Math 832: Theory of Probability

- Processes, filtrations, and stopping times
- Markov chains
- Stationary processes
- Continuous time stochastic processes
- Martingales
- Poisson and general counting processes
- Convergence in distribution
- Brownian motion
- Continuous time Markov processes
- Diffusion approximations
- ϕ-irreducibility and Harris recurrence
- Assignments

1. Processes, filtrations, and stopping times

- Stochastic processes
- Filtrations
- Stopping times

Stochastic processes

A *stochastic process* is an indexed family of random variables $\{X_\alpha, \alpha \in \mathcal{I}\}$

- State space: The set E in which X_α takes values. Usually $E \subset \mathbb{R}^d$ for some d. Always (for us), a *complete* , *separable metric space* (E, r).
- Index set: Usually, discrete time ($\mathbb{Z}$, $\mathbb{N} = \{1, 2, 3, \ldots\}$, $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$) or continuous time ($[0, \infty)$ or $(-\infty, \infty)$)
- Finite dimensional distributions:

$$\mu_{\alpha_1,\ldots,\alpha_n}(A_1 \times \cdots \times A_n) = P\{X_{\alpha_1} \in A_1, \ldots, X_{\alpha_n} \in A_n\}, \quad A_i \in \mathcal{B}(E), \tag{1.1}$$

  $\mathcal{B}(E)$ the *Borel subsets* of E.

- Kolmogorov extension theorem: If $\{\mu_{\alpha_1,\ldots,\alpha_n} \in \mathcal{P}(E^n), \alpha_i \in \mathcal{I}, n = 1, 2, \ldots\}$ is *consistent*, then there exists a probability space $(\Omega, \mathcal{F}, P)$ and $\{X_\alpha, \alpha \in \mathcal{I}\}$ defined on $(\Omega, \mathcal{F}, P)$ satisfying (1.1).

Information structure

Available information is modeled by a sub-σ-algebra of $\mathcal{F}$.

Assume that the index set is discrete or continuous *time*, $[0, \infty)$ to be specific.

- Filtration: $\{\mathcal{F}_t, t \geq 0\}$, $\mathcal{F}_t$ a sub-σ-algebra of $\mathcal{F}$. If $s \leq t$, $\mathcal{F}_s \subset \mathcal{F}_t$. $\mathcal{F}_t$ represents the information available at time t.
- Adapted process: $\{X(t) \equiv X_t, t \geq 0\}$ is $\{\mathcal{F}_t\}$-adapted if $X(t)$ is $\mathcal{F}_t$-measurable for each $t \geq 0$, that is, the state of X at time t is part of the information available at time t.
- Natural filtration for a process X: $\mathcal{F}_t^X = \sigma(X(s) : s \leq t)$. $\{\mathcal{F}_t^X\}$ is the smallest filtration for which X is adapted.

Stopping times

- Stopping time: A random variable τ with values in the index set (e.g., $[0,\infty)$) or ∞ is a $\{\mathcal{F}_t\}$-stopping time if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \in [0,\infty)$.
- The $\max$ and $\min$ of two stopping times (or any finite collection) are stopping times
- If τ is a stopping time and $c > 0$, then $\tau + c$ is a stopping time
- In discrete time, $\{\tau = n\} \in \mathcal{F}_n$ for all n if and only if $\{\tau \leq n\} \in \mathcal{F}_n$ for all n.
- In discrete time, hitting times for adapted processes are stopping times: $\tau_A = \min\{n : X_n \in A\}$

$$\{\tau_A \leq n\} = \cup_{k \leq n}\{X_k \in A\}, \quad \{\tau_A = \infty\} = \cap_k \{X_k \notin A\}$$

- In discrete time, a stopped process is adapted: If $\{X_n\}$ is adapted and τ is a stopping time, then $\{X_{n\wedge\tau}\}$ is adapted.

$$\{X_{n\wedge\tau} \in A\} = (\cup_{k<n}\{X_k \in A\} \cap \{\tau = k\}) \cup (\{X_n \in A\} \cap \{\tau \geq n\})$$

Information at a stopping time

- Information available at a stopping time τ

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \le t\} \in \mathcal{F}_t, \text{ all } t\}$$

  or in the discrete time case

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau = n\} \in \mathcal{F}_n, \text{ all } n\}$$

- $\sigma \le \tau$ implies $\mathcal{F}_\sigma \subset \mathcal{F}_\tau$

$$A \cap \{\tau \le t\} = A \cap \{\sigma \le t\} \cap \{\tau \le t\}$$

Exercise 1.1 *Show that $\mathcal{F}_\tau$ is a σ-algebra.*

Stopping times for discrete time processes

For definitness, let $\mathcal{I} = \{0, 1, 2, \ldots\}$, and let $\{X_n\}$ be $\{\mathcal{F}_n\}$-adapted.

Lemma 1.2 *If τ is a $\{\mathcal{F}_n\}$-stopping time, then $X_{m\wedge\tau}$ is $\mathcal{F}_\tau$ measurable.*

Proof.

$$\{X_{m\wedge\tau} \in A\} \cap \{\tau = n\} = \{X_{m\wedge n} \in A\} \cap \{\tau = n\} \in \mathcal{F}_n \tag{1.2}$$

□

Lemma 1.3 *Let $\mathcal{F}_n^X = \sigma(X_k : k \leq n)$ be the natural filtration for X, and let τ be a finite (that is, $\{\tau < \infty\} = \Omega$) $\{\mathcal{F}_n^X\}$-stopping time. Then $\mathcal{F}_\tau^X = \sigma(X_{k\wedge\tau} : k \geq 0)$.*

Proof. $\sigma(X_{k\wedge\tau} : k \geq 0) \subset \mathcal{F}_\tau^X$, by (1.2). Conversely, for $A \in \mathcal{F}_\tau^X$,

$$A \cap \{\tau = n\} = \{(X_0, \ldots, X_n) \in B_n\} = \{(X_{0\wedge\tau}, \ldots, X_{n\wedge\tau}) \in B_n\}$$

for some B_n. Consequently,

$$A = \cup_n \{(X_{0\wedge\tau}, \ldots, X_{n\wedge\tau}) \in B_n\} \in \sigma(X_{k\wedge\tau} : k \geq 0)$$

□

Families of processes

- Markov processes: $E[f(X(t+s))|\mathcal{F}_t] = E[f(X(t+s))|X(t)]$, all $f \in B(E)$, the bounded, measureable functions on E.
- Martingales: $E = \mathbb{R}$ and $E[X(t+s)|\mathcal{F}_t] = X(t)$
- Stationary processes: $P\{X(s+t_1) \in A_1, \ldots, X(s+t_n) \in A_n\}$ does not depend on s

2. Markov Chains

- Markov property
- Transition functions
- Strong Markov property
- Tulcea's theorem
- Optimal stopping
- Recurrence and transcience
- Stationary distributions

Markov property

$\{X_n, n \geq 0\}$ a sequence of E-valued random variables

Definition 2.1 $\{X_n\}$ *is a* Markov chain *with respect to a filtration* $\{\mathcal{F}_n\}$ *if* $\{X_n\}$ *is* $\{\mathcal{F}_n\}$*-adapted and*

$$P\{X_{n+1} \in C | \mathcal{F}_n\} = P\{X_{n+1} \in C | X_n\}, \quad C \in \mathcal{B}(E), n \geq 0,$$

or equivalently

$$E[f(X_{n+1}) | \mathcal{F}_n] = E[f(X_{n+1}) | X_n], \quad f \in B(E), n \geq 0.$$

Dynkin class theorem

Generic construction of a Markov chain

Let $F : E \times \mathbb{R} \to E$ be measurable ($F^{-1}(C) \in \mathcal{B}(E) \times \mathcal{B}(\mathbb{R})$ for each $C \in \mathcal{B}(E)$).

Let

$$X_{k+1} = F(X_k, Z_{k+1}),$$

where the $\{Z_k\}$ are iid and X_0 is independent of the $\{Z_k\}$

Lemma 2.2 $\{X_k\}$ *is a Markov chain with respect to* $\{\mathcal{F}_n\}$, $\mathcal{F}_n = \sigma(X_0, Z_1, \ldots, Z_n)$.

Proof. Let μ_Z be the distribution of Z_k and define

$$Pf(x) = \int f(F(x,z))\mu_Z(dz).$$

Then X_k is $\mathcal{F}_k$-measurable and Z_{k+1} is independent of $\mathcal{F}_k$, so

$$E[f(F(X_k, Z_{k+1}))|\mathcal{F}_k] = Pf(X_k).$$

□

Note that $\mathcal{F}_n \supset \mathcal{F}_n^X$.

conditional expectation

Transition function

$P(x,C) = P\{F(x,Z) \in C\} = \mu_Z(\{z : F(x,z) \in C\})$ is the *transition function* for the Markov chain.

$P : E \times \mathcal{B}(E) \to [0,1]$ is a *transition function* if $P(\cdot, C)$ is $\mathcal{B}(E)$-measurable for each $C \in \mathcal{B}(E)$ and $P(x, \cdot) \in \mathcal{P}(E)$ for each $x \in E$.

Note that we are considering *time homogeneous* Markov chains. We could consider

$$X_{k+1} = F_k(X_k, Z_{k+1})$$

for a sequence of functions $\{F_k\}$. The chain would then be *time inhomogeneous*.

Finite dimensional distributions

μ_{X_0} is called the *initial distribution* of the chain. The initial distribution and the transition function determine the finite dimensional distributions of the chain

$$P\{X_0 \in B_0, \ldots, X_n \in B_n\} = \int_{B_0} \mu_{X_0}(dx_0) \int_{B_1} P(x_0, dx_1) \cdots \int_{B_{n-1}} P(x_{n-1}, B_n)$$

More generally

$$E[f_0(X_0) \cdots f_n(X_n)] = \int_E \mu_{X_0}(dx_0) f_0(x_0) \int_E P(x_0, dx_1) f_1(x_1) \cdots \int_E P(x_{n-1}, dx_n) f_n(x_n)$$

and

$$E[f(X_0, \ldots, X_n)] = \int_{E \times \cdots \times E} f(x_0, \ldots, x_n) \mu_{X_0}(dx_0) P(x_0, dx_1) \cdots P(x_{n-1}, dx_n)$$

Example: FIFO queue

Let $\{(\xi_k, \eta_k)\}$ be iid with values in $[0,\infty)^2$ define

$$X_{k+1} = (X_k - \xi_{k+1})^+ + \eta_{k+1}$$

X_k is the time that the kth customer is in the system for a FIFO queue with service times $\{\eta_k\}$ and interarrival times $\{\xi_k\}$.

Note that $P : \bar{C}([0,\infty)) \to \bar{C}([0,\infty))$. Transition operators that satisfy this condition are said to have the *Feller property*.

Strong Markov property

Let τ be a stopping time with $\tau < \infty$ a.s. and consider

$$E[f(X_{\tau+1})|\mathcal{F}_\tau].$$

Let $A \in \mathcal{F}_\tau$. Then

$$\begin{aligned}\int_A f(X_{\tau+1})dP &= \sum_{n=0}^{\infty}\int_{A\cap\{\tau=n\}} f(X_{\tau+1})dP \\ &= \sum_{n=0}^{\infty}\int_{A\cap\{\tau=n\}} f(X_{n+1})dP \\ &= \sum_{n=0}^{\infty}\int_{A\cap\{\tau=n\}} Pf(X_n)dP = \int_A Pf(X_\tau)dP\end{aligned}$$

so

$$E[f(X_{\tau+1})|\mathcal{F}_\tau] = Pf(X_\tau)$$

(Note that X_τ is $\mathcal{F}_\tau$-measurable.)

Tulcea's theorem

Theorem 2.3 *For $k = 1, 2, \ldots$, let $(\Omega_k, \mathcal{F}_k)$ be a measurable space. Define $\Omega = \Omega_1 \times \Omega_2 \times \cdots$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2 \times \cdots$. Let P_1 be a probability measure on $\mathcal{F}_1$ and for $k = 2, 3, \ldots$, let $P_k : \Omega_1 \times \cdots \times \Omega_{k-1} \times \mathcal{F}_k \to [0, 1]$ be such that for each $(\omega_1, \ldots, \omega_{k-1}) \in \Omega_1 \times \cdots \times \Omega_{k-1}$, $P_k(\omega_1, \ldots, \omega_{k-1}, \cdot)$ is a probability measure on $\mathcal{F}_k$ and for each $A \in \mathcal{F}_k$, $P_k(\cdot, A)$ is a $\mathcal{F}_1 \times \cdots \times \mathcal{F}_{k-1}$-measurable function. Then there is a probability measure P on $\mathcal{F}$ such that for $A \in \mathcal{F}_1 \times \cdots \times \mathcal{F}_k$*

$$P(A \times \Omega_{k+1} \times \cdots) = \int_{\Omega_1} \cdots \int_{\Omega_k} \mathbf{1}_A(\omega_1, \ldots, \omega_k) P_k(\omega_1, \ldots, \omega_{k-1}, d\omega_k) \cdots P_1(d\omega_1)$$

Corollary 2.4 *There exists $P_x \in \mathcal{P}(E^\infty)$ such that for $C_0, C_1, \ldots, C_m \in \mathcal{B}(E)$*

$$\begin{aligned} &P_x(C_0 \times C_1 \times \cdots \times C_m \times E^\infty) \\ &\quad = \mathbf{1}_{C_0}(x) \int_{C_1} P(x, dx_1) \int_{C_2} P(x_1, dx_2) \cdots \int_{C_{m-1}} P(x_{m-2}, dx_{m-1}) P(x_{m-1}, C_m) \end{aligned}$$

For $C \in \mathcal{B}(E^{m+1})$,

$$P_x(C \times E^\infty) = \int_E P(x, dx_1) \cdots \int_E P(x_{m-1}, dx_m) \mathbf{1}_C(x, x_1, \ldots, x_m)$$

Implications of the Markov property

Note that

$$\begin{aligned} E[f_1(X_{n+1})f_2(X_{n+2})|\mathcal{F}_n] &= E[f_1(X_{n+1})E[f_2(X_{n+2})|\mathcal{F}_{n+1}]|\mathcal{F}_n] \\ &= E[f_1(X_{n+1})E[f_2(X_{n+2})|X_{n+1}]|\mathcal{F}_n] \\ &= P(f_1Pf_2)(X_n) \end{aligned}$$

and by induction

$$P\{(X_n, X_{n+1}, \cdots) \in C|\mathcal{F}_n\} = P_{X_n}(C), \tag{2.1}$$

for $C = C_0 \times C_1 \times \cdots \times C_m \times E^\infty$, $C_k \in \mathcal{B}(E)$. The Dynkin class theorem implies (2.1) holds for all $C \in \mathcal{B}(E^\infty)$.

Strong Markov property: By the same argument,

$$P\{(X_\tau, X_{\tau+1}, \cdots) \in C|\mathcal{F}_\tau\} = P_{X_\tau}(C).$$

Conditioning on $\mathcal{F}_\tau$

Lemma 2.5 *Let τ be a finite $\{\mathcal{F}_n\}$-stopping time, and let $E[|Z|] < \infty$. Then*

$$E[Z|\mathcal{F}_\tau] = \sum_{n=0}^{\infty} E[Z|\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}},$$

Proof. Let $A \in \mathcal{F}_\tau$. Then

$$E[\mathbf{1}_A \sum_{n=0}^{\infty} E[Z|\mathcal{F}_n]\mathbf{1}_{\{\tau=n\}}] = \sum_{n=0}^{\infty} E[\mathbf{1}_{A\cap\{\tau=n\}} E[Z|\mathcal{F}_n]] = \sum_{n=0}^{\infty} E[\mathbf{1}_{A\cap\{\tau=n\}} Z] = E[\mathbf{1}_A Z].$$

□

Lemma 2.6 *Let $\{Y_n\}$ be $\{\mathcal{F}_n\}$-adapted, and let τ be a finite $\{\mathcal{F}_n\}$-stopping time. If $E[|Y_n|] + E[|Y_\tau|] < \infty$, then*

$$E[Y_\tau|\mathcal{F}_n] = E[Y_{\tau\vee n}|\mathcal{F}_n]\mathbf{1}_{\{\tau\geq n\}} + Y_\tau \mathbf{1}_{\{\tau<n\}}.$$

Optimal stopping

Let $\{X_n\}$ be a $\{\mathcal{F}_n\}$-Markov chain, and let $\mathcal{S} \equiv \mathcal{S}(\{\mathcal{F}_n\})$ denote the collection of $\{\mathcal{F}_n\}$-stopping times, and let $\mathcal{S}_n = \{\tau \in \mathcal{S} : \tau \geq n\}$. The *optimal stopping problem* with reward function $u(n,x)$ is to find a stopping time τ_o satisfying

$$E[u(\tau_o, X_{\tau_o})] = V^* \equiv \sup_{\tau \in \mathcal{S}} E[u(\tau, X_\tau)]$$

To ensure the right side is finite, assume that $E[\sup_n u(n, X_n)] < \infty$ and $E[u(0, X_0)] > -\infty$. To ensure $P\{\tau_o = \infty\} = 0$, let $u(\infty, x) = -\infty$.

For more information on optimal stopping see Ferguson

Optimality equation

Suppose that $\tau_1, \tau_2 \in \mathcal{S}_n$ and $A = \{E[u(\tau_1, X_{\tau_1})|\mathcal{F}_n] > E[u(\tau_2, X_{\tau_2})|\mathcal{F}_n]\}$. Then $\tau = \tau_1 \mathbf{1}_A + \tau_2 \mathbf{1}_{A^c} \in \mathcal{S}_n$ and

$$\begin{aligned} E[u(\tau, X_\tau)|\mathcal{F}_n] &= E[u(\tau_1, X_{\tau_1})|\mathcal{F}_n]\mathbf{1}_A + E[u(\tau_2, X_{\tau_2})|\mathcal{F}_n]\mathbf{1}_{A^c} \\ &= E[u(\tau_1, X_{\tau_1})|\mathcal{F}_n] \vee E[u(\tau_2, X_{\tau_2})|\mathcal{F}_n] \end{aligned} \tag{2.2}$$

Define

$$\begin{aligned} V_n &= \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_n} E[u(\tau, X_\tau)|\mathcal{F}_n] \\ &= \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_n} E[u(\tau \vee (n+1), X_{\tau\vee(n+1)})\mathbf{1}_{\{\tau>n\}} + u(n, X_n)\mathbf{1}_{\{\tau=n\}}|\mathcal{F}_n] \\ &= \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_n} E[E[u(\tau \vee (n+1), X_{\tau\vee(n+1)})|\mathcal{F}_{n+1}]\mathbf{1}_{\{\tau>n\}} + u(n, X_n)\mathbf{1}_{\{\tau=n\}}|\mathcal{F}_n] \\ &= \operatorname*{ess\,sup}_{\tau \in \mathcal{S}_n} (E[V_{n+1}|\mathcal{F}_n]\mathbf{1}_{\{\tau>n\}} + u(n, X_n)\mathbf{1}_{\{\tau=n\}} \end{aligned}$$

It follows that

$$V_n = \max(u(n, X_n), E[V_{n+1}|\mathcal{F}_n])$$

Note that (2.2) implies that $E[V_n] = \sup_{\tau \in \mathcal{S}_n} E[u(\tau, X_\tau)]$, so $V^* = E[V_0]$.

essential supremum

Optimal stopping rule

Theorem 2.7 *Suppose that* $E[\sup_n u(n, X_n)] < \infty$ *and* $\lim_{n\to\infty} u(n, X_n) = -\infty$. *Then* $\tau_o = \min\{n : u(n, X_n) \geq V_n\}$ *is an optimal stopping rule.*

Proof.

□

Dynamic programming

Lemma 2.8 *For $n < N$, let $\mathcal{S}_n^N$ be the collection of stopping times satisfying $n \leq \tau \leq N$. Define $v_N^N(x) = u(N,x)$ and*

$$v_n^N = \max(u(n,x), Pv_{n+1}^N(x)).$$

Then for $n < N$,

$$V_n^N = \text{ess}\sup_{\tau \in \mathcal{S}_n^N} E[u(\tau, X_\tau)|\mathcal{F}_n] = v_n^N(X_n)$$

Proof. As above

$$V_n^N = \max(u(n,X_n), E[V_{n+1}^N|\mathcal{F}_n]),$$

so since $V_N^N = u(N, X_N)$,

$$V_{N-1}^N = \max(u(N-1, X_{N-1}), E[u(N,X_N)|\mathcal{F}_{N-1}]) = v_{N-1}^N(X_{N-1}),$$

and the lemma follows by induction. □

Infinite horizon

Assume $E[\sup_n u(n, X_n)] < \infty$ and $\lim_{n\to\infty} u(n, X_n) = -\infty$ a.s. Then

$$\limsup_{N\to\infty} E[u(\tau \wedge N, X_{\tau\wedge N})] \leq E[u(\tau, X_\tau)]$$

House-selling problem

Each week you pay c dollars to advertise your house, and each week you advertise, you get one offer. Suppose the offers $\{X_k\}$ are iid with a known distribution μ_X and if you reject an offer, it is gone forever. When should you sell? The problem is to maximize

$$E[X_\tau - c\tau].$$

The optimality equation becomes

$$V_n = \max(X_n - cn, E[V_{n+1}|\mathcal{F}_n]) = \max(X_n, E[V_{n+1} + cn|\mathcal{F}_n]) - cn$$

Let $V^* = \sup_{\tau\in\mathcal{S}} E[X_\tau - c\tau]$. Then $V_n = \max(X_n, V^*) - cn$, so

$$V^* = E[\max(X_1, V^*)] - c$$

which gives

$$V^* = V^*\mu_X(-\infty, V^*] + \int_{(V^*,\infty)} x\mu_X(dx) - c$$

or

$$\int_{(V^*,\infty)} (x - V^*)\mu_X(dx) = c.$$

Recurrence and transience

Assume E is countable. Let $T_y^0 = 0$ and define

$$T_y^k = \min\{n > T_y^{k-1} : X_n = y\}.$$

y is *recurrent* if $P_y\{T_y^1 < \infty\} = 1$. Otherwise y is *transient*.

Let $C_y = \{(x_0, x_1, \ldots) : x_i = y \text{ for some } i > 0\}$. Then $\{T_y^1 < \infty\} = \{X \in C_y\}$ and similarly

$$\{T_y^2 < \infty\} = \{T_y^1 < \infty, (X_{T_y^1}, X_{T_y^1+1}, \ldots) \in C_y\}$$

By the strong Markov property

$$= P_x\{T_y^1 < \infty\} P_y(C_y)$$

and more generally

$$P_x\{T_y^k < \infty\} = P_x\{T_y^1 < \infty\} P_y\{T_y^1 < \infty\}^{k-1}.$$

Consequently, if $P_y\{T_y^1 < \infty\} = 1$, then $P_y\{T_y^k < \infty\} = 1$, and if $P_y\{T_y^1 < \infty\} < 1$, then there is a last time that $X_n = y$. In particular, let $N(y) = \sum_{n=1}^{\infty} \mathbf{1}_{\{X_n = y\}}$. Then

$$E_y[N(y)] = \sum_{k=1}^{\infty} P_y\{N(y) \geq k\} = \sum_{k=1}^{\infty} P_y\{T_y^k < \infty\} = \frac{P_y\{T_y^1 < \infty\}}{1 - P_y\{T_y^1 < \infty\}}.$$

Conditions for recurrence

Theorem 2.9 *y is recurrent if and only if $E_y[N(y)] = \infty$.*

Let $\rho_{xy} = P_x\{T_y^1 < \infty\}$.

Theorem 2.10 *If x is recurrent and $P_x\{T_y^1 < \infty\} > 0$, then y is recurrent and*

$$P_y\{T_x^1 < \infty\} = 1$$

Proof. $P_x\{T_x^k = \infty,\text{ some } k\} \geq \rho_{xy}(1-\rho_{yx})$, so $\rho_{yx} = 1$. $P_x\{T_y^1 < \infty\} = \sum_{k=1}^{\infty} P_x\{T_x^{k-1} < T_y^1 < T_x^k\}$ and

$$P_x\{T_x^{k-1} < T_y^1 < T_x^k\} = P_x\{T_y^1 > T_x^1\}^{k-1} P_x\{T_y^1 < T_x^1\}.$$

Let $A_k = \{X_n = y,\text{ some } T_x^k < n < T_x^{k+1}\}$. Then $P_x(A_k) = P_x\{T_y^1 < T_x^1\}$. Consequently, since $N(y) \geq \sum_k \mathbf{1}_{A_k}$, $E_x[N(y)] = \infty$.

□

Irreducibility

X is *irreducible* if $\rho_{xy} > 0$ for all $x, y \in E$.

Lemma 2.11 *If X is irreducible, then either every state is transient or every state is recurrent.*

Conditions for recurrence/transience

$$M_n^f = \sum_{k=1}^{n} (f(X_k) - Pf(X_{k-1})) = f(X_n) - f(X_0) - \sum_{k=0}^{n-1} (Pf(X_k) - f(X_k))$$

is a martingale. Suppose $Pf = f$. Then $f(X_n)$ is a martingale. If $Pf \leq f$, $f(X_n)$ is a supermartingale.

Theorem 2.12 *Assume that the chain is irreducible. Suppose f is positive and nonconstant and that $Pf \leq f$. Then the chain is transient.*

Proof. Suppose $f(x) \neq f(y)$. Since $\lim_{n\to\infty} f(X_n)$ exists, X cannot visit both x and y infinitely often. $\square$

Conditions for transcience

Theorem 2.13 *Assume that the chain is irreducible. Suppose f is positive, $Pf(x) \leq f(x)$ for $x \notin K$, and that there exists $y \in E - K$ such that $f(y) < f(x)$ for all $x \in K$. Then the chain is transient.*

Proof. Let $\tau_K = \min\{n \geq 0 : X_n \in K\}$. Then $f(X_{n\wedge\tau_K})$ is a super martingale. Let $X_0 = y$. Since $L_f = \lim_{n\to\infty} f(X_{n\wedge\tau_K})$ exists and $E[L_f] \leq f(y)$, $P_y\{\tau_K < \infty\} < 1$. □

Let $f(x) = P_x\{\tau_K < \infty\}$. Then $f(x) = 1$ for $x \in K$ and $Pf(x) = f(x)$ for $x \notin K$. Consequently, an irreducible chain is transient if and only if there exist K and $y \notin K$ such that $f(y) = P_y\{\tau_K < \infty\} < 1$.

Conditions for recurrence

Theorem 2.14 *Assume that the chain is irreducible. If* $Pf(x) \leq f(x)$ *for* $x \notin K$, *and* $\{x : f(x) < c\}$ *is finite for each* $c > 0$, *then* $P_x\{\tau_K < \infty\} = 1$ *for all* x.

Proof. If $x \in K$, then $\tau_K = 0$. Fix $y \notin K$, and let $X_0 = y$. Then $L_f = \lim_{n\to\infty} f(X_{n\wedge\tau_K})$ exists. Since $E[L_f] \leq f(y) < \infty$, we must have $P_y\{\tau_K < \infty\} = 1$. □

Example

Let $E = \{0, 1, 2, \ldots\}$, $0 < p_i = 1 - q_i < 1$ for $i \neq 1$. Let $p(i, i+1) = p_i$ and $p(i, i-1) = q_i$, for $i > 0$, and $p_{01} = 1$. Then X is irreducible. Consider the equation $Pf(i) = f(i)$ for $i > 0$. Then

$$f(k+1) - f(k) = \frac{q_k}{p_k}(f(k) - f(k-1)),$$

so

$$f(k+1) - f(k) = \prod_{i=1}^{k} \frac{q_i}{p_i}(f(1) - f(0))$$

and

$$f(k+1) = f(1) + (f(1) - f(0)) \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{q_i}{p_i}$$

Therefore, if

$$\sum_{j=1}^{\infty} \prod_{i=1}^{j} \frac{q_i}{p_i} < \infty,$$

then X is transient.

Conversely, if

$$\sum_{j=1}^{\infty} \prod_{i=1}^{j} \frac{q_i}{p_i} = \infty,$$

let

$$f(k+1) = 1 + \sum_{j=1}^{k} \prod_{i=1}^{j} \frac{q_i}{p_i}.$$

Then $\lim_{k\to\infty} f(k) = \infty$, and $\rho_{i0} = P_i\{\tau_0 < \infty\} = 1$ for all $i > 0$. Since

$$\rho_{00} = p_{01}\rho_{10} = 1,$$

X is recurrent.

Positive recurrence

If the chain is irreducible and recurrent, then by the strong Markov property, for each $y \in E$, $\{T_y^{k+1} - T_y^k, k \geq 1\}$ are iid.

The law of large numbers then implies

$$\lim_{k\to\infty} \frac{T_y^k}{k} = E_y[T_y^1],$$

and hence

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{y\}}(X_i) = \frac{1}{E_y[T_y^1]} \equiv \pi(y).$$

If $E_y[T_y^1] < \infty$, then the y is called *positive recurrent*. Assuming irreducibility, if one state is positive recurrent, then all states are positive recurrent.

Stationary distributions

Lemma 2.15 *If the chain is irreducible and positive recurrent, then*

$$\sum_{x \in E} \pi(x) p_{xy} = \pi(y)$$

Proof. Let $f(x) = \mathbf{1}_{\{y\}}(x)$. Then

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (f(X_i) - Pf(X_{i-1})) = 0 \quad a.s.,$$

so

$$\sum_{x \in E} \pi(x) p_{xy} \leq \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} p_{X_{i-1}y} = \pi(y).$$

Summing over y, we see that equality must hold. □

Dropping the assumption that E is countable, $\pi \in \mathcal{P}(E)$ satisfying

$$\int_E P(x, A) \pi(dx) = \pi(A)$$

is called a *stationary distribution* for the chain.

Ergodicity for Markov chains

The statement that a Markov chain is *ergodic* is somewhat ambiguous. At a minimum, it means that the chain has a unique stationary distribution.

Other possibilities ($P^n f(x) = E_x[f(X_n)]$):

- There exists $\pi \in \mathcal{P}(E)$ such that for each $f \in \bar{C}(E)$ and each $x \in E$,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} P^i f(x) = \int_E f d\pi.$$

- There exists $\pi \in \mathcal{P}(E)$ such that for each $x \in E$,

$$\lim_{n\to\infty} \sup_{A\in\mathcal{B}(E)} |\frac{1}{n}\sum_{i=1}^{n} P^i(x,A) - \pi(A)| = 0.$$

- There exists $\pi \in \mathcal{P}(E)$ such that for each initial distribution

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(X_i) = \int_E f d\pi, \quad a.s., \quad f \in \bar{C}(E) \quad (\text{or } f \in B(E))$$

Stronger conditions

- There exists $\pi \in \mathcal{P}(E)$ such that for each $x \in E$,

$$\lim_{n\to\infty} P^n f(x) = \int_E f d\pi, \quad f \in \bar{C}(E) \quad (\text{or } f \in B(E)).$$

- (Uniform ergodicity) There exists $\pi \in \mathcal{P}(E)$ such that

$$\lim_{n\to\infty} \sup_{x\in E} \sup_{A\in\mathcal{B}(E)} |P^n(x,A) - \pi(A)| = 0.$$

- (Geometric ergodicity) There exists $\pi \in \mathcal{P}(E)$, $0 < \rho < 1$, and $M > 0$ such that

$$\sup_{A\in\mathcal{B}(E)} |P^n(x,A) - \pi(A)| \leq M(x)\rho^n.$$

Total variation norm

For a finite signed measure ν on $\mathcal{B}(E)$,

$$\|\nu\|_{TV} = \sup_{A \in \mathcal{B}(E)} |\nu(A)|.$$

Then $\|\mu - \nu\|_{TV}$ defines a metric on $\mathcal{P}(E)$.

