1 Analysis

1.1 Hilbert Spaces

Theorem 1.1 (Hilbert Projection). H is Hilbert space & $C \subseteq H$ closed+convex $\implies \forall x \in H\ \exists$ unique $y \in C$ s.t. $\|x-y\| = \inf_{z\in C} \|x-z\|$ (i.e., y closest to x). C is subspace $\implies \langle y-x, z\rangle = 0\ \forall z \in C$.
Theorem 1.2 (Orthogonality Principle). H is Hilbert space & $U \subseteq H$ closed subspace $\implies \forall x \in H$, $\exists$ unique $u \in U$ closest to x. $u \in U$ closest to x iff $\langle u-x, z\rangle = 0\ \forall z \in U$.

1.2 Harmonic Analysis

Definition 1.3 (DTFT). $x \in \ell^2(\mathbb{Z}) \implies \widehat{x}(\omega) = \sum_{n\in\mathbb{Z}} x(n)\mathrm{e}^{-\mathrm{i}\omega n},\ \omega \in [-\pi, \pi)$.
Theorem 1.4 (Parseval Identity). $x, y \in \ell^1(\mathbb{Z}) \implies \langle x, y\rangle_{\ell^2(\mathbb{Z})} = \langle \widehat{x}, \widehat{y}\rangle_{L^2(\mathbb{S}^1)}$.
Definition 1.5 (IDTFT). $\widehat{x} \in L^2(\mathbb{S}^1) \implies x(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \widehat{x}(\omega)\mathrm{e}^{\mathrm{i}\omega n}\,\mathrm{d}\omega,\ n \in \mathbb{Z}$
Theorem 1.6 (Convolution Theorem). $x, y \in \ell^2(\mathbb{Z})$ & $\widehat{y} \in L^\infty(\mathbb{S}^1) \implies z = x * y \in \ell^2(\mathbb{Z})$ & $\widehat{z}(\omega) = \widehat{x}(\omega)\widehat{y}(\omega)\ \forall \omega \in \mathbb{S}^1$.

