If either τ is bounded or M is uniformly integrable then $E(M_\tau \mid \mathcal{F}_\sigma) = M_\sigma$.*

Example 2.1. Brownian motion B as well as $B_t^2 - t$ are martingales.

Proposition 2.5.3. *Let τ be a stopping time such that $E\tau < \infty$. Then $EB_\tau = 0$, $EB_\tau^2 = E\tau$*

Proof.

- $B_{\tau\wedge n}^2 - (\tau \wedge n)$ is a martingale in n by the stopping time theorem.
- $EB_{\tau\wedge n}^2 = E(\tau \wedge n) \leq E\tau < \infty$. Hence $E(B_\tau^2) \leq E\tau < \infty$ by Fatou.
- $B_{\tau\wedge n}$ is also a martingale in n and is uniformly integrable because it's bounded in L^2 by the above. Hence $E(B_{\tau\wedge n}) = 0 = E(B_\tau)$ by stopping time theorem and L^1 convergence from martingale convergence theorem.
- $E(\tau \wedge n) = E(B_{\tau\wedge n}^2) = E\left(E(B_\tau \mid \mathcal{F}_{\tau\wedge n})^2\right) \leq EB_\tau^2$ by Jensen's inequality. Now use monotone convergence.

□

Let Y be an $\mathbb{R}$-valued random variable. We can ask if there is a stopping time τ such that $E\tau < \infty$ and $\mathcal{L}(B_\tau) = \mathcal{L}(Y)$. By the proposition we know that $EY = 0$, $EY^2 < \infty$ are necessary conditions. It turns out that they are already sufficient.

Example 2.2. Let Y take only two values a, b. Then the obvious choice is $\tau = \tau_{\{a,b\}}$. Since $EB_\tau = 0$ and there is only one distribution on $\{a, b\}$ with mean 0 it does the job. This is the only such stopping time with finite mean. If σ is another one then $\sigma \geq \tau$ but $E\sigma = E\tau = EY^2$.

Theorem 2.5.4 (Skorohod embedding)**.** *Let B be standard Brownian motion, Y a random variable with $EY = 0, EY^2 < \infty$. Then there exists a stopping time τ with $E\tau < \infty$ such that $\mathcal{L}(B_\tau) = \mathcal{L}(Y)$.*

Note that given $v \leq u \leq w$ there is a unique distribution on $\{v, w\}$ with mean u.

Proof (Dubin). Consider a sequence of finite subsets of $\mathbb{R}$: $S_0 = \emptyset$, $S_1 = \{0\} = \{E(Y)\}$, $S_2 = \{E(Y \mid Y < 0), E(Y \mid Y > 0)\} =: \{a, b\}$, $S_3 = \{E(Y \mid Y < a), E(Y \mid a \leq Y < 0), E(Y \mid 0 \leq Y < b), E(Y \mid b \leq Y)\}$, and so on.
More formally: Given $S_1, ..., S_n$ we let $\mathcal{F}_n = \sigma\left(\left\{x \leq Y \leq y,\ \text{for } x, y \text{ consecutive points of } T_n := \bigcup_{k\leq n} S_k \cup \{\pm\infty\}\right\}\right)$ and $S_{n+1} = \operatorname{supp}(E(Y \mid \mathcal{F}_n))$ Without loss of generality assume Y takes values in $F = \operatorname{supp}(Y)$. Let $\tau_0 = 0$ and $\tau_n = \min\{t > \tau_{n-1} : B_t \in S_{n+1}\}$.

(1) $\lim_{n\to\infty} E(Y \mid \mathcal{F}_n) = Y$ a.s.

The martingale convergence theorem applied to $M_n := E(Y \mid \mathcal{F}_n)$ gives us an a.s. limit $E(Y \mid \sigma(\bigcup_n \mathcal{F}_n))$. Thus we need to show that Y is measurable with respect to $\sigma(\bigcup_n \mathcal{F}_n)$. It suffices to show that $F \subset \overline{\bigcup_k S_k}$. Suppose $u \in F \setminus \overline{\bigcup_k S_k}$. Choose sequences $(x_n), (y_n)$ of consecutive points in $\bigcup_{k\leq n} S_k \cup \{\pm\infty\}$ such that $x_n \leq u \leq y_n$ for all n. Then $x_n \downarrow, y_n \uparrow$ and $\lim x_n = x < u < y = \lim y_n$. But $S_{n+1} \ni E(Y \mid x_n < Y \leq y_n) \to E(Y \mid x < Y \leq y)$. Hence we have $P(x < Y < y) = 0$, otherwise there would be $S_m \in (x, y)$ for some m sufficiently large. But we assumed $P(x < Y < y) > 0$ as we took $u \in F$.

(2) $E\tau_n < \infty$ for all n by induction on n.

- $E(\tau_0) = 0$
- $E\tau_{n-1} < \infty$ and $\sigma := \inf\{t \mid B_t \in S_{n+1}\}$. By the strong Markov property $E(\tau_n - \tau_{n-1} \mid \mathcal{F}_{\tau_{n-1}}) = E^{B_{\tau_{n-1}}}\sigma$. Hence $E(\tau_n - \tau_{n-1}) = E(E^{B_{\tau_{n-1}}}\sigma)$. Now $\forall u \in S_n\ \exists v, w \in S_{n+1}$ consecutive : $v \leq u \leq w$ hence the hitting time of this subset is finite.

(3) B_{τ_n} and $E(Y \mid \mathcal{F}_n)$ have the same distribution, once more by induction on n.

- $B_{\tau_0} = 0 = E(Y) = E(Y \mid \mathcal{F}_0)$
- $(B_{\tau_n} \mid B_{\tau_{n-1}} = u) \stackrel{\mathcal{L}}{=} (E(Y \mid \mathcal{F}_n) \mid E(Y \mid \mathcal{F}_n) = u)$. Let $v = \sup\{s \in S_{n+1} : s < u\}$, $w = \inf\{s \in S_{n+1} : s > u\}$. On both sides it is the unique distribution concentrated on $\{v, w\}$.

(4) Let's put everything together. Using (2), (3) and Jensen, we get $E\tau_n = EB^2_{\tau_n} = E\left(E(Y \mid \mathcal{F}_n)^2\right) \leq EY^2 < \infty$. So the monotone limit $\tau := \lim_n \tau_n$ exists and is finite a.s. with $E\tau < \infty$. Taking the limit $n \to \infty$ in (3) and combining it with (1) we win.

□

Corollary 2.5.5. *Let $(Y_i)_i$ be iid, $E(Y_i) = 0, E(Y_i^2) = 1$. Let $S_n = \sum_{i=1}^n Y_i$. Then there exist iid stopping times $(\tau_i)_i$ such that $E\tau_i = 1$ and $(S_1, S_2, ...) \stackrel{\mathcal{L}}{=} (B_{\tau_1}, B_{\tau_1+\tau_2}, ...)$.*

Proof. Construct sequences $(B_t^i), (\tau_i)$ such that $B_t^1 = B_t, \tau_1$ such that $B^1_{\tau_1} \stackrel{\mathcal{L}}{=} Y_i$. Then $B_t^2 = B^1_{\tau_1+t} - B_{\tau_1}$ is Brownian motion and is independent of $\mathcal{F}_{\tau_1}$. Choose τ_2 such that $B^2_{\tau_2} \stackrel{\mathcal{L}}{=} Y_2$. If the same construction is used then $\tau_1 \stackrel{\mathcal{L}}{=} \tau_2$. Now τ_1, B_{τ_1} are $\mathcal{F}_{\tau_1}$-measurable, hence $(\tau_1, B^1_{\tau_1})$ and $(\tau_2, B^2_{\tau_2})$ are independent so that τ_1, τ_2 are iid and $(B^1_{\tau_1}, B^2_{\tau_2}) \stackrel{\mathcal{L}}{=} (Y_1, Y_2)$. Then $(S_1, S_2) \stackrel{\mathcal{L}}{=} (B_{\tau_1}, B_{\tau_1+\tau_2})$. Iterate. □

As a consequence $\frac{S_n}{\sqrt{n}} \stackrel{\mathcal{L}}{=} \frac{B_{\tau_1+...+\tau_n}}{\sqrt{n}}$. If we had $S_n/\sqrt{n} \stackrel{\mathcal{L}}{=} B_{\frac{1}{n}(\tau_1+...+\tau_n)}$ then we would get the Central Limit Theorem.

Theorem 2.5.6. *Let (Y_i) as above. There exists a triangular array $\{\tau_{i,n} : 1 \leq i \leq n\}$ of stopping times such that:*

1. $E\tau_{i,n} = 1$
2. $\forall n : \tau_{1,n}, ..., \tau_{n,n}$ *are independent*
3. $\forall n : \left(\frac{S_k}{\sqrt{n}} : 1 \leq k \leq n\right) \stackrel{\mathcal{L}}{=} \left(B_{\frac{\tau_{1,n}+...+\tau_{k,n}}{n}} : 1 \leq k \leq n\right)$

Proof.
- $B_t^n := \sqrt{n} B_{t/n}$ is Brownian motion.
- Apply Corollary 5.5 to B^n to get $\tau_{1,n}, ..., \tau_{n,n}$
- Using Corollary 5.5 once more we get

$$\left(\frac{S_k}{\sqrt{n}} : 1 \leq k \leq n\right) \stackrel{\mathcal{L}}{=} \left(B^n_{\tau_{1,n}+...+\tau_{k,n}}/\sqrt{n} : 1 \leq k \leq n\right) \stackrel{\mathcal{L}}{=} \left(B_{\frac{\tau_{1,n}+...+\tau_{k,n}}{n}} : 1 \leq k \leq n\right).$$

□

This implies the CLT: $S_n/n \stackrel{\mathcal{L}}{=} B_{\tau_{1,n}+...+\tau_{n,n}}/n$. By the LLN: $\frac{1}{n}(\tau_{1,n} + ... + \tau_{n,n}) \xrightarrow{P} 1$ so using continuity of the paths it follows that $\frac{S_n}{\sqrt{n}} \to B_1 \sim \mathcal{N}(0, 1)$.
Given a discrete time process (Y_k) there are two ways to construct a continuous time process (Y_t):

1. Linear interpolation $(Y_t)_{t \in [0,1]}$ is a random function in $(\mathcal{C}[0, 1], \rho)$ where ρ is the sup-metric.
2. We could also work on $\mathcal{D}[0, 1] := \{f : [0, 1] \to \mathbb{R} \mid f$ is right-continuous with limits from the left$\}$ with the sup-metric ρ.

We use the second option.

Theorem 2.5.7 (Donsker's invariance principle)**.** *Let (Y_i) be iid, $EY_i = 0, EY_i^2 = 1, S_n := \sum_{k=1}^n Y_i$ and $Z_t^n := S_{\lfloor nt \rfloor}/\sqrt{n}$, $t \in [0, 1]$. Then $Z^n \xrightarrow{d} B$ in $(\mathcal{D}[0, 1], \rho)$*

Proof. Take $\{\tau_{k,n}\}_{1 \leq k \leq n}$ from the previous theorem and let $T_t^n := \frac{\tau_{1,n}+...+\tau_{\lfloor tn \rfloor,n}}{n}$ and $V_t^n := B_{T_t^n}$ for $t \in [0, 1]$. Then $(V_t^n)_t \stackrel{\mathcal{L}}{=} (Z_t^n)_t$ for all n. We now show that $V_t^n \xrightarrow{P} B_t$. For $\epsilon, \delta > 0$:

$$P\left(\rho(V^n, B) > \epsilon\right) \leq P\left(\sup_t (T_t^n - t) \geq \delta\right) + P\left(\sup_{|s-t| \leq \delta} |B_s - B_t| > \epsilon\right)$$

As $\delta \downarrow 0$ the second term converges to 0. Now let $\tau_1, \tau_2, ...$ be iid with the same distribution as the $\tau_{i,j}$. Then

$$\begin{aligned}\sup_t |T_t^n - t| &\leq \frac{1}{n} + \sup_{0 \leq k \leq n} \left|\frac{\tau_{1,n} + ... + \tau_{k,n} - k}{n}\right| \stackrel{\mathcal{L}}{=} \frac{1}{n} + \sup_{0 \leq k \leq n} \frac{k}{n} \left|\frac{\tau_1 + ... + \tau_k}{k} - 1\right| \\ &\leq \frac{1}{n} + \epsilon \sup_k \left|\frac{\tau_1 + ... + \tau_k}{k} - 1\right| + \sup_{k \geq \epsilon n} \left|\frac{\tau_1 + ... + \tau_k}{k} - 1\right|\end{aligned}$$

where the second term is a.s. bounded and the last term converges to 0 a.s. It follows that $\sup_{t \in [0,1]} |T_t^n - t| \xrightarrow{P} 0$. □

Corollary 2.5.8. *Suppose $\phi : \mathcal{D}[0,1] \to \mathbb{R}$, $P(B \in A) = 1$ for $A \subset \mathcal{D}[0,1]$ and ϕ is continuous on A. Then $\phi(Z^n) \xrightarrow{d} \phi(B)$.*

Example 2.3.

- $\phi(f) = f(1)$ gives us the CLT.
- $\phi(f) = \max_{0\le t\le 1} f(t)$. Then $\frac{\max_{0\le k\le n} S_k}{\sqrt{n}} \xrightarrow{d} \max_{0\le t\le 1} B_t \overset{\mathcal{L}}{=} |B_1|$
- etc.

3 Markov Chains

3.1 Markov Chains with finite state space in discrete time

A homogeneous Markov chain $(X_n)_{n\in\mathbb{N}}$ on a finite state space S is given by an initial distribution μ on S and a stochastic matrix P, i.e. P has dimension $|S| \times |S|$ with entries ≥ 0 and rows summing to 1, if $P(X_0 = x) = \mu(x), P(X_{n+1} = y \mid X_n = x, X_{n-1} = x_{n-1}, ..., X_0 = x_0) = P(X_{n+1} = y \mid X_n = x) = \mathrm{P}_{x,y}$. In particular

$$P(X_n = x) = \sum_{x_0,\dots,x_{n-1}} \mu(x_0)P(X_1 = x_1, ..., X_{n-1} = x_{n-1}, X_n = x) = \sum_{x_0} \mu(x_0)p^n(x_0, x) = \mu\mathrm{P}^n_{\cdot,x}$$

where we write $p^r(x, y)$ for the r-step transition probabilities. Furthermore, by definition, (X_n) has the Markov property.

Example 3.1.

- $\mathrm{P} = \left(\begin{smallmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{smallmatrix}\right)$ for $\alpha, \beta \geq 0$
- $\mathrm{P} = (a_{ij})$ where $a_{ij} = \frac{1}{2}1_{\{|i-j|=1 \mod n\}}$
- etc.

Definition 3.1.1.

- A Markov chain X is called irreducible if for all $x, y \in S$ there is an r such that $p^r(x, y) > 0$.
- For $x \in S$ we write period$(x) := \gcd\{n \in \mathbb{N} : p^n(x, x) > 0\}$.
- A chain is *aperiodic* if $\forall x \in S$: period$(x) = 1$.