Lemma 2.16 *Let $\mu, \nu \in \mathcal{P}(E)$. Then*

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sup_{f \in B(E), 0 \le f \le 1} |\int_E f d\mu - \int_E f d\nu|$$

Reversibility

Let μ be a σ-finite measure. A chain is *reversible* with respect to μ if

$$\int_E fPg d\mu = \int_E gPf d\mu.$$

In other words, P is a self-adjoint operator on $L^2(\mu)$.

If P is reversible with respect to μ, then μ is a stationary measure for P in the sense that

$$\int_E Pg d\mu = \int_E g d\mu, \quad g \in L^1(\mu).$$

If $\mu \in \mathcal{P}(E)$, then μ is a stationary distribution.

Suppose P has a density with respect to β, β σ-finite, that is,

$$P(x, dy) = p(x, y)\beta(dy).$$

Then any stationary measure is absolutely continuous with respect to β. If in addition, P is reversible with respect to $\mu(dy) = m(y)\beta(dy)$, then *detailed balance* holds:

$$m(x)p(x, y) = m(y)p(y, x).$$

Example

Let $E = \{0, 1, 2, \ldots\}$, and $p_{01} = 1$ and $p_{ii+1} = p_i = 1 - p_{ii-1} = 1 - q_i$, for $i > 0$. Then letting β be counting measure detailed balance requires

$$m_k p_k = m_{k+1} q_{k+1}.$$

Consequently, we can take $m_0 = 1$ and

$$m_k = \prod_{i=1}^{k} \frac{p_{i-1}}{q_i}, \quad k \geq 1,$$

and the chain is reversible with respect to $\mu\{i\} = m_i$. The birth and death process is postive recurrent if and only if

$$\sum_{k=1}^{\infty} \prod_{i=1}^{k} \frac{p_{i-1}}{q_i} < \infty.$$

Markov chain Monte Carlo

Markov chain Monte Carlo exploits the fact that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = \int_E f d\pi$$

under appropriate conditions on the Markov chain and stationary distribution π.

Given π, find P such that $\int_E Pf d\pi = \int_E f d\pi$. To estimate $\int_E f d\pi$, simulate $\{X_i\}$ and compute

$$\Theta_{b,n} f = \frac{1}{n-b} \sum_{i=b+1}^{n} f(X_i).$$

For b (the "burn in") sufficiently large, $\Theta_{b,n} f$ should be an approximately unbiased estimator of $\int_E f d\pi$.

Metropolis-Hastings algorithm

$\pi(dy) = \pi(y)\beta(dy)$

$Q(x, dy) = q(x, y)\beta(dy)$

Define

$$\alpha(x, y) = 1 \wedge \frac{\pi(y)q(y, x)}{\pi(x)q(x, y)} = \frac{1}{\pi(x)q(x, y)}(\pi(x)q(x, y)) \wedge \pi(y)q(y, x)),$$

where $\alpha(x, y) = 1$ if $\pi(x)q(x, y) = 0$. Given X_0, define $\{X_n\}$ recursively as follows:

Let $\{\xi_n\}$ be iid uniform $[0, 1]$. Generate Y_{n+1} so that $P\{Y_{n+1} \in A|\mathcal{F}_n^{X,Y,\xi}\} = Q(X_n, A)$ and set

$$X_{n+1} = \begin{cases} Y_{n+1} & \xi_{n+1} \leq \alpha(X_n, Y_{n+1}) \\ X_n & \xi_{n+1} > \alpha(X_n, Y_{n+1}) \end{cases}$$

Lemma 2.17 $\{X_n\}$ *is a Markov chain that is reversible with respect to* π*.*

Proof of Lemma 2.17

$$
\begin{aligned}
Pf(x) &= \int_E q(x,y)(f(y)\alpha(x,y) + f(x)(1-\alpha(x,y))\beta(dy) \\
&= \int_E \frac{1}{\pi(x)}((f(y)-f(x))(\pi(x)q(x,y)) \wedge (\pi(y)q(y,x))\beta(dy) + f(x)
\end{aligned}
$$

so

$$
\begin{aligned}
&\int_E g(x)Pf(x)\pi(x)\beta(dx) \\
&\quad = \int_E\int_E g(x)(f(y)-f(x))(\pi(x)q(x,y)) \wedge (\pi(y)q(y,x))\beta(dy)\beta(dx) \\
&\qquad\qquad + \int_E g(x)f(x)\pi(x)\beta(dx) \\
&\quad = -\frac{1}{2}\int_E\int_E (g(y)-g(x))(f(y)-f(x))(\pi(x)q(x,y)) \wedge (\pi(y)q(y,x))\beta(dx)\beta(dy) \\
&\qquad\qquad + \int_E g(x)f(x)\pi(x)\beta(dx)
\end{aligned}
$$

Reversibility follows by the symmetry in f and g.

Gibbs sampler

$E = S^d, \qquad \beta(dz)$, σ-finite on S

$\pi(dx) = \pi(x_1, \ldots, x_d)\beta(dx_1)\cdots\beta(dx_d)$

$\theta_l(x|z)$ replaces the lth component of $x \in S^d$ by $z \in S$.

$$P_l f(x) = \frac{\int_S f(\theta_l(x|z))\pi(\theta_l(x|z))\beta(dz)}{\int_S \pi(\theta_l(x|z))\beta(dz)}$$

Check that $\int_E P_l f d\pi = \int_E f d\pi$.

Deterministic scan Gibbs sampler: $P = P_1 P_2 \cdots P_d$

Random scan Gibbs sampler: $P = \frac{1}{d}\sum_{i=1}^d P_i$

Coupling

Lemma 2.18 *Let $P(x,\Gamma)$ be a transition function on E and let $\nu_{xy}(\Gamma)$ be a transition function from $E\times E$ to E. Let $\epsilon : E\times E\rightarrow[0,1]$ be $\mathcal{B}(E)\times\mathcal{B}(E)$-measurable and satisfy*

$$P(x,\Gamma)\wedge P(y,\Gamma)\geq\epsilon(x,y)\nu_{xy}(\Gamma),\qquad \Gamma\in\mathcal{B}(E)$$

Let $\{X_k\}$ and $\{Y_k\}$ be independent Markov chains with transition function P. If

$$\sum_{k=0}^{\infty}\epsilon(X_k,Y_k)=\infty\quad \text{a.s.},$$

then there exists a probability space on which is defined a Markov chain $\{(\tilde{X}_k,\tilde{Y}_k)\}$ such that $\{\tilde{X}_k\}$ has the same distribution as $\{X_k\}$, $\{\tilde{Y}_k\}$ has the same distribution as $\{Y_k\}$, and there exists a random variable $\kappa<\infty$ a.s. such that $k\geq\kappa$ implies $\tilde{X}_k=\tilde{Y}_k$.

Proof of Lemma 2.18

Proof. Assume, without loss of generality, that $\epsilon(x,x)=1$, and define

$$\begin{aligned}\tilde{P}(x,y,\Gamma_1\times\Gamma_2) &= \epsilon(x,y)\nu_{xy}(\Gamma_1\cap\Gamma_2) \\ &\quad + \frac{(P(x,\Gamma_1)-\epsilon(x,y)\nu_{xy}(\Gamma_1))(P(y,\Gamma_2)-\epsilon(x,y)\nu_{xy}(\Gamma_2))}{1-\epsilon(x,y)},\end{aligned}$$

where the second term on the right is 0 if $\epsilon(x,y)=1$. Note that if $\{(\tilde{X}_k,\tilde{Y}_k)\}$ is a Markov chain with transition function $\tilde{P}(x,y,\Gamma)$, then $\{\tilde{X}_k\}$ and $\{\tilde{Y}_k\}$ are Markov chains with transition function P. Intuitively, at the kth transition a coin is flipped which is heads with probability $\epsilon(\tilde{X}_{k-1},\tilde{Y}_{k-1})$. If heads comes up, then $\tilde{X}_k=\tilde{Y}_k$ and both have conditional distribution $\nu_{\tilde{X}_{k-1}\tilde{Y}_{k-1}}$. If tails comes up, $\tilde{X}_k$ and $\tilde{Y}_k$ are conditionally independent with conditional distribution

$$\zeta(x,y,\Gamma_1\times\Gamma_2)=\frac{(P(x,\Gamma_1)-\epsilon(x,y)\nu_{xy}(\Gamma_1))(P(y,\Gamma_2)-\epsilon(x,y)\nu_{xy}(\Gamma_2))}{(1-\epsilon(x,y))^2},$$

where $x=\tilde{X}_{k-1}$ and $y=\tilde{Y}_{k-1}$.

To see that $\tilde{X}$ and $\tilde{Y}$ eventually couple, construct a Markov chain $(\tilde{X}, \tilde{Y}, X, Y)$ such that each component is a Markov chain with transition function P, X is independent of Y, $(\tilde{X}, \tilde{Y})$ has the transition function given above, and $(X_k, Y_k) = (\tilde{X}_k, \tilde{Y}_k)$ until the coin comes up heads. The desired one-step transition function is

$$\begin{aligned}\hat{P}(x, y, x', y', \Gamma_1 \times \Gamma_2 \times \Gamma_3 \times \Gamma_4) \;&=\; \nu_{xy}(\Gamma_1 \cap \Gamma_2)(P(x, \Gamma_3)P(y, \Gamma_4) \\ &\qquad -(1 - \epsilon(x, y))\zeta(x, y, \Gamma_3 \times \Gamma_4)) \\ &\qquad +(1 - \epsilon(x, y))\zeta(x, y, (\Gamma_1 \cap \Gamma_3) \times (\Gamma_2 \cap \Gamma_4))\end{aligned}$$

if $x = x'$ and $y = y'$, and

$$\hat{P}(x, y, x', y', \Gamma_1 \times \Gamma_2 \times \Gamma_3 \times \Gamma_4) = \tilde{P}(x, y, \Gamma_1 \times \Gamma_2)P(x', \Gamma_3)P(y', \Gamma_4)$$

otherwise. Under this transition function, if $\tilde{X}_0 = X_0$ and $\tilde{Y}_0 = Y_0$, then $\tilde{X}_k = X_k$ and $\tilde{Y}_k = Y_k$ until the first time that $\tilde{X}_k = \tilde{Y}_k$. Let $\kappa = \min\{k : \tilde{X}_k = \tilde{Y}_k\}$. Then

$$P\{\kappa > k\} \leq E[\prod_{i=0}^{k-1}(1-\epsilon(\tilde{X}_i, \tilde{Y}_i))] \leq E[\prod_{i=0}^{k-1}(1-\epsilon(X_i, Y_i))] \leq E[\exp\{-\sum_{i=0}^{k-1}\epsilon(X_i, Y_i)\}] \to 0$$

as $k \to \infty$. Here the second inequality follows from the fact that, for each $i \geq 0$, either $(\tilde{X}_i, \tilde{Y}_i) = (X_i, Y_i)$ or $\epsilon(\tilde{X}_i, \tilde{Y}_i) = 1$. $\square$

3. Stationary processes

- Stationary sequences
- Measure preserving transformation
- Ergodic theorem
- Ergodicity for Markov chains
- Mean ergodic theorem
- Subadditive ergodic theorem

Stationary sequences

$\{X_n\}$ is stationary if $P\{X_{m+n} \in A_0, \ldots, X_{m+n+k} \in A_k\}$ does not depend on n for any choice of $A_0, \ldots, A_k \in \mathcal{B}(E)$.

Examples:

- iid sequence
- Markov chain with transition function $P(x, C)$ and stationary distribution π and $X_0 \sim \pi$.
- $X_{n+1} = X_n + c \bmod 1$ and X_0 uniform $[0, 1]$.
- $X_{n+1} = 2X_n \bmod 1$ and X_0 uniform $[0, 1]$

Useful facts

Theorem 3.1 *If* $\{X_k, k \geq 0\}$ *is stationary, then there exists a sequence* $\{Y_k, -\infty < k < \infty\}$ *such that* $P\{(Y_n, \ldots, Y_{n+m}) \in C\} = P\{(X_0, \ldots, X_m) \in C\}$, $-\infty < n < \infty$, $m \geq 0$, $C \in \mathcal{B}(S^{m+1})$.

Theorem 3.2 *If* $\{X_k, k \geq 0\}$ *is a stationary sequence and* $g : S^\infty \to \hat{S}$ *is measurable, then* $Z_k = g(X_k, X_{k+1}, \ldots)$ *is stationary.*

If $\{Y_k, -\infty < k < \infty\}$ *is stationary and* $g : S^{\mathbb{Z}} \to \hat{S}$, *then* $Z_k = g(\ldots, Y_{k-1}, Y_k, Y_{k+1} \ldots)$ *is stationary.*

Example

Let $\{\xi_k\}$ be iid real-valued with $E[\xi_k] = 0$ and $Var(\xi_k) < \infty$. Suppose $\sum_{k=0}^{\infty} a_k^2 < \infty$ and

$$Z_k = \sum_{l=0}^{\infty} a_l \xi_{k-l}$$

If $a_l = \rho^l$ with $|\rho| < 1$, then $Z_{k+1} = \rho Z_k + \xi_{k+1}$. (In this case, second moments aren't needed.)

Measure-preserving transformations

$\varphi : \Omega \to \Omega$ is measureable iff $\varphi^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{F}$.

A measurable transformation is measure preserving iff $P(\varphi^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$.

Lemma 3.3 *If φ is measure perserving, Z a random variable and $X_n(\omega) \equiv Z \circ \varphi^n(\omega)$, then $\{X_n\}$ is a stationary sequence.*

Proof.

$$\begin{aligned} P\{X_n \in A\} &= P\{\omega : Z \circ \varphi^n(\omega) \in A\} \\ &= P\{\omega : \varphi(\omega) \in \{\tilde{\omega} : Z \circ \varphi^{n-1}(\tilde{\omega}) \in A\}\} = P\{X_{n-1} \in A\} \end{aligned}$$

□

Conversely, $\Omega = E^\infty$ $\mathcal{F} = \mathcal{B}(E^\infty)$, P the joint distribution of a stationary sequence $\{X_n\}$. (We can identify X_n with the mapping $X_n(x_0, x_1, \ldots) = x_n$.) $\varphi(x_0, x_1, \ldots) = (x_1, x_2, \ldots)$. Then

$$P(A) = P\{(X_0, X_1 \ldots) \in A\} = P\{(X_1, X_2, \ldots) \in A\} = P(\varphi^{-1}(A))$$

Invariant sets (or almost surely invariant sets)

Let $\mathcal{I} = \{A : P(A \triangle \varphi^{-1}(A)) = 0\}$. $\mathcal{I}$ is the collection of (almost surely) *invariant sets.*

Lemma 3.4 *$\mathcal{I}$ is a σ-algebra. X is $\mathcal{I}$-measurable iff $X \circ \varphi = X$ a.s.*

Ergodicity

φ is *ergodic* if and only if $A \in \mathcal{I}$ implies $P(A) = 0$ or 1.

Lemma 3.5 *If $\{Y_k\}$ is ergodic, then $Z_k = g(\ldots, Y_{k-1}, Y_k, Y_{k+1} \ldots)$ is ergodic.*

A maximal inequality

Lemma 3.6 *Let* $\{X_n\}$ *be stationary and define* $S_k = \sum_{i=0}^{k-1} X_i$

$$M_k = \max\{0, S_1, \ldots, S_k\}.$$

Then $E[X_0 \mathbf{1}_{\{M_k>0\}}] \geq 0$*.*

Proof. If $j \leq k$, then $X_0 + 0 \vee \max_{1\leq l\leq k} \sum_{i=1}^{l} X_i \geq S_{j+1}$, so

$$X_0 \geq S_{j+1} - 0 \vee \max_{1\leq l\leq k} \sum_{i=1}^{l} X_i$$

Consequently,

$$\begin{aligned} E[X_0 \mathbf{1}_{\{M_k>0\}}] &\geq \int_{\{M_k>0\}} (\max_{1\leq l\leq k} S_l - 0 \vee \max_{1\leq l\leq k} \sum_{i=1}^{l} X_i) dP \\ &= \int_{\{M_k>0\}} (M_k - 0 \vee \max_{1\leq l\leq k} \sum_{i=1}^{l} X_i) dP \geq 0 \end{aligned}$$

□

Shift invariant sets

$C \in \mathcal{B}(E^\infty)$ is shift invariant if $(x_0, x_1, \ldots) \in C$ implies $(x_1, x_2, \ldots) \in C$.

Lemma 3.7 *If $A \in \sigma(\{X_n\})$ is invariant, then there exists a shift invariant C such that $P(A \triangle \{(X_0, X_1, \ldots) \in C\}) = 0$.*

Proof. If $A \in \sigma(\{X_n\})$, there exists $\hat{C} \in \mathcal{B}(E^\infty)$ such that $A = \{(X_0, \ldots) \in \hat{C}\}$. Define $C = \cap_n \cup_{m>n} \{x : (x_m, x_{m+1}, \ldots) \in \hat{C}\}$. Then C is shift invariant and $P(A \triangle \{(X_0, \ldots) \in C\}) = 0$. □

Lemma 3.8 *Let* $\mathcal{I}_0 = \{\{(X_0, \ldots) \in C\} : C$ shift invariant$\}$. *Then*

$$E[X_0 | \mathcal{I}_0] = E[X_n | \mathcal{I}_0]$$

Ergodic theorem

Theorem 3.9 *Let* $\{X_n\}$ *be stationary and* $E[|X_n|] < \infty$. *Then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} X_k = E[X_1|\mathcal{I}_0] \quad a.s. \text{ and in } L^1.$$

Proof. wlog assume $E[X_1|\mathcal{I}_0] = 0$. Define $\bar{X} = \limsup \frac{1}{n} S_n$ and for $\epsilon > 0$, set $D = \{\bar{X} > \epsilon\} \in \mathcal{I}_0$. Define $X_n^* = (X_n - \epsilon)\mathbf{1}_D$. Let

$$M_n^* = \max\{0, S_1^*, \ldots, S_n^*\} \quad F_n = \{M_n^* > 0\} \quad F = \cup F_n = \{\sup_k \frac{1}{k} S_k^* > 0\} = D$$

Consequently,

$$0 \leq \int_D X_0^* dP = \int_D (X_0 - \epsilon) dP = \int_D E[X_0|\mathcal{I}_0] dP - \epsilon P(D)$$

Uniform integrability implies the convergence is in L^1. □

Conditions for ergodicity

Lemma 3.10 *$\{X_n\}$ is ergodic if and only if*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} f(X_k, X_{k+1}, \ldots, X_{k+m}) = E[f(X_0, \ldots, X_m)]$$

for all bounded, measurable f on E^m and all m. (All bounded countinuous functions will also work.)

Proof. Necessity is immediate since $\mathcal{I}_f \subset \mathcal{I}_0$. Let $\mathcal{G} \subset \mathcal{B}(E^\infty)$ be the collection of $C \in \mathcal{B}(E^\infty)$ such that

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_C(X_k, \ldots) = P\{(X_0, \ldots) \in C\} \quad a.s.$$

Then $\mathcal{G}$ is a Dynkin class. ($E^\infty \in \mathcal{G}$, $A, B \in \mathcal{G}$ and $A \subset B$ implies $B - A \in \mathcal{G}$, $C_1 \subset C_2 \subset \cdots \in \mathcal{G}$ implies $\cup C_n \in \mathcal{G}$)

$C = B_1 \times B_2 \times \cdots \times B_m \times S \times S \cdots \in \mathcal{G}$. $\square$

Ergodicity for Markov chains

Lemma 3.11 *If $\{X_n\}$ is a stationary Markov chain. Then $\{X_n\}$ is ergodic if and only if*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = E[f(X_0)] \quad a.s.$$

for all bounded measurable f.

Proof. By the law of large numbers for martingales,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (f(X_k, X_{k+1}) - \int_S f(X_k, z) P(X_k, dz)) = 0,$$

and the conditions of the previous lemma follow by induction. □

Ergodicity and uniqueness of stationary distribution

Theorem 3.12 *If* $P(x, C)$ *has a unique stationary distribution* π, *then for* $X_0 \sim \pi$, $\{X_n\}$ *is ergodic.*

Proof. Suppose that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to Z \quad a.s.$$

Note that $E[Z|\mathcal{F}_0] = E[Z|X_0] \equiv h(X_0)$, but then $E[Z|\mathcal{F}_1] = h(X_1)$. Consequently, $Ph(X_0) = h(X_0)$ and

$$E[(h(X_1) - h(X_0))^2] = E[h^2(X_1)] + E[h^2(X_0)] - 2E[h(X_1)h(X_0)] = 0.$$

But by induction

$$E[Z|\mathcal{F}_n] = h(X_n) = h(X_0),$$

so $Z = h(X_0)$. Let

$$\pi_0(C) = \frac{E[\mathbf{1}_C(X_0)h(X_0)]}{E[h(X_0)]}.$$

Then

$$E_{\pi_0}[g(X_1)] = \frac{E[g(X_1)h(X_0)]}{E[h(X_0)]} = \frac{E[g(X_1)h(X_1)]}{E[h(X_1)]} = \frac{E[g(X_0)h(X_0)]}{E[h(X_0)]} = \int g d\pi_0,$$

so π_0 is a stationary distribution for $P(x, C)$ and hence must equal π. But that implies h (and hence Z) is constant a.s. π. $\square$

Irreducibility implies uniqueness

Theorem 3.13 *If E is countable and $\{X_n\}$ is an irreducible Markov chain, then there is at most one stationary distribution.*

Proof. If $\{X_n\}$ is stationary and $A \in \mathcal{B}(E)$, then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_A(X_n) = E[\mathbf{1}_A(X_0)|\mathcal{I}_0]$$

and hence

$$\{X_n \in A \text{ i.o.}\} \supset \{E[\mathbf{1}_A(X_0)|\mathcal{I}_0] > 0\}$$

Since $\{X_n = x \text{ i.o.}\}$ has probability 0 or 1, if there is a stationary distribution, then every state is recurrent. Consequently, the strong Markov property implies that the distribution of

$$Z_A = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mathbf{1}_A(X_n)$$

does not depend on the distribution of X_0, and since $E[Z_A] = \pi(A)$, there is only one stationary distribution. □

The collection of stationary distributions

Note that the collection of stationary distributions Π is convex.

Two measures μ and ν are *mutually singular* if there exists a measurable set A such that $\mu(A) = 0$ and $\nu(A^c) = 0$.

Theorem 3.14 *If π_1 and π_2 are stationary distributions with $\pi_1 \neq \pi_2$, then there exist two mutually singular stationary distributions.*

Proof. Let $\pi = \frac{1}{2}\pi_1 + \frac{1}{2}\pi_2$. Then π is a stationary distribution. Let $f \in B(E)$ satisfy $\int f d\pi_1 \neq \int f d\pi_2$. Let X^{π_1} be a Markov chain with initial distribution π_1 and X^{π_2} be a Markov chain with initial distribution π_2. Let ξ be independent of X^{π_1} and X^{π_2} and $P\{\xi = 1\} = 1 - P\{\xi = 0\} = \frac{1}{2}$. Define

$$X_n^{\pi} = \begin{cases} X_n^{\pi_1} & \xi = 1 \\ X_n^{\pi_2} & \xi = 0 \end{cases}.$$

Then X^{π} is a Markov chain with initial distribution π. Let

$$h(X_0) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k).$$

Then $\int h d\pi_1 = \int f d\pi_1 \neq \int f d\pi_2 = \int h d\pi_2$, so h is not constant a.s. π. Let

$$0 < \pi\{h > \beta\} < 1.$$

Define

$$\tilde{\pi}_1(\Gamma) = \frac{E[\mathbf{1}_\Gamma(X_0)\mathbf{1}_{\{h(X_0)>\beta\}}]}{P\{h(X_0) > \beta\}} \qquad \tilde{\pi}_2(\Gamma) = \frac{E[\mathbf{1}_\Gamma(X_0)\mathbf{1}_{\{h(X_0)\leq\beta\}}]}{P\{h(X_0) \leq \beta\}}. \tag{3.1}$$

Then $\tilde{\pi}_i$ is a stationary distribution as in the proof of Theorem 3.12, and $\tilde{\pi}_1$ and $\tilde{\pi}_2$ are mutually singular. □

Extremal stationary distributions

A stationary distribution π is *extremal* if and only if it cannot be represented as $\pi = \alpha\pi_1 + (1-\alpha)\pi_2$ for $0 < \alpha < 1$ and $\pi_1, \pi_2 \in \Pi$.

Corollary 3.15 *If π is an extremal stationary distribution, then X^π is ergodic.*

Proof. With reference to the proof of the previous theorem, if

$$h(X_0^\pi) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f(X_k^\pi)$$

is not constant, then defining $\tilde{\pi}_1$ and $\tilde{\pi}_2$ as in (3.1) and $\alpha = P\{h(X_0) > \beta\}$,

$$\pi = \alpha\tilde{\pi}_1 + (1-\alpha)\tilde{\pi}_2.$$

□

Mixing

If $\varphi : \Omega \to \Omega$ is an ergodic measure preserving transformation, then

$$\frac{1}{n}\sum_{k=1}^{n} P(A \cap \varphi^{-k}B) = E[\mathbf{1}_A \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_B \circ \varphi^k] \to P(A)P(B), \quad \forall A, B \in \mathcal{F}. \tag{3.2}$$

Note that this condition is sufficient for ergodicity also.

φ is called *mixing* if the stronger condition

$$\lim_{n\to\infty} P(A \cap \varphi^{-n}B) = P(A)P(B), \quad \forall A, B \in \mathcal{F} \tag{3.3}$$

holds.

The collection of B (A) for which (3.2) holds is a Dynkin class and similarly for (3.3).

Applications

Theorem 3.16 *$\{X_n\}$ stationary in $\mathbb{R}^d$. $S_n = \sum_{k=1}^n X_k$, R_n = number of distinct values in $\{S_1, \ldots, S_n\}$. Let $C = \{x : x_1 \neq 0, x_1 + x_2 \neq 0, \ldots\}$. Then*

$$\lim_{n\to\infty} \frac{R_n}{n} = E[\mathbf{1}_C(X_1, X_2, \ldots)|\mathcal{I}_0]$$

Proof. First, $R_n \geq \sum_{k=1}^n \mathbf{1}_C(X_k, X_{k+1}, \ldots)$, so

$$\liminf \frac{R_n}{n} \geq E[\mathbf{1}_C(X_1, X_2, \ldots)|\mathcal{I}_0].$$

$C_l = \{x : x_1 \neq 0, \ldots, x_1 + \cdots + x_l \neq 0\}$

$$R_n \leq l + \sum_{k=1}^{n-l} \mathbf{1}_{C_l}(X_k, \ldots, X_{k+l-1})$$

so

$$\limsup \frac{R_n}{n} \leq E[\mathbf{1}_{C_l}(X_1, X_2, \ldots)|\mathcal{I}_0].$$

□

Recurrence

Note that $C^c = \{x : \sum_{i=1}^k x_i = 0 \text{ some } k \geq 1\}$, $n^{-1}R_n \to 0$ a.s. implies that for $S_k = \sum_{i=1}^k X_i$, $P\{S_k = 0, \text{ some } k \geq 1\} = 1$, but then $P\{\sum_{i=m+1}^{m+k} X_i = 0, \text{ some } k \geq 1\} = 1$, so $P\{S_k = 0 \text{ i.o.}\} = 1$.

For $d = 1$,

Theorem 3.17 *If $E[X_0|\mathcal{I}_0] = 0$ a.s., then $P\{S_k = 0 \text{ i.o.}\} = 1$.*

Proof. Since $\lim_{n\to\infty} n^{-1}S_n = 0$ a.s., implies,

$$\lim_{n\to\infty} n^{-1} \max_{k\leq n} |S_k| = 0,$$

$\lim_{n\to\infty} n^{-1}R_n = 0$. □

Entropy

E finite, $\{X_n\}$ stationary and ergodic, $p(x_0, \ldots, x_n) = P\{X_0 = x_0, \ldots, X_n = x_n\}$
$p(x_n|x_{n-1}, \ldots, x_0) = P(X_n = x_n|X_{n-1} = x_{n-1}, \ldots, X_0 = x_0)$

Assume that $\{X_n\}$ is stationary for $n \in \mathbb{Z}$, and define $\mathcal{F}_n = \sigma(X_n, X_{n-1}, \ldots)$. Let

$$p(x|X_{n-1}, X_{n-2}, \ldots) = E[\mathbf{1}_{\{X_n=x\}}|\mathcal{F}_{n-1}] = \lim_{m\to\infty} p(x|X_{n-1}, \ldots X_{n-m}).$$

Then $p(X_n|X_{n-1}, X_{n-2}, \ldots)$ is stationary and

$$H = -\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \log p(X_k|X_{k-1}, X_{k-2}, \ldots)$$

exists.

Shannon-McMillan-Breiman theorem

Theorem 3.18

$$H = -\lim_{n\to\infty} \frac{1}{n} \log p(X_0, \ldots, X_{n-1}) = -\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n-1} \log p(X_k | X_{k-1}, \ldots, X_0)$$

Proof. Let

$$H_n^0 = -\frac{1}{n} \sum_{k=1}^{n-1} \log p(X_k | X_{k-1}, \ldots, X_0)$$

and

$$H_n = -\frac{1}{n} \sum_{k=0}^{n-1} \log p(X_k | X_{k-1}, X_{k-2}, \ldots).$$

Then

$$H_n - H_n^0 = \frac{1}{n} \sum_{k=1}^{n-1} \log \frac{p(X_k | X_{k-1}, \ldots, X_0)}{p(X_k | X_{k-1}, X_{k-2}, \ldots)} \to 0$$

at least in probability. □

Mean ergodic theorem

If we replace almost sure and L^1-convergence by L^2-convergence in the statement of the ergodic theorem, there is a much simpler proof. Define $TX = X \circ \varphi$, $X \in L^2(P)$ and let $H_{\mathcal{I}} = \{X \in L^2(P) : TX = X \text{ a.s.}\}$.

$E[Y|\mathcal{I}_0] = P_{\mathcal{I}}Y$ a.s., where $P_{\mathcal{I}}Y$ is the projection, in the Hilbert space sense, of Y onto $H_{\mathcal{I}}$.

$H_{\mathcal{I}}^{\perp} = \{X - TX : X \in L^2(P)\}$, so $Z \in L^2(P)$ can be written as $Z = X - TX + Y$ where $Y = P_{\mathcal{I}}Z \in H_{\mathcal{I}}$. Consequently,

$$\sum_{k=0}^{n-1} Z \circ \varphi^k = X - T^n X + nY \quad a.s.$$

It follows immediately that

$$E[|n^{-1}\sum_{k=0}^{n-1} Z \circ \varphi^k - P_{\mathcal{I}}Z|^2] \to 0.$$

Maximal ergodic theorem

For $Z \in L^1(P)$, define

$$Z^* = \sup_n \frac{1}{n} \sum_{k=1}^{n} |Z \circ \varphi^k|$$

Theorem 3.19 *There exists $A > 0$ such that for each $Z \in L^1(P)$,*

$$P\{Z^* > \alpha\} \leq \frac{A}{\alpha} E[|Z|].$$

Proof. Let $A_n = n^{-1} \sum_{k=1}^{n} |Z \circ \varphi^k|$ The maximal inequality implies

$$E[(|Z| - \alpha)\mathbf{1}_{\{\max_{1\leq k\leq n}(A_k - \alpha) > 0\}}] \geq 0$$

so

$$\alpha P\{\max_{1\leq k\leq n} A_k > \alpha\} \leq E[|Z|]$$

□

Almost sure convergence

If X and Y are bounded by constants and Y is in $H_{\mathcal{I}}$, then for $Z = X - TX + Y$,

$$\frac{1}{n}\sum_{k=0}^{n-1} Z \circ \varphi^k \to 0 \quad a.s.$$

But Z of this form is dense in $L^2(P)$ and hence in $L^1(P)$, and the maximal ergodic theorem implies almost sure convergence for all $Z \in L^1(P)$.