2 Probability Theory

2.1 Basics

Proposition 2.1. $\mathcal{B}_\mathbb{R}$ is generated by the class of half-open intervals.
Theorem 2.2. $\mathbf{P}$ is PM $\implies$ (i) $A \subseteq B \implies \mathbf{P}(A) \le \mathbf{P}(B)$; (ii) $A = \bigcup_n A_n \implies \mathbf{P}(A) \le \sum_n \mathbf{P}(A_n)$ (union bound); (iii) (A_n) ascending & $A = \bigcup_n A_n \implies \mathbf{P}(A_n) \uparrow \mathbf{P}(A)$; (iv) (A_n) descending & $A = \bigcap_n A_n \implies \mathbf{P}(A_n) \downarrow \mathbf{P}(A)$.
Theorem 2.3 (Borel-Cantelli). $\sum_n \mathbf{P}(A_n) < \infty \implies \mathbf{P}(A_n \text{ i.o.}) = 0$.
Theorem 2.4 (Borel-Cantelli Converse). (A_n) independent & $\sum_n \mathbf{P}(A_n) = \infty \implies \mathbf{P}(A_n \text{ i.o.}) = 1$.
Theorem 2.5 (Carathéodory Extension). If $\mathcal{G} \subseteq 2^\Omega$ has (i) $\emptyset, \Omega \in \mathcal{G}$, (ii) $A, B \in \mathcal{G} \implies A \cap B \in \mathcal{G}$, (iii) $A, B \in \mathcal{G} \implies \exists C_1, \ldots, C_n \in \mathcal{G}\colon A \setminus B = \bigcup_{i=1}^n C_i$ and $\exists p\colon \mathcal{G} \to [0,1]$ s.t. (i) $p(A) \ge 0\ \forall A \in \mathcal{G}$, (ii) $p(\Omega) = 1$, (iii) (A_n) disjoint and $B = \bigcup_n A_n \implies p(B) = \sum_n p(A_n)$, then $\exists$ unique PM on $\sigma(\mathcal{G})$ agreeing with p.
Theorem 2.6. $F\colon \mathbb{R} \to [0,1]$ is dist. func. of RV iff F non-decreasing and right-continuous. RV is real-valued iff $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$.
Theorem 2.7 (Fatou). (X_n) non-negative & $X = \liminf_{n\to\infty} X_n \implies \liminf_{n\to\infty} \mathbf{E}(X_n) \ge \mathbf{E}(X)$.
Theorem 2.8 (MCT). (X_n) non-negative monotone & $X = \lim_{n\to\infty} X_n \implies \lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.
Theorem 2.9. $X, Y \in L^1$ or $X, Y \ge 0 \implies$ (i) $X = Y \implies \mathbf{E}(X) = \mathbf{E}(Y)$; (ii) $X \le Y \implies \mathbf{E}(X) \le \mathbf{E}(Y)$; (iii) $a, b \in \mathbb{R} \implies \mathbf{E}(aX + bY) = a\,\mathbf{E}(X) + b\,\mathbf{E}(Y)$; (iv) $|\mathbf{E}(XY)|^2 \le \mathbf{E}\left(|X|^2\right)\mathbf{E}\left(|Y|^2\right)$ (Cauchy-Schwarz); (v) $\phi\colon \mathbb{R} \to \mathbb{R}$ convex & $\phi(X) \in L^1 \implies \phi(\mathbf{E}(X)) \le \mathbf{E}(\phi(X))$ (Jensen).
Theorem 2.10 (DCT). (X_n) RVs, $Y \in L^1$ & $|X_n| \le Y\ \forall n$ & $X = \lim_{n\to\infty} X_n \implies \lim_{n\to\infty} \mathbf{E}(X_n) = \mathbf{E}(X)$.
Theorem 2.11 (Bayes Rule). (A_n) partition Ω & $\mathbf{P}(A_n) > 0$ & $\mathbf{P}(B) > 0 \implies \mathbf{P}(A_n \mid B) = \frac{\mathbf{P}(A_n)\,\mathbf{P}(B \mid A_n)}{\sum_i \mathbf{P}(A_i)\,\mathbf{P}(B \mid A_i)}$.
Notation 2.12. $\mathcal{B}(Y) = \mathrm{Cl}_{L^2}(L^\infty \text{ func. of } Y)$; $\mathcal{L}(Y) = \mathrm{Cl}_{L^2}(\text{affine func. of } Y)$.
Theorem 2.13. $X \in L^1$, Y is RV $\implies \mathbf{E}(X \mid Y) = \mathrm{proj}_{L^2}(X; \mathcal{B}(Y))$.
Theorem 2.14 (Radon-Nikodym). $\mu \ll \lambda$ (i.e., $\lambda = 0 \implies \mu = 0$) $\implies \exists \frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \ge 0$.
Theorem 2.15 (Iterated Expectation). $X \in L^1$, Y, Z RVs $\implies \mathbf{E}(\mathbf{E}(X \mid Y, Z) \mid Y) = \mathbf{E}(X \mid Y)$.
Proposition 2.16. CE, e.g., operator $\mathbf{E}(\cdot \mid Z)$ obeys properties of $\mathbf{E}(\cdot)$.