If X is an irreducible chain then all $x \in S$ have the same period. Furthermore, if a chain is irreducible and aperiodic then $\exists r \in \mathbb{N}\, \forall x, y \in S : p^r(x, y) > 0$.

Definition 3.1.2. A distribution π on S is called *stationary* for P if $\pi\mathrm{P} = \pi$.

Let $\tau_x = \inf\{n \geq 0 \mid X_n = x\}$ and $\tau_x^+ = \inf\{n \geq 1 \mid X_n = x\}$.

Lemma 3.1.1. *If X is irreducible then $E\tau_x^+ < \infty$.*

Proof. For all x, y there exists r such that $p^r(x, y) \geq \epsilon > 0$. Then we can find $k \in \mathbb{N}, \epsilon > 0$ such that $\forall x, y \in S\ \exists r \leq k : p^r(x, y) \geq \epsilon > 0$. Now $P^x(\tau_y^+ > lk) \leq (1 - \epsilon)P^x(\tau_y^+ > (l - 1)k) \leq (1 - \epsilon)^l$. Then $E\tau_x^+ = \sum_{n=0}^{\infty} P^x(\tau_y^+ > n) \leq \sum_{l=1}^{\infty} kP^x(\tau_y^+ > (l - 1)k) < \infty$. □

Lemma 3.1.2. *Let X be irreducible. Then $\pi(x) = 1/E^x(\tau_x^+)$ is a stationary distribution.*

Proof. Let $\widetilde{\pi}(y) := \sum_{n=0}^{\infty} P^z(X_n = y, \tau_z^+ > n)$ be the expected number of visits to y before returning to z. Now

$$\begin{aligned}\widetilde{\pi}P(y) &= \sum_{x\in S}\sum_{n\geq 0} P^z(X_n = x, \tau_z^+ > n)P_{x,y} = \sum_{x\in S}\sum_{n\geq 0} Z^z(X_n = x, X_{n+1} = y, \tau_z^+ > n) \\ &= \sum_{n\geq 1} P^z(X_n = y, \tau \geq n) = \widetilde{\pi} - P^z(X_0 = y, \tau_z^+ \geq n + 1) + \sum_{n=1}^{\infty} P^z(X_n = y, \tau_z^+ = n) = \widetilde{\pi}(y).\end{aligned}$$

Hence $\pi(x) = \frac{\widetilde{\pi}(x)}{\sum_{y\in S}\widetilde{\pi}(y)} = \frac{\widetilde{\pi}(x)}{E^z\tau_z^+}$ is stationary. $\pi(z) = \frac{1}{E^z\tau_z^+}$ will follow from the uniqueness we prove below. □

We call an $h : S \to \mathbb{R}$ *harmonic* if for all $x \in S$, $h(x) = \sum_y \mathrm{P}_{x,y}h(y)$. We write h as a column vector and express this as $h = \mathrm{P}h$.

Lemma 3.1.3. *If X is irreducible and h is harmonic then h is already constant.*

Proof. Take $x_0 := \mathrm{argmax}_{x\in S}h(x)$ and $M = h(x_0)$. Now $h(x_0) = \sum_{y\in S} \mathrm{P}_{x,y}h(y)$. If $h(y) < h(x_0)$ for any y for which $\mathrm{P}_{x,y} \neq 0$ this is a contradiction. Otherwise iterate. □

Corollary 3.1.4. *If X is irreducible then the stationary distribution π is unique.*

Proof. $\dim(\ker(\mathrm{P}-I))=1$ by the previous lemma. $\square$

Definition 3.1.3. Let μ,ν be probability measures on S. Define $\|\mu-\nu\| := \max_{A\subset S}|\mu(A)-\nu(A)|$

Lemma 3.1.5. $\|\mu-\nu\| = \frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)|$.

Proof. Let $B := \{x\in S : \mu(x)\geq\nu(x)\}$. Then $\mu(A)-\nu(A)\leq\mu(A\cap B)-\nu(A\cap B)$. It follows that $\|\mu-\nu\| = \frac{1}{2}\left(\mu(B)-\nu(B)+\nu(B^c)-\mu(B^c)\right) = \frac{1}{2}\sum_{x\in S}|\mu(x)-\nu(x)|$. $\square$

We can also write it as $\|\mu-\nu\| = \sum_{\mu(x)>\nu(x)}(\mu(x)-\nu(x))$

Theorem 3.1.6 (Convergence theorem)**.** *Let X be irreducible and aperiodic, $\pi\mathrm{P}=\pi$. Then there exist $\alpha\in(0,1), c>0$ such that for all $n\geq 0$:*

$$\max_x\|p_n(x,\cdot)-\pi(\cdot)\|\leq c\alpha^n$$

Proof. From irreducibility and aperiodicity we know there exists $r>0 : \mathrm{P}^r_{x,y}>0$ for all x,y and even a $\beta\in(0,1)$ such that $\forall y\in S : \mathrm{P}^r_{x,y}\geq(1-\beta)\pi(y)$. Define Q by $\mathrm{P}^r=(1-\beta)\Pi+\beta Q$ where Π has rows given by π. If M is a stochastic matrix then $M\Pi=\Pi=\Pi M$. We see by induction on k that

$$\mathrm{P}^{r(k+1)}=\mathrm{P}^{rk}\mathrm{P}^r=(1-\beta^k)\Pi\mathrm{P}^r+(1-\beta)\beta^kQ^k\Pi+\beta^{k+1}Q^{k+1}=(1-\beta^{k+1})\Pi+\beta^{k+1}Q^{k+1}.$$

Hence for $j<r$ we have $\mathrm{P}^{rk+j}-\Pi=\beta^k(Q^k\mathrm{P}^j-\Pi)$. Take the x-row on both sides and sum the absolute values of the entries

$$\|p_{rk+j}(x,\cdot)-\pi(\cdot)\|=\sum_{y\in S}|p_{rk+j}(x,y)-\pi(y)|\leq\beta^k.$$

$\square$

Theorem 3.1.7 (Ergodic theorem for finite state Markov chains)**.** *If X is an irreducible Markov chain, μ a probability measure on S and $f:S\to\mathbb{R}$ then*

$$\frac{1}{n}\sum_{k=1}^n f(X_k)\xrightarrow{a.s.}\sum_{x\in S}\pi(x)f(x)=:E_\pi f.$$

Example 3.2. Taking $f(y)=\delta_x(y)$ the Ergodic theorem tells us that $\frac{1}{n}\sum_{k=1}^n\delta_x(X_k)\xrightarrow{\text{a.s.}}\pi(x)$.

Let $d(n)=\max_x\|p_n(x,\cdot)-\pi(\cdot)\|$ and $\overline{d}(n)=\max_{x,y}\|p_n(x,\cdot)-p_n(y,\cdot)\|$. Then $d(n)\leq\overline{d}(n)\leq 2d(n)$. The second inequality is trivial. For the first,

$$\begin{aligned}\|p_n(x,\cdot)-\pi(\cdot)\| &= \max_A\left|\sum_y\pi(y)\left(p_n(x,A)-p_n(y,A)\right)\right|\leq\sum_y\pi(y)\max_A|p_n(x,A)-p_n(y,A)|\\ &\leq\max_y\max_A|p_n(x,A)-p_n(y,A)|=\max_y\|p_n(x,\cdot)-p_n(y,\cdot)\|.\end{aligned}$$

Now take the max over x.

Lemma 3.1.8. $\overline{d}(n+m)\leq\overline{d}(n)\overline{d}(m)$.

Definition 3.1.4 (Mixing time)**.** We define $t_{\text{mix}}(\epsilon):=\min\{n:d(n)<\epsilon\}$ and $t_{\text{mix}}:=t_{\text{mix}}(1/4)$

Why do we use $1/4$? Consider

$$d(lt_{\text{mix}})\leq\overline{d}(lt_{\text{mix}})\leq\overline{d}(t_{\text{mix}})^l\leq(2d(t_{\text{mix}}))^l\leq 2^{-l}.$$

Furthermore $t_{\text{mix}}(\epsilon)\leq t_{\text{mix}}\lceil\log_2(1/\epsilon)\rceil$.

Lemma 3.1.9. *Let P be a stochastic matrix. Then*

- *If λ is ein eigenvalue for P then $|\lambda|\leq 1$.*
- *If P is irreducible then 1 has a unique eigenfunction.*
- *If P is irreducible and aperiodic then -1 is not an eigenvalue.*

Let P be reversible with respect to π, i.e. $\pi(x)p(x,y)=\pi(y)p(y,x)$. We define $\langle f,g\rangle_\pi:=\sum_{x\in S}f(x)g(x)\pi(x)$.

Lemma 3.1.10. *1. $(\mathbb{R}^S, \langle\cdot,\cdot\rangle_\pi)$ is an inner product space with an orthonormal basis consisting of eigenfunctions f_j corresponding to the real eigenvalues λ_j of* P.

2. $\mathrm{P}^n_{x,y} = \sum_j f_j(x)f_j(y)\pi(y)\lambda_j^n$*, hence* $P^n g = \sum_j \langle g, f_j\rangle_\pi f_j \lambda_j^n$

Proof. Let $A_{x,y} := \sqrt{\frac{\pi(x)}{\pi(y)}}P_{x,y}$. Then A is symmetric, hence has an ONB $\{\phi_j\}$ of eigenfunctions with real eigenvalues λ_j. $\lambda_1 = 1$ has eigenfunction $\phi_1 = \sqrt{\pi}$. Then $A = D_\pi^{1/2}\mathrm{P}D_\pi^{-1/2}$ where $D_\pi = \mathrm{diag}(\pi)$. The $f_j := D_\pi^{-1/2}\phi_j$ are the eigenfunctions of P with eigenvalues λ_j. Indeed, $Pf_j = PD_\pi^{-1/2}\phi_j = D_\pi^{-1/2}A\phi_j = D_\pi^{-1/2}\lambda_j\phi_j = \lambda_j f_j$. Furthermore $\langle f_i, f_j\rangle_\pi = \left\langle D_\pi^{1/2}f_i, D_\pi^{1/2}f_j\right\rangle = \langle\phi_i,\phi_j\rangle = \delta_{ij}$. Hence $\delta_y = \sum_j \langle\delta_y, f_j\rangle_\pi f_j = \sum_j f_j(y)\pi(y)f_j$, whence $\mathrm{P}^n_{x,y} = (\mathrm{P}^n\delta_y)(x) = \sum_j f_j(y)\pi(y)f_j(x)\lambda_j^n$. ☐

Now let's look at the spectrum of P: $1 \geq \lambda_1 \geq ... \geq \lambda_{|S|} \geq -1$ and let $\lambda_* := \max\{|\lambda| : \lambda \neq 1 \text{ is an eigenvalue}\}$. We call $1-\lambda_*$ the *absolute spectral gap* and $t_{\mathrm{rel}} := \frac{1}{1-\lambda_*}$ the *relaxation time.*

Remark. For the lazy chain $\frac{1}{2}(I+\mathrm{P})$ all eigenvalues are ≥ 0.

Note that for any f we get that $P^n f(x) \to E_\pi f$ from the convergence theorem.

Lemma 3.1.11. $Var_\pi(P^n f) \leq \lambda_*^{2n}\, Var_\pi(f)$.

For reversible, irreducible, aperiodic chains one can show that $(t_{\mathrm{rel}}-1)\log(\frac{1}{2\epsilon}) \leq t_{\mathrm{mix}}(\epsilon) \leq t_{\mathrm{rel}}\log(\frac{1}{\epsilon\min_y \pi(y)})$.

Definition 3.1.5. A *coupling* of probability measures μ,ν on S is a pair of random variables (X,Y) on $S\times S$ and joint distribution with correct margins: $P(X=x)=\mu(x),\ P(Y=y)=\nu(y)$.

Lemma 3.1.12. $\|\mu-\nu\| = \min\{P(X\neq Y) : (X,Y)$ *is a coupling of* $\mu,\nu\}$

Proof.
- "$\leq$" $\mu(A)-\nu(A) = P(X\in A)-P(Y\in A) \leq P(X\in A, Y\notin A) \leq P(X\neq Y)$.
- "$\geq$" We construct an optimal coupling: Let $q(x,x) := \mu(x)\wedge\nu(x)$ and $q(x,y)=0$ if either $q(x,x)=\mu(x)$ or $q(y,y)=\nu(y)$ and $q(x,y) = \frac{(\mu(x)-\nu(x))(\nu(y)-\mu(y))}{1-\sum_z q(z,z)}$ otherwise. Then

$$\min\{P(X\neq Y) : (X,Y) \text{ is a coupling of } \mu,\nu\} \leq P(X\neq Y)$$
$$= \sum_x \mu(x) - \left(\sum_{x:\mu(x)>\nu(x)}\nu(x) + \sum_{x:\mu(x)<\nu(x)}\mu(x)\right) = \sum_{x:\mu(x)>\nu(x)}(\mu(x)-\nu(x)) = \|\mu-\nu\|$$

☐

Proof of Lemma 3.1.8. We know that $\max_{x,y}\|p_n(x,\cdot)-p_n(y,\cdot)\| = P(X_n\neq Y_n)$ for an optimal coupling (X_n,Y_n) with respect to p_n and $X_0=x, Y_0=y$. Now $p_{n+m}(x,w) = \sum_z p_n(x,z)p_m(z,w) = E(p_m(X_n,z))$ and similarly $p_{n+m}(y,w) = E(p_m(Y_n,w))$. Then $\frac{1}{2}\sum_w |p_{n+m}(x,w)-p_{n+m}(y,w)| = \frac{1}{2}\sum_w |E(p_m(X_n,w)-p_m(Y_n,w))|$. Hence

$$\|p_{n+m}(x,\cdot)-p_{n+m}(y,\cdot)\| \leq E\left(\frac{1}{2}\sum_w |p_m(X_n,w)-p_m(Y_n,w)|\right) \leq \bar{d}(m)E(1_{\{X_n\neq Y_n\}}) \leq \bar{d}(m)\bar{d}(n)$$

because (X_n,Y_n) was the optimal coupling. ☐

Theorem 3.1.13. t_{mix}*(lazy n-cycle)* $\leq n^2$*, where the transition matrix for the n-cycle is given by* $\mathrm{P}=(a_{ij})$ *where* $a_{ij} = \frac{1}{2}1_{\{|i-j|=1 \mod n\}}$ *and the transition matrix for the lazy n-cycle is* $\frac{1}{2}(I+\mathrm{P})$.