Multiparameter ergodic theorem

Theorem 3.20 *Suppose that the joint distribution of* $\{X_{i+m,j+n}, i,j \in \mathbb{Z}\}$ *does not depend on* m *and* n*, and suppose* $E[|X_{0,0}|] < \infty$*. Define*

$$A_{n,m} = \frac{1}{nm}\sum_{i=0}^{n-1}\sum_{j=0}^{m-1} X_{i,j}.$$

Then there exists $\bar{X}$ *such that*

$$\lim_{n,m\to\infty} E[|A_{n,m} - \bar{X}|] = 0$$

Subadditive ergodic theorem

Theorem 3.21 *Suppose* $\{X_{m,n}, 0 \leq m < n\}$ *satisfies*

i) $X_{0,m} + X_{m,n} \geq X_{0,n}$

ii) For each $k = 1, 2, \ldots$*,* $\{X_{nk,(n+1)k}, n \geq 1\}$ *is stationary.*

iii) The joint distribution of $\{X_{m,m+k}, k \geq 1\}$ *does not depend on* m*.*

iv) $E[X_{0,1}^+] < \infty$*, and there exists* $\gamma_0 > -\infty$ *such the* $E[X_{0,n}] \geq \gamma_0 n$*.*

Then

a) $\lim_{n\to\infty} n^{-1}E[X_{0,n}] = \inf_m m^{-1}E[X_{0,m}] \equiv \gamma$*.*

b) $X = \lim_{n\to\infty} n^{-1}X_{0,n}$ *exists a.s. and in* L^1*.*

c) If all the stationary sequences in (ii) are ergodic, then $X = \gamma$ *a.s.*

Examples

Maximum: Let $\{Y_i\}$ be stationary with $E[|Y_i|] < \infty$, and define

$$X_{m,n} = (\max_{m<k\leq n} \sum_{i=m+1}^{k} Y_i) \vee 0.$$

Range: $\{Y_n\}$ stationary in $\mathbb{R}^d$. $S_n = \sum_{k=1}^n Y_k$, $X_{m,n}$ = number of distinct values in $\{S_{m+1}, \ldots, S_n\}$. ($R_n = X_{0,n}$.)

Longest common subsequences: $\{(X_i, Y_i)\}$ stationary.

$$L_{m,n} = \max\{K : \exists m < i_1 < i_2 < \cdots < i_K \leq n, m < j_1 < j_2 < \cdots < j_K \leq n, X_{i_k} = Y_{j_k}\}$$

Random permutations

Let Σ_n be the collection of all permutations of $(1, 2, \ldots, n)$, and let $Z^n = (Z_1^n, \ldots, Z_n^n)$ be a uniform draw from this set

For example, if $\{\xi_i, 1 \leq i \leq n\}$ are iid uniform, then we can let Z_k^n be the index l such that $\xi_l = \xi_{(k)}^n$, the kth order statistic.

Let

$$L_n = \max\{K : i_1 < i_2 < \cdots < i_K \leq n, Z_{i_1}^n < \cdots < Z_{i_K}^n\}$$

Poisson construction

Let η be a Poisson random measure on $[0,\infty) \times [0,\infty)$ with mean Lebesgue mean measure.

Let $\tau(n) = \inf\{t : \eta([0,t] \times [0,t]) \geq n+1\}$

Order the points (X_k^n, Y_k^n) in the square so that $X_1^n < X_2^n < \cdots < X_n^n$. Then $\xi_k^n = \tau(n)^{-1} Y_k^n$ are iid uniform $[0,1]$.

Consequently, L_n is the length of the longest (in the sense of number of points connected) increasing path in the square $[0,\tau(n)) \times [0,\tau(n))$.

Let $R_{s,t}$ be the length of the longest (in the same sense) increasing path in the square $[s,t) \times [s,t)$. Then $R_{0,s} + R_{s,t} \leq R_{0,t}$.

4. Continuous time stochastic processes

- Measurability for stochastic processes
- Stopping times
- A process observed at a stopping time
- Right continuous processes are progressive
- Approximation of a stopping time by discrete stopping times
- Right-continuous filtrations

Measurability for stochastic processes

A stochastic process is an indexed family of random variables, but if the index set is $[0,\infty)$, then we may want to know more about $X(t,\omega)$ than that it is a measurable function of ω for each t. For example, for a $\mathbb{R}$-valued process X, when are

$$\int_a^b X(s,\omega)ds \quad \text{and} \quad X(\tau(\omega),\omega)$$

random variables?

X is *measurable* if $(t,\omega)\in[0,\infty)\times\Omega\rightarrow X(t,\omega)\in E$ is $\mathcal{B}([0,\infty))\times\mathcal{F}$-measurable.

Lemma 4.1 *If X is measurable and $\int_a^b|X(s,\omega)|ds<\infty$, then $\int_a^b X(s,\omega)ds$ is a random variable.*

If, in addition, τ is a nonnegative random variable, then $X(\tau(\omega),\omega)$ is a random variable.

Proof. The first part is a standard result for measurable functions on a product space. Verify the result for $X(s, \omega) = \mathbf{1}_A(s)\mathbf{1}_B(\omega)$, $A \in \mathcal{B}[0, \infty)$, $B \in \mathcal{F}$ and apply the Dynkin class theorem to extend the result to $\mathbf{1}_C$, $C \in \mathcal{B}[0, \infty) \times \mathcal{F}$.

If τ is a nonnegative random variable, then $\omega \in \Omega \to (\tau(\omega), \omega) \in [0, \infty) \times \Omega$ is measurable. Consequently, $X(\tau(\omega), \omega)$ is the composition of two measurble functions. □

Measurability continued

A stochastic process X is $\{\mathcal{F}_t\}$-adapted if for all $t \geq 0$, $X(t)$ is $\mathcal{F}_t$-measurable.

If X is measurable and adapted, the restriction of X to $[0,t] \times \Omega$ is $\mathcal{B}[0,t] \times \mathcal{F}$-measurable, but it may not be $\mathcal{B}[0,t] \times \mathcal{F}_t$-measurable.

X is *progressive* if for each $t \geq 0$, $(s,\omega) \in [0,t] \times \Omega \to X(s,\omega) \in E$ is $\mathcal{B}[0,t] \times \mathcal{F}_t$-measurable.

Let

$$\mathcal{W} = \{A \in \mathcal{B}[0,\infty) \times \mathcal{F} : A \cap [0,t] \times \Omega \in \mathcal{B}[0,t] \times \mathcal{F}_t, t \geq 0\}.$$

Then $\mathcal{W}$ is a σ-algebra and X is progressive if and only if $(s,\omega) \to X(s,\omega)$ is $\mathcal{W}$-measurable.

Since pointwise limits of measurable functions are measurable, pointwise limits of progressive processes are progressive.

Stopping times

Let $\{\mathcal{F}_t\}$ be a filtration. τ is a $\mathcal{F}_t$-*stopping time* if and only if $\{\tau \leq t\} \in \mathcal{F}_t$ for each $t \geq 0$.

If τ is a stopping time, $\mathcal{F}_\tau \equiv \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, t \geq 0\}$.

If τ_1 and τ_2 are stopping times with $\tau_1 \leq \tau_2$, then $\mathcal{F}_{\tau_1} \subset \mathcal{F}_{\tau_2}$.

If τ_1 and τ_2 are stopping times then τ_1 and $\tau_1 \wedge \tau_2$ are $\mathcal{F}_{\tau_1}$-measurable.

A process observed at a stopping time

If X is measurable and τ is a stopping time, then $X(\tau(\omega),\omega)$ is a random variable.

Lemma 4.2 *If τ is a stopping time and X is progressive, then $X(\tau)$ is $\mathcal{F}_\tau$-measurable.*

Proof. $\omega\in\Omega\to(\tau(\omega)\wedge t,\omega)\in[0,t]\times\Omega$ is measurable as a mapping from $(\Omega,\mathcal{F}_t)$ to $([0,t]\times\Omega,\mathcal{B}[0,t]\times\mathcal{F}_t)$. Consequently, $\omega\to X(\tau(\omega)\wedge t,\omega)$ is $\mathcal{F}_t$-measurable, and

$$\{X(\tau)\in A\}\cap\{\tau\le t\}=\{X(\tau\wedge t)\in A\}\cap\{\tau\le t\}\in\mathcal{F}_t.$$

□

Right continuous processes

Most of the processes you know are either continous (e.g., Brownian motion) or right continuous (e.g., Poisson process).

Lemma 4.3 *If X is right continuous and adapted, then X is progressive.*

Proof. If X is adapted, then

$$(s,\omega) \in [0,t] \times \Omega \to Y_n(s,\omega) \equiv X(\frac{[ns]+1}{n} \wedge t, \omega) = \sum_k X(\frac{k+1}{n} \wedge t, \omega) \mathbf{1}_{[\frac{k}{n}, \frac{k+1}{n})}(s)$$

is $\mathcal{B}[0,t] \times \mathcal{F}_t$-measurable. By the right continuity of X, $Y_n(s,\omega) \to X(s,\omega)$ on $[0,t] \times \mathcal{F}_t$, so $(s,\omega) \in [0,t] \times \Omega \to X(s,\omega)$ is $[0,t] \times \mathcal{F}_t$-measurable and X is progressive. □

More on stopping times

Lemma 4.4 *Let* τ *be a nonnegative random variable. If* $\{\tau < t\} \in \mathcal{F}_t$*,* $t \geq 0$*, then there exists a sequence of stopping times* $\tau_n \geq \tau$ *such that* $\lim_{n\to\infty} \tau_n = \tau$*.*

Proof. Define

$$\tau_n = \frac{k+1}{2^n} \quad \text{on} \quad \{\frac{k}{2^n} \leq \tau < \frac{k+1}{2^n}\}. \tag{4.1}$$

Then $\tau_n > \tau$ on $\{\tau < \infty\}$, and

$$\{\tau_n \leq t\} = \{\tau_n \leq \frac{[2^n t]}{2^n}\} = \{\tau < \frac{[2^n t]}{2^n}\} \in \mathcal{F}_t.$$

□

Example: Optional sampling theorem

For a discrete time $\{\mathcal{F}_n\}$-martingale $\{M_n\}$, the optional sampling theorem states that if τ_1 and τ_2 are stopping times, then

$$E[M_{n\wedge\tau_2}|\mathcal{F}_{\tau_1}] = M_{n\wedge\tau_1\wedge\tau_2}.$$

Suppose M is a right-continuous $\{\mathcal{F}_t\}$-martingale. For $t \geq 0$, let $t_n = \frac{[2^n t]+1}{2^n}$. The restriction of M to $\{\frac{k}{2^n}, k = 0, 1, 2, \ldots\}$ gives a discrete-time martingale, so defining $\tau_{i,n}$ as in (4.1),

$$E[M(t_n \wedge \tau_{2,n})|\mathcal{F}_{\tau_{1,n}}] = M(t_n \wedge \tau_{1,n} \wedge \tau_{2,n})$$

and

$$E[M(t_n \wedge \tau_{2,n})|\mathcal{F}_{\tau_1}] = E[M(t_n \wedge \tau_{1,n} \wedge \tau_{2,n})|\mathcal{F}_{\tau_1}].$$

By the right continuity of M and the fact that $\{M(t_n\wedge\tau_{2,n}), M(t_n\wedge\tau_{1,n}\wedge\tau_{2,n}), n \geq 1\}$ is uniformly integrable (why?),

$$E[M(t \wedge \tau_2)|\mathcal{F}_{\tau_1}] = E[M(t \wedge \tau_1 \wedge \tau_2)|\mathcal{F}_{\tau_1}] = M(t \wedge \tau_1 \wedge \tau_2).$$

Right continuous filtrations

If $\mathcal{F}_t = \mathcal{F}_{t+} \equiv \cap_{s>t}\mathcal{F}_s$, $t \geq 0$, the filtration is *right continuous*.

If $\{\mathcal{F}_t\}$ is right continuous, then τ is a stopping time if and only if $\{\tau < t\} \in \mathcal{F}_t$, $t \geq 0$.

If $\{\mathcal{F}_t\}$ is right continuous and $\{\tau_n\}$ are stopping times, then $\inf_n \tau_n$ is a stopping time, since

$$\{\inf_n \tau_n < t\} = \cup_n\{\tau_n < t\}.$$

Example: First entrance time of an open set

Let X be a right-continuous, $\{\mathcal{F}_t\}$-adapted process, and let $O \subset E$ be open. Define

$$\tau = \inf\{t \geq 0 : X(t) \in O\} \qquad \tau_n = \min\{\frac{k}{2^n} : X(\frac{k}{2^n}) \in O\}.$$

Then τ_n is an $\{\mathcal{F}_t\}$-stopping time and $\tau = \inf_n \tau_n$. Consequently, τ is an $\{\mathcal{F}_{t+}\}$-stopping time but may not be an $\{\mathcal{F}_t\}$-stopping time.

Projections from product spaces

For $A \subset \mathbb{R}^2$, define $\pi_1 A = \{x : \exists (x,y) \in A\}$. If $A \in \mathcal{B}(\mathbb{R}^2)$, then $\pi_1 A$ need not be in $\mathcal{B}(\mathbb{R})$.

$\Gamma_1 = \{A \in \mathcal{B}(\mathbb{R}^2) : \pi_1 A \in \mathcal{B}(\mathbb{R})\}$ is not a Dynkin class.

Γ_1 is closed under countable unions but not intersections or complements.

Projections onto complete probability spaces

Theorem 4.5 *Let* $(\Omega, \mathcal{F}, P)$ *be a complete probability space, and let* S *be a locally compact, separable metric space. Suppose* $A \in \mathcal{B}(S) \times \mathcal{F}$*. Then* $\pi_\Omega A \in \mathcal{F}$*.*

Proof. See Theorem T32 of Dellacherie (1972). □

Debut theorem

A filtration $\{\mathcal{F}_t\}$ is *complete* if $\mathcal{F}_0$ contains all subsets of sets of probability zero.

Theorem 4.6 *Let $(\Omega, \mathcal{F}, P)$ be a complete probability space, and let $\{\mathcal{F}_t\}$ be a complete, right-continuous filtration. If A is progressive, then $\tau(\omega) = \inf\{t : (t, \omega) \in A\}$ is a $\{\mathcal{F}_t\}$-stopping time.*

Proof. By the right-continuity of $\{\mathcal{F}_t\}$, we only need to verify that $\{\tau < t\} \in \mathcal{F}_t$. But since $(\Omega, \mathcal{F}_t, P)$ is a complete probability space and

$$\{\omega : \tau(\omega) < t\} = \pi_\Omega(A \cap [0, t) \times \Omega),$$

by Theorem 4.5, $\{\tau < t\} \in \mathcal{F}_t$. □

Further notions of measurability

If X is right continuous and adapted, then X is progressive.

Consequently, $\mathcal{O} = \sigma(X : X \text{ right continuous and adapted}) \subset \mathcal{B}[0,\infty) \times \mathcal{F}$ is a sub-σ-algebra of $\mathcal{W}$. $\mathcal{O}$ is the σ-algebra of *optional* sets.

Similarly, $\mathcal{P} = \sigma(X : X \text{ continuous and adapted}) \subset \mathcal{B}[0,\infty) \times \mathcal{F}$ is the σ-algebra of *predictable* sets.

Clearly, $\mathcal{P} \subset \mathcal{O} \subset \mathcal{W}$.

5. Martingales

- Definitions
- Optional sampling theorem
- Doob's inequalities
- Upcrossing inequality
- Martingale convergence theorem
- Martingales and finance

Definitions

Let X be a $\{\mathcal{F}_t\}$-adapted process.

X is a *submartingale* if

$$E[X(t+s)|\mathcal{F}_t] \geq X(t), \quad t, s, \geq 0.$$

X is a *supermartingale* if

$$E[X(t+s)|\mathcal{F}_t] \leq X(t), \quad t, s, \geq 0.$$

X is a *martingale* if

$$E[X(t+s)|\mathcal{F}_t] = X(t), \quad t, s, \geq 0.$$

Applications of Jensen's inequality

If φ is convex, Y is a martingale, and $E[|\varphi(Y(t))|] < \infty$, $t \geq 0$, then $X(t) = \varphi(Y(t))$ is a submartingale.

If φ is convex and nondecreasing, Y is a submartingale, and $E[|\varphi(Y(t))|] < \infty$, $t \geq 0$, then $X(t) = \varphi(Y(t))$ is a submartingale. In particular, if Y is a submartingale, then $X(t) = Y(t) \vee c$ is a submartingale.

Optional sampling theorem

Lemma 5.1 *Let X be a right-continuous submartingale, τ_1 a stopping time assuming values in a countable set $t_1 < t_2 < \cdots$ and τ_2 a stopping time assuming values in the finite set $t_1 < \cdots < t_m$. Then*

$$E[X(\tau_2)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2)$$

Proof. Recall that

$$E[X(\tau_2)|\mathcal{F}_{\tau_1}] = \sum_{t_i} E[X(\tau_2)|\mathcal{F}_{t_i}]\mathbf{1}_{\{\tau_1 = t_i\}}.$$

Then for $i \geq m$, $E[X(\tau_2)|\mathcal{F}_{t_i}] = X(\tau_2)$, and

$$\begin{aligned} E[X(\tau_2)|\mathcal{F}_{t_{m-1}}] &= E[\mathbf{1}_{\{\tau_2 = t_m\}} X(t_m) + \mathbf{1}_{\{\tau_2 \leq t_{m-1}\}} X(\tau_2 \wedge t_{m-1})|\mathcal{F}_{t_{m-1}}] \\ &\geq \mathbf{1}_{\{\tau_2 = t_m\}} X(t_{m-1}) + \mathbf{1}_{\{\tau_2 \leq t_{m-1}\}} X(\tau_2 \wedge t_{m-1}) \\ &= X(\tau_2 \wedge t_{m-1}), \end{aligned}$$

so by induction on m,

$$E[X(\tau_2)|\mathcal{F}_{t_i}] \geq X(\tau_2 \wedge t_i)$$

and the lemma follows. □

Theorem 5.2 *Let X be a right-continuous submartingale, and τ_1 and τ_2 be stopping times. Then*

$$E[X(\tau_2 \wedge t)|\mathcal{F}_{\tau_1}] \geq X(\tau_1 \wedge \tau_2 \wedge t) \tag{5.1}$$

Proof. Taking $\tau_{1,n}$ and $\tau_{2,n}$ as in the optional sampling theorem example, and using the fact that $X(t) \vee c$ is a submartingale,

$$E[X(\tau_{2,n} \wedge t) \vee c|\mathcal{F}_{\tau_{1,n}}] \geq X(\tau_{1,n} \wedge \tau_{2,n} \wedge t) \vee c.$$

Since $E[X(t) \vee c|\mathcal{F}_{\tau_{2,n} \wedge t}] \geq X(\tau_{2,n} \wedge t) \vee c \geq c$, $\{X(\tau_{2,n \wedge t}) \vee c\}$ is uniformly integrable, passing to the limit gives (5.1). □

Doob's inequalities

Theorem 5.3 *Let X be a right-continuous submartingale. Then for each $c>0$ and $T>0$,*

$$P\{\sup_{t\leq T} X(t)\geq c\}\leq c^{-1}E[X^+(T)],$$

$$P\{\inf_{t\leq T} X(t)\leq -c\}\leq c^{-1}(E[X^+(T)]-E[X(0)])$$

and for $\alpha>1$,

$$E[\sup_{t\leq T} X^+(t)^\alpha]\leq\left(\frac{\alpha}{\alpha-1}\right)^\alpha E[X^+(T)^\alpha].$$

Proof. Let $\tau=\inf\{t: X(t)>c\}$. Then

$$\{\sup_{t\leq T} X(t)>c\}\subset\{\tau\leq T\}\subset\{\sup_{t\leq T} X(t)\geq c\},$$

and

$$cP\{\tau\leq T\}\leq E[X^+(\tau\wedge T)]\leq E[X^+(T)],$$

or more precisely,

$$cP\{\sup_{t\leq T} X^+(t)>c\}\leq E[X^+(T)\mathbf{1}_{\{\tau\leq T\}}].$$

Setting $Z = \sup_{t \leq T} X^+(t)$, for nondecreasing, absolutely continuous φ with $\varphi(0) = 0$ and $\psi(z) = \int_0^z \varphi'(x)x^{-1}dx$,

$$\begin{aligned} E[\varphi(Z \wedge \beta)] &= \int_0^\beta \varphi'(x)P\{Z > x\}dx \\ &\leq \int_0^\beta \varphi'(x)x^{-1}E[X^+(T)\mathbf{1}_{\{Z \geq x\}}]dx \\ &= E[X^+(T)\psi(Z \wedge \beta)]. \end{aligned}$$

If $\varphi(x) = x^\alpha$, $\psi(x) = \frac{\alpha}{\alpha-1}x^{\alpha-1}$, and the result follows by Hölder's inequality. $\square$

Upcrossing inequality

For $a < b$, let $\tau_1 = \inf\{t : X(t) \leq a\}$, and for $k = 1, 2, \ldots$, $\sigma_k = \inf\{t > \tau_k : X(t) \geq b\}$ and $\tau_{k+1} = \inf\{t > \sigma_k : X(t) \leq a\}$.

$$U(a,b,T) = \max\{k : \sigma_k \leq T\}.$$

If X is a submartingale,

$$\begin{aligned} 0 &\leq E[\sum_{k=1}^{\infty}(X(\tau_{k+1} \wedge T) - X(\sigma_k \wedge T))] \\ &= E[\sum_{k=1}^{U(a,b,T)}(X(\tau_{k+1} \wedge T) - X(\sigma_k \wedge T))] \\ &= E[-\sum_{k=2}^{U(a,b,T)}(X(\sigma_k \wedge T) - X(\tau_k \wedge T))] \\ &\qquad + E[X(\tau_{U(a,b,T)+1} \wedge T) - a - (X(\sigma_1 \wedge T) - a)] \\ &\leq E[-(b-a)U(a,b,T) + (X(T) - a)^+] \end{aligned}$$

so

$$E[U(a,b,T)] \leq \frac{E[(X(T)-a)^+]}{b-a}.$$

Martingale convergence theorem

Theorem 5.4 *Let X be a right-continuous submartingale. Then $\lim_{s\to t-} X(s)$ exists a.s. If $\sup E[X^+(t)] < \infty$, then $\lim_{t\to\infty} X(t)$ exists a.s.*

Reverse martingale convergence theorem

Theorem 5.5 *Suppose the submartingale is defined for $-\infty < t < \infty$ and $\inf_t E[X(t)] > -\infty$. Then $\lim_{t\to-\infty} X(t)$ exists a.s.*

Model of a market

Consider financial activity over a time interval $[0, T]$ modeled by a probability space $(\Omega, \mathcal{F}, P)$.

Assume that there is a "fair casino" or market which is *complete* in the sense that at time 0, for each event $A \in \mathcal{F}$, a price $Q(A) \geq 0$ is fixed for a bet or a contract that pays one dollar at time T if and only if A occurs.

Assume that the market is *frictionless* in that an investor can either buy or sell the contract at the same price and that it is *liquid* in that there is always a buyer or seller available. Also assume that $Q(\Omega) < \infty$.

An investor can construct a *portfolio* by buying or selling a variety of contracts (possibly countably many) in arbitrary multiples.

No arbitrage condition

If a_i is the "quantity" of a contract for A_i ($a_i < 0$ corresponds to selling the contract), then the payoff at time T is

$$\sum_i a_i \mathbf{1}_{A_i}.$$

Require $\sum_i |a_i| Q(A_i) < \infty$ (only a finite amount of money changes hands) so that the initial cost of the portfolio is (unambiguously)

$$\sum_i a_i Q(A_i).$$

The market has *no arbitrage* if no combination (buying and selling) of countably many policies with a net cost of zero results in a positive profit at no risk.

That is, if $\sum |a_i| Q(A_i) < \infty$,

$$\sum_i a_i Q(A_i) = 0, \text{ and } \sum_i a_i \mathbf{1}_{A_i} \geq 0 \quad a.s.,$$

then

$$\sum_i a_i \mathbf{1}_{A_i} = 0 \quad a.s.$$

Consequences of the no arbitrage condition

Lemma 5.6 *Assume that there is no arbitrage. If $P(A) = 0$, then $Q(A) = 0$. If $Q(A) = 0$, then $P(A) = 0$.*

Proof. Suppose $P(A) = 0$ and $Q(A) > 0$. Buy one unit of Ω and sell $Q(\Omega)/Q(A)$ units of A.

$$\text{Cost} = Q(\Omega) - \frac{Q(\Omega)}{Q(A)} Q(A) = 0$$

$$\text{Payoff} = 1 - \frac{Q(\Omega)}{Q(A)} \mathbf{1}_A = 1 \quad a.s.$$

which contradicts the no arbitrage assumption.

Now suppose $Q(A) = 0$. Buy one unit of A. The cost of the portfolio is $Q(A) = 0$ and the payoff is $\mathbf{1}_A \geq 0$. So by the no arbitrage assumption, $\mathbf{1}_A = 0$ a.s., that is, $P(A) = 0$. □

Price monotonicity

Lemma 5.7 *If there is no arbitrage and $A \subset B$, then $Q(A) \leq Q(B)$, with strict inequality if $P(A) < P(B)$.*

Proof. Suppose $P(B) > 0$ (otherwise $Q(A) = Q(B) = 0$) and $Q(B) \leq Q(A)$. Buy one unit of B and sell $Q(B)/Q(A)$ units of A.

$$\text{Cost} = Q(B) - \frac{Q(B)}{Q(A)} Q(A) = 0$$

$$\text{Payoff} = \mathbf{1}_B - \frac{Q(B)}{Q(A)} \mathbf{1}_A = \mathbf{1}_{B-A} + (1 - \frac{Q(B)}{Q(A)}) \mathbf{1}_A \geq 0,$$

Payoff $= 0$ a.s. implies $Q(B) = Q(A)$ and $P(B - A) = 0$. □

Q must be a measure

Theorem 5.8 *If there is no arbitrage, Q must be a measure on $\mathcal{F}$.*

Proof. $A_1, A_2, \ldots$ disjoint and $A = \cup_{i=1}^{\infty} A_i$. Assume $P(A_i) > 0$ for some i. (Otherwise, $Q(A) = Q(A_i) = 0$.)

Let $\rho \equiv \sum_i Q(A_i)$, and buy one unit of A and sell $Q(A)/\rho$ units of A_i for each i.

$$\text{Cost} = Q(A) - \frac{Q(A)}{\rho} \sum_i Q(A_i) = 0$$

$$\text{Payoff} = \mathbf{1}_A - \frac{Q(A)}{\rho} \sum_i \mathbf{1}_{A_i} = (1 - \frac{Q(A)}{\rho})\mathbf{1}_A.$$

If $Q(A) \le \rho$, then $Q(A) = \rho$.

If $Q(A) \ge \rho$, sell one unit of A and buy $Q(A)/\rho$ units of A_i. □

Theorem 5.9 *If there is no arbitrage, $Q << P$ and $P << Q$. (P and Q are equivalent measures.)*

Proof. The result follows from Lemma 5.6. □

Pricing general payoffs

If X and Y are random variables satisfying $X \leq Y$ a.s., then no arbitrage should mean

$$Q(X) \leq Q(Y).$$

It follows that for any Q-integrable X, the price of X is

$$Q(X) = \int X dQ.$$

Assets that can be traded at intermediate times

$\{\mathcal{F}_t\}$ represents the information available at time t.

$B(t)$ is the price at time t of a bond that is worth \$1 at time T (e.g. $B(t) = e^{-r(T-t)}$), that is, at any time $0 \leq t \leq T$, $B(t)$ is the price of a contract that pays exactly \$1 at time T.

Note that $B(0) = Q(\Omega)$

Define $\hat{Q}(A) = Q(A)/B(0)$.

Martingale properties of tradeable assets

Let $X(t)$ be the price at time t of another tradeable asset.

For any stopping time $\tau \leq T$, we can buy one unit of the asset at time 0, sell the asset at time τ and use the money received ($X(\tau)$) to buy $X(\tau)/B(\tau)$ units of the bond. Since the payoff for this strategy is $X(\tau)/B(\tau)$, we must have

$$X(0) = \int \frac{X(\tau)}{B(\tau)} dQ = \int \frac{B(0)X(\tau)}{B(\tau)} d\hat{Q}.$$

Lemma 5.10 *If $E[Z(\tau)] = E[Z(0)]$ for all bounded stopping times τ, then Z is a martingale.*

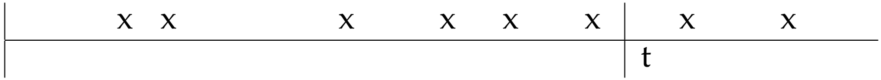
Corollary 5.11 *If X is the price of a tradeable asset, then X/B is a martingale on $(\Omega, \mathcal{F}, \hat{Q})$.*

6. Poisson and general counting processes

- Poisson process
- Martingale properties of the Poisson process
- Strong Markov property for the Poisson process
- General counting processes
- Intensities
- Counting processes as time changes of Poisson processes
- Martingale characterizations of a counting process
- Multivariate counting processes

Poisson process

A Poisson process is a model for a series for random observations occurring in time. For example, the process could model the arrivals of customers in a bank, the arrivals of telephone calls at a switch, or the counts registered by radiation detection equipment.



Let $N(t)$ denote the number of observations by time t. In the figure above, $N(t) = 6$. Note that for $t < s$, $N(s) - N(t)$ is the number of observations in the time interval $(t, s]$. We make the following assumptions about the model.

1) Observations occur one at a time.

2) Numbers of observations in disjoint time intervals are independent random variables, i.e., if $t_0 < t_1 < \cdots < t_m$, then $N(t_k) - N(t_{k-1})$, $k = 1, \ldots, m$ are independent random variables.

3) The distribution of $N(t + a) - N(t)$ does not depend on t.