2.2 Limit Theorems

Theorem 2.17 (WLLN). $(X_n) \in L^1$ i.i.d. $\implies \lim^p_{n\to\infty} \frac{S_n}{n} = \mathbf{E}(X_1)$.
Theorem 2.18 (SLLN). $(X_n) \in L^1$ i.i.d. $\implies \lim_{n\to\infty} \frac{S_n}{n} = \mathbf{E}(X_1)$.
Theorem 2.19 (CLT). $(X_n) \in L^2$ i.i.d. $\implies \lim^d_{n\to\infty} \frac{S_n - n\,\mathbf{E}(X_1)}{\sqrt{n\,\mathrm{Var}(X_1)}} = \mathcal{N}(0,1)$.
Theorem 2.20. $\lim_{n\to\infty} X_n = X \implies \lim^p_{n\to\infty} X_n = X \implies \lim^d_{n\to\infty} X_n = X$.
Theorem 2.21 (Skorohod). (F_n), F dist. func., $\lim_{n\to\infty} F_n(x) = F(x)\ \forall x \in C_F \implies \exists (X_n), X\colon X_n \sim F_n,\ X \sim F,\ \lim_{n\to\infty} X_n = X$.
Theorem 2.22 (Helly). (X_n) RVs $\implies \exists (X_{n_k}),\ X\colon \lim^d_{k\to\infty} X_{n_k} = X$.
Theorem 2.23. $\lim^d_{n\to\infty} X_n = X \iff \lim_{n\to\infty} \mathbf{E}(g(X_n)) = \mathbf{E}(g(X))\ \forall g \in L^\infty \cap C^0$.
Theorem 2.24 (Lévy's Continuity). $\lim^d_{n\to\infty} X_n = X \iff \lim_{n\to\infty} \phi_{X_n}(t) = \phi_X(t)\ \forall x$.
Theorem 2.25. $X \perp\!\!\!\perp Y \iff \phi_{(X,Y)}(s,t) = \phi_X(s)\phi_Y(t)$.

2.3 Gaussians

Theorem 2.26. Distribution of jointly Gaussian random variables is defined by the mean/covariance. E.g., if X, Y are JG then $\operatorname{Cov}(X,Y)=0 \iff X \perp\!\!\!\perp Y$.

Theorem 2.27. X, Y JG & $\Sigma_Y > 0 \implies X = \mu_X + \Sigma_{XY}\Sigma_Y^{-1}(Y-\mu_Y) + V$, $V \sim \mathcal{N}\big(0, \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}\big)$ & $V \perp\!\!\!\perp Y$.

Definition 2.28. (X_t) is Gaussian process (GP) iff all FDDs are jointly Gaussian.

Definition 2.29. (X_t) is Brownian motion (BM) if: (i) it has independent stationary increments; (ii) it is zero-mean GP with $X_0 = 0$; (iii) a.s. continuous sample paths. Standard BM has $\operatorname{Var}(X_1) = 1$.

Proposition 2.30. BM is zero-mean GP with a.s. continuous sample paths and $\operatorname{Cov}(X_s, X_t) = \operatorname{Var}(X_1) \cdot (s \wedge t)$.

2.4 Linear Estimation

Theorem 2.31. $Y = (Y_i) \in L^2 \implies \forall X \in L^2\ \exists \mathbf{L}(X \mid Y) \in \mathcal{L}(Y)\colon\ \mathbf{L}(X \mid Y) = \operatorname{proj}_{L^2}(X; \mathcal{L}(Y))$.

Theorem 2.32 (Vector BLE). $X \in L^2, Y = (Y_1, \dots, Y_n) \in L^2$ & $\Sigma_Y > 0 \implies \mathbf{L}(X \mid Y) = \mathbf{E}(X) + \Sigma_{XY}\Sigma_Y^{-1}(Y - \mathbf{E}(Y))$ & $\|X - \mathbf{L}(X \mid Y)\|_{L^2}^2 = \Sigma_X - \Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX}$.

Theorem 2.33 (Kalman Filter). State space model $X_{n+1} = a_n X_n + U_n; Y_n = X_n + V_n$ with $\operatorname{Var}(X_0) = P_0$ & $\operatorname{Var}(U_n) = Q_n$ & $\operatorname{Var}(V_n) = N_n$ and $U_i \perp\!\!\!\perp V_j \perp\!\!\!\perp X_0$, initialize $\widehat{X}_{0|-1} = 0$, $\sigma_{0|-1}^2 = P_0$, iterate $\widehat{X}_{n+1|n} = a_n \widehat{X}_{n|n-1} + k_n(Y_n - \widehat{X}_{n|n-1})$ where $k_n = a_n \sigma_{n|n-1}^2(\sigma_{n|n-1}^2 + N_n)^{-1}$ is Kalman gain, and $\sigma_{n+1|n}^2 = \operatorname{Var}\Big(X_{n+1} - \widehat{X}_{n+1|n}\Big) = a_n(a_n - k_n)\sigma_{n|n-1}^2 + Q_n$.