Proof. Use X_n, Y_n coupled lazy walks on the n-cube. Before $\tau = \min\{n : X_n = Y_n\}$ let $P(X_{n+1}\neq X_n, Y_{n+1}=Y_n) = P(X_{n+1}=X_n, Y_{n+1}\neq Y_n) = 1/2$ with equal probabilities to go left and right. Now $D_n := X_n - Y_n$ is a simple symmetric random walk on $\{0,...,n\}$. Then $\tau = \min\{t : D_t\in\{0,n\}\}$. After τ move X_n, Y_n together. Then $k = E^k(D_0) = E^k(D_\tau) = nP^k(D_\tau = n)$. Since $(D_n^2 - n)_n$ is also a martingale we get $k^2 = E^k(D_0^2 - 0) = E^k(D_\tau^2 - \tau) = n^2P^k(D_\tau = n) - E^k(\tau)$. Hence $E^k(\tau) = k(n-k) \leq \frac{n^2}{4}$ for all k. Furthermore $d(t) \leq \bar{d}(t) \leq \max_{x,y}P^{x,y}(X_t\neq Y_t) = \max_k P^k(D_\tau > t) \leq \max_k E^k(\tau)/t \leq \frac{n^2}{4t}$. If $t=n^2$ then $d(n^2)\leq 1/4$, whence the claim follows. ☐

3.1.1 The symmetric group and card shufflings

Let $S = \mathfrak{S}_N$ be the symmetric group of $[N] = \{1,..,n\}$. Shuffling cards is then the process of achieving a uniform distribution on S. The idea now is to choose a random transposition and shuffle in that way. However, this is 2-periodic thanks to the parity of permutations.
Instead, for τ a transposition, let

$$\mu(\mathrm{id}) = \frac{1}{N}, \quad \mu(\tau) = \frac{2}{N^2},$$

We now choose two positions L_n, R_n uniformly from $[N]$ and swap at these positions: If $X_n\in S$ then $P(X_{n+1} = \sigma\circ\sigma' \mid X_n = \sigma') = \mu(\sigma)$ and $P(X_0 = \mathrm{id}) = 1$.

Proposition 3.1.14. *For the chain above, $t_{mix} \leq \frac{2}{3}\pi^2 N^2$*

Proof. Strategy: Choose a *card* $X_n \in [N]$ and *position* $Y_n \in [N]$ independently and uniformly and swap cards at position Y_n and card X_n. Note that this gives us the same chain as above.
Start with two decks. Use the same $(X_n), (Y_n)$ for both decks. Let a_n be the number of cards at the same location in both decks. There are three possibilities for the n-th step:

- If position of card X_n and the card at position Y_n agree in both decks then $a_n = a_{n-1}$.
- If one is different and the other not then $a_n = a_{n-1}$ as well.
- If both card X_n at different position and card at position Y_n are different then $a_n - a_{n-1} \in \{1, 2, 3\}$.

Using the lemma below we get that $d(n) \leq P(X_n \neq Y_n) = P(\tau > n) \leq \frac{1}{n}E(\tau) \leq \frac{\pi^2 N^2}{6n}$. Now set $n = \frac{2}{3}\pi^2 N^2$. □

Lemma 3.1.15. *Let $\tau = \min\{n : a_n = N\}$. Then $E(\tau) \leq \frac{\pi^2}{6}N^2$*

Proof. Decompose $\tau = \tau_1 + \tau_2 + ... + \tau_N$ where τ_i is the first time that $\{a_n = i\}$ after $\{a_n = i-1\}$, that is $\tau_i = \inf\{m : a_{n+m} \geq i, \tau_{i-1} = n\}$. Then $\tau_0 = 0$ and jumps can occur. Given $\{a_n = i\}$ then $N - i$ cards are not aligned. Hence $\tau_{i+1} \mid a_n = i \sim \text{Geo}\left((\frac{N-i}{N})^2\right)$. Then $E(\tau_{i+1} \mid a_n = i) = \left(\frac{N}{N-i}\right)^2$, whence $E(\tau) \leq \sum_{i=0}^{N-1} \frac{N^2}{(N-i)^2} \leq \frac{\pi^2}{6}N^2$ □

3.1.2 Markov Chain Monte Carlo (MCMC)

Suppose you want to sample from a finite yet very complicated distribution.
Idea: Construct a Markov chain with its stationary distribution equal to the distribution we want to sample from.

Example 3.3 (Ising model on a finite graph)**.** Let $G = (V, E)$ be a finite graph and $S = \{-1, 1\}^V$. Let $\sigma \in S$ be a configuration and $\sigma(v)$ the spin at $v \in V$. Let $H(\sigma) = -\sum_{\{v,w\}\in E} \sigma(v)\sigma(w)$. Define the Boltzmann distribution by $\mu(\sigma) = \frac{1}{Z(\beta)} \exp(-\beta H(\sigma))$ where $Z(\beta) = \sum_{\sigma \in S} \exp(-\beta H(\sigma))$. Since S can be large already for small graphs G, it's very hard to compute $Z(\beta)$.
We use Glauber dynamics:

- Choose sites uniformly iid
- Update state there subject to everything else staying fixed

Set

$$\mu(\sigma(w) = 1 \mid \sigma(v), v \neq w) = \frac{\exp(-\beta \sum_{u:\{u,w\}\in E} \sigma(u))}{\exp(-\beta \sum_{u':\{u',w\}\in E} \sigma(u)) + \exp(-\beta \sum_{u:\{u,w\}\in E} -\sigma(u))}$$

The Glauber chain is given by

$$\mathrm{P}_{\sigma,\sigma'} = \frac{1}{V} \sum_{w \in V} \frac{\exp(-\beta\sigma'(w) \sum_{u:\{u,w\}\in E} \sigma'(u))}{\exp(-\beta \sum_{u':\{u',w\}\in E} \sigma'(u')) + \exp(\beta \sum_{u:\{u,w\}\in E} \sigma'(u))} 1_{\{\sigma(v)=\sigma'(v) \text{ for } w\neq v\}}.$$

3.2 Markov chains on countable state spaces

Example 3.4 (Random walk on the lattice $\mathbb{Z}^d$)**.** Let $S = \mathbb{Z}^d$ and $P(X_0 = 0) = 1$. Let $P(X_{n+1} = y \mid X_n = x) = \frac{1}{2d} 1_{\{|x-y|=1\}}$. Then $(X_n)_{n\geq 0}$ is a Markov chain.
Let $\tau^+ = \inf\{n \geq 1 \mid X_n = 0\}$

Definition 3.2.1. A random walk is called *recurrent* if $P(\tau^+ < \infty) = 1$, otherwise it is called *transient*.

Theorem 3.2.1 (Pólya)**.** *The random walk on $\mathbb{Z}^d$ is recurrent iff $d \leq 2$.*

Define Green's function $G(x) = \sum_{n=0}^{\infty} P(X_n = x)$, which is the expected number of visits to x.

Theorem 3.2.2. *$(X_n)_n$ is recurrent iff $G(0) = \infty$.*

Proof. Using the Markov property,

$$P(X_n = 0) = \sum_{i=1}^{n} P(\tau^+ = i) P(X_{n-i} = 0).$$

Consider the generating functions $G_Z := \sum_{n=0}^{\infty} Z^n P(X_n = 0)$, $F_Z := \sum_{n=0}^{\infty} Z^n P(\tau^+ = n)$ for $Z \in [0, 1)$:

$$G_Z = 1 + \sum_{n=1}^{\infty} \sum_{i=1}^{n} Z^{i+(n-i)} P(\tau^+ = i) P(X_{n-i} = 0) = 1 + G_Z F_Z,$$

so that $F_Z = 1 - \frac{1}{G_Z}$, whence

$$P(\tau^+ < \infty) = F_1 = \lim_{Z\uparrow 1} F_Z = 1 - \frac{1}{\lim_{Z\uparrow 1} G_Z} = 1 - \frac{1}{G(0)}.$$

$\square$

Pólya. For $d = 1$:

$$G(0) = \sum_n P(X_n = 0) = \sum_n \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \sum_n \frac{(2n)!}{(n!)^2 2^{2n}} \sim \sum_n \frac{(2n)^{2n} e^{2n}\sqrt{2\pi 2n}}{e^{2n} n^{2n} 2^{2n} 2\pi n} \sim \sum_n \frac{1}{\sqrt{\pi n}}.$$

For $d = 2$: Consider $Y_n = (U_n, V_n) := (A_n + B_n, A_n - B_n)$. The coordinates of Y are independent so that we get

$$P(X_{2n} = 0) = P(Y_{2n} = 0) = P(U_{2n} = 0)P(V_{2n} = 0) \sim \frac{1}{\pi n}.$$

For $d \geq 3$: Let's look at the characteristic function $\phi(k) = E(e^{ik\cdot X})$. This is $(2\pi)^d$-periodic, so we may consider $k \in [-\pi, \pi)^d$. Then

$$\phi(k) = \sum_{x\in\mathbb{Z}^d} e^{ik\cdot x} P(X_1 = x) = \frac{1}{2d}(e^{ik_1} + e^{-ik_1} + ... + e^{ik_d} + e^{-ik_d}) = \frac{1}{d}\sum_{j=1}^d \cos(k_j)$$

Now use the inversion formula: $P(X_n = x) = \int_{[-\pi,\pi)^d} \frac{1}{(2\pi)^d} e^{-ik\cdot x}\phi(k)^n dk$. For $z \in [0, 1)$:

$$G_Z(0) := \sum_{n=0}^{\infty} Z^n P(X_n = 0) = \sum_n \frac{1}{(2\pi)^d}\int Z^n \phi(k)^n dk = \frac{1}{(2\pi)^d}\int \frac{1}{1 - z\phi(k)} dk.$$

Taking $Z \uparrow 1$ we get $G_1(0) < \infty$ iff $\int_{[-\pi,\pi)^d} \frac{1}{1-\phi(k)} dk < \infty$. Use $\frac{2}{\pi^2 d} \leq \sum_{j=1}^d k_j^2 \leq 1 - \phi(k) \leq \frac{1}{2d}\sum_{j=1}^d k_j^2$. Now the integral is finite iff $\int_{[-\pi,\pi)^d} \|k\|_2^{-2} dk < \infty$ $\square$

Let $\alpha := E^1(\tau_0^+) = 1 + E^0(\tau_0^+)$. Then $\alpha = \frac{1}{2} + \frac{1}{2}(E^2(\tau_0^+) + 1)$, whence $\alpha = 1 + \alpha$, i.e. $\alpha = \infty$.

Definition 3.2.2. A state $x \in S$ is *null recurrent* if $P^x(\tau_x^+ < \infty) = 1$ but $E^x(\tau_x^+) = \infty$. It is *positive recurrent* if $E^x(\tau_x^+) < \infty$.

3.3 Markov chains in continuous time

Let S be a countable state space.

Definition 3.3.1 (Continuous Markov chain)**.** Let Ω be the set of S-valued cádlág functions on $[0, \infty)$ and $X_t(\omega) = \omega(t)$ for $t \geq 0$, $\theta_s(\omega)(t) = \omega(s+t)$ for $s, t \geq 0$ and $\mathcal{F}$ the smallest σ-field such that $\omega \mapsto \omega(t)$ is measurable for all $t \geq 0$. Then a continuous time Markov chain is given by

- $\{P^x : x \in S\}$ a family of probability measures on $(\Omega, \mathcal{F})$
- A right-continuous filtration $(\mathcal{F}_t)_{t\geq 0}$ such that $\mathcal{F}_t \subset \mathcal{F}$, X_t is adapted to $\mathcal{F}_t$

such that $P^x(X_0 = x) = 1$ and $E^x(g \circ \theta_s \mid \mathcal{F}_s) = E^{X_s} g$ P^x-a.s. for all bounded measurable g.

Definition 3.3.2 (Transition function)**.** $\{p_t(x, y) : t \geq 0,\, x, y \in S\}$ such that $p_t(x, y) \geq 0$ and $\sum_y p_t(x, y) = 1$ as well as $\lim_{t\downarrow 0} p_t(x, x) = p_0(x, x) = 1$ for all x, y and for which the Chapman-Kolmogoroff equation $p_{s+t}(x, y) = \sum_{z\in S} p_s(x, z)p_t(z, y)$ holds, is called *transition function.*

Given a transition function, we can construct a consistent family of probability measures by

$$P^x(X_{t_1} = x_1, ..., X_{t_n} = x_n) = p_{t_1}(x, x_1)p_{t_2-t_1}(x_1, x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1}, x_n).$$

Note that the Chapman-Kolmogoroff equations are necessary for this family to be consistent. We aim to get an infinitesimal description of p_t:

$$q(x, y) = \frac{dp_t}{dt}(x, y)\,|_{t=0}$$

Definition 3.3.3 (Q-matrix)**.** $\{q(x, y) : x, y \in S\}$ with $q(x, y) \geq 0$ if $x \neq y$ and $\sum_{y\in S} q(x, y) = 0$ for all $x \in S$. Further, set $c(x) := -q(x, x) \geq 0$

Example 3.5 (Poisson process)**.** Let $S = \mathbb{Z}$. After an exponentially distributed waiting time jump up by 1. The Q-matrix is given by $Q = (q_{xy} := -\lambda\delta_{xx} + \lambda\delta_{x,x+1})$

Example 3.6. Take a discrete time Markov chain given by a stochastic matrix P. Define X_t as follows: Take a sequence of iid Exp(1)-distributed random variables, wait at x for one of these, then jump according to P. The Markov property holds for $\mathcal{F}_t := \sigma(X_u : u \le t)$ because the exponential distribution is memoryless.
Let $T_1, T_2, ...$ be iid Exp(1). Let $N_t := \max\left\{k : \sum_{i=1}^k T_i \le t\right\}$.

$$p_t(x,y) := P^x(X_t = y) = \sum_{k=0}^{\infty} P^x(X_t = y, N_t = k) = \sum_{k=0}^{\infty} P^x(X_t = y \mid N_t = k)P^x(N_t = k) = \sum_{k=0}^{\infty} p_k(x,y)e^{-t}\frac{t^k}{k!}$$

hence $P_t = \sum e^{-t}\frac{t^k}{k!}P^k$ and $P_0 = I$.
Claim: $Q = P - I$.

$$\frac{d}{dt}p_t(x,y)\,|_{t=0} = \frac{d}{dt}e^{-t}P^0_{x,y} + \frac{d}{dt}e^{-t}\sum_{k=1}^{\infty}\frac{t^k}{k!}p_k(x,y)\,|_{t=0} = ... = I_{x,y} + P_{x,y}$$

Lemma 3.3.1. *Let S be finite and Q a Q-matrix. Then the transition function $p_t(x,y)$ is given by $P_t = e^{tQ}$.*

Proof.
1. P_t is well-defined due to submultiplicativity of the operator norm.
2. P_t is a stochastic matrix.
3. $P_{s+t} = P_s \cdot P_t$ whence the Chapman-Kolmogoroff equation follows.

□

Example 3.7 (Birth- and death chain)**.** (X_t) a Markov chain on $S = \mathbb{N}_0$. $q(k,k-1) = \rho_k, q(k,k-1) = \lambda_k, q(k,k) = -\rho_k - \lambda_k$, e.g. for $\rho_k = k\rho, \lambda_k = k\lambda$. In particular, $c(k)$ need not be bounded.