Characterization of a Poisson process

Theorem 6.1 *Under assumptions 1), 2), and 3), there is a constant* $\lambda > 0$ *such that, for* $t < s$, $N(s) - N(t)$ *is Poisson distributed with parameter* $\lambda(s-t)$, *that is,*

$$P\{N(s) - N(t) = k\} = \frac{(\lambda(s-t))^k}{k!} e^{-\lambda(s-t)}.$$

Proof. Let $N_n(t)$ be the number of time intervals $(\frac{k}{n}, \frac{k+1}{n}]$, $k = 0, \ldots, [nt]$ that contain at least one observation. Then $N_n(t)$ is binomially distributed with parameters n and $p_n = P\{N(\frac{1}{n}) > 0\}$. Then

$$P\{N_n(1) = 0\} = (1 - p_n)^n \leq P\{N(1) = 0\} \leq (1 - p_n)^{n-1}$$

and $np_n \to \lambda \equiv -\log P\{N(1) = 0\}$, and the rest follows by standard Poisson approximation of the binomial. □

Interarrival times

Let S_k be the time of the kth observation. Then

$$P\{S_k \leq t\} = P\{N(t) \geq k\} = 1 - \sum_{i=0}^{k-1} \frac{(\lambda t)^i}{i!} e^{-\lambda t}, \quad t \geq 0.$$

Differentiating to obtain the probability density function gives

$$f_{S_k}(t) = \begin{cases} \frac{1}{(k-1)!}\lambda(\lambda t)^{k-1} e^{-\lambda t} & t \geq 0 \\ 0 & t < 0 \end{cases}$$

Theorem 6.2 *Let $T_1 = S_1$ and for $k > 1$, $T_k = S_k - S_{k-1}$. Then $T_1, T_2, \ldots$ are independent and exponentially distributed with parameter λ.*

Martingale properties of the Poisson process

Theorem 6.3 *(Watanabe) If N is a Poisson process with parameter λ, then $N(t) - \lambda t$ is a martingale. Conversely, if N is a counting process and $N(t) - \lambda t$ is a martingale, then N is a Poisson process.*

Proof.

$$\begin{aligned} & E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t] \\ & \quad = 1 + \sum_{k=0}^{n-1} E[(e^{i\theta(N(s_{k+1})-N(s_k)} - 1 - (e^{i\theta} - 1)(N(s_{k+1}) - N(s_k))e^{i\theta(N(s_k)-N(t))}|\mathcal{F}_t] \\ & \qquad + \sum_{k=0}^{n-1} \lambda(s_{k+1} - s_k)(e^{i\theta} - 1)E[e^{i\theta(N(s_k)-N(t))}|\mathcal{F}_t] \end{aligned}$$

The first term converges to zero by the dominated convergence theorem, so we have

$$E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t] = 1 + \lambda(e^{i\theta} - 1)\int_0^r E[e^{i\theta(N(t+s)-N(t))}|\mathcal{F}_t]ds$$

and $E[e^{i\theta(N(t+r)-N(t))}|\mathcal{F}_t] = e^{\lambda(e^{i\theta}-1)t}$. (See Exercise 5.) □

Strong Markov property

A Poisson process N is *compatible* with a filtration $\{\mathcal{F}_t\}$, if N is $\{\mathcal{F}_t\}$-adapted and $N(t+\cdot)-N(t)$ is independent of $\mathcal{F}_t$ for every $t \geq 0$.

Lemma 6.4 *Let N be a Poisson process with parameter $\lambda > 0$ that is compatible with $\{\mathcal{F}_t\}$, and let τ be a $\{\mathcal{F}_t\}$-stopping time such that $\tau < \infty$ a.s. Define $N_\tau(t) = N(\tau+t) - N(\tau)$. Then N_τ is a Poisson process that is independent of $\mathcal{F}_\tau$ and compatible with $\{\mathcal{F}_{\tau+t}\}$.*

Proof. Let $M(t) = N(t) - \lambda t$. By the optional sampling theorem,

$$E[M((\tau+t+r)\wedge T)|\mathcal{F}_{\tau+t}] = M((\tau+t)\wedge T),$$

so

$$E[N((\tau+t+r)\wedge T) - N((\tau+t)\wedge T)|\mathcal{F}_{\tau+t}] = \lambda((\tau+t+r)\wedge T - (\tau+t)\wedge T).$$

By the monotone convergence theorem

$$E[N(\tau+t+r) - N(\tau+t)|\mathcal{F}_{\tau+t}] = \lambda r$$

which gives the lemma. □

General counting processes

N is a counting process if $N(0) = 0$, N is right continuous, and N is constant except for jumps of $+1$.

N is determined by its jump times $0 < \sigma_1 < \sigma_2 < \cdots$. If N is adapted to $\mathcal{F}_t$, then the σ_k are $\mathcal{F}_t$-stopping times.

Intensity for a counting process

If N is a Poisson process with parameter λ and N is compatible with $\{\mathcal{F}_t\}$, then

$$P\{N(t+\Delta t) > N(t)|\mathcal{F}_t\} = 1 - e^{-\lambda\Delta t} \approx \lambda\Delta t.$$

For a general counting process N, at least intuitively, a nonnegative, $\{\mathcal{F}_t\}$-adapted stochastic process $\lambda(\cdot)$ is an $\{\mathcal{F}_t\}$-intensity for N if

$$P\{N(t+\Delta t) > N(t)|\mathcal{F}_t\} \approx E[\int_t^{t+\Delta t} \lambda(s)ds|\mathcal{F}_t] \approx \lambda(t)\Delta t.$$

Let σ_n be the nth jump time of N.

Definition 6.5 *λ is an $\{\mathcal{F}_t\}$-*intensity *for N if and only if for each $n = 1, 2, \ldots$.*

$$N(t \wedge \sigma_n) - \int_0^{t\wedge\sigma_n} \lambda(s)ds$$

is a $\{\mathcal{F}_t\}$-martingale.

Modeling with intensities

Let Z be a stochastic process (cadlag, E-valued for simplicity) that models "external noise." Let $D^c[0,\infty)$ denote the space of counting paths (zero at time zero and constant except for jumps of $+1$).

Condition 6.6

$$\lambda : [0,\infty) \times D_E[0,\infty) \times D^c[0,\infty) \to [0,\infty)$$

is measurable and satisfies $\lambda(t,z,v) = \lambda(t,z^t,v^t)$*, where* $z^t(s) = z(s \wedge t)$ *(*λ *is* nonanticipating*), and*

$$\int_0^t \lambda(s,z,v)ds < \infty$$

for all $z \in D_E[0,\infty)$ *and* $v \in D^c[0,\infty)$*.*

Let Y be a unit Poisson process that is $\{\mathcal{F}_t\}$-compatible and assume that $Z(s)$ is $\mathcal{F}_0$-measurable for every $s \geq 0$. (In particular, Z is independent of Y.) Consider

$$N(t) = Y(\int_0^t \lambda(s,Z,N)ds). \tag{6.1}$$

Solution of the stochastic equation

Theorem 6.7 *There exists a unique solution of (6.1) up to* $\lim_{n\to\infty}\sigma_n$, $\tau(t)=\int_0^t\lambda(s,Z,N)ds$ *is a* $\{\mathcal{F}_u\}$*-stopping time, and for each* $n=1,2,\ldots$,

$$N(t\wedge\sigma_n)-\int_0^{t\wedge\sigma_n}\lambda(s,Z,N)ds$$

is a $\{\mathcal{F}_{\tau(t)}\}$*-martingale.*

Proof. Existence and uniqueness follows by solving from one jump to the next. Let $Y^r(u) = Y(r \wedge u)$ and let

$$N^r(t) = Y^r(\int_0^t \lambda(s, Z, N^r)ds).$$

Then $N^r(t) = N(t)$, if $\tau(t) = \int_0^t \lambda(s, Z, N)ds \leq r$. Consequently,

$$\{\tau(t) \leq r\} = \{\int_0^t \lambda(s, Z, N^r)ds \leq r\} \in \mathcal{F}_r,$$

as is $\{\tau(t \wedge \sigma_n) \leq r\}$. By the optional sampling theorem

$$E[M(\tau((t+v) \wedge \sigma_n) \wedge T)|\mathcal{F}_{\tau(t)}] = M(\tau((t+v) \wedge \sigma_n)) \wedge \tau(t) \wedge T) = M(\tau(t \wedge \sigma_n) \wedge T).$$

We can let $T \to \infty$ by the monotone convergence argument used in the proof of the strong Markov property for Poisson processes. □

Martingale problems for counting processes

Definition 6.8 *Let Z be a cadlag, E-valued stochastic process, and let λ satisfy Condition 6.6. A counting process N is a solution of the* martingale problem *for (λ, Z) if*

$$N(t \wedge \sigma_n) - \int_0^{t\wedge\sigma_n} \lambda(s, Z, N) ds$$

is a martingale with respect to the filtration

$$\mathcal{F}_t = \sigma(N(s), Z(r) : s \leq t, r \geq 0)$$

Theorem 6.9 *If N is a solution of the martingale problem for (λ, Z), then N has the same distribution as the solution of the stochastic equation (6.1).*

Proof. Suppose λ is an intensity for a counting process N and $\int_0^\infty \lambda(s)ds = \infty$ a.s. Let $\gamma(u)$ satisfy

$$\gamma(u) = \inf\{t : \int_0^t \lambda(s)ds \geq u\}.$$

Then, since $\gamma(u+v) \geq \gamma(u)$,

$$E[N(\gamma(u+v)\wedge\sigma_n\wedge T) - \int_0^{\gamma(u+v)\wedge\sigma_n\wedge T} \lambda(s)ds|\mathcal{F}_{\gamma(u)}] = N(\gamma(u)\wedge\sigma_n\wedge T) - \int_0^{\gamma(u)\wedge\sigma_n\wedge T} \lambda(s)ds.$$

The monotone convergence argument lets us send T and n to infinity. We then have

$$E[N(\gamma(u+v)) - (u+v)|\mathcal{F}_{\gamma(u)}] = N(\gamma(u)) - u,$$

so $Y(u) = N(\gamma(u))$ is a Poisson process. But $\gamma(\tau(t)) = t$, so (6.1) is satisfied.

If $\int_0^\infty \lambda(s)ds < \infty$ with positive probability, then let Y^* be a unit Poisson process that is independent of $\mathcal{F}_t$ for all $t \geq 0$ and consider $N^\epsilon(t) = N(t) + Y^*(\epsilon t)$. N^ϵ has intensity $\lambda(t) + \epsilon$, and Y^ϵ, obtained as above, converges to

$$Y(u) = \begin{cases} N(\gamma(u)) & u < \tau(\infty) \\ N(\infty) + Y^*(u - \tau(\infty)) & u \geq \tau(\infty) \end{cases}$$

(except at points of discontinuity). □

Multivariate counting processes

$D_d^c[0,\infty)$: The collection of d-dimensional counting paths

Condition 6.10 $\lambda_k : [0,\infty) \times D_d^c[0,\infty) \times D_E[0,\infty) \to [0,\infty)$, *measurable and nonanticipating with*

$$\int_0^t \sum_k \lambda_k(s,z,v)ds < \infty, \quad v \in D_d^c[0,\infty), z \in D_E[0,\infty).$$

Z cadlag, E-valued and independent of independent Poisson processes $Y_1, \ldots, Y_d$.

$$N_k(t) = Y_k\Big(\int_0^t \lambda_k(s,Z,N)ds\Big), \tag{6.2}$$

where $N = (N_1,\ldots,N_d)$. Existence and uniqueness holds (including for $d=\infty$) and

$$N_k(t\wedge\sigma_n) - \int_0^{t\wedge\sigma_n} \lambda_k(s,Z,N)ds$$

is a martingale for $\sigma_n = \inf\{t : \sum_k N_k(t) \geq n\}$, but what is the correct filtration?

Multiparameter optional sampling theorem

$\mathcal{I}$ is a directed set with partial ordering $t \leq s$. If $t_1, t_2 \in \mathcal{I}$, there exists $t_3 \in \mathcal{I}$ such that $t_1 \leq t_3$ and $t_2 \leq t_3$.

$\{\mathcal{F}_t, t \in \mathcal{I}\}$, $s \leq t$ implies $\mathcal{F}_s \subset \mathcal{F}_t$.

A stochastic process $X(t)$ indexed by $\mathcal{I}$ is a martingale if and only if for $s \leq t$,

$$E[X(t)|\mathcal{F}_s] = X(s).$$

An $\mathcal{I}$ valued random variable is a stopping time if and only if $\{\tau \leq t\} \in \mathcal{F}_t$, $t \in \mathcal{I}$.

$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, t \in \mathcal{I}\}$

Lemma 6.11 *Let X be a martingale and let τ_1 and τ_2 be stopping times assuming countably many values and satisfying $\tau_1 \leq \tau_2$ a.s. If there exists a sequence $\{T_m\} \subset \mathcal{I}$ such that $\lim_{m\to\infty} P\{\tau_2 \leq T_m\} = 1$, $\lim_{m\to\infty} E[|X(T_m)|\mathbf{1}_{\{\tau_2 \leq T_m\}^c}] = 0$, and $E[|X(\tau_2)|] < \infty$, then*

$$E[X(\tau_2)|\mathcal{F}_{\tau_1}] = X(\tau_1)$$

Proof. Define

$$\tau_i^m = \begin{cases} \tau_i & \text{on } \{\tau_i \le T_m\} \\ T_m & \text{on } \{\tau_i \le T_m\}^c \end{cases}$$

Then τ_i^m is a stopping time, since

$$\begin{aligned} \{\tau_i^m \le t\} &= (\{\tau_i^m \le t\} \cap \{\tau_i \le T_m\}) \cup (\{\tau_i^m \le t\} \cap \{\tau_i \le T_m\}^c \\ &= (\cup_{s\in\Gamma, s\le t, s\le T_m} \{\tau_i = s\}) \cup (\{T_m \le t\} \cap \{\tau_i \le T_m\}^c \end{aligned}$$

Let $\Gamma \subset \mathcal{I}$ be countable and satisfy $P\{\tau_i \in \Gamma\} = 1$ and $\{T_m\} \subset \Gamma$. For $A \in \mathcal{F}_{\tau_1}$,

$$\begin{aligned} \int_{A\cap\{\tau_1^m = t\}} X(\tau_2^m)dP &= \sum_{s\in\Gamma, s\le T_m} \int_{A\cap\{\tau_1^m=t\}\cap\{\tau_2^m=s\}} X(s)dP \\ &= \sum_{s\in\Gamma, s\le T_m} \int_{A\cap\{\tau_1^m=t\}\cap\{\tau_2^m=s\}} X(T_m)dP \\ &= \int_{A\cap\{\tau_1^m=t\}} X(T_m)dP \\ &= \int_{A\cap\{\tau_1^m=t\}} X(t)dP = \int_{A\cap\{\tau_1^m=t\}} X(\tau_1^m)dP \end{aligned}$$

□

Multiple time change

$\mathcal{I} = [0,\infty)^d$, $u \in \mathcal{I}$, $\mathcal{F}_u = \sigma(Y_k(s_k) : s_k \le u_k, k = 1,\ldots,d)$. Then

$$M_k(u) \equiv Y_k(u_k) - u_k$$

is a $\{\mathcal{F}_u\}$-martingale. For

$$N_k(t) = Y_k(\int_0^t \lambda_k(s,Z,N)ds),$$

define $\tau_k(t) = \int_0^t \lambda_k(s,Z,N)ds$ and $\tau(t) = (\tau_1(t),\ldots,\tau_d(t))$. Then $\tau(t)$ is a $\{\mathcal{F}_u\}$-stopping time.

Lemma 6.12 *Let $\mathcal{G}_t = \mathcal{F}_{\tau(t)}$. If σ is a $\{\mathcal{G}_t\}$-stopping time, then $\tau(\sigma)$ is a $\{\mathcal{F}_u\}$-stopping time.*

Approximation by discrete stopping times

Lemma 6.13 *If τ is a $\{\mathcal{F}_u\}$-stopping time, then $\tau^{(n)}$ defined by*

$$\tau_k^{(n)} = \frac{[\tau_k 2^n] + 1}{2^n}$$

is a $\{\mathcal{F}_u\}$-stopping time.

Proof.

$$\{\tau^{(n)} \leq u\} = \cap_k \{\tau_k^{(n)} \leq u_k\} = \cap_k \{[\tau_k 2^n] + 1 \leq [u_k 2^n]\} = \cap_k \{\tau_k < \frac{[u_k 2^n]}{2^n}\}$$

□

Note that $\tau_k^{(n)}$ decreases to τ_k.

Martingale problems for multivariate counting processes

Let $\sigma_n = \inf\{t : \sum_k N_k(t) \geq n\}$.

Theorem 6.14 *Let Condition 6.10 hold. For $n = 1, 2, \ldots$, there exists a unique solution of (6.2) up to σ_n, $\tau_k(t) = \int_0^t \lambda_k(s, Z, N)ds$ defines a $\{\mathcal{F}_u\}$-stopping time, and*

$$N_k(t \wedge \sigma_n) - \int_0^{t\wedge\sigma_n} \lambda_k(s, Z, N)ds$$

is a $\{\mathcal{F}_{\tau(t)}\}$-martingale.

Definition 6.15 *Let Z be a cadlag, E-valued stochastic process, and let $\lambda = (\lambda_1, \ldots, \lambda_d)$ satisfy Condition 6.10. A multivariate counting process N is a solution of the* martingale problem *for (λ, Z) if for each k,*

$$N_k(t \wedge \sigma_n) - \int_0^{t\wedge\sigma_n} \lambda_k(s, Z, N)ds$$

is a martingale with respect to the filtration

$$\mathcal{G}_t = \sigma(N(s), Z(r) : s \leq t, r \geq 0)$$

Existence and uniqueness for the martingale problem

Theorem 6.16 *Let Z be a cadlag, E-valued stochastic process, and let $\lambda = (\lambda_1, \ldots, \lambda_d)$ satisfy Condition 6.10. Then there exists a unique solution of the martingale problem for (λ, Z).*

Continuous time Markov chains

Let X be a Markov chain with values in $\mathbb{Z}^d$. Let $N_l(t)$ be the number of jumps with $X(s) - X(s-) = l$ up to time t. Then

$$X(t) = X(0) + \sum_l l N_l(t).$$

Define $\beta_l(k) = q_{k,k+l}$, $q_{k,k+l}$ is the usual intensity for a transition from k to $k+l$. Then

$$X(t) = X(0) + \sum_l l Y_l(\int_0^t \beta_l(X(s))ds).$$

7. Convergence in distribution

- Prohorov metric
- Weak convergence
- Skorohod representation theorem
- Continuous mapping theorem
- Prohorov theorem
- Skorohod topology

Prohorov metric

(S, d) a metric space

$$\begin{aligned}\rho(\mu, \nu) &\equiv \inf\{\epsilon > 0 : \mu(F) \leq \nu(F^\epsilon) + \epsilon, F \in \mathcal{B}(S)\} \\ &= \inf\{\epsilon > 0 : \mu(F) \leq \nu(F^\epsilon) + \epsilon, F \in \mathcal{C}(S)\}\end{aligned}$$

The equality follows from the fact that $F^\epsilon = \bar{F}^\epsilon$.

If $\rho(\mu_n, \mu) \to 0$, then

$$\limsup_{n\to\infty} \mu_n(F) \leq \mu(\cap_\epsilon F^\epsilon) = \mu(\bar{F})$$

which is equivalent to

$$\liminf \mu_n(G) \geq \mu(G), \quad \text{all open } G.$$

It follows that

$$\mu(A^o) \leq \liminf \mu_n(A) \leq \limsup \mu_n(A) \leq \mu(\bar{A})$$

If $P_{X_n} = \mu_n$ and $P_X = \mu$, then for f bounded, continuous and nonnegative

$$E[f(X_n)] = \int_0^{\|f\|} P\{f(X_n) > z\}dz = \int_0^{\|f\|} P\{f(X_n) \geq z\}dz.$$

Since $\{x : f(x) > z\}$ is open and $\{x : f(x) \geq z\}$ is closed, $E[f(X_n)] \to E[f(X)]$.

Weak convergence and convergence in the Prohorov metric

Lemma 7.1 *If* $E[f(X_n)] \to E[f(X)]$ *for all bounded continous* f*, then* $\rho(\mu_n, \mu) \to 0$.

Proof. Let $\{x_i\}$ be dense in S. For $\epsilon > 0$, select N such that $\mu(\cup_{i=1}^N B_\epsilon(x_i)) \geq 1 - \epsilon$. For $I \subset \{1, \ldots, N\}$, let

$$f_I(x) = (1 - d(x, \cup_{i \in I} B_\epsilon(x_i))/\epsilon) \vee 0.$$

Let n satisfy

$$\max_{I \subset \{1,\ldots,N\}} |E[f_I(X_n)] - E[f_I(X)]| \leq \epsilon$$

For F closed, let

$$F_{0,\epsilon} = \cup\{B_\epsilon(x_i) : i \leq N, B_\epsilon(x_i) \cap F \neq \emptyset\} \subset F^\epsilon.$$

Then

$$\mu(F) \leq \mu(F_{0,\epsilon}) + \epsilon \leq E[f_I(X)] + \epsilon \leq E[f_I(X_n)] + 2\epsilon \leq \mu_n(F^\epsilon) + 2\epsilon,$$

so $\rho(\mu, \mu_n) \leq 2\epsilon$. □

Skorohod representation

Theorem 7.2 *Let (S, d) be complete and separable. If $\mu_n, \mu \in \mathcal{P}(S)$ and $\mu_n \Rightarrow \mu$, then there exists a probability space $(\Omega, \mathcal{F}, P)$ and random variables X_n, X such that $\mu_{X_n} = \mu_n$ and $\mu_X = \mu$ and $X_n \to X$ a.s.*

More precisely, there exist $H : \mathcal{P}(S) \times [0, 1] \to S$ such that if ξ is uniform $[0, 1]$, then $P\{H(\mu, \xi) \in \Gamma\} = \mu(\Gamma)$ for all $\mu \in \mathcal{P}(S)$ and $\mu_n \Rightarrow \mu$ implies $H(\mu_n, \xi) \to H(\mu, \xi)$ a.s.

See Blackwell and Dubins (1983).

Continuous mapping theorem

Theorem 7.3 *Suppose* $\{X_n\}$ *is a sequence of* S*-valued random variables and* $X_n \Rightarrow X$. *Let* $F : S \rightarrow \hat{S}$ *and* $C_F = \{x \in S : F \text{ is continuous at } x\}$. *Suppose that*

$$P\{X \in C_F\} = 1.$$

Then $F(X_n) \Rightarrow F(X)$.

Donsker invariance principle

$\xi_1, \xi_2, \ldots$ iid $E[\xi] = 0$, $Var(\xi) = \sigma^2 < \infty$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \xi_i$$

If $\{\xi_i^2, 1 \leq i \leq n\}$ are uniformly integrable, then

$$P\{\max_{i \leq n} |\xi_i| > \sqrt{n}\epsilon\} \leq \sum_{i=1}^{n} P\{|\xi_i| \geq \sqrt{n}\epsilon\} \leq \frac{1}{n\epsilon^2} \sum_{i=1}^{n} E[\xi_i^2 \mathbf{1}_{\{|\xi_i| \geq \sqrt{n}\epsilon\}}] \to 0$$

Let $\hat{X}_n$ be the linear interpolation of X_n, so $\hat{X}_n$ has values in $C[0, \infty)$

Empirical distribution function

Let $\xi_1, \xi_2, \ldots$ be iid with distribution function F. Define

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{(-\infty,t]}(\xi_i)$$

Then $F_n \to F$ uniformly in t. Define

$$B_n^F(t) = \sqrt{n}(F_n(t) - F(t))$$

and let $B_n(t)$ denote the uniform [0,1] case. B_n^F has the same distribution as $B_n(F(\cdot))$.

$(B_n(t_1), \ldots, B_n(t_m)) \Rightarrow (B(t_1), \ldots, B(t_m))$ where $(B(t_1), \ldots, B(t_m))$ is jointly Gaussian with mean zero and covariance given by

$$E[B_n(t)B_n(s)] = E[(\mathbf{1}_{[0,t]}(\xi) - t)(\mathbf{1}_{[0,s]}(\xi) - s)] = t \wedge s - ts$$

Wright-Fisher

Let $\{Y_k\}$ be a Markov chain with state space $\{\frac{k}{N} : 0 \leq k \leq N\}$

$$P\{Y_{k+1} = \frac{l}{N} | Y_k = x\} = \binom{N}{l} x^l (1-x)^{N-l}$$

Note that $E[Y_{k+1}|Y_k] = E[Y_{k+1}|\mathcal{F}_k] = Y_k$, and

$$E[(Y_{k+m} - Y_k)^2] = \sum_{i=0}^{m-1} E[(Y_{k+i+1} - Y_{k+i})^2] = \sum_{i=0}^{m-1} \frac{1}{N} E[Y_i(1-Y_i)]$$

$$X_N(t) = Y_{[Nt]}$$

X_N is a martingale as is

$$X_N^2(t) - \sum_{i=0}^{[Nt]-1} \frac{1}{N} Y_i(1-Y_i) = X_N^2(t) - \int_0^{[Nt]/N} X_N(s)(1 - X_N(s))ds$$

Topological proof of convergence

- Prove relative compactness of $\{\mu_n\}$
- Prove that there is at most one limit point.

Claim: The three examples are relatively compact (tight).

The limits for the first two are characterized by their finite dimensional distributions.

The limit for the third is characterized by its martingale properties.

To carry out a topological proof of convergence, we need to characterize compact subsets of $\mathcal{P}(S)$.

A metric for convergence in probability

For X, Y, S-valued random variables, let

$$\gamma(X,Y) = \inf\{\epsilon > 0 : P\{d(X,Y) > \epsilon\} < \epsilon\}.$$

Claim: γ is a metric on the space of S-valued random variable on $(\Omega, \mathcal{F}, P)$.

$\lim_{n\to\infty} \gamma(X_n, X)$ if and only if $X_n \to X$ in probability.

Note: Almost sure convergence is not metrizable.

Probabilistic interpretation of ρ

Lemma 7.4 *Let $\rho(\mu,\nu) < \epsilon$. Then there exist random variables X and Y such that $\mu_X = \mu$ and $\mu_Y = \nu$ and*

$$P\{d(X,Y) \geq \epsilon\} \leq \epsilon$$

Specifically,

$$\rho(\mu,\nu) = \inf\{\gamma(X,Y) : \mu_X = \mu, \mu_Y = \nu\}$$

Remark 7.5 *Note that the converse is straight forward since*

$$P\{X \in F\} \leq P\{Y \in F^\epsilon\} + P\{d(X,Y) \geq \epsilon\}.$$

Completeness and separability

Lemma 7.6 *$(\mathcal{P}(S), \rho)$ is complete iff (S, d) is complete. $(\mathcal{P}(S), \rho)$ is separable iff (S, d) is separable.*

Proof. Suppose $\rho(\mu_n, \mu_m) \to 0$. There exists a subsequence such that $\rho(\mu_{n_k}, \mu_{n_{k+1}}) \leq 2^{-k}$ and hence joint distributions $\mu_{X_k, X_{k+1}}$ with $\mu_{X_k} = \mu_{n_k}$ and $P\{d(X_k, X_{k+1}) \geq 2^{-k}\} \leq 2^{-k}$. By Tulcea's theorem, $\{X_k\}$ on a single probability space. Then

$$P\{\sup_{m>n} d(X_n, X_m) \geq 2^{-(n+1)}\} \leq 2^{-(n+1)},$$

and the completeness of S implies X_n converges a.s.

If S is separable and $\{x_k\}$ is dense in S, then $\{\sum_{k=1}^n p_k \delta_{x_k} : \sum p_k = 1, p_k \text{ rational}\}$ is dense in $\mathcal{P}(S)$. $\square$

Total boundedness

K is *totally bounded* if and only if for each $\epsilon > 0$, there exist $x_1, x_2, \ldots, x_n$ such that $K \subset \cup_{i=1}^n B_\epsilon(x_i)$.

Lemma 7.7 *A set K is compact if and only if it is complete and totally bounded.*

Proof. Total boundedness follows from compactness by the definition of compactness. Total boundedness and completeness imply sequential compactness which in turn implies compactness. □

Prohorov's theorem

Theorem 7.8 $\{\mu_\alpha\} \subset \mathcal{P}(S)$ *is relatively compact in the topology generated by the Prohorov metric if and only if for each* $\epsilon > 0$*, there exists a compact* $K_\epsilon \subset S$ *such that*

$$\inf_\alpha \mu_\alpha(K_\epsilon) \geq 1 - \epsilon. \text{ [tightness]}$$

Proof. Suppose $\{\mu_\alpha\}$ is tight. Let $x_1, \ldots, x_n$ satisfy $K_\epsilon \subset \cup_{i=1}^n B_\epsilon(x_i)$ and $x_0 \in S$. Select $m \geq n/\epsilon$, and let $\Gamma_m = \{\nu : \nu = \sum_{i=0}^n \frac{k_i}{m}\delta_{x_i}\}$. Let $E_1 = B_\epsilon(x_1)$ and $E_i = B_\epsilon(x_i) \cap (\cup_{j=1}^{i-1} B_\epsilon(x_j))^c$. Define

$$\nu_\alpha = \sum_{i=1}^n \frac{[m\mu_\alpha(E_i)]}{m}\delta_{x_i} + (1 - \sum_{i=1}^n \frac{[m\mu_\alpha(E_i)]}{m})\delta_{x_0}$$

Then

$$\mu_\alpha(F) \leq \mu_\alpha(\cup_{F\cap E_i \neq \emptyset} E_i) + \epsilon \leq \sum_{F\cap E_i \neq \emptyset} \frac{[m\mu_\alpha(E_i)]}{m} + \frac{n}{m} + \epsilon \leq \nu_\alpha(F^{2\epsilon}) + 2\epsilon$$

and $\rho(\mu_\alpha, \nu_\alpha) \leq 2\epsilon$. Consequently, $\{\mu_\alpha\}$ is totally bounded. □

Arzela-Ascoli Theorem

The following is a special case of the Arzela-Ascoli theorem.

Lemma 7.9 *$K \subset C_{\mathbb{R}^d}[0,1]$ is relatively compact if and only if $\sup_{x \in K} |x(0)| < \infty$ and $\lim_{\delta \to 0} \sup_{|s-t| \le \delta} |x(s) - x(t)| = 0$.*

Proof. The proof can be found in http://www.math.byu.edu/~klkuttle/lecturenotes641.pdf □

Corollary 7.10 *Let $c, \eta_k, \delta_k > 0$ and $\eta_k, \delta_k \to 0$. Then*

$$K_{c,\{(\eta_k,\delta_k)\}} \equiv \{x \in C[0,1] : |x(0)| \le c, \sup_{|s-t| \le \delta_k} |x(s) - x(t)| \le \eta_k, k = 1, 2, \ldots\}$$

is compact.

Tightness for $S = C[0,1]$

Theorem 7.11 *$\{X_\alpha\}$ is relatively compact in distribution if and only if for each $\epsilon, \eta > 0$, there exist $c, \delta > 0$ such that*

$$\sup_\alpha P\{|X(0)| \geq c\} \leq \epsilon$$

and

$$\sup_\alpha P\{\sup_{|s-t|\leq\delta} |X_\alpha(s) - X_\alpha(t)| \geq \eta\} \leq \epsilon.$$

Proof. Let $\eta_k > 0$, $\eta_k \to 0$. For $\epsilon > 0$, select $c_\epsilon > 0$ so that $\sup_\alpha P\{|X(0)| \geq c\} \leq \epsilon/2$ and $\delta_k > 0$ so that

$$\sup_\alpha P\{\sup_{|s-t|\leq\delta_k} |X_\alpha(s) - X_\alpha(t)| \geq \eta_k\} \leq 2^{-(k+1)}\epsilon.$$

Then $P\{X_\alpha \notin K_{c,\{(\eta_k,\delta_k)\}}\} \leq \epsilon$. □

Kolmogrov criterion

Theorem 7.12 *Let* $\{X_\alpha\}$ *be process in* $C_{\mathbb{R}^d}[0,1]$. *Suppose that there exist* $C > 0$, $\beta > 0$, *and* $\theta > 1$ *such that*

$$\sup_\alpha \sup_{|t-s|\leq\delta} E[|X_\alpha(t) - X_\beta(s)|^\beta \wedge 1] \leq C\delta^\theta.$$

Then $\{X_\alpha\}$ *is relatively compact in distribution.*

Chaining argument

Suppose $\delta < 2^{-(n+1)}$ and $|t-s| \leq \delta$. Let $t_m = 2^{-m}[2^m t]$ and $s_m = 2^{-m}[2^m s]$. Then $|t_{m+1} - t_m| \leq 2^{-(m+1)}$, $|s_{m+1} - s_m| \leq 2^{-(m+1)}$, $|s_n - t_n| \leq 2^{-n}$, $\lim_{m\to\infty} t_m = t$, $\lim_{m\to\infty} s_m = s$, and for $x \in C[0,1]$,

$$|x(t) - x(s)| \leq |x(t_n) - x(s_n)| + \sum_{m=n}^{\infty}(|x(t_{m+1}) - x(t_m)| + |x(s_{m+1}) - x(s_m)|).$$

Define $\eta_m^\alpha = \sum_{k=0}^{2^m-1} |X_\alpha(2^{-m}(k+1)) - X_\alpha(2^{-m}k)|^\beta \wedge 1$.