2.5 WSS Estimation

Definition 2.34. WSS process X has regular covariance (RC) iff (i) $R_{XX} \in \ell^2(\mathbb{Z})$, and (ii) $\inf S_{XX} > 0$ & $\sup S_{XX} < \infty$.

Definition 2.35. $I \subseteq \mathbb{Z}, Y$ is zero-mean WSS RC, $Z \in L^2 \implies \exists h \in \ell^2(\mathbb{Z})\colon\ \mathbf{L}(X \mid Y_I) = \sum_{i \in I} h_i Y_i$, h unique on I.

Theorem 2.36 (Wiener-Hopf). For $I \subseteq \mathbb{Z}$, Y is zero-mean WSS RC, $h \in \ell^2(\mathbb{Z})$ s.t. $\mathbf{L}(X \mid Y_I) = \sum_{i \in I} h_i Y_i$ satisfies $\langle X, Y_n \rangle_{L^2} = (R_{YY} * h)(n)$ for $n \in I$ and $h(n) = 0$ for $n \notin I$.

Corollary 2.37. Y zero-mean WSS RC $\implies h \in \ell^2(\mathbb{Z})$ s.t. $\mathbf{L}(X \mid Y) = \sum_{i \in \mathbb{Z}} h_i Y_i$ has frequency response $H(\omega) = \frac{S_{YX}(\omega)}{S_{YY}(\omega)}$ $\forall \omega \in \mathbb{S}^1$ & $\|X - \mathbf{L}(X \mid Y)\|_{L^2}^2 = \operatorname{Var}(X) - \frac{1}{2\pi}\int_{\mathbb{S}^1} \frac{|S_{YX}(\omega)|}{S_{YY}(\omega)}\,\mathrm{d}\omega$.

Theorem 2.38 (Noncausal Wiener Filter). X, Y zero-mean JWSS, Y RC, $\mathbf{L}(X_n \mid Y) = \sum_{i \in \mathbb{Z}} h_i Y_{n-i} \implies h$ has frequency response $H(\omega) = \frac{S_{YX}^*(\omega)}{S_{YY}(\omega)}$ $\forall \omega \in \mathbb{S}^1$ & $\|X_n - \mathbf{L}(X_n \mid Y)\|_{L^2}^2 = \frac{1}{2\pi}\int_{\mathbb{S}^1}\left(S_{XX}(\omega) - \frac{|S_{YX}(\omega)|^2}{S_{YY}(\omega)}\right)\mathrm{d}\omega$.

3 Random Processes

3.1 DTMC

Proposition 3.1. $A \subseteq \mathbb{N}$, $\gcd(A) = d$, A closed under addition $\implies \exists N\colon kd \in A\ \forall k \geq N$.

Notation 3.2. $T_j := \inf\{n \in \mathbb{N}\colon X_n = j\}$. $\mu_{jj} := \mathbf{E}(T_j \mid X_0 = j)$. $N_j(n) := |\{k \in [n]\colon X_k = j\}|$.

Definition 3.3 (Markov Property). $X_s - X_t - X_u\ \forall s < t < u$.

Definition 3.4 (Strong Markov Property). $X_s - X_\tau - X_u\ \forall s < \tau < u$, for all finite stopping times τ.

Theorem 3.5 (Chapman-Kolmogorov). $m, n \geq 0$ & $i, j \in \mathcal{S} \implies P_{ij}^{m+n} = \sum_{k \in \mathcal{S}} P_{ik}^m P_{kj}^n$.

Proposition 3.6. Periodicity, transience, positive recurrence, and null recurrence are class properties.

Lemma 3.7. State i is recurrent iff $\sum_{n \in \mathbb{N}} P_{ii}^n = +\infty$.

Corollary 3.8. If $i \leftrightarrow j$ and j is recurrent, then $\mathbf{P}(T_j < \infty \mid X_0 = i) = 1$.

Theorem 3.9 (DTMC SLLN). (X_n) DTMC, $X_0 = i$, $i \leftrightarrow j \implies \lim_{n\to\infty} \frac{N_j(n)}{n} = \frac{1}{\mu_{jj}}$.

Corollary 3.10. (X_n) irreducible DTMC $\implies \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P_{ij}^k = \frac{1}{\mu_{jj}}$.