Theorem 3.3.2. *Let (X_t) be a Markov chain and $p_t(x,y) := P^x(X_t = y)$. Then*

1. *$\{p_t(x,y) : x,y \in S, t \ge 0\}$ is a transition function*
2. *It determines the measures $\{P^x : x \in S\}$ uniquely*

Proof.
1. We first show $\lim_{t\downarrow 0} p_t(x,x) = 1$. Let $T := \inf\{t > 0 : X_t \ne X_0\} > 0$ P-a.s. for all $x \in S$ because the paths are right-continuous. Since $p_t(x,x) \ge P^x(T > t)$ for all $t > 0$, we get $\lim_{t\downarrow 0} p_t(x,x) = 1$. For Chapman-Kolmogoroff, use the Markov property with $g = 1_{\{X_t = y\}}$:
$$E^x(P^x(X_{s+t} = y \mid \mathcal{F}_s)) = E^x(P^{X_s}(X_t = y)) = E^x(p_t(X_s, y)),$$
so that $p_{s+t}(x,y) = \sum_z p_s(x,z)p_t(z,y)$
2. By the Markov property:
$$P^x(X_{t_1} = x_1, X_{t_2} = x_2, ..., X_{t_n} = x_n) = p_{t_1}(x,x_1)p_{t_2-t_1}(x_1,x_2)\cdots p_{t_n-t_{n-1}}(x_{n-1},x_n)$$
This determines the finite dimensional marginals completely, hence also the process.

□

Definition 3.3.4. A state $x \in S$ is *absorbing* if $p_t(x,x) = 1$ for all t and *instantaneous* if $c(x) = \infty$.

Heuristically, we have the *Kolmogoroff backward equation* $\frac{d}{ds}p_{s+t}(x,y) = \sum_{z\in S}\frac{d}{ds}p_s(x,z)\,|_{s=0}\,p_t(z,y)$ as well as the *Kolmogoroff forward equation* $\frac{d}{dt}p_{s+t}(x,y) = \sum_{z\in S}p_s(x,z)\frac{d}{dt}p_t(z,y)\,|_{t=0}$.

Lemma 3.3.3.
1. *$\forall t \ge 0\, \forall x \in S : p_t(x,x) > 0$*
2. *$\exists t > 0 : p_t(x,x) = 1 \Rightarrow \forall t > 0 : p_t(x,x) = 1$*
3. *$t \mapsto p_t(x,y)$ is uniformly continuous. In particular $|p_t(x,y) - p_s(x,y)| \le 1 - p_{|t-s|}(x,x)$*

Proof.
1. $\lim_{t\downarrow 0} p_t(x,x) = 1$ so the claim is is clear for small t. Now use Chapman-Kolmogoroff.
2. $p_{s+t}(x,x) \le p_s(x,x)p_t(x,x) + (1 - p_s(x,x)) = 1 - p_s(x,x)(1 - p_t(x,x))$ so that $p_{s+t}(x,x) = 1 \Rightarrow p_t(x,x) = 1$, whence $\{t : p_t(x,x) = 1\} \supset [0,\epsilon)$. Use Chapman-Kolmogoroff again.
3. Now write $p_{s+t}(x,y) - p_t(x,y) = p_t(x,y)(p_s(x,x) - 1) + \sum_{z\ne x} p_s(x,z)p_t(z,y) =: T_1 + T_2$. $|T_1| \le 1 - p_s(x,x), |T_2| \le 1 - p_s(x,x)$. Since $T_1 \le 0$, $T_2 \ge 0$, the claim follows.

□

Theorem 3.3.4. *Let $(p_t(x,y))_{x,y\in S,t>0}$ be a transition function. Then*

1. $c(x) = -q(x,x) = -\frac{d}{dt}p_t(x,x)|_{t=0} \in [0,\infty]$ *exists and* $p_t(x,x) \geq e^{-c(x)t}$

2. $c(x) < \infty \Rightarrow \forall y \neq x : q(x,y) = \frac{d}{dt}p_t(x,y)|_{t=0} \in [0,\infty)$ *exists and* $\sum_y q(x,y) \leq 0$

3. $c(x) < \infty$ *and* $\sum_y q(x,y) = 0$ *then* $\forall y : t \mapsto p_t(x,y) \in \mathcal{C}^1[0,\infty)$ *and* $\frac{d}{dt}p_t(x,y) = \sum_z q(x,z)p_t(z,y)$

Proof. $f(t) = -\log p_t(x,x)$ is continuous and subadditive.

1. In particular, using Fekete's lemma, $c(x) = \lim_{t\downarrow 0}\frac{f(t)}{t} = \inf_t \frac{f(t)}{t}$ exists and satisfies $f(t) \leq c(x)t$

2. $1-p_t(x,x) \leq 1-e^{-c(x)t} \leq c(x)t$ so that $\sum_{y:y\neq x} p_t(x,y)/t \leq c(x)$ so that $q(x,y) := \limsup_{t\downarrow 0} p_t(x,y)/t \leq c(x) < \infty$. We get $\forall \delta > 0\, \forall n \in \mathbb{N} : p_{n\delta}(x,y) \geq \sum_{k=0}^{n-1} p_\delta^k(x,x)p_\delta(x,y)p_{(n-k-1)\delta}(y,y)$. Using $p_t(x,x) \geq e^{-c(x)t}$ we get:
$$\frac{p_{n\delta}(x,y)}{n\delta} \geq e^{-c(x)n\delta}\frac{p_\delta(x,y)}{\delta}\inf_{0\leq s\leq n\delta} p_s(y,y).$$
Choose $n\to\infty, \delta \to 0$ so that $n\delta \to t$: Then
$$p_t(x,y)/t \geq q(x,y)e^{-c(x)t}\inf_{0\leq s\leq t} p_s(y,y)$$
so that $\liminf_{t\downarrow 0} p_t(x,y)/t \geq q(x,y)$.

3. We have
$$\frac{1}{s}\left(p_{t+s}(x,y) - p_t(x,y)\right) - \sum_z q(x,z)p_t(z,y) = \sum_z \left(\frac{1}{s}(p_s(x,z) - p_0(x,z)) - q(x,z)\right)p_t(z,y).$$
For any $T \subset S, |T| < \infty$ and $x \in T$ we get
$$\sum_{z\notin T}\left|\frac{p_s(x,z)}{s} - q(x,z)\right|p_t(z,y) \leq \sum_{z\notin T}\frac{p_s(x,z)}{s} + \sum_{z\notin T} q(x,z) = \frac{1}{s}\left(1-\sum_{z\in T}p_s(x,z)\right) - \sum_{z\in T}q(x,z) \xrightarrow{s\downarrow 0} -2\sum_{z\in T}q(x,z)$$
The right-hand side $\to 0$ as $T \uparrow S$ because $c(x) < \infty$ and $\sum_y q(x,y) = 0$, hence the right-derivative is continuous and has the required form. Furthermore any continuous function with continuous right derivative is already differentiable.

□

Let Q be a Q-matrix. Define the transition probability for a discrete time chain; if $c(x) = 0$, take $p(x,y) := 1_{\{x=y\}}$, if $c(x) > 0$, take $p(x,y) := \frac{q(x,y)}{c(x)}1_{x\neq y}$. Note that indeed $p(x,y) \geq 0$ and $\sum_y p(x,y) = 1$. Consider the discrete time Markov chain $(Z_n)_n$ on S with this transition probability. Call it the *embedded discrete time chain.* Let $\tau_0, \tau_1, ...$ be random variables whose conditional distribution (given $Z_0, Z_1, ...$) is $\tau_k \sim \mathrm{Exp}(c(Z_k))$ and $\tau_k = 0$ if $c(Z_k) = \infty$. The finite dimensional marginals are
$$P^x(Z_0 = x, Z_1 = x_1, ..., Z_m = x_m, \tau_0 > t_0, ..., \tau_m > t_m) = p(x,x_1)p(x_1,x_2)\cdots p(x_{m-1},x_m)e^{-c(x_0)t_0}\cdots e^{-c(x_m)t_m}.$$
Let $N(t) = \min\{m : \tau_0 + ... + \tau_m > t\}$. Hence $N(t) = 0$ for an interval of length τ_0, then $N(t) = 1$. Finally $\widetilde{X}_t := Z_{N_t}$ on $\{N(t) < \infty\}$. $\widetilde{X}$ has right-continuous paths, waits at x an $\mathrm{Exp}(c(x))$-distributed time, then jumps to y with probability $p(x,y)$. The only trouble is jumping infinitely many times in a finite time.

Theorem 3.3.5. *The following are equivalent:*

1. *The Kolmogorov Backward equation has a unique solution which is a transition function.*

2. $P(N(t) < \infty) = 1$ *for all* $t \geq 0$.

3. $\sum_n \tau_n = \infty$ *P-a.s.*

4. $\sum_n \frac{1}{c(Z_n)} = \infty$ *P-a.s.*

Corollary 3.3.6. *If either*

1. $\sup_{x\in S} c(x) < \infty$, *or*

2. *the embedded discrete-time chain is irreducible and recurrent,*

then condition 4. holds in the above theorem.

Proof note. 1. $\exists \epsilon > 0 \forall x \in S : c(x) < \frac{1}{\epsilon} \Rightarrow \frac{1}{c(x)} > \epsilon \Rightarrow \sum_n \frac{1}{c(Z_n)} = \infty$.

2. Recurrence $\Rightarrow \exists x : \{Z_n = x\}$ infinitely often, so that $\frac{1}{c(x)}$ occurs infinitely often in the sum.

□

4 Feller Processes

Let S be a compact or locally compact space. Let $C_0(S) := \{f \in C(S) : f \text{ vanishes at } \infty\}$. Note that if S is compact then $C_0(S) = C(S)$. Endowing $C_0(S)$ with $\|f\| := \sup_{x\in S} |f(x)|$ makes $C_0(S)$ a seperable Banach space. Let $\Omega = D[0,\infty) = \{f : [0,\infty) \to S \text{ cádlág}\}$. Let $X_t(\omega) = \omega(t)$, $(\theta_s(\omega))(t) = \omega(s+t)$, $\mathcal{F}$ the σ-field that makes all $t \mapsto \omega(t)$ measurable.

Definition 4.0.1 (Feller process)**.** A *Feller process* $(X_t)_{t\geq 0}$ is given by

- $\{P^x : x \in S\}$ probability measures on $(\Omega, \mathcal{F})$
- A right-continuous filtration $(\mathcal{F}_t)_{t\geq 0}$ to which X is adapted

such that $P^x(X_0 = x) = 1$ and $\forall f \in C_0(S), t \geq 0 : x \mapsto E^x f(X_t) \in C_0(S)$ as well as $\forall g : \Omega \to \mathbb{R}$ bounded and measurable, $x \in S$: $E^x\left(g \circ \theta_s \mid \mathcal{F}_s\right) = E^{X_s} g$ P^x-a.s.

Theorem 4.0.1 (Strong Markov Property)**.** *Let (X_t) be a Feller process and $Y : [0,\infty) \times \Omega \to \mathbb{R}$ be a bounded and jointly measurable. Let τ be a stopping time with respect to $(\mathcal{F}_t)$. Then $\forall x \in S$:*

$$E^x\left(Y_\tau \circ \theta_\tau \mid \mathcal{F}_\tau\right) = E^{X_\tau}(Y_\tau) \quad P^x\text{-a.s. on } \{\tau < \infty\}$$

Remark. There are only three ingredients needed:

- The Markov property
- Right-continuous paths
- $y \mapsto E^y Y$ is continuous for special Y.

Definition 4.0.2 (Transition semigroup)**.** A *probability semigroup* is a family of continuous linear operators $(T_t)_{t\geq 0}$ on $C_0(S)$ such that

1) $T_0 = \mathrm{id}$

2) $\lim_{t\downarrow 0} T_t f = f$, which we call *strong continuity*

3) $T_{s+t} = T_s T_t$, the *semigroup property*; in particular $T_t T_s = T_s T_t$

4) $T_t f \geq 0$ if $f \geq 0$

5) $T_t 1 = 1$ if S is compact, and otherwise $\exists (f_n)_n \in C_0(S)^{\mathbb{N}}$ such that $\sup_n \|f_n\| < \infty$ and $T_t f_n \to 1$ pointwise for all $t \geq 0$.

Claim $\|T_t f\| \leq \|f\|$ for all $f \in C_0(S)$

Proof for S compact. Let $g := \|f\| - f \geq 0$ and $g \in C_0(S)$. Then $T_t g = \|f\| - T_t f \geq 0$ so that $T_t f \leq \|f\|$ pointwise. □

Claim $t \mapsto T_t f$ is continuous on $[0,\infty)$.

Proof. Let $t_n \downarrow t$. Then

$$\|T_{t_n} f - T_t f\| \leq \|(T_{t_n - t} - I)\, f\|$$

□

Example 4.1.

- Let X be a Brownian motion, $S = \mathbb{R}$. Then $T_t f(x) := E^x\left(f(X_t)\right)$ defines a transition semigroup.
- For S countable, $p_t(x,y)$ a transition function, $T_t f(x) := \sum_{y\in S} p_t(x,y) f(y)$. If S is finite, then $(T_t)_t$ is a transition semigroup iff $\lim_{|x|\to\infty} p_t(x,y) = 0$ for all $t > 0$ and $y \in \mathbb{Z}$.

Definition 4.0.3 (Resolvent)**.** We write $U_\alpha f := \int_0^\infty e^{-\alpha t} T_t f dt$ for $\alpha > 0, f \in C_0(S)$ and call it the *resolvent* associated with the transition semigroup.

Clearly $U_\alpha f$ is a well-defined, linear operator with operator norm $\frac{1}{\alpha}$ and $\lim_{\alpha\to\infty} \alpha U_\alpha f = f$.

Claim $U_\alpha - U_\beta = (\beta - \alpha) U_\alpha U_\beta$ for $\alpha, \beta > 0$, known as the *resolvent equation.*

Proof.

$$U_\alpha U_\beta = \int_0^\infty e^{-\alpha t} T_t U_\beta f dt = \int_0^\infty e^{-\alpha t} \int_0^\infty e^{-\beta s} T_{t+s} f ds dt = \int_0^\infty T_r f \int_0^r e^{-\alpha t - \beta(r-t)} dt dr = \int_0^\infty T_r f \frac{e^{-\alpha r} - e^{-\beta r}}{\beta - \alpha} dr$$

□

For S countable let $\mathcal{L}f(x) := \sum_y q(x,y)f(y) = \sum_y q(x,y)\left(f(y) - f(x)\right)$. Write $D(\mathcal{L})$ for the domain of $\mathcal{L}$ and $R(\mathcal{L})$ for its range.

Definition 4.0.4. A probability generator $\mathcal{L}$ is a linear operator on $C_0(S)$ such that

1) $D(\mathcal{L})$ is dense in $C_0(S)$

2) $f \in D(\mathcal{L}), \lambda \geq 0$ then for $g = f - \lambda \mathcal{L} f$ we have $\inf_S f \geq \inf_S g$

3) For λ small enough we have $R(I - \lambda\mathcal{L}) = C_0(S)$

4) For S compact we have $1 \in D(\mathcal{L})$ and $\mathcal{L}1 = 0$. If S is locally compact then for λ sufficiently small there exist $f_n \in D(\mathcal{L}), g_n := f_n - \lambda\mathcal{L}f_n$ such that $\sup_n \|g_n\| < \infty$ and both $f_n, g_n \overset{n\to\infty}{\longrightarrow} 1$ pointwise.

Claim For $f \in D(\mathcal{L}), \lambda \geq 0, g = f - \lambda\mathcal{L}f$ we have $\|f\| \leq \|g\|$.