Then for $\delta < 2^{-(n+1)}$,

$$\sup_{|t-s|\leq\delta} |X_\alpha(t) - X_\alpha(s)| \wedge 1 \leq 2\sum_{m=n}^{\infty}(\eta_m^\alpha)^{1/\beta}.$$

Consequently,

$$E[\sup_{|t-s|\leq\delta} |X_\alpha(t) - X_\alpha(s)| \wedge 1] \leq 2E[\sum_{m=n}^{\infty}(\eta_m^\alpha)^{1/\beta}] \leq 2C^{1/\beta}\left(\sum_{m=n}^{\infty} 2^{-m(\theta-1)}\right)^{1/\beta}.$$

As $\delta \to 0$, we can let $n \to \infty$ and the right side goes to zero.

Donsker's invariance principle

$\xi_1, \xi_2, \ldots$ iid $E[\xi] = 0$, $Var(\xi) = \sigma^2 < \infty$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \xi_i$$

Let

$$\hat{X}_n(t) = X_n(t) + \frac{nt - [nt]}{\sqrt{n}} \xi_{[nt]+1}.$$

Assuing $t > s$, Let $\gamma = E[\xi^4]$. Then, assuming $[nt] > [ns]$

$$\begin{aligned}
&E[(\hat{X}_n(t) - \hat{X}_n(s))^4] \\
&\quad = \frac{\gamma}{n^2}\Big((nt - [nt])^4 + [nt] - [ns] - 1 + ([ns] + 1 - ns)^4\Big) \\
&\qquad + \frac{12\sigma^2}{n^2}\Big(((nt - [nt])^2 + ([ns] + 1 - ns)^2)([nt] - [ns] - 1) \\
&\qquad\qquad + ([ns] + 1 - ns)^2 (nt - [nt])^2 + 2^{-1}([nt] - [ns] - 1)([nt] - [ns] - 2)\Big)
\end{aligned}$$

Conditions for convergence in $C[0,1]$

Lemma 7.13 *If $X_n \Rightarrow X$, then $(X_n(t_1),\ldots,X_n(t_m)) \Rightarrow (X(t_1),\ldots,X(t_m))$, $0 \leq t_1 < \cdots < t_m \leq 1$. If $\mu_1, \mu_2 \in \mathcal{P}(C[0,1])$ have the same finite dimensional distributions, then $\mu_1 = \mu_2$.*

Proof. $\pi_t : x \in C[0,1] \to x(t)$ is continuous, so the first part follows. The second follows from the fact that $\mathcal{B}(C[0,1]) = \sigma(\pi_t, 0 \leq t \leq 1)$ $(B_r(y) = \cap_{t\in\mathbb{Q}\cap[0,1]}\{y : |\pi_t(x) - \pi_t(y)| \leq r\}$. $\mu_1 = \mu_2$ on $\sigma(\pi_t, 0 \leq t \leq 1)$ by the Dynkin-class theorem. □

Theorem 7.14 *Suppose $\{X_n\}$ is relatively compact in distribution in $C[0,1]$ and*

$$(X_n(t_1),\ldots,X_n(t_m)) \Rightarrow (X(t_1),\ldots,X(t_m)), \qquad 0 \leq t_1 < \cdots < t_m \leq 1.$$

Then X has a continuous version and $X_n \Rightarrow X$.

Poisson approximation

Suppose that for each n, $\{\xi_k^n\}$ is a Bernoulli sequence with $np_n \to \lambda$, and define

$$X_n(t) = \sum_{k=1}^{[nt]} \xi_k^n.$$

"Clearly" $X_n \Rightarrow X$ where X is a Poisson process with parameter λ, but in what sense. Assuming the Skorhod representation theorem applies, $\sup_{t \le T} |X(t) - X_n(t)|$ does not converge to zero.

Skorohod topology on $D_E[0,\infty)$

(E,r) complete, separable metric space

$D_E[0,\infty)$ space of cadlag, E-valued functions

$x_n \to x \in D_E[0,\infty)$ in the Skorohod (J_1) topology if and only if there exist strictly increasing λ_n mapping $[0,\infty)$ onto $[0,\infty)$ such that for each $T>0$,

$$\lim_{n\to\infty} \sup_{t\le T}(|\lambda_n(t)-t| + r(x_n\circ\lambda_n(t), x(t))) = 0.$$

The Skorohod topology is metrizable so that $D_E[0,\infty)$ is a complete, separable metric space.

Note that $\mathbf{1}_{[1+\frac{1}{n},\infty)} \to \mathbf{1}_{[1,\infty)}$ in $D_{\mathbb{R}}[0,\infty)$, but $(\mathbf{1}_{[1+\frac{1}{n},\infty)}, \mathbf{1}_{[1,\infty)})$ does *not* converge in $D_{\mathbb{R}^2}[0,\infty)$. (It does converge in $D_{\mathbb{R}}[0,\infty)\times D_{\mathbb{R}}[0,\infty)$).

Conditions for tightness

$S_0^n(T)$ collection of discrete $\{\mathcal{F}_t^n\}$-stopping times $\qquad q(x,y) = 1 \wedge r(x,y)$

Theorem 7.15 *Suppose that for $t \in \mathcal{T}_0$, a dense subset of $[0,\infty)$, $\{X_n(t)\}$ is tight. Then the following are equivalent.*

a) $\{X_n\}$ is tight in $D_E[0,\infty)$.

b) (Kurtz) For $T > 0$, there exist $\beta > 0$ and random variables $\gamma_n(\delta, T)$ such that for $0 \leq t \leq T$, $0 \leq u \leq \delta$, and $0 \leq v \leq t \wedge \delta$

$$E[q^\beta(X_n(t+u), X_n(t)) \wedge q^\beta(X_n(t), X_n(t-v))|\mathcal{F}_t^n] \leq E[\gamma_n(\delta,T)|\mathcal{F}_t^n]$$

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E[\gamma_n(\delta,T)] = 0,$$

and

$$\lim_{\delta\to 0}\limsup_{n\to\infty} E[q^\beta(X_n(\delta), X_n(0))] = 0. \qquad (7.1)$$

c) (Aldous) Condition (7.1) holds, and for each $T > 0$, there exists $\beta > 0$ such that

$$C_n(\delta,T) \equiv \sup_{\tau\in S_0^n(T)} \sup_{u\leq\delta} E[\sup_{v\leq\delta\wedge\tau} q^\beta(X_n(\tau+u), X_n(\tau)) \wedge q^\beta(X_n(\tau), X_n(\tau-v))]$$

satisfies $\lim_{\delta\to 0}\limsup_{n\to\infty} C_n(\delta,T) = 0$.

Example

$\eta_1, \eta_2, \ldots$ iid, $E[\eta_i] = 0$, $\sigma^2 = E[\eta_i^2] < \infty$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \eta_i$$

Then

$$E[(X_n(t+u) - X_n(t))^2 | \mathcal{F}_t^{X_n}] = \frac{[n(t+u)] - [nt]}{n} \sigma^2 \leq (\delta + \frac{1}{n}) \sigma^2$$

for $u \leq \delta$.

Uniqueness of limit

Theorem 7.16 *If* $\{X_n\}$ *is tight in* $D_E[0,\infty)$ *and*

$$(X_n(t_1),\ldots,X_n(t_k)) \Rightarrow (X(t_1),\ldots,X(t_k))$$

for $t_1,\ldots,t_k \in \mathcal{T}_0$, $\mathcal{T}_0$ *dense in* $[0,\infty)$, *then* $X_n \Rightarrow X$.

For the example, this condition follows from the central limit theorem.

Some continuous functions

$F_1 : x \in D_E[0,\infty) \to y \in D_{\mathbb{R}}[0,\infty)$, $y(t) = \sup_{s\le t} r(x(s).x(s-))$

$F_2 : x \in D_{\mathbb{R}}[0,\infty) \to y \in D_{\mathbb{R}}[0,\infty)$, $y(t) = \sup_{s\le t} x(s)$

$G_t : x \in D_{\mathbb{R}}[0,\infty) \to \mathbb{R}$, $G_t(x) = \sup_{s\le t} x(s)$ is not continuous. (Exercise: Identify the continuity set for G_t.)

If $f : E \to \hat{E}$ is continuous, then
$H_f : x \in D_E[0,\infty) \to y \in D_{\hat{E}}[0,\infty)$, $y(t) = f(x(t))$ is continuous, but
$G_{f,t} : x \in D_E[0,\infty) \to y \in \hat{E}$, $y(t) = f(x(t))$ is not continuous.

Compact uniform topology

$$d_u(x, y) = \int_0^\infty e^{-t} \sup_{s \le t} 1 \wedge r(x(s), y(s))\, dt$$

defines a metric on $D_E[0, \infty)$, but $D_E[0, \infty)$ is not separable under d_u.

However:

Lemma 7.17 *Suppose $x_n \to x$ in the Skorohod topology. Then $F_1(x_n) \to 0$ if and only if x is continuous, and if x is continuous, $d_u(x_n, x) \to 0$. In particular, $C_E[0, \infty)$ is closed in the Skorohod topology, and the restriction of the Skorohod topology to $C_E[0, \infty)$ is the compact uniform topology.*

Other conditions

Let $\{X_\alpha\}$ be processes with sample paths in $D_E[0,\infty)$.

The *compact containment* condition holds if and only if for each $T, \epsilon > 0$, there exists a compact set $K_{\epsilon,T} \subset E$

$$\inf_\alpha P\{X_\alpha(t) \in K_{\epsilon,T}\} \geq 1 - \epsilon.$$

Let $\mathcal{C} \subset \bar{C}(E)$ be linear and separate points.

Theorem 7.18 *$\{X_\alpha\}$ is relatively compact in $D_E[0,\infty)$ if and only if the compact containment condition holds and for each $f \in \mathcal{C}$, $\{f \circ X_\alpha\}$ is relatively compact in $D_\mathbb{R}[0,\infty)$.*

8. Brownian motion

- Construction by Donsker invariance
- Markov property
- Transition density and heat semigroup
- Strong Markov property
- Sample path properties
- Lévy characterization
- Martingale central limit theorem

Construction by Donsker invariance

$\xi_1, \xi_2, \ldots$ iid $E[\xi] = 0$, $Var(\xi) = 1$

$$X_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} \xi_i$$

Then $X_n \Rightarrow W$, standard Browian motion.

W is continuous

W has independent increments

$E[W(t)] = 0$, $\quad Var(W(t)) = t$, $\quad Cov(W(t), W(s)) = t \wedge s$

W is a martingale.

Markov property

$X(t) = X(0) + W(t)$, $X(0)$ independent of W.

$$T(t)f(x) \equiv E[f(x + W(t))] = \int_{-\infty}^{\infty} f(y) \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} dy$$

$$E[f(X(t+s))|\mathcal{F}_t^X] = E[f(X(t) + W(t+s) - W(t))|\mathcal{F}_t^X] = T(s)f(X(t))$$

Transition density

The transity density is

$$p(t,x,y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}}$$

which satisfies the Chapman-Kolmogorov equation

$$p(t+s,x,y) = \int_{\mathbb{R}} p(t,x,z)p(s,z,y)dz$$

Note that

$$\frac{\partial}{\partial t} T(t)f(x) = \frac{1}{2}\frac{d^2}{dx^2} T(t)f(x)$$

Right continuous filtration

$$E[f(X(t+s))|\mathcal{F}^X_{t+}] = \lim_{h\to 0} E[f(X(t+s))|\mathcal{F}^X_{t+h}] = \lim_{h\to 0} T(s-h)f(X(t+h)) = T(s)f(X(t))$$

Lemma 8.1 *If Z is bounded and measurable with respect to $\sigma(X(0), W(s), s \geq 0)$, then*

$$E[Z|\mathcal{F}^X_t] = E[Z|\mathcal{F}^X_{t+}] \quad a.s.$$

Proof. Consider

$$E[\prod_i f_i(X(t_i))|\mathcal{F}^X_{t+}]$$

and apply the Dynkin-class theorem. □

Corollary 8.2 *Let $\bar{\mathcal{F}}^X_t$ be the completion of $\mathcal{F}^X_t$. Then $\bar{\mathcal{F}}^X_t = \bar{\mathcal{F}}^X_{t+}$.*

Proof. If $C \in \mathcal{F}^X_{t+}$, then $E[\mathbf{1}_C|\mathcal{F}^X_t] = \mathbf{1}_C$ a.s. Consequently, setting

$$C^o = \{E[\mathbf{1}_C|\mathcal{F}^X_t] = 1\} \qquad P(C^o \triangle C) = 0$$

□

Strong Markov Property

Prove first for discrete stopping times

$$E[f(X(\tau + t))|\mathcal{F}_\tau] = T(t)f(X(\tau)$$

Every stopping time is the limit of a decreasing sequence of discrete stopping times

If $\gamma \geq 0$ is $\mathcal{F}_\tau$-measurable, then

$$E[f(X(\tau + \gamma))|\mathcal{F}_\tau] = T(\gamma)f(X(\tau)$$

Reflection principle

$$P\{\sup_{s\le t} W(s) > c\} = 2P\{W(t) > c\}$$

Samplepath properties

Finite, nonzero quadratic variation

$$\lim \sum (W(t_{i+1}) - W(t_i))^2 = t.$$

Brownian paths are nowhere differentiable (Theorem 1.8)

Law of the Iterated Logarithm

$$\limsup_{t\to\infty} \frac{W(t)}{\sqrt{2t\log\log t}} = 1$$

$\hat{W}(t) = tW(1/t)$ is Brownian motion. $Var(\hat{W}(t)) = t^2\frac{1}{t} = t$ Therefore

$$\limsup_{t\to 0} \frac{W(1/t)}{\sqrt{2t^{-1}\log\log 1/t}} = \limsup_{t\to 0} \frac{\hat{W}(t)}{\sqrt{2t\log\log 1/t}} = 1$$

Consequently,

$$\limsup_{h\to 0} \frac{W(t+h)-W(t)}{\sqrt{2h\log\log 1/h}} = 1$$

The tail of the normal distribution

Lemma 8.3

$$\int_a^\infty e^{-\frac{x^2}{2}}\,dx < a^{-1}e^{-\frac{a^2}{2}} = \int_a^\infty (1+x^{-2})e^{-\frac{x^2}{2}}\,dx$$
$$< (1+a^{-2})\int_a^\infty e^{-\frac{x^2}{2}}\,dx$$

Proof. Differentiate

$$\frac{d}{da}a^{-1}e^{-\frac{a^2}{2}} = -(a^{-2}+1)e^{-\frac{a^2}{2}}$$

□

Modulus of continuity

Theorem 8.4 *Let* $h(t) = \sqrt{2t \log 1/t}$. *Then*

$$P\{\lim_{\epsilon\to 0} \sup_{t_1,t_2\in[0,1],|t_1-t_2|\le\epsilon} \frac{|W(t_1)-W(t_2)|}{h(|t_1-t_2|)} = 1\} = 1$$

Proof.

$$P\{\max_{k\le 2^n}(W(k2^{-n}) - W((k-1)2^{-n})) \le (1-\delta)h(2^{-n})\} = (1-I)^{2^n} < e^{-2^n I}$$

for

$$I = \int_{(1-\delta)\sqrt{2\log 2^n}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}dx > C\frac{1}{\sqrt{n}} e^{-(1-\delta)^2 \log 2^n} > \frac{C}{\sqrt{n}} 2^{-(1-\delta)^2 n}$$

so $2^n I > 2^{n\delta}$ for n sufficiently large and Borel-Cantelli implies

$$P\{\limsup_{n\to\infty} \max_{k\le 2^n}(W(k2^{-n}) - W((k-1)2^{-n}))/h(2^{-n}) \ge 1\} = 1.$$

For $\delta > 0$ and $\epsilon > \frac{1+\delta}{1-\delta} - 1$

$$P\{\max_{0<k\le 2^{n\delta}, 0\le i\le 2^n - 2^{n\delta}} \frac{|W((i+k)2^{-n}) - W(i2^{-n})|}{h(k2^{-n})} \ge (1+\epsilon)\}$$

$$\leq \sum 2(1-\Phi((1+\epsilon)\sqrt{2\log(2^n/k)}))$$

$$\leq C\sum \frac{1}{(1+\epsilon)\sqrt{2\log(2^n/k))}} e^{-2(1+\epsilon)^2\log(2^n/k))}$$

$$\leq C\frac{1}{\sqrt{n}}2^{n(1+\delta)}2^{-2n(1-\delta)(1+\epsilon)^2}$$

and the right side is a term in a convergent series. Consequently, for almost every ω, there exists $N(\omega)$ such that $n > N(\omega)$ and $0 < k \leq 2^{n\delta}, 0 \leq i \leq 2^n - 2^{n\delta}$ implies

$$|W((i+k)2^{-n}) - W(i2^{-n})| \leq (1+\epsilon)h(k2^{-n})$$

If $|t_1 - t_2| \leq 2^{-(N(\omega)+1)(1-\delta)}$,

$$|W(t_1) - W(t_2)| \leq |W([2^{N(\omega)}t_1]2^{-N(\omega)}) - W([2^{N(\omega)}t_2]2^{-N(\omega)})|$$

$$\sum_{n\geq N(\omega)} |W([2^n t_1]2^{-n}) - W([2^{n+1}t_1]2^{-(n+1)})| + \sum_{n\geq N(\omega)} |W([2^n t_2]2^{-n}) - W([2^{n+1}t_2]2^{-(n+1)})| +$$

so

$$|W(t_1) - W(t_2)| \leq h(|[2^{N(\omega)}t_1] - [2^{N(\omega)}t_2]|)$$

□

Quadratic variation for continuous martingales

Lemma 8.5 *Let M be a (local) square integrable martingale. Then*

$$[M]_t = \lim_{\max|t_{i+1}-t_i|\to 0} \sum_i (M(t_{i+1}\wedge t) - M(t_i \wedge t))^2$$

exists in probability and $M^2 - [M]$ is a local martingale.

Remark 8.6 *Any local martingale with bounded jumps in a local square integrable martingale.*

Proof. Assume that M is a square integrable martingale (otherwise, consider the stopped martingale). Let $\tau_c = \inf\{t : |M(t)| \geq c\}$, and replace M be $M(\cdot \wedge \tau_c)$.

Suppose $\{s_i\}$ is a refinement of $\{t_i\}$. Then

$$\begin{aligned} Z(t) &= \sum_i (M(t_{i+1}\wedge t) - M(t_i \wedge t))^2 - \sum_i (M(s_{j+1}\wedge t) - M(s_j \wedge t))^2 \\ &= \sum_i 2 \sum_{j<k\in\Gamma_i} (M(s_{k+1}\wedge t) - M(s_k \wedge t))(M(s_{j+1}) - M(s_j)) \end{aligned}$$

Each term in the sum is a martingale, so Z is a martingale. Let $\gamma(s) = \max\{t_i : t_i < s\}$. Then

$$Z(t) = 2\sum_j (M(s_{j+1}\wedge t) - M(s_j\wedge t))(M(s_j) - M(\gamma(s_{j+1}))).$$

Note that if $s_j \geq \tau_c$, then the jth term is zero, and if $s_j < \tau_c$, the jth term is bounded by

$$4c|M(s_{j+1}\wedge t) - M(s_j\wedge t)|.$$

In particular,

$$E[Z(t)^2] \leq 16c^2 E[M(t)^2].$$

Then, if $\max|t_{i+1} - t_i| \leq \delta$,

$$E[Z(t\wedge\beta_{\delta,\epsilon})^2] \leq \epsilon^2 E[M(t)^2].$$

For $\epsilon_n \to 0$, there exist $\delta_n \to 0$ such that $\beta_n \equiv \beta_{\delta_n,\epsilon_n} \to \infty$ and hence if $\max |t_{i+1}^n - t_i^n| \leq \delta_n$

$$\sum_i (M(t_{i+1}^n \wedge t \wedge \beta_n) - M(t_i^n \wedge t \wedge \beta_n))^2 \xrightarrow{L_1} [M]_t$$

□

Lévy characterization

Theorem 8.7 *Let M be a continuous local martingale with $[M]_t = t$. Then M is a standard Brownian motion*

Proof. For each $c > 0$, $E[M(t \wedge \tau_c)^2] = E[t \wedge \tau_c]$, and by Fatou, $E[M(t)^2] < \infty$. Then $M(t \wedge \tau_c)^2 \leq \sup_{s \leq t} M(s)^2$, so by Doob's inequality and the dominated convergence theorem $E[M(t \wedge \tau_c)^2] \to E[M(t)^2] = t$. It follows that

$$\sum_i (M(t_{i+1} \wedge t) - M(t_i \wedge t))^2 \stackrel{L_1}{\to} t.$$

$$\begin{aligned} E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] \\ = 1 + \sum_{k=0}^{n-1} E\Big[\Big(e^{i\theta(M(s_{k+1})-M(s_k)} - 1 - i\theta(M(s_{k+1}) - M(s_k)) \\ + \frac{1}{2}\theta^2(M(s_{k+1}) - M(s_k))^2\Big) e^{i\theta(M(s_k)-M(t))}|\mathcal{F}_t] \\ -\frac{1}{2}\theta^2 \sum_{k=0}^{n-1} (s_{k+1} - s_k) E[e^{i\theta(M(s_k)-M(t))}|\mathcal{F}_t] \end{aligned}$$

The first term converges to zero by the dominated convergence theorem, so we have

$$E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] = 1 - \frac{1}{2}\theta^2 \int_0^r E[e^{i\theta(M(t+s)-M(t))}|\mathcal{F}_t]ds$$

and $E[e^{i\theta(M(t+r)-M(t))}|\mathcal{F}_t] = e^{-\frac{\theta^2 r}{2}}$. □

Limits of martingales are martingales

Lemma 8.8 *Suppose that for* $n = 1, 2, \ldots$, M_n *is a cadlag martingale,* $M_n \Rightarrow M$ *and for each* $t \geq 0$, $\{M_n(t)\}$ *is uniformly integrable. Then* M *is a martingale.*

Proof. There exists a countable set D such that if $(t_1, \ldots, t_m) \in [0, \infty) - D$, then

$$(M_n(t_1), \ldots, M_n(t_m)) \Rightarrow (M(t_1), \ldots, M(t_m)).$$

If $f_i \in \bar{C}(\mathbb{R})$, then $\{M_n(t_m) \prod_i f_i(M_n(t_i))\}$ is uniformly integrable and converges in distribution to $M(t_m) \prod_i f_i(M(t_i))$. It follows that for $0 \leq t_1 < t_2 < \cdots < t_{m+1}$, $t_i \notin D$,

$$0 = \lim_{n\to\infty} E[(M_n(t_{m+1}) - M_n(t_m)) \prod_i f_i(M_n(t_i))] = E[(M(t_{m+1}) - M(t_m)) \prod_i f_i(M(t_i))]$$

By the right continuity of M, the right side is zero for all t_i, and hence M is a $\{\mathcal{F}_t^M\}$-martingale. □

Martingale central limit theorem

Theorem 8.9 *Let $\{M_n\}$ be a sequence of martingales. Suppose that*

$$\lim_{n\to\infty} E[\sup_{s\leq t} |M_n(s) - M_n(s-)|] = 0$$

and

$$[M_n]_t \to c(t) \tag{8.1}$$

for each $t > 0$, where $c(t)$ is continuous and deterministic. Then $M_n \Rightarrow M = W \circ c$.

Remark 8.10 *If*

$$\lim_{n\to\infty} E[|[M_n]_t - c(t)|] = 0, \qquad \forall t \geq 0, \tag{8.2}$$

then by the continuity of c, the conditions hold. If (8.1) holds and $\lim_{n\to\infty} E[[M_n]_t] = c(t)$ for each $t \geq 0$, then (8.2) holds by the dominated convergence theorem.

Proof.(Assuming (8.2).) For $0 \leq u \leq \delta$, $s \leq t$,

$$E[(M_n(s+u) - M_n(s))^2|\mathcal{F}_s^n] = E[[M_n]_{s+u} - [M_n]_s|\mathcal{F}_s^n] \leq E[\sup_{s\leq t}([M_n]_{s+\delta} - [M_n]_s)|\mathcal{F}_s^n],$$

so by the tightness criterion, $\{M_n\}$ is relatively compact. □

Example 8.11 *If* $M_n \Rightarrow W \circ c$*, then*

$$P\{\sup_{s\leq t} M_n(s) \leq x\} \to P\{\sup_{s\leq t} W(c(s)) \leq x\} = P\{\sup_{u\leq c(t)} W(u) \leq x\}.$$

Corollary 8.12 *(Donsker's invariance principle.) Let ξ_k be iid with mean zero and variance σ^2. Let*

$$M_n(t) = \frac{1}{\sqrt{n}} \sum_{k=1}^{[nt]} \xi_k.$$

Then M_n is a martingale for every n, and $M_n \Rightarrow \sigma W$.

Proof. Since M_n is a finite variation process, we have

$$\begin{aligned} [M_n]_t &= \sum_{s \le t} (\Delta M_n(s))^2 \\ &= \frac{1}{n} \sum_{k=1}^{[nt]} \xi_k^2 \\ &= \frac{[nt]}{n[nt]} \sum_{k=1}^{[nt]} \xi_k^2 \to t\sigma^2. \end{aligned}$$

where the limit holds by the law of large numbers. Note that the convergence is in L_1, and $M_n \Rightarrow W(\sigma^2 \cdot)$. □

Corollary 8.13 *(CLT for renewal processes.) Let ξ_k be iid, positive and have mean μ and variance σ^2. Let*

$$N(t) = \max\{k : \sum_{i=1}^{k} \xi_i \leq t\}.$$

Then

$$Z_n(t) \equiv \frac{N(nt) - nt/\mu}{\sqrt{n}} \Rightarrow W(\frac{t\sigma^2}{\mu^3}).$$

Proof. The renewal theorem states that

$$E[|\frac{N(t)}{t} - \frac{1}{\mu}|] \to 0$$

and

$$\frac{N(t)}{t} \to \frac{1}{\mu}, \qquad a.s.$$

Let $S_k = \sum_{i=1}^{k} \xi_i$, $M(k) = S_k - \mu k$ and $\mathcal{F}_k = \sigma\{\xi_1, \ldots, \xi_k\}$. Then M is a $\{\mathcal{F}_k\}$-martingale and $N(t)+1$ is a $\{\mathcal{F}_k\}$-stopping time. By the optional sampling theorem $M(N(t)+1)$ is a martingale with respect to the filtration $\{\mathcal{F}_{N(t)+1}\}$.

Note that

$$\begin{aligned} M_n(t) &= -M(N(nt)+1)/(\mu\sqrt{n}) \\ &= \frac{N(nt)+1}{\sqrt{n}} - \frac{S_{N(nt)+1}-nt}{\mu\sqrt{n}} - \frac{nt}{\mu\sqrt{n}} \\ &= \frac{N(nt)-nt/\mu}{\sqrt{n}} + \frac{1}{\sqrt{n}} - \frac{1}{\mu\sqrt{n}}(S_{N(nt)+1}-nt) . \end{aligned}$$

So asymptotically Z_n behaves like M_n, which is a martingale for each n.

$$[M_n]_t = \frac{1}{\mu^2 n}\sum_{1}^{N(nt)+1}|\xi_k-\mu|^2 \xrightarrow{L_1} \frac{t\sigma^2}{\mu^3}.$$

□

Corollary 8.14 *Let $N(t)$ be a Poisson process with parameter λ and*

$$X(t) = \int_0^t (-1)^{N(s)} ds.$$

Define $X_n(t) = \frac{X(nt)}{\sqrt{n}}$. Then $X_n \Rightarrow \frac{1}{\sqrt{\lambda}} W$.

Proof. Note that

$$(-1)^{N(t)} = 1 - 2\int_0^t (-1)^{N(s-)} dN(s) = 1 - 2M(t) - 2\lambda \int_0^t (-1)^{N(s)} ds,$$

where

$$M(t) = \int_0^t (-1)^{N(s-)} d(N(s) - \lambda s)$$

is a martingale. Thus

$$X_n(t) = \frac{X(nt)}{\sqrt{n}} = \frac{1 - (-1)^{N(nt)}}{2\lambda\sqrt{n}} - \frac{M(nt)}{\lambda\sqrt{n}}.$$

$$[M_n]_t = N(nt)/(n\lambda^2) \to \frac{t}{\lambda}.$$

□

Multidimensional case

Theorem 8.15 *(Multidimensional Martingale CLT). Let $\{M_n\}$ be a sequence of $\mathbb{R}^d$-valued martingales. Suppose*

$$\lim_{n\to\infty} E[\sup_{s\leq t} |M_n(s) - M_n(s-)|] = 0$$

and

$$[M_n^i, M_n^j]_t \to c_{i,j}(t)$$

for all $t \geq 0$ where, $C = ((c_{i,j}))$ is deterministic and continuous. Then $M_n \Rightarrow M$, where M is Gaussian with independent increments and $E[M(t)M(t)^T] = C(t)$.

Remark 8.16 *Note that $C(t) - C(s)$ is nonnegative definite for $t \geq s \geq 0$. If C is differentiable, then the derivative will also be nonnegative definite and will have a nonnegative definite square root. Suppose $C(t) = \sigma(t)^2$ where σ is symmetric. Then M can be written as*

$$M(t) = \int_0^t \sigma(s) dW(s)$$

where W is d-dimensional standard Brownian motion.

9. Continuous-time Markov processes

- Markov processes corresponding to an operator semigroup
- Markov processes: Martingale problems
- Markov processes: Stability and stationary distributions

Markov processes: Semigroups

$\{T(t) : B(E) \to B(E), t \geq 0\}$ is an operator semigroup if $T(t)T(s)f = T(t+s)f$

X is a *Markov process* with operator semigroup $\{T(t)\}$ if and only if

$$E[f(X(t+s))|\mathcal{F}_t^X] = T(s)f(X(t)), \quad t, s \geq 0, f \in B(E).$$

$$\begin{aligned} T(s+r)f(X(t)) &= E[f(X(t+s+r))|\mathcal{F}_t^X] \\ &= E[E[f(X(t+s+r))|\mathcal{F}_{t+s}^X]|\mathcal{F}_t^X] \\ &= E[T(r)f(X(t+s))|\mathcal{F}_t^X] \\ &= T(s)T(r)f(X(t)) \end{aligned}$$

Lemma 9.1 *If X is a Markov processe corresponding to $\{T(t)\}$, then the finite dimensional distributions of X are determined by $\{T(t)\}$ and the distribution of $X(0)$.*

Proof. For $0 \leq t_1 \leq t_2$,

$$\begin{aligned} E[f_1(X(t_1))f_2(X(t_2))] &= E[f_1(X(t_1))T(t_2-t_1)f_2(X(t_1))] \\ &= E[T(t_1)[f_1T(t_2-t_1)f_2](X(0))] \end{aligned}$$

□

Semigroup generators

f is in the domain of the *strong generator* of the semigroup if there exists $g \in B(E)$ such that

$$\lim_{t \to 0+} \|g - \frac{T(t)f - f}{t}\| = 0.$$

Then $Af \equiv g$.

f is in the domain of the *weak generator* $\tilde{A}$, if $\sup_t \|t^{-1}(T(t)f - f)\| < \infty$, and there exists $g \in B(E)$ such that

$$\lim_{t \to 0+} \frac{T(t)f(x) - f(x)}{t} = g(x) \equiv \tilde{A}f(x), \quad x \in E.$$

See Dynkin (1965).