Theorem 3.11 (Balance Equations). π is SD iff $\pi_j = \sum_i \pi_i p_{ij}\ \forall j \in \mathcal{S}$.

Definition 3.12. State i is positive recurrent iff $\mu_{ii} < \infty$, null recurrent iff recurrent and $\mu_{ii} = \infty$, and transient otherwise.

Theorem 3.13. Irreducible DTMC satisfies exactly one: (i) all states are transient or null recurrent, then, no SD exists and $\lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P_{ij}^k = 0\ \forall i, j$; (ii) all states are positive recurrent, then $\exists$ unique SD and $\pi_j = 1/\mu_{jj} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^n P_{ij}^k\ \forall i, j$.

Theorem 3.14. (X_n) irreducible aperiodic positive recurrent DTMC $\implies \lim_{n\to\infty} \sum_j \big|P_{ij}^n - \pi_j\big| = 0$, i.e., $P_{ij}^n \to \pi_j\ \forall i, j$. If periodic, $\lim_{n\to\infty} P_{jj}^{nd} = d\pi_j\ \forall j$. If transient/null recurrent (not necessarily aperiodic), $P_{ij}^n \to 0\ \forall i, j$.

Definition 3.15 (DTMC Reversibility). $\pi_i p_{ij} = \pi_j p_{ji}\ \forall i, j$.

Theorem 3.16 (DBEs). If $\exists \pi$ s.t. $\pi_i p_{ij} = \pi_j p_{ji}\ \forall i, j$ then DTMC reversible with SD π.

Definition 3.17 (Total Variation Distance). $\|\mu - \nu\|_{\mathrm{TV}} := \sup_{A \in \mathcal{F}} |\mu(A) - \nu(A)|$.

Definition 3.18 (Spectral Gap). (X_n) reversible DTMC, spectral gap is $\gamma := 1 - \lambda_2$, where $\lambda_2 \geq 0$ is the smallest number for which $\operatorname{Var}_\pi(Pf) \leq \lambda_2 \operatorname{Var}_\pi(f)$; here $(Pf)(i) = \sum_j p_{ij} f(j)$ and $\operatorname{Var}_\pi(f) = \operatorname{Var}_{X\sim\pi}(f(X))$.

Theorem 3.19. (X_n) reversible DTMC, spectral gap $\gamma \implies \|P^n_{i\cdot} - \pi\|^2_{\mathrm{TV}} \leq \frac{(1-\gamma)^n}{\pi_i}$ $\forall n \in \mathbb{N}, i \in \mathcal{S}$. DTMC is irreducible, aperiodic, finite state space $\implies \gamma > 0$.

3.2 Martingales

Definition 3.20 (Martingale). (X_n) RVs, (M_n) is (sub/super) martingale w.r.t. (X_n) if (M_n) adapted to (X_n), each $M_n \in L^1$, and $\mathbf{E}(M_{n+1} \mid X^n) = M_n$ $(\geq/\leq)$.

Proposition 3.21. (M_n) submartingale, $m > n \implies \mathbf{E}(M_m \mid X^n) \geq M_n$.

Proposition 3.22. (M_n) submartingale, T stopping time $\implies (M_{T\wedge n})$ submartingale.

Proposition 3.23. (M_n) submartingale, T stopping time $\implies \mathbf{E}(M_0) \leq \mathbf{E}(M_{T\wedge n}) \leq \mathbf{E}(M_n)$.

Theorem 3.24 (Optional Stopping Theorem). (M_n) submartingale, T stopping time; if $\exists k < \infty$ s.t. $T \leq k$ or $|M_n| \leq k$ $\forall n$ & $T < \infty$ or $\mathbf{E}(T) \in L^\infty$ & $|M_n - M_{n-1}| \leq k$ $\forall n$, then $\mathbf{E}(M_0) \leq \mathbf{E}(M_T)$.