Proof. We have $\inf_S g \leq \inf_S f \leq \sup_S f \leq \sup_S g$ by property 2. $\square$

Hence $(I - \lambda\mathcal{L})^{-1}$ exists and is a contraction for λ sufficiently small.

Example 4.2.

- $S = \mathbb{R}$ and $D(\mathcal{L}) = \left\{f \in C_0(S), f \in C^1\right\}$, $\mathcal{L}f = f'$ is a generator.
- For S finite and $q(x,y)$ a Q-matrix, $\mathcal{L}f(x) = \sum_y q(x,y)f(y)$ is a probability generator.

Remark. Often $\mathcal{L}$ is second order differential operator. Consider the PDE $\frac{\partial}{\partial t}u(t,x) = \mathcal{L}u(t,x)$ where $\mathcal{L}$ acts only on x with initial condition $u(0,x) = f(x)$. Under some mild conditions, the solution is given by $u(t,x) = T_t f(x)$.

Theorem 4.0.2. *For any Feller process on S, $T_t f(x) := E^x f(X_t)$ defines a probability semigroup on $C_0(S)$.*

Proof. 1), 4), 5) follow from construction. 3) uses the Markov property:

$$T_{t+s}f = E^x f(X_{s+t}) = E^x\left(E^x\left(f(X_{s+t}) \mid \mathcal{F}_s\right)\right) = E^x\left(E^{X_s} f(X_t)\right) = E^x\left(T_t f(X_s)\right) = T_s T_t f$$

For 2) note first that the pointwise convergence $T_t f(x) \overset{t\downarrow 0}{\longrightarrow} f(x)$ follows from right-continuity of the paths and the continuity of f. Now for the uniform continuity:
First, we obtain the pointwise resolvent equation from the pointwise continuity.
Second, let $U_\alpha : C_0(S) \to C_0(S)$. Now let $R = R(U_\alpha)$. This is independent of α: Let $f \in R$ so that $f = U_\alpha g$ for some g. Then $U_\beta f = U_\beta U_\alpha g = \frac{1}{\alpha-\beta}(U_\beta g - U_\alpha g)$, hence $f = (\alpha - \beta)U_\beta(U_\alpha g - g)$. Now let $f = U_\alpha g$. Then

$$T_t f = \int_0^\infty e^{-\alpha s} T_{s+t} g ds = \int_t^\infty e^{-\alpha(r-t)} T_r g dr \overset{t\downarrow 0}{\longrightarrow} U_\alpha g = f \;\text{ uniformly}$$

Fourth, the contraction property implies $\lim_{t\downarrow 0} |T_t f - f| = 0$ for all $f \in \overline{R}$. Finally, as for any linear subspace of a Banach space the strong closure equals the weak closure and $\alpha U_\alpha f \to f$ pointwise it follows that $R = C_0(S)$. $\square$

Theorem 4.0.3. *Let $(T_t)_{t\geq 0}$ be a probability semigroup. Then $\mathcal{L}f = \lim_{t\downarrow 0}\frac{T_t f - f}{t}$ having the domain $D(\mathcal{L}) = \{f \in C_0(S) : \mathcal{L}f$ is (strongly) convergent$\}$ is a probability generator. Moreover,*

1. *For all $g \in C_0(S)$, $\alpha > 0$*

$$f = \alpha U_\alpha g \;\textit{iff}\; \left(f \in \mathcal{D}(\mathcal{L}) \textit{ and } f - \frac{1}{\alpha}\mathcal{L}f = g\right)$$

2. *If $f \in \mathcal{D}(\mathcal{L})$ then $T_t f \in \mathcal{D}(\mathcal{L})$, $t \mapsto T_t f$ is C^1 and*

$$\frac{d}{dt}T_t f = T_t\mathcal{L}f = \mathcal{L}T_t f$$

Proof. Suppose that $f = \alpha U_\alpha g$ for $g \in C_0(S)$. Then

$$\begin{aligned}
\frac{1}{t}(T_t f - f) &= \frac{\alpha}{t}T_t\int_0^\infty e^{-\alpha r}T_r g dr - \frac{\alpha}{t}\int_0^\infty e^{-\alpha s}T_s g ds = \frac{\alpha}{t}\int_t^\infty e^{-\alpha(s-t)}T_s g ds - \frac{\alpha}{t}\int_0^\infty e^{-\alpha s}T_s g ds \\
&= \frac{\alpha}{t}\left(e^{\alpha t} - 1\right)\int_t^\infty e^{-\alpha s}T_s g ds - \frac{\alpha}{t}\int_0^t e^{-\alpha s}T_s g ds \overset{t\downarrow 0}{\longrightarrow} \alpha^2 U_\alpha g - \alpha g = \alpha(f - g) \;\text{ uniformly}
\end{aligned}$$

This proves "$\Rightarrow$" in 1. as well as 3) in the definition of the probability generator. If $\alpha U_\alpha g \in \mathcal{D}(\mathcal{L})$ then $\alpha U_\alpha g \overset{\alpha\to\infty}{\longrightarrow} g$, hence $\mathcal{D}(\mathcal{L})$ is dense in $C_0(S)$, whence 1) in the definition follows. Now for $t > 0$, $f \in \mathcal{D}(\mathcal{L})$ and $g_t := \left(1 + \frac{\lambda}{t}\right)f - \frac{\lambda}{t}T_t f = f - \lambda\frac{T_t f - f}{t}$ we get $\lim_{t\downarrow 0} g_t = f - \lambda\mathcal{L}f$ and $\left(1 + \frac{\lambda}{t}\right)\inf_x f(x) \geq \frac{\lambda}{t}\inf_x T_t f(x) + \inf_x g_t(x) \geq \frac{\lambda}{t}\inf_x f(x) + \inf_x g_t(x)$. Hence we get 2) in the definition.

Now for "$\Leftarrow$" in 1. suppose that $\left(I - \frac{1}{\alpha}\mathcal{L}\right) f = g$ with $f \in \mathcal{D}(\mathcal{L})$. By "$\Rightarrow$" in 1. we get $h := \alpha U_\alpha g \Rightarrow h - \frac{1}{\alpha}\mathcal{L}h = g$. Now since $f - h \in \mathcal{D}(\mathcal{L})$ we get $\|f - h\| \le \|g - g\| = 0$ whence $f = h$. 4) of the definition is clear if S is compact. Otherwise suppose that $(g_n)_n$ is a sequence in $C_0(S)$ and $\sup_n \|g_n\| < \infty$ as well as $g_n \to 1, T_t g_n \to 1$ pointwise. Define $f_n \in \mathcal{D}(\mathcal{L})$ by $g_n = (I - \lambda\mathcal{L}) f_n$, i.e. $f_n = \alpha U_\alpha g_n$ by 1. Since $T_t g_n \to 1$ pointwise, then $f_n \to 1$ pointwise. For 2.:

$$\frac{d}{dt}T_t f = \lim_{s\downarrow 0} \frac{T_{s+t}f - T_t f}{s} = \lim_{s\downarrow 0} \frac{T_s\left(T_t f\right) - T_t f}{s}$$

provided that any of the limits exist. The middle one does and equals $T_t\mathcal{L}f$, which is continuous in t. Then the right one exists as well and equals $\mathcal{L}T_t f$. □

Lemma 4.0.4. *For $f \in C_0(S)$, $t > 0$, $\mathcal{L}f = \lim_{t\downarrow 0} \frac{T_t f - f}{t}$ we have $\lim_{n\to\infty} \left(I + \frac{t}{n}\mathcal{L}\right)^{-n} f = T_t f$.*

Proof. We know $\left(I - \frac{1}{\alpha}\mathcal{L}\right)^{-n} f = \alpha^n U_\alpha^n f = \int_0^\infty \frac{\alpha^n s^{n-1}}{(n-1)!} e^{-\alpha s} T_s f ds$. Hence $\left(I - \frac{t}{n}\mathcal{L}\right)^{-n} f = E\left(T_{\frac{\xi_1+\ldots+\xi_n}{n}t} f\right)$ where $\xi_1, \ldots, \xi_n$ are iid Exp(1)-distributed. Now if $f \in D(\mathcal{L})$ then $\|T_t f - T_s f\| \le \|\mathcal{L}f\| \, |t - s|$. Therefore

$$\left\|\left(I - \frac{t}{n}\mathcal{L}\right)^{-n} f - T_t f\right\| \le t \|\mathcal{L}f\| \, E\left|\frac{\xi_1 + \ldots + \xi_n}{n} - 1\right| \overset{n\to\infty}{\longrightarrow} 0$$

Now both operators on the left-hand side are contractions, so that we can approximate all $f \in C_0(S)$. □

For $\epsilon > 0$ define $\mathcal{L}_\epsilon := \mathcal{L}\left(I - \epsilon\mathcal{L}\right)^{-1}$, which is well defined since $R(I - \epsilon\mathcal{L}) = C_0(S)$ for $\epsilon > 0$ sufficiently small. Further $f - \epsilon\mathcal{L}f = g$ iff $f = \left(I - \epsilon\mathcal{L}\right)^{-1} g$ which implies $\|\mathcal{L}_\epsilon g\| = \|\mathcal{L}f\| \le \frac{\|f\| + \|g\|}{\epsilon} \le \frac{2}{\epsilon}\|g\|$, so that $\mathcal{L}_\epsilon$ is also bounded. Write $T_{\epsilon,t} := e^{t\mathcal{L}_\epsilon} := \sum_k \frac{t^k \mathcal{L}_\epsilon^k}{k!}$, which exists as strong limit.

Lemma 4.0.5. *If $\mathcal{L}$ is a bounded operator then $R(I - \lambda\mathcal{L}) = C_0(S)$ for λ small enough.*

Proof. For g find f with $f - \lambda\mathcal{L}f = g$ by $f = \sum_k \lambda^n \mathcal{L}^n g$ which converges if $\lambda \|\mathcal{L}\| < 1$. □

Lemma 4.0.6. *1. For $f \in C_0(S)$ we have $\left(I - \epsilon\mathcal{L}\right)^{-1} f - \epsilon\mathcal{L}_\epsilon f = f$*

2. $\mathcal{L}_\epsilon$ is a probability generator; the associated semigroup $(T_{\epsilon,t})_t$ has $\mathcal{L}_\epsilon$ as its corresponding generator.

Proof. 1. $\left(I - \epsilon\mathcal{L}\right)^{-1} f - \epsilon\mathcal{L}(I - \epsilon\mathcal{L})^{-1} f = \left(I - \epsilon\mathcal{L}\right)\left(I - \epsilon\mathcal{L}\right)^{-1} f = f$.

2. Indeed

- $D(\mathcal{L}_\epsilon) = C_0(S)$
- $f = \lambda\mathcal{L}_\epsilon f = g$ implies that $\inf f \ge \inf g$ so that $f - \lambda\left(\frac{1}{\epsilon}\left(I - \epsilon\mathcal{L}\right)^{-1} f - \frac{1}{\epsilon} f\right) = g$, i.e. $\left(1 + \frac{\lambda}{\epsilon}\right) f - \frac{\lambda}{\epsilon}\left(I - \epsilon\mathcal{L}\right)^{-1} f = g$, or $\frac{\epsilon}{\epsilon+\lambda}\left(I - \epsilon\mathcal{L}\right)^{-1} f(x) + \frac{\epsilon}{\epsilon+\lambda} g(x) = f(x)$. It follows $\frac{\lambda}{\epsilon+\lambda}\inf_x f(x) + \frac{\epsilon}{\epsilon+\lambda}\inf_x g(x) \le \frac{\lambda}{\lambda+\epsilon}\inf_x \left(I - \epsilon\mathcal{L}\right)^{-1} f(x) + \frac{\epsilon}{\epsilon+\lambda}\inf_x g(x) \le \inf_x f(x)$
- $R\left(I - \lambda\mathcal{L}_\epsilon\right) = C_0(S)$ for $\lambda > 0$ small.
- If S is compact then $\mathcal{L}_\epsilon 1 = \mathcal{L}\left(I - \epsilon\mathcal{L}\right)^{-1} 1$ but $\left(I - \epsilon\mathcal{L}\right)^{-1} 1 = 1$ because $f - \epsilon\mathcal{L}f = 1$ is solved by $f = 1$ and $\mathcal{L}_\epsilon 1 = \mathcal{L}1 = 0$.

□

Theorem 4.0.7. *For $f \in C_0(S)$, $T_t f := \lim_{\epsilon\downarrow 0} T_{\epsilon,t} f$ exists in the uniform sense on bounded t-intervals. It defines a semigroup whose generator is $\mathcal{L}$.*

Proof. **Step 1:** $\mathcal{L}_\epsilon$ and $\mathcal{L}_\delta$ commute: We know $\left(I - \epsilon\mathcal{L}\right)^{-1}\left(I - \delta\mathcal{L}\right)^{-1} f = g$ iff $f = g - (\epsilon + \delta)\mathcal{L}g + \epsilon\delta\mathcal{L}^2 g$ which is symmetric in ϵ, δ.

Step 2: We have $(T_{\epsilon,t} - T_{\delta,t})f = \int_0^t \frac{d}{ds} T_{\epsilon,s} T_{\delta,t-s} f ds = \int_0^t T_{\epsilon,s} T_{\delta,t-s}\left(\mathcal{L}_\epsilon - \delta\mathcal{L}\right) f ds$, so that

$$\|(T_{\epsilon,t} - T_{\delta,t})f\| \le t \|T_{\epsilon,s} T_{\delta,t-s}\left(\mathcal{L}_\epsilon - \mathcal{L}_\delta\right) f\| \le t \|\left(\mathcal{L}_\epsilon - \mathcal{L}_\delta\right) f\|$$

Step 3: Let $f \in D(\mathcal{L})$. Then $\left(I - \epsilon\mathcal{L}\right)^{-1} f - f = \epsilon\left(I - \epsilon\mathcal{L}\right)^{-1}\mathcal{L}f$

$$\left\|\left(I - \epsilon\mathcal{L}\right)^{-1} f - f\right\| \le \epsilon \|\mathcal{L}f\|$$

In particular $\lim_{\epsilon\downarrow 0}\left(I - \epsilon\mathcal{L}\right)^{-1} f = f$ in the strong sense, so that $\lim_{\epsilon\downarrow 0}\mathcal{L}_\epsilon f = \mathcal{L}f$.

Step 4: From Step 2, 3 we obtain that $\lim_{\epsilon\downarrow 0} T_{\epsilon,t}$ exists in the strong sense on bounded t-intervals, using that $C_0(S)$ is complete. The semigroup properties of $T_{\epsilon,t}$ carry over to T_t:

$$\|T_0 f - f\| \leq \|(T_0 - T_{\epsilon,0})f\| + \|T_{\epsilon,0}f - f\| \xrightarrow{\epsilon\downarrow 0} 0$$

and similarly for the other properties.

Step 5: Check that T_t has $\mathcal{L}$ as its generator: Let $f \in D(\mathcal{L})$ Then

$$T_t f - f \xleftarrow{\epsilon\downarrow 0} T_{\epsilon,t}f - f = \int_0^t \frac{d}{ds} T_{\epsilon,s} f ds = \int_0^t T_{\epsilon,s}\mathcal{L}_\epsilon f ds \xrightarrow{\epsilon\downarrow 0} \int_0^t T_s \mathcal{L}_\epsilon f ds,$$

whence $\lim_{t\downarrow 0} \frac{T_t f - f}{t} = \mathcal{L}f$ so that $\mathcal{L}$ is an extension of the generator according to Theorem 4.0.3.