The *full generator* $\hat{A}$ is

$$\hat{A} = \{(f, g) \in B(E) \times B(E) : T(t)f = f + \int_0^t T(s)g ds$$

$A \subset \tilde{A} \subset \hat{A}$.

Martingale properties

Lemma 9.2 *If X is a progressive Markov process corresponding to $\{T(t)\}$ and $(f,g)\in\hat{A}$, then*

$$M_f(t)=f(X(t))-f(X(0))-\int_0^t g(X(s))ds$$

is a martingale.

Proof.

$$E[M_f(t+r)-M_f(t)|\mathcal{F}_t]=T(r)f(X(t))-f(X(t))-\int_t^{t+r}T(s-t)g(X(t))ds=0$$

□

Dynkin's identity

Change of notation: Simply write $\hat{A}f$ for g, if $(f,g) \in \hat{A}$.

The optional sampling theorem implies

$$E[f(X(t \wedge \tau))] = E[f(X(0))] + E[\int_0^{t\wedge\tau} \hat{A}f(X(s))ds].$$

Assume D is open and X is right continuous. Let $\tau_D = \inf\{t : X(t) \notin D\}$. Write E_x for expectations under the condition that $X(0) = x$.

Suppose f is bounded and continuous, $\hat{A}f = 0$, and $\tau_D < \infty$ a.s. Then

$$f(x) = E_x[f(X(\tau_D))].$$

If f is bounded and continuous, $\hat{A}f(x) = -1$, $x \in D$, and $f(y) = 0$, $y \notin D$, then

$$f(x) = E_x[\tau_D]$$

Exit distributions in one dimension

For a one-dimensional diffusion process

$$Lf(x) = \frac{1}{2}a(x)f''(x) + b(x)f'(x).$$

Find f such that $Lf(x) = 0$ (i.e., solve the linear first order differential equation for f'). Then $f(X(t))$ is a local martingale.

Fix $a < b$, and define $\tau = \inf\{t : X(t) \notin (a,b)\}$. If $\sup_{a<x<b} |f(x)| < \infty$, then

$$E_x[f(X(t\wedge\tau))] = f(x).$$

Moreover, if $\tau < \infty$ a.s.

$$E_x[f(X(\tau))] = f(x).$$

Hence

$$f(a)P_x(X(\tau) = a) + f(b)P_x(X(\tau) = b) = f(x),$$

and therefore the probability of exiting the interval at the right endpoint is given by

$$P_x(X(\tau) = b) = \frac{f(x) - f(a)}{f(b) - f(a)} \tag{9.1}$$

Exit time

To find conditions under which $P_x(\tau < \infty) = 1$, or more precisely, under which $E_x[\tau] < \infty$, solve $Lg(x) = -1$. Then

$$g(X(t)) - g((X(0)) - t,$$

is a local martingale and $C = \sup_{a<x<b} |g(x)| < \infty$,

$$E_x[g(X(t \wedge \tau))] = g(x) + E_x[t \wedge \tau]$$

and $2C \geq E[t \wedge \tau]$, so $2C \geq E[\tau]$, which implies $\tau < \infty$ a.s. By (9.1),

$$\begin{aligned} E_x[\tau] &= E_x[g(X(\tau))] - g(x) \\ &= g(b)\frac{f(x) - f(a)}{f(b) - f(a)} + g(a)\frac{f(b) - f(x)}{f(b) - f(a)} - g(x) \end{aligned}$$

Strongly continuous contraction semigroup

Semigroups associated with Markov processes are *contraction* semigroups, i.e.,

$$\|T(t)f\| \leq \|f\|, \quad f \in B(E).$$

Let $L_0 = \{f \in B(E) : \lim_{t\to 0+} \|T(t)f - f\| = 0$. Then

- $\mathcal{D}(A)$ is dense in L_0.
- $\|\lambda f - Af\| \geq \lambda \|f\|, \quad f \in \mathcal{D}(A)$, $\lambda > 0$.
- $\mathcal{R}(\lambda - A) = L_0$, $\forall \lambda > 0$.

The resolvent

Lemma 9.3 *For* $\lambda > 0$ *and* $h \in L_0$*,*

$$(\lambda - A)^{-1}h = \int_0^\infty e^{-\lambda t}T(t)hdt$$

Proof. Let $f = \int_0^\infty e^{-\lambda t}T(t)hdt$. Then

$$\begin{aligned} r^{-1}(T(r)f - f) &= r^{-1}(\int_0^\infty e^{-\lambda t}T(t+r)hdt - \int_0^\infty e^{-\lambda t}T(t)hdt) \\ &= r^{-1}(e^{\lambda r}\int_r^\infty e^{-\lambda t}T(t)hdt - \int_0^\infty e^{-\lambda t}T(t)hdt) \\ &\to \lambda f - h \end{aligned}$$

□

Hille-Yosida theorem

Theorem 9.4 *The closure of A is the generator of a strongly continuous contraction semigroup on L_0 if and only if*

- $\mathcal{D}(A)$ *is dense in* L_0.
- $\|\lambda f - Af\| \geq \lambda \|f\|, \quad f \in \mathcal{D}(A), \lambda > 0.$
- $\mathcal{R}(\lambda - A)$ *is dense in* L_0.

Proof. Necessity is discussed above. Assuming A is closed (otherwise, replace A by its closure), the conditions imply $\mathcal{R}(\lambda - A) = L_0$ and the semigroup is obtained by

$$T(t)f = \lim_{n\to\infty} (I - \frac{1}{n}A)^{-[nt]} f.$$

(One must show that the right side is Cauchy.) □

Probabilistic interpretation of the limit

If $T(t)$ is given by a transition function, then

$$(I - \frac{1}{n}A)^{-1} f(x) = E_x[f(X(\frac{1}{n}\Delta))],$$

where Δ is a unit exponential independent of X, and

$$(I - \frac{1}{n}A)^{-[nt]} f(x) = E_x[f(X(\frac{1}{n}\sum_{i=1}^{[nt]}\Delta_i))]$$

The resolvent for the full generator

Lemma 9.5 *Suppose* $T(t) : B(E) \to B(E)$ *is given by a transition function,* $T(t)f(x) = \int_E f(y)P(t,x,dy)$. *For* $h \in B(E)$, *define*

$$f(x) = \int_0^\infty e^{-\lambda t} T(t)h(x)dt.$$

Then $(f, \lambda f - h) \in \hat{A}$.

Proof.

$$\begin{aligned}
\int_0^t T(s)(\lambda f - h)ds &= \lambda \int_0^t \int_0^\infty e^{-\lambda u} T(s+u)hduds - \int_0^t T(s)hds \\
&= \lambda \int_0^t e^{\lambda s} \int_s^\infty e^{-\lambda u} T(u)hduds - \int_0^t T(s)hds \\
&= e^{\lambda t} \int_t^\infty e^{-\lambda u} T(u)hdu - \int_0^\infty e^{-\lambda u} T(u)hdu \\
&= T(t)f - f
\end{aligned}$$

□

A convergence lemma

Lemma 9.6 *Let E be compact and suppose $\{f_k\} \subset C(E)$ separates points. If $\{x_n\}$ satisfies $\lim_{n\to\infty} f_k(x_n)$ exists for every f_k, then $\lim_{n\to\infty} x_n$ exists.*

Proof. If x and x' are limit points of $\{x_n\}$, we must have $f_k(x) = f_k(x')$ for all k. But then $x = x'$, since $\{f_k\}$ separates points. □

Feller processes

Lemma 9.7 *Assume E is compact, $T(t) : C(E) \to C(E)$, and $\lim_{t\to 0} T(t)f(x) = f(x)$, $x \in E$, $f \in C(E)$. If X is a Markov process corresponding to $\{T(t)\}$, then X has a modification with cadlag sample paths.*

Proof. For $h \in C(E)$, $f = R_\lambda h \equiv \int_0^\infty e^{-\lambda t} T(t) h dt \in C(E)$, so setting $g = \lambda f - h$,

$$f(X(t)) - f(X(0)) - \int_0^t g(X(s))ds$$

is a martingale. By the upcrossing inequality, there exists a set $\Omega_f \subset \Omega$ with $P(\Omega_f) = 1$ such that for $\omega \in \Omega_f$, $\lim_{s\to t+, s\in\mathbb{Q}} f(X(s,\omega))$ exists for each $t \geq 0$ and $\lim_{s\to t-, s\in\mathbb{Q}} f(X(s,\omega))$ exists for each $t > 0$.

Suppose $\{h_k, k \geq 1\} \subset C(E)$ is dense. Then $\{R_\lambda h_k : \lambda \in \mathbb{Q} \cap (0,\infty), k \geq 1\}$ separates points in E. $\square$

Markov processes: Martingale problems

E state space (a complete, separable metric space)

A generator (a linear operator with domain and range in $B(E)$

$\mu \in \mathcal{P}(E)$

X is a solution of the martingale problem for (A, μ) if and only if $\mu = PX(0)^{-1}$ and there exists a filtration $\{\mathcal{F}_t\}$ such that

$$f(X(t)) - \int_0^t Af(X(s))ds$$

is an $\{\mathcal{F}_t\}$-martingale for each $f \in \mathcal{D}(A)$

Examples of generators

Standard Brownian motion ($E = \mathbb{R}^d$)

$$Af = \frac{1}{2}\Delta f, \quad \mathcal{D}(A) = C_c^2(\mathbb{R}^d)$$

Poisson process ($E = \{0, 1, 2 \ldots\}$, $\mathcal{D}(A) = B(E)$)

$$Af(k) = \lambda(f(k+1) - f(k))$$

Pure jump process (E arbitrary)

$$Af(x) = \lambda(x) \int_E (f(y) - f(x))\mu(x, dy)$$

Diffusion ($E = \mathbb{R}^d$, $\mathcal{D}(A) = C_c^2(\mathbb{R}^d)$)

$$Af(x) = \frac{1}{2} \sum_{i,j} a_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_i b_i(x) \frac{\partial}{\partial x_i} f(x) \tag{9.2}$$

Conditions for the martingale property

Lemma 9.8 *For* $(f,g) \in A$, $h_1,\ldots,h_m \in \bar{C}(E)$, *and* $t_1 \leq t_2 \leq \cdots \leq t_{m+1}$, *let*

$$\eta(Y) \equiv \eta(Y,(f,g),\{h_i\},\{t_i\}) = (f(Y(t_{m+1}) - f(Y(t_m)) - \int_{t_m}^{t_{m+1}} g(Y(s)ds) \prod_{i=1}^{m} h_i(Y(t_i)).$$

Then Y *is a solution of the martingale problem for* A *if and only if* $E[\eta(Y)] = 0$ *for all such* η.

The assertion that Y is a solution of the martingale problem for A is an assertion about the finite dimensional distributions of Y.

Uniqueness and the Markov property

Theorem 9.9 *If any two solutions of the martingale problem for A satisfying $PX_1(0)^{-1} = PX_2(0)^{-1}$ also satisfy $PX_1(t)^{-1} = PX_2(t)^{-1}$ for all $t \geq 0$, then the f.d.d. of a solution X are uniquely determined by $PX(0)^{-1}$*

If X is a solution of the MGP for A and $Y_a(t) = X(a+t)$, then Y_a is a solution of the MGP for A.

Markov property

Theorem 9.10 *Suppose the conclusion of Theorem 9.9 holds. If X is a solution of the martingale problem for A with respect to a filtration $\{\mathcal{F}_t\}$, then X is Markov with respect to $\{\mathcal{F}_t\}$.*

Proof. Assuming that $P(F) > 0$, let $F \in \mathcal{F}_r$ and define

$$P_1(B) = \frac{E[\mathbf{1}_F E[\mathbf{1}_B | \mathcal{F}_r]]}{P(F)}, \quad P_2(B) = \frac{E[\mathbf{1}_F E[\mathbf{1}_B | X(r)]]}{P(F)}.$$

Define $Y(t) = X(r+t)$. Then

$$P_1\{Y(0) \in \Gamma\} = \frac{E[\mathbf{1}_F E[\mathbf{1}_{\{Y(0)\in\Gamma\}} | \mathcal{F}_r]]}{P(F)} = \frac{E[\mathbf{1}_F E[\mathbf{1}_{\{X(r)\in\Gamma\}} | \mathcal{F}_r]]}{P(F)} = P_2\{Y(0) \in \Gamma\}$$

Check the $E^{P_1}[\eta(Y)] = E^{P_2}[\eta(Y)] = 0$ for all $\eta(Y)$ as in Lemma 9.8. Therefore

$$E[\mathbf{1}_F E[f(X(r+t))|\mathcal{F}_r]] = P(F)E^{P_1}[f(Y(t))] = P(F)E^{P_2}[f(Y(t))] = E[\mathbf{1}_F E[f(X(r+t))|X(r$$

□

Cadlag versions

Lemma 9.11 *Suppose E is compact and $A \subset \bar{C}(E) \times B(E)$. If $\mathcal{D}(A)$ is separating, then any solution of the martingale problem for A has a cadlag modification.*

Quasi-left continuity

X is *quasi-left continuous* if and only if for each sequence of stopping times $\tau_1 \leq \tau_2 \leq \cdots$ such that $\tau \equiv \lim_{n\to\infty} \tau_n < \infty$ a.s.,

$$\lim_{n\to\infty} X(\tau_n) = X(\tau) \quad a.s.$$

Lemma 9.12 *Let $A \subset \bar{C}(E) \times B(E)$, and suppose that $\mathcal{D}(A)$ is separating. Let X be a cadlag solution of the martingale problems for A. Then X is quasi-left continuous*

Proof. For $(f,g) \in A$,

$$\begin{aligned} \lim_{n\to\infty} f(X(\tau_n \wedge t) &= \lim_{n\to\infty} E[f(X(X(\tau \wedge t)) - \int_{\tau_n \wedge t}^{\tau \wedge t} g(X(s))ds | \mathcal{F}_{\tau_n}] \\ &= E[f(X(\tau \wedge t))| \vee_n \mathcal{F}_{\tau_n}] \,. \end{aligned}$$

See Exercise 10. □

Continuity of diffusion process

Lemma 9.13 *Suppose $E = \mathbb{R}^d$ and*

$$Af(x) = \frac{1}{2}\sum_{i,j} a_{ij}(x)\frac{\partial^2}{\partial x_i \partial x_j} f(x) + \sum_i b_i(x)\frac{\partial}{\partial x_i} f(x), \quad \mathcal{D}(A) = C_c^2(\mathbb{R}^d).$$

If X is a solution of the martingale problem for A, then X has a modification that is cadlag in $\mathbb{R}^d \cup \{\infty\}$. If X is cadlag, then X is continuous.

Proof. The existence of a cadlag modification follows by Lemma 9.11. To show continuity, it is enough to show that for $f \in C_c^\infty(\mathbb{R}^d)$, $f \circ X$ is continuous. To show $f \circ X$ is continuous, it is enough to show

$$\lim_{\max|t_{i+1}-t_i| \to 0} \sum (f(X(t_{i+1} \wedge t) - f(X(t_i \wedge t)))^4 = 0.$$

¿From the martingale properties,

$$E[(f(X(t+h)) - f(X(t)))^4]$$
$$= \int_t^{t+h} E\Big[Af^4(X(s)) - 4f(X(t))Af^3(X(s))$$
$$+6f^2(X(t))Af^2(X(s)) - 4f^3(X(t))Af(X(s))\Big]ds$$

Check that

$$Af^4(x) - 4f(x)Af^3(x) + 6f^2(x)Af^2(x) - 4f^3(x)Af(x) = 0$$

□

Markov processes: Stability and stationary distributions

- Extension of martingale properties
- Moment estimates
- Stationary distributions

Extension of martingale properties

Lemma 9.14 *Suppose* X *is a solution of the martingale problem for* A*,* $\{(f_n, g_n)\} \subset A$*,* $\inf_{x,n} f_n(x) > -\infty$*,* $\sup_{x,n} g_n(x) < \infty$*,* $f_n(x) \to f(x)$*,* $g_n(x) \to g(x)$*,* $x \in E$*. Then*

$$Z_f(t) = f(X(t)) - f(X(0)) - \int_0^t g(X(s))ds$$

is a supermartingale.

Extension for diffusion processes

Let L be the differential operator that defines A in (9.2) for $f \in C_c^2(\mathbb{R}^d)$.

Lemma 9.15 *Suppose* $f \in C^2(\mathbb{R}^d)$, $\inf_x f(x) > -\infty$ *and* $\sup_x Lf(x) < \infty$. *Then*

$$f(X(t)) - f(X(0)) - \int_0^t Lf(X(s))ds$$

is a supermartingale

Proof. For each r, there exist $f_r \in C_c^2(\mathbb{R}^d)$ such that $f(x) = f_r(x)$ for $|x| \leq r$. Consequently, defining $\tau_r = \inf\{t : |X(t)| \geq r\}$,

$$f(X(t \wedge \tau_r)) - f(X(0)) - \int_0^{t \wedge \tau_r} Lf(X(s))ds$$

is a martingale. Letting $r \to \infty$, the lemma follows by Fatou's lemma. (We assumed here that $\tau_r \to \infty$ for $r \to \infty$.) □

Moment estimates

Lemma 9.16 *Suppose A is given by (9.2) and*

$$g(x)=\sum_i a_{ii}(x)+2x\cdot b(x)\leq K_1+K_2|x|^2.$$

If X is a solution of the martingale problem for A, and $E[|X(0)|^2]<\infty$, then $E[|X(t)|^2]<\infty$ for all $t>0$.

Proof. Taking $f(x)=|x|^2$, $Lf=\sum_i a_{ii}(x)+2x\cdot b(x)$, and

$$\begin{aligned} E[|X(t\wedge\tau_r)|^2] &= E[|X(0)|^2]+E[\int_0^{t\wedge\tau_r} g(X(s))ds] \\ &\leq E[|X(0)|^2]+\int_0^t (K_1+K_2E[|X(s\wedge\tau_r)|^2)ds \end{aligned}$$

so $E[|X(t\wedge\tau_r|^2]\leq(E[|X(0)|^2]+K_1t)e^{K_2t}$. □

Lemma 9.17 *If X is a solution of the martingale problem for A, $f \in \mathcal{D}(A)$, and $\gamma : [0,\infty) \to \mathbb{R}$ continuously differentiable, then*

$$\gamma(t)f(X(t)) - \int_0^t (\gamma'(s)f(X(s)) + \gamma(s)Af(X(s)))ds$$

is a martingale

Proof.

$$\begin{aligned} &E[\gamma(t+r)f(X(t+r)) - \gamma(t)f(X(t))|\mathcal{F}_t] \\ &\qquad = E[\sum \gamma(t_{i+1})f(X(t_{i+1})) - \gamma(t_i)f(X(t_i))|\mathcal{F}_t] \\ &\qquad = E[\sum \gamma(t_{i+1})\int_{t_i}^{t_{i+1}} Af(X(s))ds + (\gamma(t_{i+1}) - \gamma(t_i))f(X(t_i))|\mathcal{F}_t] \end{aligned}$$

□

Lemma 9.18 *Suppose A is given by (9.2) and*

$$g(x) = \sum_i a_{ii}(x) + 2x \cdot b(x) \leq K_1 - K_2|x|^2, \quad K_1, K_2 > 0.$$

If X is a solution of the martingale problem for A, and $E[|X(0)|^2] < \infty$, then $\sup_t E[|X(t)|^2] < \infty$.

Proof.

$$Z(t) = |X(t)|^2 e^{K_2 t} - \int_0^t K_1 e^{K_2 s} ds$$

is a supermartingale, so

$$E[|X(t)|^2] e^{K_2 t} \leq E[|X(0)|^2] + \frac{K_1}{K_2}(e^{K_2 t} - 1)$$

□

Stationary distributions

$\mu \in \mathcal{P}(E)$ is a *stationary distribution* for A is there is a solution of the martingale problem for (A, μ) that is a stationary process.

Lemma 9.19 *If μ is a stationary distribution for A, then*

$$\int_E Af d\mu = 0, \quad f \in \mathcal{D}(A).$$

Proof. If X is a stationary solution, then

$$\begin{aligned} 0 &= E[f(X(t)) - f(X(0)) - \int_0^t Af(X(s))ds] \\ &= \langle f, \mu\rangle - \langle f, \mu\rangle - \int_0^t \langle Af, \mu\rangle ds \end{aligned}$$

□

Stationary distributions for semigroups

Lemma 9.20 *If A is the generator for a semigroup on $L \subset B(E)$, $\mu \in \mathcal{P}(E)$, and $\langle Af, \mu \rangle = 0$, then $\langle T(t)f, \mu \rangle = \langle f, \mu \rangle$, $f \in L$.*

Proof. If $f \in \mathcal{D}(A)$, then $T(t)f \in \mathcal{D}(A)$ and $AT(t)f = T(t)Af$. Consequently,

$$\langle T(t)f, \mu \rangle = \langle f, \mu \rangle + \int_0^t \langle AT(s)f, \mu \rangle ds = \langle f, \mu \rangle.$$

Since $\mathcal{D}(A)$ is dense in L, the identity extends to all $f \in L$. □

Note: The assertion that A generates the semigroup requires verification of the range condition in the Hille-Yosdia theorem.

Conditions on the generator

$A \subset B(E) \times B(E)$ is a *pre-generator* if A is dissipative (if A is linear, $\|\lambda f - Af\| \geq \lambda\|f\|$, $\lambda > 0$, $f \in \mathcal{D}(A)$) and there are sequences of functions $\mu_n : E \to \mathcal{P}(E)$ and $\lambda_n : E \to [0, \infty)$ such that for each $(f, g) \in A$

$$g(x) = \lim_{n\to\infty} \lambda_n(x)(\int_E (f(y) - f(x))\mu_n(x, dy)$$

for each $x \in E$.

A is *bp-separable* if there exists a countable subset $\{g_k\} \subset \mathcal{D}(A) \cap \bar{C}(E)$ such that the graph of A is contained in the bounded, pointwise closure of $\{(g_k, Ag_k)\}$.

i) $A : \mathcal{D}(A) \subset \overline{C}(E) \to C(E)$, $1 \in \mathcal{D}(A)$, and $A1 = 0$.

ii) There exist $\psi \in C(E)$, $\psi \geq 1$, and constants a_f, $f \in \mathcal{D}(A)$, such that

$$|Af(x)| \leq a_f \psi(x), \qquad x \in E.$$

iii) Defining $A_0 = \{(f, \psi^{-1}Af) : f \in \mathcal{D}(A)\}$, A_0 is bp-separable and a pre-generator.

iv) $\mathcal{D}(A)$ is closed under multiplication and separates points.

Echeverria's theorem

Theorem 9.21 *Suppose that A satisfies the condtions above. Let $\mu \in \mathcal{P}(E)$ satisfy*

$$\int_E \psi d\mu < \infty$$

and

$$\int_E Af d\mu = 0, \quad f \in \mathcal{D}(A).$$

Then there exists a stationary solution of the martingale problem for (A, μ).

Example: Diffusion processes

Let $d = 1$. Integrating by parts, we have

$$\int_{-\infty}^{\infty} p(x)\left(\frac{1}{2}a(x)f''(x) + b(x)f'(x)\right)dx$$
$$= \frac{1}{2}p(x)a(x)f'(x)\Big|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} f'(x)\left(\frac{1}{2}\frac{d}{dx}(a(x)p(x)) - b(x)p(x)\right)dx.$$

The first term is zero, and integrating by parts again

$$\int_{-\infty}^{\infty} f(x)\frac{d}{dx}\left(\frac{1}{2}\frac{d}{dx}(a(x)p(x)) - b(x)p(x)\right)dx$$

so solve

$$\frac{d}{dx}\underbrace{\left(\frac{1}{2}\frac{d}{dx}(a(x)p(x)) - b(x)p(x)\right)}_{\substack{\text{this is a constant:}\\ \text{let the constant be 0}}} = 0,$$

so

$$\frac{1}{2}\frac{d}{dx}(a(x)p(x)) = b(x)p(x).$$

Applying the integrating factor $\exp(-\int_0^x 2b(z)/a(z)dz)$ to get a perfect differential, we have

$$\frac{1}{2}e^{-\int_0^x \frac{2b(z)}{a(z)}dz}\frac{d}{dx}(a(x)p(x)) - b(x)e^{-\int_0^x \frac{2b(z)}{a(z)}dz}p(x) = 0$$

$$a(x)e^{-\int_0^x \frac{2b(z)}{a(z)}dz}p(x) = C$$

$$p(x) = \frac{C}{a(x)}e^{\int_0^x \frac{2b(z)}{a(z)}dz}.$$

Assume $a(x) > 0$ for all x. The condition for the existence of a stationary distribution is

$$\int_{-\infty}^{\infty} \frac{1}{a(x)}e^{\int_0^x \frac{2b(z)}{a(z)}dz}dx < \infty.$$

Example: Spatial birth and death processes

Let $\nu \in \mathcal{M}_f(S)$

$$Af(\eta) = \int_S (f(\eta + \delta_y) - f(\eta))\nu(dy) + \int_S (f(\eta - \delta_x) - f(\eta))\eta(dx)$$

for

$$f \in \mathcal{D}(A) = \{e^{-\langle h,\eta \rangle} : \inf_x h(x) > 0\}.$$

The stationary distribution

Let ξ be a Poisson random measure with mean measure ν. Then

$$E[\int_S h(\xi-\delta_x,x)\xi(dx)]=E[\int_S h(\xi,x)\nu(dx)]. \tag{9.3}$$

Consequently,

$$E[\int_S (f(\xi-\delta_x)-f(\xi))\xi(dx)]=E[\int_S f(\xi)\nu(dx)]-E[\int_S f(\xi+\delta_x)\nu(dx)]$$

so

$$\int_{\mathcal{N}(S)} Af(\eta)\mu_\nu^0(d\eta)=0.$$

where μ_ν^0 is the distribution of the Poisson random measure with mean measure ν.

Moment lemma for Poisson random measures

Let ξ be a Poisson random measure on E with nonatomic mean measure ν. Let $\{E_k^n\}$ be a sequence of partitions of E with $\max\text{diam}(E_k^n) \to 0$, and let $x_k^n \in E_k^n$. Then for bounded, continuous, $h : \mathcal{N}(E) \times E \to \mathbb{R}$ and $F \in \mathcal{B}(E)$ with $\nu(F) < \infty$,

$$\begin{aligned}
E[\int_F h(\xi - \delta_x, x)\xi(dx)] &= \lim_{n\to\infty} \sum_k E[h(\xi - \xi_{E_k^n \cap F}, x_k^n)\xi(E_k^n \cap F)] \\
&= \lim_{n\to\infty} \sum_k E[h(\xi - \xi_{E_k^n \cap F}, x_k^n)\nu(E_k^n \cap F)] \\
&= \int_F E[h(\xi, x)]\nu(dx)
\end{aligned}$$

Existence of stationary distributions: Feller case

Lemma 9.22 *Suppose $\{T(t)\}$ is a Feller semigroup corresponding to a Markov process X. Let $\nu_t = PX^{-1}(t)$, and define*

$$\mu_t = \frac{1}{t}\int_0^t \nu_s ds$$

If $\{\mu_t\}$ is tight, then any limit point is a stationary measure for $\{T(t)\}$.

Proof. Suppose μ_{t_n} converges weakly to μ_∞. Then for $f \in \bar{C}(E)$,

$$\begin{aligned}
\langle T(r)f, \mu_\infty\rangle &= \lim_{n\to\infty}\langle T(r)f, \mu_{t_n}\rangle \\
&= \lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}\langle T(r)f, \nu_s\rangle ds \\
&= \lim_{n\to\infty}\frac{1}{t_n}\int_0^{t_n}\langle f, \nu_{s+r}\rangle ds \\
&= \lim_{n\to\infty}\frac{1}{t_n}\int_r^{r+t_n}\langle f, \nu_s\rangle = \langle f, \mu_\infty\rangle
\end{aligned}$$

□

Existence of stationary distributions: Generators

Lemma 9.23 *Suppose $A \subset \bar{C}(E) \times \bar{C}(E)$ and A satisfies the conditions of Theorem 9.21. Suppose f, g satisfy the conditions of Lemma 9.14 and that $K_a = \{x : g(x) \geq -a\}$ is compact for each $a > 0$. Then there exists a stationary distribution*

Proof. Assume that $E[f(X(0))] < \infty$. Then

$$E[f(X(t))] \leq E[f(X(0))] + E[\int_0^t g(X(s))ds]$$

and letting $C_1 = \sup_x g(x)$ and $C_2 = \inf_x f(x)$,

$$\begin{aligned} aE[\int_0^t \mathbf{1}_{K_a^c}(X(s))] - C_1 E\int_0^t \mathbf{1}_{K_a}(X(s))ds] &\leq [E[\int_0^t (-g)(X(s))ds] \\ &\leq E[f(X(0)] - C_2 \end{aligned}$$

and

$$\mu_t(K_a^c) \leq \frac{E[f(X(0))] - C_2}{ta} + \frac{C_1}{a},$$

so $\{\mu_t\}$ is tight. Since $t^{-1}(\langle f, \nu_t\rangle - \langle f, \nu_0\rangle) - \langle Af, \mu_t\rangle = 0$, and $Af \in \bar{C}(E)$, any limit point of $\{\mu_t\}$ will satisfy $\langle Af, \mu_\infty\rangle = 0$. □

Example

Let $Af(x) = \frac{1}{2}\sum a_{ij}(x)\partial_i\partial_j f(x) + \sum_i b_i(x)\partial_i f(x)$, and let $f(x) = |x|^2$. Then

$$g(x) = \sum a_{ii}(x) + 2b(x)\cdot x$$

so if $\lim_{|x|\to\infty} \sum a_{ii}(x) + 2b(x)\cdot x = -\infty$ and the a_{ij} and b_i are continuous, there exists a stationary distribution.

Birth and death processes: Stationary distribution

Want

$$\sum_k \pi_k Af(k) = 0, \quad f \text{ with finite support.}$$

For $f = \delta_k$,

$$\pi_{k+1}\mu_{k+1} + \pi_{k-1}\lambda_{k-1} - \pi_k(\lambda_k + \mu_k) = 0,$$

which is implied by $\pi_k\mu_k = \pi_{k-1}\lambda_{k-1}$, $k = 1, 2, \ldots$. Consequently,

$$\pi_k = \pi_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$$

Example: Spatial birth and death processes

Let $\nu \in \mathcal{M}_f(S)$

$$Af(\eta) = \int_S \beta(y,\eta)(f(\eta+\delta_y) - f(\eta))\nu(dy) + \int_S \delta(x,\eta)(f(\eta-\delta_x) - f(\eta))\eta(dx)$$

where β and δ are continuous. Let $f(\eta) = e^{\alpha|\eta|}$. Then

$$g(\eta) = e^{\alpha|\eta|}\left(\int_S (e^{\alpha}-1)\beta(y,\eta)\nu(dy) - \int_S (1-e^{-\alpha})\delta(x,\eta)\eta(dx)\right).$$

Suppose

$$\begin{aligned} \int_S \beta(y,\eta)\nu(dy) &\leq \bar{\beta}(|\eta|) \\ \int_S \delta(x,\eta)\eta(dx) &\geq \underline{\delta}(|\eta|) \end{aligned}$$

If

$$\sum_k \prod_{i=0}^{k-1} \frac{\bar{\beta}(i)}{\underline{\delta}(i+1)} < \infty,$$

then there is a unique stationary distribution for A.

MCMC for spatial point processes

Consider the class of spatial point processes specified through a density (Radon-Nikodym derivative) with respect to a Poisson point process with mean measure ν, that is, the distribution of the point process is given by

$$\mu_{\nu,H}(d\eta) = \frac{1}{Z_{\nu,H}} e^{-H(\eta)} \mu_\nu^0(d\eta), \tag{9.4}$$

where $H(\eta)$ is referred to as the *energy function*, $Z_{\nu,H}$ is a normalizing constant, and μ_ν^0 is the law of a Poisson process with mean measure ν.

Assuming $Z_{\nu,H}$ exists, $\mu_{\nu,H}$ is a probability measure on $\mathcal{S} = \{\eta \in \mathcal{N}(S); H(\eta) < \infty\}$.

H is *hereditary* in the sense of Ripley (1977), if $H(\eta) < \infty$ and $\tilde{\eta} \subset \eta$ implies $H(\tilde{\eta}) < \infty$.

Conditions to be a stationary distribution

Suppose that $\beta(x,\eta) > 0$ if $H(\eta+\delta_x) < \infty$ and that β and δ satisfy

$$\beta(x,\eta)e^{-H(\eta)} = \delta(x,\eta+\delta_x)e^{-H(\eta+\delta_x)}. \tag{9.5}$$

This equation is a *detailed balance condition* which ensures that births from η to $\eta+\delta_x$ match deaths from $\eta+\delta_x$ to η and that the process is time-reversible with (9.4) as its stationary distribution. Since

$$\begin{aligned} Af(\eta) = \int_S \delta(y,\eta+\delta_y)e^{H(\eta)-H(\eta+\delta_y)}(f(\eta+\delta_y)-f(\eta))\nu(dy) \\ + \int_S \delta(x,\eta)(f(\eta-\delta_x)-f(\eta))\eta(dx), \end{aligned}$$

the Poisson identity implies

$$\int Af(\eta)\mu_{\nu,H}(d\eta) = \frac{1}{Z_{\nu,H}}\int Af(\eta)e^{-H(\eta)}\mu_\nu^0(d\eta) = 0.$$

(9.5) holds for any pair of birth and death rates such that

$$\frac{\beta(x,\eta)}{\delta(x,\eta+\delta_x)} = \exp\{-H(\eta+\delta_x)+H(\eta)\}.$$

Pairwise interaction point processes

Take $\delta(x,\eta)=1$, that is, whenever a point is added to the configuration, it lives an exponential length of time independently of the configuration of the process.

$$\begin{aligned} H_\rho(\eta) &= \sum_{i<j}\rho(x_i,x_j) \\ &= \frac{1}{2}[\int\int\rho(x,y)\eta(dx)\eta(dy)-\int\rho(x,x)\eta(dx)]\} \end{aligned}$$

Then $\beta(x,\eta)=\exp\{-\int\rho(x,y)\eta(dy)\}$ and

$$Af(\eta)=\int e^{-\int\rho(x,y)\eta(dy)}(f(\eta+\delta_x)-f(\eta))dx+\int(f(\eta-\delta_x)-f(\eta))\eta(dx).$$

10. Diffusion approximations

- Convergence of generators
- Limits of martingales should be martingales
- Tightness based on generator estimates
- Diffusion approximations
- Heavy traffic limits for queueing models

Convergence of generators

$\{A_n\}$ a sequence of generators for Markov processes with state space E.

Convergence condition: For each $(f, g) \in A \subset \bar{C}(E) \times \bar{C}(E)$, there exist $(f_n, g_n) \in A_n$, $n = 1, 2, \ldots$, such that $\sup_n(\|f_n\| + \|g_n\|) < \infty$ and

$$f_n \to f, \quad g_n \to g \quad \text{uniformly on compact subsets of } E$$

Limits of martingales should be martingales

Lemma 10.1 *Assume that the Convergence Condition holds. Suppose $\{(X_n, Z_n)\}$ is relatively compact in $D_{E\times E'}[0,\infty)$ and X_n is a solution of the martingale problem for A_n with respect to the filtration $\{\mathcal{F}_t^{X_n,Z_n}\}$. If (X, Z) is a limit point of $\{(X_n, Z_n)\}$, then X is a solution of the martingale problem for A with respect to $\{\mathcal{F}_t^{X,Z}\}$.*

Proof. Suppose $(X_n, Z_n) \Rightarrow (X, Z)$. Let $T_d = \{t : P\{(X(t), Z(t)) \neq (X(t-), Z(t-))\} > 0\}$. ($T_d$ is countable.) Then for $\{t_i\} \subset T_d^c$, $h_i \in \bar{C}(E \times E')$, and $(f_n, g_n) \to (f, g)$ as in the convergence condition,

$$(f_n(X_n(t_{m+1})) - f_n(X_n(t_m)) - \int_{t_m}^{t_{m+1}} g_n(X_n(s))ds) \prod h_i(X_n(t_i), Z_n(t_i))$$

converges in distribution to

$$(f(X(t_{m+1})) - f(X(t_m)) - \int_{t_m}^{t_{m+1}} g(X(s))ds) \prod h_i(X(t_i), Z(t_i))$$

which by the martingale properties of X_n and the boundedness of $\{(f_n, g_n)\}$ must have expectation zero. □

Tightness based on generator estimates

Suppose $\mathcal{D}(A)$ is closed under multiplication. Then for $f \in \mathcal{D}(A)$ and $\tau_K^n = \inf\{s > 0 : X_n(s) \notin K\}$,

$$\begin{aligned} &E[(f(X_n(t+u)) - f(X_n(t)))^2|\mathcal{F}_t^n] \\ &\quad = E[f^2(X_n(t+u)) - f^2(X_n(t))|\mathcal{F}_t^n] - 2f(X_n(t))E[f(X_n(t+u)) - f(X_n(t))|\mathcal{F}_t^n] \\ &\quad \leq 2\sup_{x\in K}|f^2(x) - \tilde{f}_n(x)| + C\sup_{x\in K}|f(x) - f_n(x)| + CE[\mathbf{1}_{\{\tau_K^n \leq t+u\}}|\mathcal{F}_t^n] \\ &\qquad\qquad + E[\int_t^{t+u}(\tilde{g}_n(X_n(s)) + C|g_n(X_n(s))|)ds|\mathcal{F}_t^n] \end{aligned}$$

Consequently, if the convergence condition holds and there exists a sequence of compact sets K_m such that

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\{\tau_{K_m}^n \leq T\} = 0,$$

$\{X_n\}$ is relatively compact.

Diffusion approximations

For $n = 1, 2, \ldots$, let $\{Y_k^n, k \geq 0\}$ be a Markov chain in $\mathbb{R}^d$ with transition function $\mu_n(x, dy)$. Suppose

$$\lim_{n\to\infty} n \int (y-x)\mu_n(x,dy) = b(x), \qquad \lim_{n\to\infty} n \int (y-x)(y-x)^T \mu_n(x,dy) = a(x)$$

uniformly on compact $K \subset \mathbb{R}^d$, and

$$\lim_{n\to\infty} \sup_x n \int |x-y|^3 \mu_n(x,dy) = 0.$$

Let

$$X_n(t) = Y_{[nt]}^n.$$

Define $A_n f(x) = n(\int f(y)\mu_n(x,dy) - f(x))$

$$f(X_n(t)) - f(X_n(0)) - \sum_{k=0}^{[nt]-1} (\mu_n f(Y_k^n) - f(Y_k^n) = f(X_n(t)) - f(X_n(0)) - \int_0^{\frac{[nt]}{n}} A_n f(X_n(s))ds$$

is a martingale, and for $f \in C_2^c(\mathbb{R}^d)$

$$\lim_{n\to\infty} A_n f(x) = \frac{1}{2} \sum_{i,j} a_{ij}(x)\partial_i\partial_j f(x) + \sum_i b_i(x)\partial_i f(x).$$

Heavy traffic limits for queueing models

Queueing model with Poisson arrivals and exponential service times:

$$Bf(k) = \lambda(f(k+1) - f(k)) + \mu \mathbf{1}_{\{k>0\}}(f(k-1) - f(k))$$

Suppose $\sqrt{n}(\mu_n - \lambda_n) \to c$ and $\lambda_n \to \lambda$. Define $X_n(t) = \frac{Q(nt)}{\sqrt{n}}$, and

$$\begin{aligned} A_n f(x) &= n\lambda_n(f(x + \frac{1}{\sqrt{n}}) - f(x)) + n\mu_n \mathbf{1}_{\{x>0\}}(f(x - \frac{1}{\sqrt{n}}) - f(x)) \\ &= \sqrt{n}(\lambda_n - \mu_n)f'(x) + \frac{1}{2}(\lambda_n + \mu_n)f''(x) + O(\frac{1}{\sqrt{n}}) \\ &\quad - n\mu_n \mathbf{1}_{\{x=0\}}(f(x - \frac{1}{\sqrt{n}}) - f(x)) \end{aligned}$$

If $f'(0) = 0$,

$$\lim_{n\to\infty} A_n f(x) = \lambda f''(x) - cf'(x) - \frac{\lambda}{2}\mathbf{1}_{\{x=0\}} f''(0)$$

Let $f_n(x) = f(x) + \frac{1}{\sqrt{n}}h(x)$. Then

$$\lim_{n\to\infty} A_n f_n(x) = \lambda f''(x) - cf'(x) - \frac{\lambda}{2}\mathbf{1}_{\{x=0\}} f''(0) + \lambda \mathbf{1}_{\{x=0\}} h'(0).$$

Quality control schemes

$Y_1, Y_2, \ldots$ process measurements in $\mathbb{R}^d$

a target mean, that is, we want $E[Y_k] = a$

CUSUM (cumulative sum) procedures

Page: For $d = 1$, $K^- < a < K^+$,

$$S^H_{k+1} = \max(0, S^H_k + Y_{k+1} - K^+)$$
$$S^L_{k+1} = \min(0, S^L_k + Y_{k+1} - K^-)$$

Crosier: Two-sided procedure

$$S_{k+1} = (S_k + Y_{k+1} - a) \times 0 \vee (1 - \frac{K}{|S_k + Y_{k+1} - a|})$$

If the Y_k are independent, then the recursions give Markov chains.

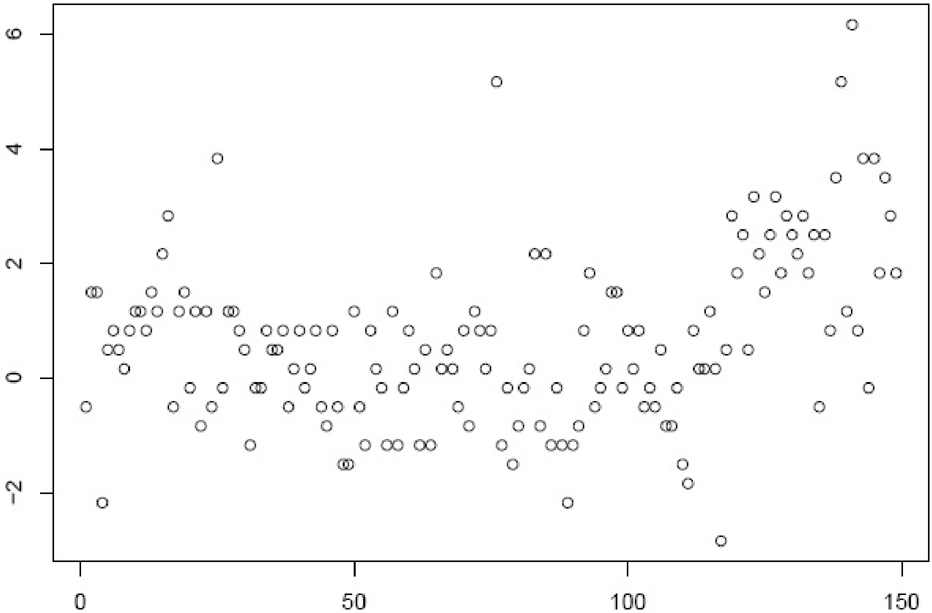


Figure 1: Triglyceride data

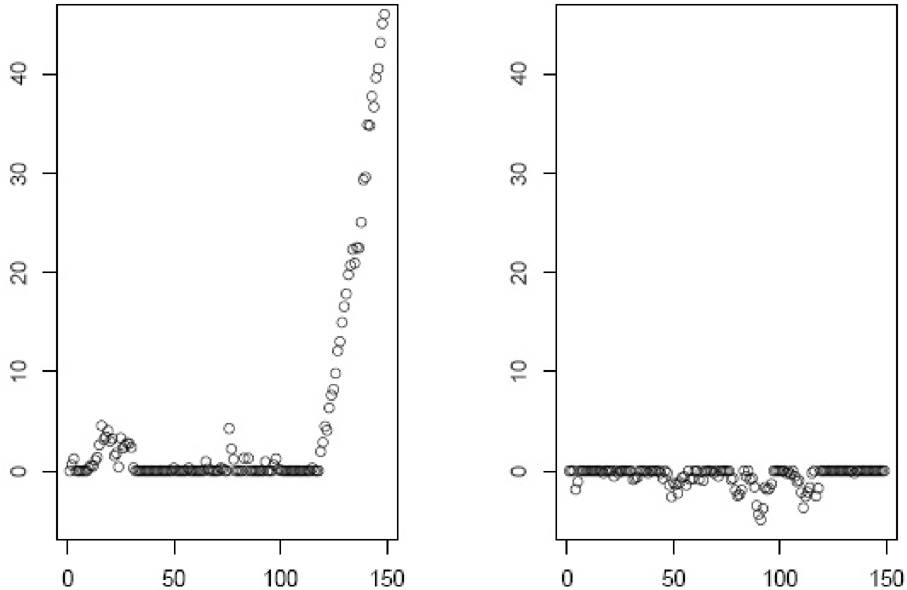


Figure 2: Page's CUSUM statistics

Scaling limit (think CLT)

Assume $a = 0$, $E[Y_k] = \frac{c}{\sqrt{n}}$, and replace K^+ by $\frac{K^+}{\sqrt{n}}$. Define $X_n^H(t) = \frac{1}{\sqrt{n}} S_{[nt]}^H$, so

$$X_n^H(t+n^{-1}) = \max(0, X_n^H(t) + \frac{1}{\sqrt{n}} Y_{[nt]+1} - \frac{K^+}{n})$$

$$\begin{aligned}
A_n f(x) &= n(E[f(0 \vee (x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n} - \frac{K^+}{n}))] - f(x)) \\
&\approx E[(-\sqrt{n}x) \vee (\xi + \frac{c-K^+}{\sqrt{n}}))]\sqrt{n} f'(x) \\
&\quad + \frac{1}{2} E[((-\sqrt{n}x) \vee (\xi + \frac{c-K^+}{\sqrt{n}})))^2] f''(x) \\
&= E[(-\sqrt{n}x - \frac{c-K^+}{\sqrt{n}}) \vee \xi]\sqrt{n} f'(x) + (c-K^+) f'(x) \\
&\quad + \frac{1}{2} E[((-\sqrt{n}x) \vee (\xi + \frac{c-K^+}{\sqrt{n}})))^2] f''(x)
\end{aligned}$$

Estimates

For simplicity, assume $Y_k^n = \xi_k + \frac{c}{\sqrt{n}}$ for iid ξ_k with $E[\xi_k] = 0$ and $Var(\xi_k) < \infty$. More generally, one could assume that $\{|Y_k^n|^2\}$ are uniformly integrable.

Lemma 10.2 *If $E[\xi^2] < \infty$, then*

$$\lim_{c\to\infty} c^2 P\{|\xi| \geq c\} = 0 \qquad \lim_{c\to\infty} cE[|\xi|\mathbf{1}_{\{|\xi|\geq c\}}] = 0.$$

If, in addition, $E[\xi] = 0$, then $\lim_{c\to\infty} cE[(-c)\vee\xi] = 0$.

Proof. Note that

$$E[|\xi|^2\mathbf{1}_{\{|\xi|\geq c\}}] \geq cE[|\xi|\mathbf{1}_{\{|\xi|\geq c\}}] \geq c^2 P\{|\xi| \geq c\}.$$

If $E[\xi] = 0$, then

$$cE[(-c)\vee\xi] = -c^2 P\{\xi \leq -c\} + cE[\xi\mathbf{1}_{\{\xi>-c\}}] = -c^2 P\{\xi \leq -c\} - cE[\xi\mathbf{1}_{\{\xi\leq -c\}}]$$

□

Note that if $E[\xi] = 0$ and $E[\xi^2] < \infty$, then for $z > 0$,

$$\lim_{n\to\infty} \sqrt{n}E[\xi \vee (-\sqrt{n}z)] = \lim_{n\to\infty} (-nzP\{\xi < -\sqrt{n}z\} - \sqrt{n}E[\xi\mathbf{1}_{\{\xi<-\sqrt{n}z\}}] = 0$$

Boundary condition

If $f'(0) = 0$. Then

$$E[(-\sqrt{n}x) \vee (\xi + \frac{c - K^+}{\sqrt{n}}))]\sqrt{n}f'(x) \leq E[(-\sqrt{n}x) \vee (\xi + \frac{c - K^+}{\sqrt{n}}))]\sqrt{n}x\|f''\|,$$

which converges to zero if $\sqrt{n}x \to \infty$. If $\sqrt{n}x \to u$, then

$$E[(-\sqrt{n}x) \vee (\xi + \frac{c - K^+}{\sqrt{n}}))]\sqrt{n}f'(x) \to E[(-u) \vee \xi]uf''(0).$$

Claim: If $f'(0) = 0$, then

$$\sup_x |E[(-\sqrt{n}x) \vee (\xi + \frac{c - K^+}{\sqrt{n}}))]\sqrt{n}f'(x)| < \infty.$$

and for $\sqrt{n}\epsilon_n \to \infty$

$$\sup_{x \geq \epsilon_n} |E[(-\sqrt{n}x) \vee (\xi + \frac{c - K^+}{\sqrt{n}}))]\sqrt{n}f'(x)| \to 0.$$

Need to show

$$E\int_0^t \mathbf{1}_{\{X_n^H(s) \leq \epsilon_n\}} ds] \to 0.$$

Scaling the two-sided procedure

Recall

$$S_{k+1} = (S_k + Y_{k+1} - a) \times 0 \vee (1 - \frac{K}{|S_k + Y_{k+1} - a|})$$

Again, assume $a = 0$, $Y_{k+1} = \xi_{k+1} + \frac{c}{\sqrt{n}}$, and replace K by $\frac{K}{\sqrt{n}}$. Then

$$X_n(t + n^{-1}) = (X_n(t) + \frac{1}{\sqrt{n}} Y_{[nt]+1}) \times 0 \vee \left(1 - \frac{K}{n|X_n(t) + \frac{1}{\sqrt{n}} Y_{[nt]+1}|}\right)$$

and

$$\begin{aligned} A_n f(x) &= n(E[f((x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n})(1 - (n^{-1}K|x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n}|^{-1}) \wedge 1) - f(x)] \\ &\approx cf'(x) - E[\frac{x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n}}{|x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n}|}(K \wedge (n|x + \frac{1}{\sqrt{n}}\xi + \frac{c}{n}|)] f'(x) + \frac{1}{2}E[\xi^2] f''(x) \end{aligned}$$

11. ϕ-irreducibility and Harris recurrence

- Uniqueness of stationary distributions
- Ergodicity under uniqueness assumption
- Example: Spatial birth and death processes
- Standard assumptions for Markov processes (Borel right processes)
- Generator conditions
- ϕ-irreducibility
- Equivalent condition

Renewal conditions

Lemma 11.1 *Suppose $A \subset B(E) \times B(E)$, and X is strong Markov solution of the martingale problem for A. Let $z \in E$, and let $\tau_1 = \inf\{t : X(t) = z\}$ and $\tau_{k+1} = \inf\{t > \tau_k : X(t) = z\}$. Suppose $\tau_1 < \infty$ a.s. and $E[\tau_{k+1} - \tau_k] < \infty$. Then there is a stationary distribution for A. If $\tau_1 < \infty$ a.s. for all initial distributions, then the stationary distribution is unique.*

Proof. We have

$$\lim_{n\to\infty} \frac{1}{\tau_n - \tau_1} \int_{\tau_1}^{\tau_n} f(X(s))ds = \frac{E[\int_{\tau_k}^{\tau_{k+1}} f(X(s))ds]}{E[\tau_{k+1} - \tau_k]} \equiv \int f d\mu$$

If X corresponds to a semigroup $\{T(t)\}$, then

$$\frac{1}{t}\int_{\mathbf{0}}^{t} \int T(s) f d\nu ds = E[\frac{1}{t}\int_0^t f(X(s))ds] \to \int f d\mu$$

□

Birth and death processes: Recurrence

For $\lambda_k > 0$, $k = 0, 1, \ldots$, $\mu_0 = 0$, and $\mu_k > 0$, $k = 1, 2, \ldots$, consider

$$Af(k) = \lambda_k(f(k+1) - f(k)) + \mu_k(f(k-1) - f(k)) = 0, \quad k \geq 1$$

Then

$$f(k+1) - f(k) = \frac{\mu_k}{\lambda_k}(f(k) - f(k-1) = (f(0) - f(1)) \sum_{l=1}^{k} \prod_{i=1}^{l} \frac{\mu_i}{\lambda_i}, \quad k \geq 1,$$

and

$$f(k) = f(1) + (f(1) - f(0)) \sum_{l=1}^{k-1} \prod_{i=1}^{l} \frac{\mu_i}{\lambda_i}$$

If $f(k) \to \infty$, then process hits zero with probability one. If the limit is finite, the process is transient.

Example: Spatial birth and death processes

Let $\nu \in \mathcal{M}_f(S)$

$$Af(\eta) = \int_S \beta(y,\eta)(f(\eta+\delta_y) - f(\eta))\nu(dy) + \int_S \delta(x,\eta)(f(\eta-\delta_x) - f(\eta))\eta(dx)$$

where β and δ are continuous. Suppose

$$\beta(y,\eta) \le \lambda_{|\eta|}, \quad \delta(x,\eta) \ge \mu_{|\eta|}$$

Then for $f(\eta) = f(|\eta|)$ from above

$$E[f(|Z((t+s)\wedge\tau_0)|) - f(|Z(t\wedge\tau_0)|)|\mathcal{F}_t] \le 0$$

Markov processes: Transition functions

E a complete, separable metric space and $E_0 \in \mathcal{B}(E)$. (This assumption is essentially equivalent to the assumption that E_0 is a *Lusin space*.)

Definition 11.2 $P(t, x, \Gamma)$ is a time-homogeneous, Markov *transition function* on E_0, if

a) For each $\Gamma \in \mathcal{B}(E_0)$, $(t, x) \in [0, \infty) \times E_0 \to P(t, x, \Gamma)$ is $\mathcal{B}([0, \infty)) \times \mathcal{B}(E_0)$-measurable.

b) For each $(t, x) \in [0, \infty) \times E_0$, $P(t, x, \cdot) \in \mathcal{P}(E_0)$.

c) (The Chapman-Kolmogorov Equation) For all $t, s \geq 0$, $x \in E_0$, and $\Gamma \in \mathcal{B}(E_0)$,

$$P(t + s, x.\Gamma) = \int_{E_0} P(t, y, \Gamma) P(s, x, dy).$$

Markov processes: The semigroup

Define

$$T(t)f(x) \equiv \int_E f(y)P(t,x,dy), \quad f \in B(E_0), t \geq 0,$$

and note that $\{T(t)\}$ defines a semigroup of operators on $B(E_0)$, that is, $T(t+s)f = T(t)T(s)f$, $s,t \geq 0$. We will refer to $\{T(t)\}$ as a *transition semigroup*.

Definition 11.3 X is a *Markov process* with transition semigroup $\{T(t)\}$ if and only if there exists a filtration $\{\mathcal{F}_t\}$ such that X is adapted to $\{\mathcal{F}_t\}$ and

$$E[f(X(t+s))|\mathcal{F}_t] = T(s)f(X(t)) \ a.s. \quad \forall t, s \geq 0, f \in B(E_0). \tag{11.1}$$

X is *strong Markov* if for each $\{\mathcal{F}_t\}$-stopping time τ,

$$E[f(X(\tau+s))|\mathcal{F}_\tau] = T(s)f(X(\tau)) \ a.s. \quad \forall t, s \geq 0, f \in B(E_0).$$

Markov processes: Basic conditions

For each $\mu \in \mathcal{P}(E_0)$, let X_μ be a Markov process with respect to a filtration $\{\mathcal{F}_t^\mu\}$ with semigroup $\{T(t)\}$ and $P\{X_\mu(0) \in C\} = \mu(C)$, $C \in \mathcal{B}(E_0)$. (μ is the *initial distribution* for X_μ.) If $\mu = \delta_x$, we write X_x.

We assume the following basic conditions on X_μ and $\{\mathcal{F}_t^\mu\}$.

Condition 11.4

a) X_μ *is right continuous.*

b) X_μ *is strong Markov with respect to the filtration* $\{\mathcal{F}_t^\mu\}$.

c) $\{\mathcal{F}_t^\mu\}$ *is complete and right continuous.*

σ-algebra on the right continuous functions

Let $R_{E_0}[0,\infty)$ denote the collection of right-continuous, E_0-valued functions. Let $\mathcal{S}_{E_0}$ be the σ-algebra generated by the evaluation functions $t \rightarrow x(t)$.

- $D_{E_0}[0,\infty) \in \mathcal{S}_{E_0}$
- For $y \in R_{E_0}[0,\infty)$ $\{x : \sup_{0\leq s\leq t} r(x(s),y(s)) \leq \epsilon\} \in \mathcal{S}_{E_0}$
- For each closed $F \subset E_0$, $\{x : x(s) \in F, s \leq t\} \in \mathcal{S}_{E_0}$.

$$P\{X_\mu \in C\} = \int_{E_0} P\{X_x \in C\}\mu(dx), \quad C \in \mathcal{S}_{E_0}$$

$\mathcal{S}_{E_0}$-measurability

Lemma 11.5 *Let*

$$d(x,y) = \int_0^\infty e^{-t} \sup_{s \le t} r(x(s), y(s))dt$$

Then $\mathcal{S}_{E_0}$ *is the* σ*-algebra generated by* $B_\delta(y)$, $y \in R_{E_0}[0,\infty)$.

Proof. Let

$$B_\delta^n(y) = \{x : \int_0^\infty e^{-t} \sup_{s \le t} r(x(\frac{[ns]}{n}), y(\frac{[ns]}{n}))dt \le \delta\}.$$

Then $B_\delta^n(y) \in \sigma(s \to x(s) : s = \frac{k}{n}, k = 0, 1, \ldots) \subset \mathcal{S}_{E_0}$ and $B_\delta(y) = \cap_n B_\delta^n(y)$. □

Markov processes: Sufficient conditions

Lemma 11.6 *Let E_0 be compact, and let $A \subset C(E_0) \times B(E_0)$. Suppose that for each $\mu \in \mathcal{P}(E_{\mathbf{0}})$ there exists a unique solution X_μ of the martingale problems for (A, μ). Then X_μ has a modification satisfying Condition 11.4.*

ϕ-irreducibility

For $B \in \mathcal{B}(E_0)$, let $\tau_B = \inf\{t : X(t) \in B\}$, and let $\phi \in \mathcal{P}(E_0)$. $\{T(t)\}$ is ϕ-irreducible if $\phi(B) > 0$ implies $P_x\{\tau_B < \infty\} > 0$ for all $x \in E_0$.

Lemma 11.7 *Suppose $\{T(t)\}$ is ϕ-irreducible, and define*

$$\psi(B) = E_\phi[\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt] = \int_{E_0} E_x[\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt]\phi(dx). \tag{11.2}$$

If $\psi(B) > 0$, then $P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt > 0\} > 0$ for every $x \in E_0$.

Proof. Let $\Gamma = \{x : P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt > \epsilon\} > \delta\}$. There exist $\epsilon > 0$ and $\delta > 0$ such that $\phi(\Gamma) > 0$. There exists compact $K \subset \Gamma$ such that $\phi(K) > 0$. Therefore, for every $x \in E_0$, $P_x\{\tau_K < \infty\} > 0$.

Note that

$$\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt \geq e^{-\tau_K}\int_0^\infty e^{-t}\mathbf{1}_B(X(\tau_K+t))dt.$$

By the strong Markov property,

$$\begin{aligned}
&P_x\{\tau_K<\infty, \int_0^\infty e^{-t}\mathbf{1}_B(X(\tau_K+t))dt>\epsilon\} \\
&\quad = E_x[\mathbf{1}_{\{\tau_K<\infty\}}P_{X(\tau_K)}\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt>\epsilon\} \\
&\quad \geq \delta P_x\{\tau_K<\infty\}
\end{aligned}$$

and $P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt>\epsilon e^{-\tau_K}\} > \delta P_x\{\tau_K<\infty\}$. □

Equivalent notions of irreducibility

Let $\psi \in \mathcal{P}(E_0)$. Suppose that $\psi(B) > 0$ implies $P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(s))ds > 0\} > 0$ for every x. Then $\{T(t)\}$ is ψ-irreducible.

Lemma 11.8 *If $\{T(t)\}$ is ϕ_1-irreducible and ϕ_2-irreducible and ψ_1 and ψ_2 are defined as in (11.2), then ψ_1 and ψ_2 are equivalent (mutually absolutely continuous) measures.*

Proof. If $\psi_1(B) > 0$, then $\int E_x[\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt]\phi_2(dx) = \psi_2(B) > 0$. □

Ergodicity and ϕ-irreducibility

Lemma 11.9 *Suppose π is the unique stationary distribution for $\{T(t)\}$, and $\{T(t)\}$ is ϕ-irreducible. If $\pi(B) > 0$, then $\phi(\{x : P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt = 0\} = 1\}) = 0$, and hence, $\pi(B) > 0$ implies $\psi(B) > 0$.*

Proof. Suppose not. Let $K \subset \{x : P_x\{\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt = 0\} = 1\}$ be compact. Then

$$\int_0^\infty \mathbf{1}_B(X_\pi(s))ds \leq \tau_K \tag{11.3}$$

But uniqueness of the stationary distribution implies X_π is ergodic and $\pi(B) > 0$ implies the integral on the left of (11.3) is infinite $a.s. - \pi$. Consequently, $P_\pi\{\tau_K < \infty\} = 0$, and hence $\phi(K) = 0$. □

Equivalence of stationary distribution

Theorem 11.10 *Suppose $\{T(t)\}$ is ϕ-irreducible and ψ is defined as above. If π is a stationary distribution for $\{T(t)\}$, then π and ψ are equivalent measures and π is the unique stationary distribution.*

Proof. If $\psi(B) > 0$, then

$$\pi(B) = \int E_x[\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt]\pi(dx) > 0.$$

Consequently, $\psi << \pi$.