□

Corollary 4.0.8. $\lim_{\epsilon\downarrow 0} \sum_{n=0}^\infty \frac{t^n}{n!} \left(\mathcal{L}\left(I - \epsilon\mathcal{L}\right)^{-1}\right)^n = \lim_{n\to\infty} \left(I - \frac{t}{n}\mathcal{L}\right)^n$

Theorem 4.0.9. *If $(T_t)_{t\geq 0}$ is a probability semigroup on $C_0(S)$, then there exists a Feller process (X_t) such that $E^x f(X_t) = T_t f(x)$ for $x \in S, f \in C_0(S), t \geq 0$.*

Proposition 4.0.10. *Suppose $(M_t)_{t\in\mathbb{Q}^+}$ is a uniformly bounded sub-/supermartingale. Then a.s. $\lim_{s\to t, s\in\mathbb{Q}^+} M_s$ exists.*

Proof. Use the upcrossing lemma. □

Proof of theorem. **Step 1:** Define the finite dimensional distributions as follows: The one-dimensional marginals are given by the theorem. For the two-dimensional marginals, take $s \leq t$,

$$E^x f(X_s) g(X_t) = E^x \left(f(X_s) E^{X_s}\left(g(X_{t-s})\right)\right) = T_s\left(f(\cdot) T_{t-s} g(\cdot)\right)(x).$$

Higher-dimensional marginals are constructed inductively.

Step 2: Apply the Kolmogorov consistency theorem for *rational* $t \in \mathbb{Q}^+$, which yields a process $(Y_t)_{t\in\mathbb{Q}^+}$ with $Y_0 = x$ P^x-a.s.

Step 3: Let $0 \leq f \in C_0(S)$.

$$e^{-\alpha t} T_t U_\alpha f = \int_t^\infty e^{-\alpha s} T_s f ds \leq U_\alpha f$$

so that

$$E^x\left(e^{-\alpha t} U_\alpha f(Y_t)\right) \leq U_\alpha f(Y_0).$$

Hence $e^{-\alpha t} U_\alpha f(Y_t)$ is a bounded supermartingale, so that by the proposition the left- and right limits of $U_\alpha f(Y_s)$ exist a.s. for all $s \in [0, \infty)$.

Step 4: Any α and f contained an exceptional set. Therefore we take $\alpha \in \mathbb{N}$ and f in a countable dense subset of $C_0(S)$. Thus we obtain the left- and right limits everywhere because $\alpha U_\alpha f \to f$ and $C_0(S)$ is seperable.

Step 5: Define $X_t := \lim_{s\downarrow t, s\in\mathbb{Q}^+} Y_s$, which is cádlág.

- $P^x(X_0 = x) = 1$ by construction.
- For the Feller property, note $x \mapsto E^x f(X_t) = T_t f(x) \in C_0(S)$.
- For the Markov property proceed as for Brownian motion.

□

Claim: Feller processes are quasi-left continuous, i.e. $(\tau_n)_{n\in\mathbb{N}}$ are stopping times and $\tau_n \uparrow \tau$ then $X_{\tau_n} \to X_\tau$ on $\{\tau < \infty\}$.

Definition 4.0.5 (Diffusion process)**.** A diffusion process is a Feller process with continuous paths.

Definition 4.0.6. An operator $\mathcal{L}$ is closed if its graph $\{(f, \mathcal{L}f) : f \in D(\mathcal{L})\}$ is a closed subset in $C_0(S) \times C_0(S)$. $\overline{\mathcal{L}}$ is the closure of $\mathcal{L}$ if its graph is $\overline{\text{graph}\,(\mathcal{L})} = \text{graph}(\overline{\mathcal{L}})$.

Note that not every operator has a closure.

Lemma 4.0.11. *1. Assume that $\mathcal{L}$ satisfies 1), 2) in the definition of a probability generator. Then so does $\overline{\mathcal{L}}$.*

2. If $\mathcal{L}$ satisfies 1), 2), 3) then $\mathcal{L}$ is closed.

3. *If $\mathcal{L}$ satisfies 2), 3) then $R(I-\lambda\mathcal{L}) = C_0(S)$ for* all $\lambda > 0$.

4. *If $\mathcal{L}$ is closed and satisfies 2) then $R(I-\lambda\mathcal{L})$ is a closed subset of $C_0(S)$.*

Definition 4.0.7. $D \subset D(\mathcal{L})$ is called *core* of $\mathcal{L}$ if $\mathcal{L}$ is the closure of $\mathcal{L}|_D$.

Theorem 4.0.12. *Let (X_t) be a Feller process with generator $\mathcal{L}$. For each $f \in D(\mathcal{L})$,*

$$M_t := f(X_t) - \int_0^t \mathcal{L}f(X_s)ds$$

is a $(P^x, \mathcal{F})$-martingale for each $x \in S$.

Proof. Recall that $\frac{d}{dt}T_t f = T_t\mathcal{L}f = \mathcal{L}T_t f$. Hence

$$E^x(M_t) = T_t f(x) - \int_0^t T_s\mathcal{L}f(x)ds = T_t f(x) - \int_0^t \frac{d}{ds}T_s f(x)ds = f(x)$$

which is finite. For $s < t$ we get

$$\begin{aligned}
E^x\left(M_t \mid \mathcal{F}_s\right) &= E^x\left(f(X_{t-s}) \circ \theta_s \mid \mathcal{F}_s\right) - \int_0^s \mathcal{L}f(X_u)du - E^x\left(\int_0^{t-s} \mathcal{L}f(X_u) \circ \theta_s du \mid \mathcal{F}_s\right) \\
&= E^{X_s} f(X_{t-s}) - \int_0^s \mathcal{L}f(X_u)du - E^{X_s}\left(\int_0^{t-s} \mathcal{L}f(X_u)du\right) \\
&= E^{X_s}\left(M_{t-s}\right) - \int_0^s \mathcal{L}f(X_u)du = f(X_s) - \int_0^s \mathcal{L}f(X_u)du = M_s
\end{aligned}$$

$\square$

We know that if X_t is Brownian motion then it generates $\mathcal{L}f = \frac{1}{2}f''$. If $Y_t = X_{ct}$ for $c > 0$ then

$$\lim_{t\downarrow 0} \frac{E^x(f(Y_t)) - f(x)}{t} = \frac{c}{2}f''(x)$$

Hence $\mathcal{L}f = \frac{c}{2}f''$ is the generator of time-changed Brownian motion.

4.1 Wright-Fisher Diffusion

Let there be N individuals with genotypes aa, aA, AA and total numbers $N_1 + N_2 + N_3 = N$. The next generation has $(\widetilde{N}_1, \widetilde{N}_2, \widetilde{N}_3)$ trinomially distributed individuals with parameters $(1-x)^2, 2x(1-x), x^2$ where $x = \frac{N_2+2N_3}{2N}$. Let X_n be the proportion of As in the n-th generation. Write

$$\mathcal{L}_N f(x) = E^x f(X_t) - f(x) = \sum_{k=0}^{2N} \binom{2N}{k} x^k (1-x)^{2N-k} \left(f(\frac{k}{2N}) - f(x)\right)$$

If $f : [0,1] \to \mathbb{R}$ is C^2 then

$$f(\frac{k}{2N}) - f(x) = f'(x)\left(\frac{k}{2N} - x\right) + \frac{1}{2}f''(x)\left(\frac{k}{2N} - x\right)^2 + o\left(\left(\frac{k}{2N} - x\right)^2\right)$$

Hence $\lim_{N\to\infty} 2N\mathcal{L}_N f(x) = \frac{1}{2}$. Now consider $\mathcal{L}f(x) = \frac{1}{2}x(1-x)f''(x)$ for (at least) polynomials f in $C([0,1])$.

Theorem 4.1.1.

1. *The closure of $\mathcal{L}$ is a probability generator.*

2. *The Feller process (X_t) associated with $\overline{\mathcal{L}}$ is a diffusion process.*

3. *For $\tau := \inf\{t \geq 0 : X_t \in \{0,1\}\}$ we have*

$$E_x(\tau) = -2x\log(x) - 2(1-x)\log(1-x)$$

and

$$P^x(X_\tau = 1) = x, \quad E^x\left(\int_0^\infty X_t(1-X_t)dt\right) = x(1-x)$$

Proof.

1. $\mathcal{L}$ maps polynomials of degree n to polynomials of degree $\leq n$. We need to check properties 1)-4) of a probability generator.

   (a) Polynomials are dense in $C([0,1])$.

(b) Let f be a polynomial and $f - \lambda \mathcal{L} f = g$ for some $\lambda > 0$. If f has a minimum in $x_0 \in [0,1]$, clearly $\mathcal{L}f(x_0) \geq 0$ so that $\min_x f = f(x_0) \geq g(x_0) \geq \min_x g(x)$.

(c) Let $g = \sum_{k=0}^n a_k x^k$ be some poynomial and consider $f - \lambda \mathcal{L} f = g$. Assume that $f(x) = \sum_{k=0}^n b_k x^k$. Then we get

$$b_k - \frac{\lambda}{2}\left(k(k+1)b_{k+1} - (k-1)kb_k\right) = a_k$$

with $b_{n+1} = 0$. We can solve these equations recursively. Then $R(I - \lambda \mathcal{L})$ contains all polynomials and is dense. Now use the Lemma.

(d) is obvious.

2. $D(\mathcal{L}) \supset C^2[0,1]$ because we can approximate every $f \in C^2[0,1]$ by polynomials f_n such that $f_n \to f, f_n' \to f', f_n'' \to f''$. Using the previous theorem with $f(X) = X$, $\mathcal{L}f = 0$ we get that X is a (uniformly bounded) Martingale, hence it has a limit X_∞ P^x-a.s. and $E^x X_t = x$, hence $P^x(X_\tau = 1) = P^x(X_\infty = 1) = x$. Now using the same theorem again with $f(x) = x(1-x)$ we get $\mathcal{L}f = -x(1-x)$ and $Z_t = X_t(1-X_t) + \int_0^t X_s(1-X_s)ds$ is a non-negative martingale. Therefore $Z_\infty := \lim_{t\to\infty} Z_t$ exists a.s. and $Z_\infty \geq 0$ a.s. as well as $E^x\left(\int_0^\infty X_s(1-X_s)ds\right) = x(1-x)$.
Now we verify the continuity criterion. Let $y \in [0,1]$ be fixed and $f(x) = (x-y)^2$ so that $\mathcal{L}f = x(1-x)$ and $(X_t - y)^2 - \int_0^t X_s(1-X_s)ds$ is a martingale. Then $E^y(X_t - y)^2 = \int_0^t E^y X_s(1-X_s)ds \leq \frac{t}{4}$. Now $f(x) = (x-y)^4$ so that $\mathcal{L}f = 6x(1-x)(x-y)^2$. Then $(X_t-y)^4 - 6\int_0^t X_s(1-X_s)(X_s-y)^2 ds$ is a martingale and $E^y(X_t - y)^4 = 6\int_0^t E^y X_s(1-X_s)(X_s-y)^2 ds \leq \frac{3}{2}\int_0^t E^y(X_s-y)^2 ds \leq \frac{3}{4}\left(\frac{t}{4}\right)^2$. Now for $s < t$:

$$E^y(X_t - X_s)^4 = E^y\left(E^y\left((X_t - X_s)^4 \mid \mathcal{F}_s\right)\right) = E^y\left(E^{X_s}\left((X_{t-s} - X_0)^4\right)\right) \leq \frac{3}{16}(t-s)^2$$

so that the paths are continuous.

3. Let $f(x) = 2x\log x + 2(1-x)\log(1-x)$. We want to show $E^x\tau = -f(x)$. We have $f''(x) = \frac{2}{x} + \frac{2}{1-x}, x \in (0,1)$ so that $\frac{1}{2}x(1-x)f''(x) = 1$. Problem: $f \notin D(\mathcal{L})$ because otherwise $f(X_t) - t$ would be a martingale in contradiction to the martingale convergence theorem. Therefore consider $f_\epsilon \in C^2[0,1]$ such that $f_\epsilon(x) = f(x)$ for all $x \in [\epsilon, 1-\epsilon]$ and extend to $[0,1]$ such that $f_\epsilon \in C^2$. Now let $\tau_\epsilon := \inf\{t : X_t \in \{\epsilon, 1-\epsilon\}\}$. Then $f_\epsilon(X_t) - \int_0^t \mathcal{L}f_\epsilon(X_s)ds$ is a martingale, hence

$$f(X_{\tau_\epsilon \wedge \tau}) - (\tau_\epsilon \wedge t)$$

is a P^x-martingale for $x \in [\epsilon, 1-\epsilon]$. Hence $E^x(f(X_{\tau_\epsilon})) - E_x(\tau_\epsilon) = f(x)$. Now $\epsilon \downarrow 0$.

□

4.2 Brownian motion on $[0,\infty)$

1. Brownian motion with absorption $\tau = \inf\{s \geq 0 : B_s = 0\}$ and $X_t^{\text{abs}} := B_t 1_{\tau > t}$. For $f \in C_0[0,\infty)$ consider the odd extension $f_o(x) = f(x)1_{x\geq 0} + (2f(0) - f(-x))1_{x<0}$. Then

$$E^x(f_o(B_t)1_{\tau\leq t}) = E^x(f_o(-B_t)1_{\tau\leq t}) = \frac{1}{2}E^x\left((f_o(B_t) + f_o(-B_t))1_{\tau\leq t}\right) = f(0)P(\tau \leq t)$$

For $x \geq 0$:

$$\left(T_t^{\text{abs}}f\right)(x) = E^x\left(f(X_t^{\text{abs}})1_{\tau\leq t}\right) + E^x\left(f(X_t^{\text{abs}})1_{\tau>t}\right) = f(0)P(\tau \leq t) + E^x(f(B_t)1_{\tau>t})$$

Now $f_o \in C_0(\mathbb{R})$ iff $f(0) = 0$. Furthermore $f_0''(x) = f''(x)1_{x>0} - f''(x)1_{x<0}$ so we require $f''(0) = 0$. Hence $\mathcal{L}^{\text{abs}}f = \frac{1}{2}f''$ on $D(\mathcal{L}^{\text{abs}}) = \{f \in C_0[0,\infty) : f', f'' \in C_0, f(0) = 0 = f''(0)\}$.

2. Brownian motion with reflection $X_t^{\text{refl}} := |B_t|$. For $f \in C_0[0,\infty)$ we set $f_e(x) = f(|x|)$. We have

$$T_t^{\text{refl}}f(x) = E^x(f(|B_t|)) = E^x(f_e(B_t))$$

Hence $f \in D(\mathcal{L}^{\text{refl}})$ iff $f_e \in D(\mathcal{L}^{\text{refl}})$, and $\mathcal{L}^{\text{refl}}f = \frac{1}{2}f''$, $D(\mathcal{L}^{\text{refl}}) = \{f \in C_0[0,\infty) : f', f'' \in C_0[0,\infty), f'(0) = 0\}$.

Remark. Consider $Af = \frac{1}{2}f''$ and $D(A) = \{f \in C_0[0,\infty) : f', f'' \in C_0[0,\infty), f'(0) = f''(0) = 0\}$. This is not a generator because one generator cannot extend another generator.

4.3 The Feynman-Kac formula

Consider the partial differential equation

$$\begin{aligned} \partial_t u(t,x) &= \mathcal{L}u(t,x) + \xi(x)\cdot u(t,x) \\ u(0,x) &= f(x) \end{aligned} \tag{1}$$

which is solved by $u(t,x) = T_t f(x) = E^x f(X_t)$ for $\xi \equiv 0$.

Theorem 4.3.1. *Let (X_t) be a Feller process with $(T_t), \mathcal{L}$ and $f \in D(\mathcal{L}), \xi \in C_0(S)$. Define*

$$u(t,x) = E^x\left(f(X_t)\exp\left(\int_0^t \xi(X_s)ds\right)\right)$$

Then $u(t,\cdot) \in D(\mathcal{L})$ and $u(t,x)$ solves (1).

Proof. The initial condition is clearly satisfied. Furthermore $u(t,\cdot) \in C_0(S)$ because it is a uniform limit of the continuous function $E^x\left(f(X_t)\exp\left(\frac{t}{n}\sum_{i=1}^n \xi(X_{it/n})\right)\right)$. Now we check the differential equation, letting $I(s,t) := \int_s^t \xi(X_r)dr$. Then we write

$$\begin{aligned} u(t+\epsilon,x) - u(t,x) &= E^x(\xi_1+\xi_2+\xi_3) \\ \xi_1 &= f(X_{t+\epsilon} - f(X_t))e^{I(0,t)}\left(e^{I(t,t+\epsilon)} - 1\right) \\ \xi_2 &= (f(X_{t+\epsilon}) - f(X_t))e^{I(0,t)} \\ \xi_3 &= f(X_t)e^{I(0,t)}\left(e^{I(t,t+\epsilon)} - 1\right) \end{aligned}$$

We have, uniformly in x:

$$\begin{aligned} E^x|\xi_1| &\le E^x|f(X_{t+\epsilon}) - f(X_t)|\, e^{t\|\xi\|}\left(e^{\epsilon\|\xi\|} - 1\right) = o(\epsilon) \\ E^x\xi_2 &= E^x\left(f(X_{t+\epsilon}) - f(X_t)e^{I(0,t)}\right) = E^x\left(T_\epsilon f(X_t) - f(X_t)e^{I(0,t)}\right) \\ \lim_{\epsilon\to 0}\frac{u(t+\epsilon,x) - u(t,x)}{\epsilon} &= E^x\mathcal{L}fe^{I(0,t)} + E^x\left(f(X_t)\xi(X_t)e^{I(0,t)}\right) \end{aligned}$$

The right-hand side is continuous in t, s othat u is differential with respect to t. However, it is not in the desired form. We condition therefore on $\mathcal{F}_\epsilon$ and then use the Markov property to obtain

$$u(t+\epsilon,x) = E^x\left(u(t,X_\epsilon)e^{I(0,\epsilon)}\right)$$

so that

$$u(t+\epsilon,x) - u(t,x) = E^x\left(u(t,X_\epsilon)\left(e^{I(0,\epsilon)} - 1\right)\right) + (T_\epsilon u(t,x) - u(t,x))$$

Divide by ϵ and let $\epsilon \to 0$ and see that a) the limit on the left-hand side exists by an earlier calculation b) the limit of the first term on the right-hand side is $\xi(x)u(t,x)$ and c) the limit of the second term on the right-hand side also exists and converges to $\mathcal{L}u(t,x)$. □

4.4 Parabolic Anderson model

Let $\mathcal{L} = \Delta$ on $S = \mathbb{Z}^d$ or $\mathbb{R}^d$ and $\xi(x)$ random. On $\mathbb{Z}^d$ we let $\Delta f(x) = \sum_{y:|y-x|=1}(f(y) - f(x))$.

Theorem 4.4.1 (Two cities theorem)**.** *$P(\xi < x) = 1 - x^{-\alpha}$. Then there exist two $\mathbb{Z}^d$-valued processes Z^1, Z^2 such that*

$$\frac{u(t,Z_t^1) + u(t,Z_t^2)}{\sum_x u(t,x)} \overset{t\to\infty}{\longrightarrow} 1$$

4.5 Lévy Processes

Definition 4.5.1. A Lévy-process is a Feller process with stationary independent increments.

Definition 4.5.2. A random variable Z is called *infinitely divisible* if for all $n \in \mathbb{N}$ there exist iid random variables $Z_{1,n}, \dots, Z_{n,n}$ such that $\sum_{i=1}^n Z_{i,n} \overset{d}{=} Z$.
Equivalently, μ is infinitely divisible if for all $n \in \mathbb{N}$ there exists μ_n such that $\mu = (\mu_n)^{*n}$.

Let $\psi(n) = -\log E\left(e^{inZ}\right)$ be the *characteristic exponent.* For $\phi_t(\theta) = e^{t\psi(\theta)}$ we have $\psi(\theta) = \frac{d}{dt}\mid_{t=0}\phi_t(\theta) = \lim_{n\to\infty} n\left(\phi_{1/n}(\theta) - 1\right)$.

Lemma 4.5.1. *For (ϕ_n) sequence of characteristic functions, $\phi_n(\theta) \to 1$ for all θ, the following are equivalent*

1. $\lim_{n\to\infty} \phi_n(\theta)^n = \phi(\theta)$ *exists for all θ and is continuous in* 0.

2. $\lim_{n\to\infty} n\,(\phi_n(\theta) - 1) =: \psi(\theta)$ *exists for all θ and is continuous in* 0.

If either of the above holds then $\phi(\theta) = e^{\psi(\theta)}$ is a characteristic function.

Corollary 4.5.2. 1. *A characteristic function ϕ of an inifinitely divisible distribution fulfills 1. with $\phi_n = \phi^{1/n}$ so that $\phi(\theta) \neq 0$ for all θ. Thus there is a unique representation $\phi = e^{\psi}$ where ψ is continuous and $\psi(0) = 0$. In particular, the ϕ_n are uniquely determined to be $\phi_n = e^{\psi/n}$.*

2. *Under the assumptions of the Lemma, $\phi^r := e^{r\psi}$ is a characteristic function.*

3. *Infinitely divisible distributions are closed under weak limits.*

Proof. 1. Lots of complex analysis!

2. This follows from $e^{r\psi} = \lim e^{rn(\phi_n - 1)}$ being continuous limit of characteristic functions to compount Poisson distributions.

3. Indeed, if $\widetilde{\phi}_n$ is infinitely divisible and $\widetilde{\phi}_n \to \phi$ then apply 1. of the Lemma to $(\widetilde{\phi}_n^{1/n})^n = \widetilde{\phi}_n \to \phi$. Use the previous point of the corollary.

□

Theorem 4.5.3. *A probability measure is infinitely divisible iff it is weakly approximable by compound Poisson distributions.*

Proof. "$\Leftarrow$" follows from the above, since compound Poisson processes are infinitely divisible.
"$\Rightarrow$" If ϕ is infinitely divisible then $\phi = \lim(\phi^{1/n})^n = \lim e^{n(\phi^{1/n}-1)}$. □

Theorem 4.5.4 (Lévy-Khinchin formula)**.** *The law μ is infinitely divisible with characteristic exponent ψ if and only if $\psi(\theta) = -ia\theta + \frac{1}{2}\sigma^2\theta^2 + \int_{\mathbb{R}-0} \left(1 - e^{i\theta x} + i\theta x 1_{|x|\leq 1}\right) \pi(dx)$ for $a \in \mathbb{R}, \sigma \geq 0$ and π on $\mathbb{R} - 0$ with $\int \left(|x|^2 \wedge 1\right) \pi(dx) < \infty$. Hence any inifinitely divisible distribution is characterised by the triplet (a, σ, π). π is called Lévy-measure.*

Proof. "$\Leftarrow$" ϕ given by e^{ψ} of this form is a characteristic function. If ψ is of the given form, then so is ψ/n.
"$\Rightarrow$" Choose a sequence of compound Poisson distributions such that $P_n \xrightarrow{w} P$ with characteristic function

$$\phi_n(\theta) = e^{\lambda_n \int (e^{i\theta x} - 1)\nu_n(dx)} =: e^{\psi_n(\theta)} \to \phi(\theta) =: e^{\psi(\theta)}$$

where $\psi_n(\theta) = \int g_\theta(x)\lambda_n\nu_n(dx) + ia_n\theta$ where $a_n = \int x1_{|x|<1}\lambda_n\nu_n(dx)$. Does $\lambda_n\nu_n$ converge? Consider

$$\begin{aligned}
\overline{\psi}_n(\theta) &:= \psi_n(\theta) - \frac{1}{2}\int_{-1}^{1} \psi_n(s)ds \\
&= \int e^{i\theta x}\lambda_n\nu_n(dx) - \lambda_n - \frac{1}{2}\int_{\theta-1}^{\theta+1}\int e^{isx}\lambda_n\nu_n(dx)ds + \lambda_n \\
&= \int e^{i\theta x}\left(1 - \left(\frac{1}{2}\int_{-1}^{1} e^{isx}ds\right)\lambda_n\nu_n(dx)\right) = \int e^{i\theta x}\left(1 - \frac{\sin(x)}{x}\right)\lambda_n\nu_n(dx)
\end{aligned}$$

Now using $1 - \frac{\sin(x)}{x} \approx \frac{x^2}{6}$ for small x. Since $\phi_n \to \phi$ uniformly on compact intervals (by Lévy's continuity theorem) it follows that $\psi_n \to \psi$ uniformly on compact intervals and thus also $\overline{\psi}_n \to \overline{\psi}$ converges pointwise.
Furthermore, using $\phi(-\theta) = \overline{\phi(\theta)}$, i.e. $\mathrm{Im}\psi$ being odd, as well as $|\phi(\theta)| \leq 1$, i.e. $\mathrm{Re}\psi \leq 0$, it follows that $\overline{\psi}(0) = -\frac{1}{2}\int_{-1}^{1}\mathrm{Re}\psi(s)ds \geq 0$ with equality iff $\mathrm{Re}\psi \equiv 0$ on $[-1, 1]$. But since $\left|\int e^{i\theta x}P(dx)\right| = 1$ for all $|\theta| \leq 1$ it follows that $\mathrm{supp}P \subset a_0(\theta) + \frac{2\pi}{\theta}\mathbb{Z}$ for all θ, hence $P = \delta_0$, so that $(a, 0, 0)$ is a suitable triple. There, assume $\overline{\psi}(0) > 0$. Then $\overline{\psi}_n(0) > 0$ for sufficiently large n, hence $\overline{\psi}_n(\theta)/\overline{\psi}_n(0)$ is characteristic function of $\widetilde{\nu}_n(dx) := \frac{\lambda_n}{\overline{\psi}_n(0)}h(x)\nu_n(dx)$. Since $\overline{\psi}_n \to \overline{\psi}$, $\widetilde{\nu}_n \xrightarrow{w} \widetilde{\nu}$ with characteristic function $\overline{\psi}(\theta)/\overline{\psi}(0)$. Hence

$$\int g_\theta(x)\lambda_n\nu_n(dx) \sim \int f_\theta(x)\widetilde{\nu}_n(dx) \to \int f_\theta(x)\widetilde{\nu}(dx)$$

where $f_\theta(x) = \frac{\overline{\psi}(0)g_\theta(x)}{h(x)}$ where f is bounded and continuous with $f_\theta(0) = -3\overline{\psi}(0)\theta^2$. Since $\psi_n \to \psi$, we also have $a := \lim_{n\to\infty} a_n$ existing so that

$$\psi(\theta) = \lim_{n\to\infty}\psi_n(\theta) = ia\theta + \int f_\theta(x)\widetilde{\nu}(dx) = ia\theta - 3\overline{\psi}(0)\widetilde{\nu}(\{0\})\theta^2 + \int g_\theta(x)\pi(dx)$$

where $\pi(dx) := 1_{x\neq 0}\frac{\overline{\psi}(0)}{h(x)}\widetilde{\nu}(dx)$ satisfies $\int(1 \wedge x^2)\pi(dx) < \infty$. □

From the definition of Lévy process we see that the law of X_t must be infinitely divisible. Define $\psi_t(\theta) = -\log E\left(e^{i\theta X_t}\right)$ for $\theta \in \mathbb{R}, t \geq 0$. In particular, $m\psi_1(\theta) = \psi_m(\theta) = n\psi_{m/n}(\theta)$ so that for $t \in \mathbb{Q}^+$ we have $\psi_t(\theta) = t\psi_1(\theta)$. This holds for all $t \in \mathbb{R}$ because the paths are right-continuous, so that every Lévy-process has $E\left(e^{i\theta X_t}\right) = e^{-t\psi(\theta)}$.

Theorem 4.5.5. *Suppose $a \in \mathbb{R}, \sigma \geq 0, \pi$ a measure on $\mathbb{R} - 0$ such that $\int \left(|x|^2 \wedge 1\right) \pi(dx) < \infty$. Define $\psi(\theta)$ as in the Lévy-Khinchin formula. Then there is a Lévy-process which satisfies $E(e^{i\theta X_t}) = e^{-t\psi(\theta)}$.*

5 Spin Systems

Let V be a countable space, $S = \{0,1\}^V$. Note that S is compact in the product topology. We let $c(x,\eta) \geq 0$ be a uniformly bounded function $V \times S \to \mathbb{R}$ such that $c(x,\cdot)$ is continuous for each $x \in V$. We let

$$\eta_x(z) = \begin{cases} \eta(z), & z \neq x \\ 1-\eta(z), & z = x \end{cases}$$

Aim: Define a Feller process such that at each time only one state is changed, e.g. η becomes η_x at rate $c(x,\eta)$. We consider $\mathcal{L}f(\eta) = \sum_{x\in V} c(x,\eta)\left(f(\eta_x) - f(\eta)\right)$ defined on $D = \left\{f \in C(S) : \|\|f\|\| := \sum_{x \in V} \sup_{\eta\in S} |f(\eta_x) - f(\eta)| < \infty\right\}$.

Claim. If $f \in D$ then f is Lipschitz continuous with respect to some norm that generates the product topology. Note that the product topology is generated by $d_\alpha(\eta,\xi) = \sum_{x\in V} \alpha(x)\,|\eta(x) - \xi(x)|$ where $\alpha > 0$ and $\sum_{x \in V} \alpha(x) < \infty$.

Proof. If $f \in D$ then $\alpha_f(x) := \sup_{\eta\in S} |f(\eta_x) - f(\eta)|$ and f is Lipschitz with respect to the metric induced by α_f. Indeed, Letting $I = \{x \in V : \xi(x) \neq \eta(x)\}$ we obtain

$$|f(\eta) - f(\xi)| \leq \sum_{i=1}^{|I|} \left|f(\xi^i) - f(\xi)^{i-1}\right|$$

where $\xi^0 = \eta$ and $\left(\xi^i\right)_{i\in\mathbb{N}}$ is a sequence of single changes. Then

$$|f(\eta) - f(\xi)| \leq \sum_{x\in I} \alpha_f(x) = d_{\alpha_f}(\xi,\eta)$$

□

Claim. For all $f \in D$ we have $\mathcal{L}f \in C(S)$.