If there were more than one stationary distribution, there would be two mutually singular stationary distributions. (See Lemma 3.14.) But if $\pi_1(B) = \pi_2(B^c) = 0$, then $\psi(B) \vee \psi(B^c) > 0$ implies a contradiction.

By Lemma 11.9 $\pi(B) > 0$ implies $\psi(B) > 0$ so $\pi << \psi$. □

Harris recurrence

Definition 11.11 $\{T(t)\}$ is *Harris recurrent*, if there exists $\psi \in \mathcal{P}(E_0)$ such that $\psi(B) > 0$ implies

$$P_x\{\int_0^\infty \mathbf{1}_B(X(t))dt = \infty\} = 1, \quad \forall x \in E_0.$$

Equivalent definition

Theorem 11.12 *$\{T(t)\}$ is Harris recurrent if and only if there exists $\phi \in \mathcal{P}(E_0)$ such that $\phi(B) > 0$ implies $P_x\{\tau_B < \infty\} = 1$ for all $x \in E_0$.*

Proof. If $\{T(t)\}$ is Harris recurrent then $\phi = \psi$ has the desired property. Conversely, if $\phi(B) > 0$ implies $P_x\{\tau_B < \infty\} = 1$ for all $x \in E_0$, then ψ defined by (11.2) satisfies the condition in the definition of Harris recurrence.

In particular, as in the proof of Lemma 11.7, there exist $\epsilon, \delta > 0$, compact $K \subset E_0$ with $\phi(K) > 0$, and $t_0 > 0$ such that

$$P_x\{\int_0^{t_0} \mathbf{1}_B(X(s))ds \geq \epsilon\} \geq \delta, \quad x \in K.$$

For $\mu \in \mathcal{P}(E_0)$, define

$$\tau_1 = \inf\{t > 0 : X_\mu(t) \in K\}, \quad \tau_{n+1} = \inf\{t > \tau_n + t_0 : X_\mu(t) \in K\}.$$

Then $\tau_n < \infty$ a.s., for every n, and by the right continuity of X_μ and the compactness of K, $X_\mu(\tau_n) \in K$ a.s. Consequently, by the strong Markov property, $P\{\int_{\tau_n}^{\tau_n+t_0} \mathbf{1}_B(X_\mu(s))ds \geq \epsilon | \mathcal{F}_{\tau_n}\} \geq \delta$. It follows that $\int_{\tau_n}^{\tau_{n+1}} \mathbf{1}_B(X_\mu(s))ds \geq \epsilon$ infinitely often, so

$$\int_0^\infty \mathbf{1}_B(X_\mu(s))ds = \infty \quad a.s. \tag{11.4}$$

□

Example: Workload process

Poisson arrivals at rate one. Single server, FIFO (first in first out) queue.

$$V(t) = V(0) + \sum_{k=1}^{N(t)} \xi_k - \int_0^t \mathbf{1}_{\{V(s)>0\}} ds$$

$$Af(v) = \lambda \int_0^\infty (f(v+z) - f(z))\mu_\xi(dz) - \mathbf{1}_{(0,\infty)}(v) f'(v)$$

Diffusions

Relationship between discrete and continuous time

Let

$$R_1(x,B)=\int_0^\infty e^{-t}T(t)\mathbf{1}_B(x)dt=E_x[\int_0^\infty e^{-t}\mathbf{1}_B(X(t))dt].$$

Then R_1 is a transition function on E_0. The corresponding discrete-time Markov chain can be obtained by

$$Y_k=X(\sum_{i=1}^k\Delta_i),$$

where $\{\Delta_i\}$ are iid unit exponential random variables, independent of X. Clearly, if π is a stationary distribution for $\{T(t)\}$ it is a stationary distribution for R_1.

Lemma 11.13 *If π is a stationary distribution for R_1, then π is a stationary distribution for $\{T(t)\}$.*

Proof. Let A be the full generator for $\{T(t)\}$. Then $R_1f=(I-A)^{-1}f$. Setting $g=(I-A)^{-1}f$, $(I-A)g=f$. But

$$\int gd\pi=\int R_1fd\pi=\int fd\pi,$$

so $\int Ag d\pi = 0$. Consequently, π is a stationary distribution for $\{T(t)\}$. $\square$

“Petite” sets

With reference to Lemma 2.18.

$C \in \mathcal{B}(E_0)$ is *petite* if there is a probability measure $\nu \in \mathcal{P}(E_0)$ and $\epsilon > 0$ such that

$$R_1(x, \cdot) \geq \epsilon\nu, \quad x \in C.$$

Conventions and caveats

State spaces are always complete, separable metric spaces (sometimes called *Polish spaces*), usually denoted (E, r).

All probability spaces are complete.

All identities involving conditional expectations (or conditional probabilities) only hold almost surely (even when I don't say so).

If the filtration $\{\mathcal{F}_t\}$ involved is obvious, I will say adapted, rather than $\{\mathcal{F}_t\}$-adapted, stopping time, rather than $\{\mathcal{F}_t\}$-stopping time, etc.

All processes are *cadlag* (right continuous with left limits at each $t > 0$), unless otherwise noted.

A process is real-valued if that is the only way the formula makes sense.

Assignments

1. Durrett Problems 5.1.6 and 5.1.8 due 1/26/06
2. Durrett Problems 5.3.4 and 5.4.1 due 2/07/06
3. Durrett Problem 6.3.4
4. Exercises 1 and 2 due 2/21/06
5. Exercises 3 and 4 due 2/28/06
6. Exercises 5 through 8
7. Exercise 9 and Durrett Problems 7.2.2 and 7.51

Exercises

1. Let E be the space of permutations of the positive integers. Let $p_k > 0$, $k = 1, 2, \ldots$ and $\sum p_k = 1$. Let $\{\xi_n\}$ be iid with $P\{\xi_n = k\} = p_k$ and let X_0 be an E-valued random variable independent of $\{\xi_n\}$. Let $\{X_n\}$ be the Markov chain in which if $\xi_{n+1} = k$, X_{n+1} is obtained from X_n by moving k to the beginning of the permutation and leaving the order of the other elements unchanged. Write

$$X_n = (X_n^1, X_n^2, X_n^3, \ldots)$$

   (a) For $n > 0$, what is $P\{X_n^1 = k | X_0\}$, $P\{X_n^1 = k, X_n^2 = l | X_0\}$?

   (b) Find a stationary distribution for this Markov chain and show that it is unique.

2. Let $\{\xi_n\}$ be iid with $P\{\xi_n = \frac{1}{2^k}\} = p_k > 0$, for $k = 1, 2, \ldots$. Let $E = [0, 1)$ and $X_{n+1} = X_n + \xi_{n+1} \bmod 1$. Show that this Markov chain has a unique stationary distribution.

3. Let X be $\{\mathcal{F}_t\}$-progressive. Suppose that $E[X(\tau)] = E[X(0)]$ for every $\{\mathcal{F}_t\}$-stopping time τ. Show that X is a $\{\mathcal{F}_t\}$-martingale.

4. Let $0 = \tau_0 < \tau_1 < \cdots$ be stopping times satisfying $\lim_{k\to\infty} \tau_k = \infty$, and for $k = 0, 1, 2, \ldots$, let $\xi_k \in \mathcal{F}_{\tau_k}$. Define

$$X(t) = \sum_{k=0}^{\infty} \xi_k \mathbf{1}_{[\tau_k, \tau_{k+1})}(t).$$

   Show that X is adapted.

   Example: Let X be a cadlag adapted process and let $\epsilon > 0$. Define $\tau_0^\epsilon = 0$ and for $k = 0, 1, 2, \ldots$,

$$\tau_{k+1}^\epsilon = \inf\{t > \tau_k^\epsilon : |X(t) - X(\tau_k^\epsilon)| \vee |X(t-) - X(\tau_k^\epsilon)| \geq \epsilon\}.$$

Note that the τ_K^ϵ are stopping times, by Problem 1. Define

$$X^\epsilon(t) = \sum_{k=0}^{\infty} X(\tau_k^\epsilon)\mathbf{1}_{[\tau_k^\epsilon,\tau_{k+1}^\epsilon)}(t).$$

Then X^ϵ is a piecewise constant, adapted process satisfying

$$\sup_t |X(t) - X^\epsilon(t)| \leq \epsilon.$$

5. Show that $E[f(X)|\mathcal{D}] = E[f(X)]$ for all bounded continuous fucntions (all bounded measurable functions) if and only if X is independent of $\mathcal{D}$.
6. Let N be a Poisson process with parameter λ, and let $X_1, X_2, \ldots$ be a sequence of Bernoulli trials with parameter p. Assume that the X_k are independent of N, and define

$$M(t) = \sum_{k=1}^{N(t)} X_k.$$

   (a) What is $P\{M(t) = k|N(t) = n\}$?
   (b) What is the distribution of $M(t)$?
   (c) For $t < s$, calculate the probability that $P\{N(t) = 1, N(s) = 1\}$.
   (d) Give an event in terms of S_1 and S_2 that is equivalent to the event $\{N(t) = 1, N(s) = 1\}$, and use the calculation in the previous part to calculate the joint density function for S_1 and S_2.
   (e) For $k \geq 1$, find the conditional density of S_1 given that $N(t) = k$. (Hint: First calculate $P\{S_1 \leq s, N(t) = k\}$ for $s \leq t$.)

7. Verify tightness (relative compactness) for the renormalized empirical distribuiton B_n.
8. Functional convergence and the continuous mapping theorem enable one to obtain convergence for many interesting quantities; however, the continuity properties of the quantities of interest need to be checked.
   (a) Show that $F : x \in C[0,1] \to \sup_{0\leq t\leq 1} x(s) \in \mathbb{R}$ is continuous.
   (b) Let $\tau_c : x \in C[0,\infty) \to \inf\{t : x(t) \geq c\} \in [0,\infty]$ and Let $\tau_c^0 : x \in C[0,\infty) \to \inf\{t : x(t) > c\} \in [0,\infty]$. Describe the points of continuity for τ_c and τ_c^0.
9. Consider the Markov chain with transition matrix

$$\begin{bmatrix} 1-\alpha & \alpha & 0 \\ \gamma & 1-2\gamma & \gamma \\ 0 & \alpha & 1-\alpha \end{bmatrix}$$

   Derive the maximum likelihood estimators for α and γ and apply the martingale central limit theorem to show asymptotic normality.
10. Let X and Y S-valued random variables defined on $(\Omega,\mathcal{F},P)$, and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra. Suppose that $M \subset \bar{C}(S)$ is separating and

$$E[f(X)|\mathcal{G}] = f(Y) \quad a.s.$$

   for every $f \in M$. Show that $X = Y$ a.s.

Glossary

Complete. We say that a metric space (E, r) is *complete* if every Cauchy sequence in it converges.

Conditional expectation. Let $\mathcal{D} \subset \mathcal{F}$ and $E[|X|] < \infty$. Then $E[X|\mathcal{D}]$ is the, essentially unique, $\mathcal{D}$-measurable random variable satisfying

$$\int_D X dP = \int_D E[X|\mathcal{D}] dP, \quad \forall D \in \mathcal{D}.$$

Consistent. Assume we have an arbitrary state space $(E, \mathcal{B})$ and an index set I. For each nonempty subset $J \subset I$ we denote by E^J the product set $\prod_{t \in J} E$, and we define $\mathcal{B}^J$ to be the product-σ-algebra $\otimes_{t \in J} \mathcal{B}$. Obviously, if $J \subset H \subset I$ then there is a projection map

$$p_J^H : E^H \to E^J.$$

If for every two such subsets J and H we have

$$P_J = p_J^H (P_H)$$

then the family $(P_J)_{\emptyset \neq J \subset H}$ is called *consistent*.

Closure of an operator. Let L be a Banach space and $A \subset L \times L$. The closure $\bar{A}$

of A is the collection of $(f,g) \in L \times L$ such that there exist $(f_n, g_n) \in A$ satisfying $\lim_{n\to\infty} f_n = f$ and $\lim_{n\to\infty} g_n = g$. If $\bar{A} = A$, then A is *closed*.

Separable. A metric space (E, r) is called *separable* if it contains a countable dense subset; that is, a set with a countable number of elements whose closure is the entire space. Standard example: $\mathbf{R}$, whose countable dense subset is $\mathbf{Q}$.

Separating set A collection of function $M \subset \bar{C}(S)$ is *separating* is $\mu, \nu \in \mathcal{M}_f(S)$ and $\int g d\nu = \int g d\mu$, $g \in M$, implies that $\mu = \nu$.

12. Technical lemmas

- Caratheodary extension theorem
- Dynkin class theorem
- Essential supremum
- Martingale convergence theorem
- Kronecker's lemma
- Law of large numbers for martingales
- Geometric rates
- Uniform integrability
- Dominated convergence theorem
- Metric spaces

Caratheodary extension theorem

Theorem 12.1 *Let M be a set, and let $\mathcal{A}$ be an algebra of subsets of M. If μ is a σ-finite measure on $\mathcal{A}$, then there exists a unique extension of μ to a measure on $\sigma(\mathcal{A})$.*

Dynkin class theorem

A collection $\mathcal{D}$ of subsets of Ω is a *Dynkin class* if $\Omega \in \mathcal{D}$, $A, B \in \mathcal{D}$ and $A \subset B$ imply $B - A \in \mathcal{D}$, and $\{A_n\} \subset \mathcal{D}$ with $A_1 \subset A_2 \subset \cdots$ imply $\cup_n A_n \in \mathcal{D}$.

Theorem 12.2 *Let $\mathcal{S}$ be a collection of subsets of Ω such that $A, B \in \mathcal{S}$ implies $A \cap B \in \mathcal{S}$. If $\mathcal{D}$ is a Dynkin class with $\mathcal{S} \subset \mathcal{D}$, then $\sigma(\mathcal{S}) \subset \mathcal{D}$.*

Essential Supremum

Let $\{Z_\alpha, \alpha \in \mathcal{I}\}$ be a collection of random variables. Note that if $\mathcal{I}$ is uncountable, $\sup_{\alpha\in\mathcal{I}} Z_\alpha$ may not be a random variable; however, we have the following:

Lemma 12.3 *There exists a random variable $\bar{Z}$ such that $P\{Z_\alpha \le \bar{Z}\} = 1$ for each $\alpha \in \mathcal{I}$ and there exist $\alpha_k, k = 1, 2, \ldots$ such that $\bar{Z} = \sup_k Z_{\alpha_k}$.*

Proof. Without loss of generality, we can assume $0 < Z_\alpha < 1$. (Otherwise, replace Z_α by $\frac{1}{1+e^{-Z_\alpha}}$.) Let $C = \sup\{E[Z_{\alpha_1} \vee \cdots \vee Z_{\alpha_m}], \alpha_1, \ldots, \alpha_m \in \mathcal{I}, m = 1, 2, \ldots\}$. Then there exist $(\alpha_1^n, \ldots, \alpha_{m_n}^n)$ such that

$$C = \lim_{n\to\infty} E[Z_{\alpha_1^n} \vee \cdots \vee Z_{\alpha_{m_n}^n}].$$

Define $\bar{Z} = \sup\{Z_{\alpha_i^n}, 1 \le i \le m_n, n = 1, 2, \ldots\}$, and note that $C = E[\bar{Z}]$ and $C = E[\bar{Z} \vee Z_\alpha]$ for each $\alpha \in \mathcal{I}$. Consequently, $P\{Z_\alpha \le \bar{Z}\} = 1$. □

Martingale convergence theorem

Theorem 12.4 *Suppose* $\{X_n\}$ *is a submartingale and* $\sup_n E[|X_n|] < \infty$. *Then* $\lim_{n\to\infty} X_n$ *exists a.s.*

Kronecker's lemma

Lemma 12.5 *Let* $\{A_n\}$ *and* $\{Y_n\}$ *be sequences of random variables where* $A_0 > 0$ *and* $A_{n+1} \geq A_n$, $n = 0, 1, 2, \ldots$. *Define* $R_n = \sum_{k=1}^{n} \frac{1}{A_{k-1}}(Y_k - Y_{k-1})$. *and suppose that* $\lim_{n\to\infty} A_n = \infty$ *and that* $\lim_{n\to\infty} R_n$ *exists a.s. Then,* $\lim_{n\to\infty} \frac{Y_n}{A_n} = 0 \quad a.s.$

Proof.

$$\begin{aligned} A_n R_n &= \sum_{k=1}^{n}(A_k R_k - A_{k-1}R_{k-1}) = \sum_{k=1}^{n} R_{k-1}(A_k - A_{k-1}) + \sum_{k=1}^{n} A_k(R_k - R_{k-1}) \\ &= Y_n - Y_0 + \sum_{k=1}^{n} R_{k-1}(A_k - A_{k-1}) + \sum_{k=1}^{n} \frac{1}{A_{k-1}}(Y_k - Y_{k-1})(A_k - A_{k-1}) \end{aligned}$$

and

$$\frac{Y_n}{A_n} = \frac{Y_0}{A_n} + R_n - \frac{1}{A_n}\sum_{k=1}^{n} R_{k-1}(A_k - A_{k-1}) - \frac{1}{A_n}\sum_{k=1}^{n}\frac{1}{A_{k-1}}(Y_k - Y_{k-1})(A_k - A_{k-1})$$

□

Law of large numbers for martingales

Lemma 12.6 *Suppose $\{A_n\}$ is as in Lemma 12.5 and is adapted to $\{\mathcal{F}_n\}$, and suppose $\{M_n\}$ is a $\{\mathcal{F}_n\}$-martingale such that for each $\{\mathcal{F}_n\}$-stopping time τ, $E[(M_\tau - M_{\tau-1})^2 \mathbf{1}_{\{\tau<\infty\}}] < \infty$. If*

$$\sum_{k=1}^{\infty} \frac{1}{A_{k-1}^2}(M_k - M_{k-1})^2 < \infty \quad a.s.,$$

then $\lim_{n\to\infty} \frac{M_n}{A_n} = 0$ *a.s.*

Proof. Without loss of generality, we can assume that $A_n \geq 1$. Let

$$\tau_c = \min\{n : \sum_{k=1}^{n} \frac{1}{A_{k-1}^2}(M_k - M_{k-1})^2 \geq c\}.$$

Then

$$\sum_{k=1}^{\infty} \frac{1}{A_{k-1}^2}(M_{k\wedge\tau_c} - M_{(k-1)\wedge\tau_c})^2 \leq c + (M_{\tau_c} - M_{\tau_c-1})^2 \mathbf{1}_{\{\tau_c<\infty\}}.$$

It follows that $R_n^c = \sum_{k=1}^{n} \frac{1}{A_{k-1}}(M_{k\wedge\tau_c} - M_{(k-1)\wedge\tau_c})$ converges a.s. and hence, by Lemma 12.5, that $\lim_{n\to\infty} \frac{M_{n\wedge\tau_c}}{A_n} = 0$. □

Geometric convergence

Lemma 12.7 *Let $\{M_n\}$ be a martingale with $|M_{n+1} - M_n| \leq c$ a.s. for each n and $M_0 = 0$. Then for each $\epsilon > 0$, there exist C and η such that*

$$P\{\frac{1}{n}|M_n| \geq \epsilon\} \leq Ce^{-n\eta}.$$

Proof. Let $\hat{\varphi}(x) = e^{-x} + e^{x}$ and $\varphi(x) = e^{x} - 1 - x$. Then, setting $X_k = M_k - M_{k-1}$

$$\begin{aligned} E[\hat{\varphi}(aM_n)] &= 2 + \sum_{k=1}^{n} E[\hat{\varphi}(aM_k) - \hat{\varphi}(aM_{k-1})] \\ &= 2 + \sum_{k=1}^{n} E[\exp\{aM_{k-1}\}\varphi(aX_k) + \exp\{-aM_{k-1}\}\varphi(-aX_k)] \\ &\leq 2 + \sum_{k=1}^{n} \varphi(ac) E[\hat{\varphi}(aM_{k-1})], \end{aligned}$$

and hence

$$E[\hat{\varphi}(aM_n)] \leq 2e^{n\varphi(ac)}.$$

Consequently,

$$P\{\sup_{k\leq n}\frac{1}{n}|M_k|\geq\epsilon\}\leq\frac{E[\hat{\varphi}(aM_n)]}{\hat{\varphi}(an\epsilon)}\leq 2e^{n(\varphi(ac)-a\epsilon)}.$$

Then $\eta=\sup_a(a\epsilon-\varphi(ac))>0$, and the lemma follows. □

Uniform integrability

Lemma 12.8 *If X is integrable, then for $\epsilon > 0$ there exists a $K > 0$ such that*

$$\int_{\{|X|>K\}} |X|dP < \epsilon.$$

Proof. $\lim_{K\to\infty} |X|\mathbf{1}_{\{|X|>K\}} = 0$ a.s. □

Lemma 12.9 *If X is integrable, then for $\epsilon > 0$ there exists a $\delta > 0$ such that $P(F) < \delta$ implies $\int_F |X|dP < \epsilon$.*

Proof. Let $F_n = \{|X| \geq n\}$. Then $nP(F_n) \leq E[|X|\mathbf{1}_{F_n}] \to 0$. Select n so that $E[|X|\mathbf{1}_{F_n}] \leq \epsilon/2$, and let $\delta = \frac{\epsilon}{2n}$. Then $P(F) < \delta$ implies

$$\int_F |X|dP \leq \int_{F_n} |X|dP + \int_{F_n^c \cap F} |X|dP < \frac{\epsilon}{2} + n\delta = \epsilon$$

□

Theorem 12.10 *Let* $\{X_\alpha\}$ *be a collection of integrable random variables. The following are equivalent:*

a) $\sup E[|X_\alpha|] < \infty$ *and for* $\epsilon > 0$ *there exists* $\delta > 0$ *such that* $P(F) < \delta$ *implies* $\sup_\alpha \int_F |X_\alpha| dP < \epsilon$*.*

b) $\lim_{K\to\infty} \sup_\alpha E[|X_\alpha| \mathbf{1}_{\{|X_\alpha|>K\}}] = 0$*.*

c) $\lim_{K\to\infty} \sup_\alpha E[|X_\alpha| - |X_\alpha| \wedge K] = 0$

d) There exists a convex function φ *with* $\lim_{|x|\to\infty} \frac{\varphi(x)}{|x|} = \infty$ *such that* $\sup_\alpha E[\varphi(|X_\alpha|)] < \infty$*.*

Proof. a) implies b) follows by

$$P\{|X_\alpha| > K\} \leq \frac{E[|X_\alpha|]}{K}$$

b) implies d): Select N_k such that

$$\sum_{k=1}^{\infty} k \sup_\alpha E[\mathbf{1}_{\{|X_\alpha| > N_k\}} |X_\alpha|] < \infty$$

Define $\varphi(0) = 0$ and

$$\varphi'(x) = k, \qquad N_k \leq x < N_{k+1}.$$

Recall that $E[\varphi(|X|)] = \int_0^\infty \varphi'(x) P\{|X| > x\} dx$, so

$$E[\varphi(|X_\alpha|)] = \sum_{k=1}^{\infty} k \int_{N_k}^{N_{k+1}} P\{|X_\alpha| > x\} dx \leq \sum_{k=1}^{\infty} k \sup_\alpha E[\mathbf{1}_{\{|X_\alpha| > N_k\}} |X_\alpha|].$$

d) implies b): $E[\mathbf{1}_{\{|X_\alpha| > K\}} |X_\alpha|] < \frac{E[\varphi(|X_\alpha|)]}{\varphi(K)/K}$

b) implies a): $\int_F |X_\alpha| dP \leq P(F) K + E[\mathbf{1}_{\{|X_\alpha| > K\}} |X_\alpha|]$.

To see that (b) is equivalent to (c), observe that

$$E[|X_\alpha| - |X_\alpha| \wedge K] \le E[|X_\alpha|\mathbf{1}_{\{|X_\alpha|>K\}}] \le 2E[|X_\alpha| - |X_\alpha| \wedge \frac{K}{2}]$$

□

Uniformly integrable families

- For X integrable, $\Gamma = \{E[X|\mathcal{D}] : \mathcal{D} \subset \mathcal{F}\}$
- For $X_1, X_2, \ldots$ integrable and identically distributed

$$\Gamma = \{\frac{X_1 + \cdots + X_n}{n} : n = 1, 2, \ldots\}$$

- For $Y \geq 0$ integrable, $\Gamma = \{X : |X| \leq Y\}$.

Uniform integrability and L^1 convergence

Theorem 12.11 *$X_n \to X$ in L^1 iff $X_n \to X$ in probability and $\{X_n\}$ is uniformly integrable.*

Proof. If $X_n \to X$ in L^1, then

$$\lim_{n\to\infty} E[|X_n| - |X_n| \wedge K] = E[|X| - |X| \wedge K]$$

and Part (c) of Theorem 12.10 follows from the fact that

$$\lim_{K\to\infty} E[|X| - |X| \wedge K] = \lim_{K\to\infty} E[|X_n| - |X_n| \wedge K] = 0.$$

□

Measurable functions

Let $(M_i, \mathcal{M}_i)$ be measurable spaces.

$f : M_1 \to M_2$ is *measurable* if $f^{-1}(A) = \{x \in M_1 : f(x) \in A\} \in \mathcal{M}_1$ for each $A \in \mathcal{M}_2$.

Lemma 12.12 *If $f : M_1 \to M_2$ and $g : M_2 \to M_3$ are measurable, then $g \circ f : M_1 \to M_3$ is measurable.*

Dominated convergence theorem

Theorem 12.13 *Let $X_n \to X$ and $Y_n \to Y$ in probability. Suppose that $|X_n| \leq Y_n$ a.s. and $E[Y_n|\mathcal{D}] \to E[Y|\mathcal{D}]$ in probability. Then*

$$E[X_n|\mathcal{D}] \to E[X|\mathcal{D}] \qquad \text{in probability}$$

Proof. A sequence converges in probability iff every subsequence has a further subsequence that converges a.s., so we may as well assume almost sure convergence. Let $D_{m,c} = \{\sup_{n \geq m} E[Y_n|\mathcal{D}] \leq c\}$. Then

$$E[Y_n \mathbf{1}_{D_{m,c}}|\mathcal{D}] = E[Y_n|\mathcal{D}]\mathbf{1}_{D_{m,c}} \stackrel{L_1}{\to} E[Y|\mathcal{D}]\mathbf{1}_{D_{m,c}} = E[Y\mathbf{1}_{D_{m,c}}|\mathcal{D}].$$

Consequently, $E[Y_n\mathbf{1}_{D_{m,c}}] \to E[Y\mathbf{1}_{D_{m,c}}]$, so $Y_n\mathbf{1}_{D_{m,c}} \to Y\mathbf{1}_{D_{m,c}}$ in L_1 by the ordinary dominated convergence theorem. It follows that $X_n\mathbf{1}_{D_{m,c}} \to X\mathbf{1}_{D_{m,c}}$ in L_1 and hence

$$E[X_n|\mathcal{D}]\mathbf{1}_{D_{m,c}} = E[X_n\mathbf{1}_{D_{m,c}}|\mathcal{D}] \stackrel{L_1}{\to} E[X\mathbf{1}_{D_{m,c}}|\mathcal{D}] = E[X|\mathcal{D}]\mathbf{1}_{D_{m,c}}.$$

Since m and c are arbitrary, the lemma follows. □

Metric spaces

$d : S \times S \to [0,\infty)$ is a *metric* on S if and only if $d(x,y) = d(y,x)$, $d(x,y) = 0$ if and only if $x = y$, and $d(x,y) \leq d(x,z) + d(z,y)$.

If d is a metric then $d \wedge 1$ is a metric.

Examples

- $\mathbb{R}^m$ $\quad d(x,y) = |x - y|$
- $C[0,1]$ $\quad d(x,y) = \sup_{0 \leq t \leq 1} |x(t) - y(t)|$
- $C[0,\infty)$ $\quad d(x,y) = \int_0^\infty e^{-t} \sup_{s \leq t} 1 \wedge |x(s) - y(s)|\, dt$

Sequential compactness

$K \subset S$ is *sequentially compact* if every sequence $\{x_n\} \subset K$ has a convergent subsequence with limit in K.

Lemma 12.14 *If (S, d) is a metric space, then $K \subset S$ is compact if and only if K is sequentially compact.*

Proof. Suppose K is compact. Let $\{x_n\} \subset K$. If x is not a limit point of $\{x_n\}$, then there exists $\epsilon_x > 0$ such $\max\{n : x_n \in B_{\epsilon_x}(x)\} < \infty$. If $\{x_n\}$ has no limit points, then $\{B_{\epsilon_x}(x), x \in K\}$ is an open cover of K. The existence of a finite subcover contradicts the definition of ϵ_x.

If K is sequentially compact, and $\{U_\alpha\}$ is an open cover of K. Let $x_1 \in K$ and $\epsilon_1 > \frac{1}{2}\sup_\alpha \sup\{r : B_r(x_1) \subset U_\alpha\}$ and define recursively, $x_{k+1} \in K \cap (\cup_{l=1}^k B_{\epsilon_l}(x_l))$ and $\epsilon_{k+1} > \frac{1}{2}\sup_\alpha \sup\{r : B_r(x_{k+1}) \subset U_\alpha\}$. (If x_{k+1} does not exist, then there is a finite subcover in $\{U_\alpha\}$.) By sequential compactness, $\{x_k\}$ has a limit point x and $x \notin B_{\epsilon_k}(x_k)$ for any k. But setting $\epsilon = \frac{1}{2}\sup_\alpha \sup\{r : B_r(x) \subset U_\alpha\}$, $\epsilon_k > \epsilon - d(x, x_k)$, so if $d(x, x_k) < \epsilon/2$, $x \in B_{\epsilon_k}(x_k)$. □

Completeness

A metric space (S, d) is complete if and only if every Cauchy sequence has a limit.

Completeness depends on the metric, not the topology: For example

$$r(x, y) = |\frac{x}{1 + |x|} - \frac{y}{1 + |y|}|$$

is a metric giving the usual topology on the real line, but $\mathbb{R}$ is not complete under this metric.

References

Fergusan, Thomas S. *Optimal Stopping and Applications.* Electronic text.

http://www.math.ucla.edu/~tom/Stopping/Contents.html

Kelly, Frank. *Reversibility and Stochastic Networks.* Wiley, Chichester, 1979, reprinted 1987, 1994.

http://www.statslab.cam.ac.uk/~frank/rsn.html

Meyn, Sean and Tweedie, Richard. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

http://decision.csl.uiuc.edu/~meyn/pages/book.html

Roberts, Gareth O. and Rosenthal, Jeffrey S. General state space Markov chains and MCMC algorithms. *Probab. Surv.* 1 (2004), 20–71 (electronic).

http://www.i-journals.org/ps/viewarticle.php?id=15&layout=abstract

Dellacherie, Claude. *Capacités et processus stochastiques.* Springer-Verlag, 1972.

Dynkin, E. B. *Markov Processes, I,II.* Springer-Verlag, 1965.

Blackwell, David; Dubins, Lester E. An extension of Skorohod's almost sure representation theorem. *Proc. Amer. Math. Soc.* 89 (1983), no. 4, 691–692.

Ripley, B. D. Modelling spatial patterns. With discussion. *J. Roy. Statist. Soc. Ser. B* 39 (1977), no. 2, 172–212.