Proof. $\forall \epsilon > 0\, \exists N_\epsilon > 0$ such that

$$c(x_i,\eta)\,|f(\eta_{x_i}) - f(\eta)| \leq \alpha_f(x_i) < \epsilon$$

for $i \geq N_\epsilon$ because $\|\|f\|\| < \epsilon$. Hence

$$|\mathcal{L}f(\eta) - \mathcal{L}f(\eta_{x_i})| < 2\epsilon,\ i > N_\epsilon.$$

□

Let $\overline{\mathcal{L}}$ be the closure of $\mathcal{L}$. We need to check the conditions for the generator.

1) Use Stone-Weierstrass: D is an algebra of continuous functions that separate points on a compact space.

2) Suppose $f \in D$ and $\lambda \geq 0, f - \lambda\mathcal{L}f = g$. Since S is compact, we have a minimum η of f. Hence, using $\mathcal{L}f(\eta) \geq 0$,

$$\inf_{\xi\in S} f(\xi) = f(\eta) \geq g(\eta) \geq \inf_{\xi\in S} g(\xi)$$

4) S is compact so that $1 \in C(S)$ and $\mathcal{L}1 \equiv 0$.

3) Let $\epsilon := \inf_{u\in V,\eta\in S}\left(c(u,\eta) + c(u,\eta_u)\right)$ and $a(x,u) := \sup_{\eta\in S} |c(x,\eta_u) - c(x,\eta)|$. For $\alpha \in \ell^1(V)$ we define $\Gamma\alpha(u) := \sum_{x\in V: x\neq u} \alpha(x)a(x,u)$, and we want to show that Γ is an operator on $\ell^1(V)$. Γ is well-defined if $M := \sup_{x\in V}\sum_{u\neq x} a(x,u) < \infty$ and then $\|\Gamma\| = M$. For $f \in C(S)$ and $x \in V$ we define $\Delta_f(x) := \sup_{\eta\in S} |f(\eta_x) - f(\eta)|$

Lemma 5.0.1. *Assume either $f \in D$ or f continuous and $c(x,\cdot) = 0$ except for finitely many x. Then, if $f - \mathcal{L}f = g \in D, \lambda > 0, \lambda M < 1 + \lambda\epsilon$ then $\Delta_f \leq \left((1-\lambda\epsilon)I - \lambda\Gamma\right)^{-1}\Delta_g$ where the inverse is defined via $\left((1-\lambda\epsilon)I - \lambda\Gamma\right)^{-1}\alpha = \frac{1}{1+\lambda\epsilon}\sum_{k=0}^{\infty}\left(\frac{\lambda}{1+\lambda\epsilon}\right)^k \Gamma^k\alpha$.*

Theorem 5.0.2. *Let $M < \infty$. Then $\overline{\mathcal{L}}$ is a probability generator and $\Delta_{T_t f} \leq e^{-\epsilon t}e^{t\Gamma}\Delta_f$*

Proof. $\overline{\mathcal{L}}$ satisfies properties 1),2),4) because $\mathcal{L}$ does so. For 3) take (V_n) to be an increasing sequence of finite subsets $V_n \subset V$ such that $\bigcup_n V_n = V$. Set $\mathcal{L}_n f(\eta) = \sum_{x\in V_n} c(x,\eta)\left(f(\eta_x) - f(\eta)\right)$ for $f \in D$. $\mathcal{L}_n$ is a *bounded* operator, so that $R(I - \lambda\mathcal{L}_n) = C(S)$ for *all* $\lambda > 0$. For a given $g \in D, f_n \in C(S)$ such that $f_n - \lambda\mathcal{L}_n f_n = g$, $\mathcal{L}_n$ satisfies the condition of the previous Lemma and for λ small we get $\lambda M < 1 + \lambda\epsilon$. Define $g_n = f_n - \lambda\mathcal{L} f_n \in R(I - \lambda\mathcal{L})$. Set $K := \sup_{x\in V, \eta\in S} c(x,\eta)$. Then

$$\|g_n - g\| = \lambda \left\|(\mathcal{L} - \mathcal{L}_n) f\right\| \leq \lambda\mathcal{L} \sum_{x\notin V_n} \Delta_{f_n}(x) \leq \lambda K \sum_{x\notin V_n} \left((1+\lambda\epsilon) I - \lambda\mathcal{L}\right)^{-1} \Delta_g(x)$$

Since $\Delta_g \in \ell^1(S)$ the right-hand side goes to zero and $g_n \to g$. We conclude that $g \in \overline{R(I - \lambda\mathcal{L})} \supset D$ and $R(I - \lambda\mathcal{L})$ is dense in $C(S)$. Again, this tells us that $R(I - \lambda\overline{\mathcal{L}})$ is closed, hence equal to $C(S)$. □

5.1 Ergodicity of spin systems

Definition 5.1.1. μ is called stationary for a Feller process (X_t) if for all $f \in C_0(S), t \geq 0$:

$$\int T_t f d\mu = \int f d\mu$$

Equivalently, for all $f \in \text{core}(\mathcal{L})$:

$$\int \mathcal{L} f d\mu = 0$$

Theorem 5.1.1. *If S is compact, then $I := \{\mu \in \mathbb{P}(S) : \mu \text{ is invariant for } (X_t)_t\} \neq \emptyset$.*

Definition 5.1.2 (Ergodicity)**.** A spin system is ergodic if $|I| = 1$. Equivalently:

$$\forall \nu \in \mathbb{P}(S) : \nu T_t \xrightarrow{w} \mu$$

Example 5.1 (Voter model for $\mathbb{Z}^d$). Let $c(x,\eta) = \frac{1}{2d}\sum_{y:|x-y|=1} 1_{\{\eta(y)\neq\eta(x)\}}$ and $a(x,u) = \begin{cases} 0, & \text{if } |x-u| > 1 \\ \frac{1}{2d}, & \text{if } |x-u| = 1 \end{cases}$ so that $M = 1$, and the process exists. It is not ergodic, since $\delta_i(\eta) = 1$ if $\eta(x) = i$ for all x and 0 otherwise are both invariant.

Theorem 5.1.2. *If $\epsilon > M$ then X is ergodic.*

Proof. Let $\eta, \xi \in S$ and change η into ξ pointwise:

$$\eta^0 = \eta,\ \eta^i = \eta^{i-1}_{x_i}$$

so that $\xi = \lim_{i\to\infty} \eta^i$. Now let $f \in C(S)$, so that

$$|f(\eta) - f(\xi)| \leq \sum_{i=1}^{\infty} \left|f(\eta^i) - f(\eta^{i-1})\right| \leq \sum_{x\in V} \Delta_f(x)$$

By Theorem 5.0.2 we obtain that $|||T_t f||| \leq e^{(M-\epsilon)t} |||f|||$, so that

$$\sup_{\eta,\xi} |T_t f(\eta) - T_t f(\eta)| \leq e^{(M-\epsilon)t} |||f|||$$

Letting $\mu \in I$, $\nu \in \mathbb{P}(S), f \in D$, we obtain

$$\left|\int f d\mu - \int f d(\nu T_t)\right| = \left|\int_{S\times S} \left(T_t f(\eta) - T_t f(\xi)\right)(\mu\otimes\nu)(d\eta d\xi)\right| \leq e^{(M-\epsilon)t} |||f||| \overset{t\to\infty}{\longrightarrow} 0$$

Now since D is dense in $C(S)$, the claim follows. □

Example 5.2 (Noisy voter model)**.** $c(x,\eta) = \sum_y p(x,y) 1_{\eta(x)\neq\eta(y)} + \beta 1_{\eta(x)=0} + \gamma 1_{\eta(x)=1}$ where $\beta, \gamma \geq 0$.

Example 5.3 (Contact process)**.** V graph of bounded degree, and write $x \sim y$ if x and y are neighbors. Set

$$c(x,\eta) = \begin{cases} 1, & \text{if } \eta(x) = 1 \\ \lambda \cdot |\{y \sim x : \eta(y) = 1\}|, & \text{if } \eta(x) = 0 \end{cases}$$

Interpretation: $x \in V$ individuals of a population, $\eta(x) = 1$ if x is infected and $\eta(x) = 0$ if x is healthy. $\delta_{\underline{0}}$ is an invariant measure. Infected people get healthy at rate 1 but infect their neighbors at rate $\lambda > 0$. Are there other invariant measures?
No, if $\epsilon = \inf_{u,\eta}\left(c(u,\eta) + c(u,\eta_u)\right) = 1 > \lambda \max\text{degree}(V) = \sup_x \sum_{u\neq x} \sup_\eta |c(x,\eta) - c(x,\eta_u)| = M$.

Example 5.4 (Stochastic Ising Model). $V = \mathbb{Z}^d, \beta > 0$ inverse temperature, $c(x,\eta) = \exp\left(-\beta \sum_{y:y\sim x}(2\eta(x)-1)(2\eta(y)-1)\right)$. Interpretation: $2\eta(x)-1 \in \{-1,1\}$ are spins. Neighboring atoms prefer to align their spin values (in particular when β is large, i.e. temperature low). We have $\epsilon = 2, M = 2de^{2d\beta}(1-e^{-2\beta})$. Feller process (η_t) is hence well-defined and has a unique invariant measure if β is small enough. In fact, the following hold:

- For $d = 1$, it is ergodic for all β.
- For $d \geq 1$ it is ergodic iff $\beta < \beta_c$.

5.2 Attractive Spin Systems and Coupling

Definition 5.2.1. A coupling of (η_t) and (ξ_t) on S is a process $(\widetilde{\eta}_t, \widetilde{\xi}_t)$ on $S \times S$ with $\widetilde{\eta}_t \overset{\mathcal{L}}{=} \eta_t$ and $\widetilde{\xi}_t \overset{\mathcal{L}}{=} \xi_t$.

Lemma 5.2.1. *Let (η_t) have rate $c_1(x,\eta)$ and (ξ_t) have rates $c_2(x,\xi)$. If $\eta \leq \xi$ implies $c_1(x,\eta) \leq c_2(x,\xi)$ for $\eta(x) = \xi(x) = 0$ and $c_1(x,\eta) \geq c_2(x,\xi)$ for $\eta(x) = \xi(x) = 1$. Then there is a coupling such that*

$$\forall \eta \leq \xi : P^{(\eta,\xi)}\left(\forall t \geq 0 : \eta_t \leq \xi_t\right) = 1$$

Proof. We give rates for $(\eta_t, \xi_t)_t$ on the space $\{(0,0),(0,1),(1,1)\}$:

$$(0,0) \to \begin{cases} (1,1), & \text{with rate } c_1(x,\eta) \\ (0,1), & \text{with rate } c_2(x,\xi) - c_1(x,\eta) \end{cases}$$
$$(0,1) \to \begin{cases} (0,0), & \text{with rate } c_2(x,\xi) \\ (1,1), & \text{with rate } c_1(x,\eta) \end{cases}$$
$$(1,1) \to \begin{cases} (0,0), & \text{with rate } c_2(x,\xi) \\ (0,1), & \text{with rate } c_1(x,\eta) - c_2(x,\xi) \end{cases}$$

□

Definition 5.2.2 (Attractive spin system). A spin system is called *attractive* if

$$\eta \leq \xi \Rightarrow \begin{cases} c(x,\eta) \leq c(x,\xi) & \text{if } \eta(x) = \xi(x) = 0 \\ c(x,\eta) \geq c(x,\xi) & \text{if } \eta(x) = \xi(x) = 1 \end{cases}$$

(Noisy) voter models, contact processes and the stochastic Ising model are all attractive.

Corollary 5.2.2. *For any attractive spin system there is a coupling of two copies (η_t, ξ_t) started in $\eta \leq \xi$ such that $P^{(\eta,\xi)}(\forall t : \eta_t \leq \xi_t) = 1$.*

Proof. Lemma for $c_1 = c_2$. □

Definition 5.2.3. Function $f \in C(S)$ is called increasing if $\eta \leq \xi \Rightarrow f(\eta) \leq f(\xi)$. Denote the set consisting of these functions f by $C^{\uparrow}(S)$. For $\mu, \nu \in \mathbb{P}(S)$ write $\mu \preceq \nu :\Leftrightarrow \int f d\mu \leq \int f d\nu$ for all $f \in C^{\uparrow}(s)$.

Lemma 5.2.3. *Let (η_t) be an attractive spin system, and (T_t) its corresponding semi-group. Then*

1. *$f \in C^{\uparrow} \Rightarrow T_t f \in C^{\uparrow}$*
2. *$\mu \preceq \nu \Rightarrow \mu T_t \preceq \nu T_t$*

Proof. 1. Let $\eta \leq \xi$. Then

$$T_t f(\eta) = E^{\eta} f(\eta_t) = E^{(\eta,\xi)}\left(f_1(\nu_t, \xi_t)\right) \leq E^{(\eta,\xi)}\left(f_2(\eta_t,\xi_t)\right) = E^{\xi}(f(\eta_t)) = (T_t f)(\xi)$$

2. For $f \in C^{\uparrow}$ we have

$$\int f d\mu \leq \int f d\nu \Rightarrow \int T_t f d\mu \leq \int T_t f d\nu \Leftrightarrow \int f d(\mu T_t) \leq \int f d(\nu T_t)$$

□

Theorem 5.2.4. *Let (η_t) be an attractive spin system*

1. *$\delta_{\underline{0}} T_s \preceq \delta_{\underline{0}} T_t$ and $\delta_{\underline{1}} > \delta_{\underline{1}}$ for $s \leq t$.*
2. *$\underline{\nu} := \lim_{t\to\infty} \delta_{\underline{0}} T_t$ exists and is invariant*
   $\overline{\nu} := \lim_{t\to\infty} \delta_{\underline{1}} T_t$ exists and is invariant
3. *$\forall \mu \in \mathbb{P}(S)\, \forall s \geq 0 : \delta_{\underline{0}} T_s \preceq \mu T_s \preceq \delta_{\underline{1}} T_s$*

4. *If* $\nu := \lim_{k\to\infty} \mu T_{t_k}$ *for* $t_k \uparrow \infty$ *exists then* $\underline{\nu} \preceq \nu \preceq \overline{\nu}$.

Proof.

1. $\delta_{\underline{0}} \preceq \delta_{\underline{0}} T_r \Rightarrow \delta_{\underline{0}} T_t \preceq \delta_{\underline{0}} T_{r+t}$
2. For each $t_k \uparrow \infty$ $\delta_{\underline{0}} T_{t_k}$ has a limit point, which are the same due to 1.
3. $\delta_{\underline{0}} \preceq \mu \preceq \delta_{\underline{1}}$ for all μ, hence also $\delta_{\underline{0}} T_t \preceq \mu T_t \preceq \delta_{\underline{1}} T_t$
4. Follows from 3.

□

Corollary 5.2.5. *An attractive spin system is ergodic iff* $\underline{\nu} = \overline{\nu}$.