Theorem 3.25 (Wald's Identity). (Y_n) adapted to (X_n), $Y_{n+1} \perp\!\!\!\perp X^n$ $\forall n$, $\sup_n \mathbf{E}(|Y_n|) < \infty$, $\mathbf{E}(Y_n) = \mu$ $\forall n$, $T \in L^1$ is stopping time w.r.t. $(X_n) \implies \mathbf{E}\left(\sum_{i=1}^T Y_i\right) = \mu\,\mathbf{E}(T)$.

Theorem 3.26 (Doob). (M_n) submartingale, $\alpha > 0, n \in \mathbb{N} \implies \mathbf{P}(\max_{0\leq k\leq n} M_k \geq \alpha) \leq \frac{\mathbf{E}\left(M_n^+\right)}{\alpha}$.

Theorem 3.27 (Azuma-Hoeffding). (M_n) submartingale w.r.t. (X_n), $\forall k\ \exists L_k, U_k$ predictable s.t. $L_k \leq M_{k-1} - M_k \leq U_k$ and $U_k - L_k \leq c_k$ for some $c_k \geq 0 \implies \forall n \in \mathbb{N}, t > 0\colon \mathbf{P}(M_n \leq M_0 - t) \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n c_k^2}\right)$.

Lemma 3.28 (Hoeffding). Y RV s.t. $\mathbf{E}(Y) \leq 0$, $Y \in [a,b]$, $\alpha > 0 \implies \mathbf{E}\left(\mathrm{e}^{\alpha Y}\right) \leq \exp\left(\frac{\alpha^2(b-a)^2}{8}\right)$.

Theorem 3.29 (McDiarmid). $X_1, \ldots, X_n$ independent, $\exists c_k$ s.t. f varies by at most c_k by only changing k^{th} argument $\implies \forall t > 0\colon \mathbf{P}(f(X_1,\ldots,X_n) \geq \mathbf{E}(f(X_1,\ldots,X_n)) + t) \leq \exp\left(-\frac{2t^2}{\sum_{k=1}^n c_k^2}\right)$.

Theorem 3.30 (Convergence). (M_n) (sub/super) martingale, $\sup_n \mathbf{E}(|M_n|) < \infty \implies M := \lim_{n\to\infty} M_n$ exists, L^1.

Corollary 3.31. (M_n) (sub/super) martingale, $\exists Y \in L^1\colon |M_n| \leq Y$ $\forall n \implies M$ exists, $\mathbf{E}(M) = \lim_{n\to\infty} \mathbf{E}(M_n)$.

3.3 Poisson Processes

Theorem 3.32 (Memoryless). Exponential distribution is memoryless: $\mathbf{P}(T > t+s \mid T > t) = \mathbf{P}(T > s)$.

Theorem 3.33. $X_i \overset{\perp\!\!\!\perp}{\sim} \operatorname{Exp}(\lambda_i) \implies \min_{1\leq i\leq n} X_i \sim \operatorname{Exp}(\sum_{i=1}^n \lambda_i)$.

Definition 3.34 (Counting Process). (N_t) is CP iff $N_t \in \mathbb{N}_0$ $\forall t$, sample paths non-decreasing & right-continuous.

Proposition 3.35. (N_t) is $\mathrm{PP}(\lambda) \implies \forall t\colon N_t \sim \operatorname{Pois}(\lambda t)$.

Theorem 3.36. CP (N_t) is $\mathrm{PP}(\lambda)$ iff $N_0 = 0$, $N_t \sim \operatorname{Pois}(\lambda t)$ $\forall t$, and has independent/stationary increments.

Theorem 3.37. (N_t) is PP with arrival times $(T_i) \implies (T_1, \ldots, T_n) \mid \{N_t = n\} \sim (tU_{(1)}, \ldots, tU_{(n)})$.

Theorem 3.38 (Thinning). (N_t) is $\mathrm{PP}(\lambda)$, arrival labeled i with prob. p_i ind. $\implies$ arrivals of class i are ind. $\mathrm{PP}(\lambda p_i)$.

Theorem 3.39 (Merging). (N_t^i) are ind. $\mathrm{PP}(\lambda_i) \implies (\sum_i N_t^i)$ is $\mathrm{PP}(\sum_i \lambda_i)$.

Definition 3.40 (Inhomogeneous). CP (N_t) is $\mathrm{IPP}(\lambda)$ if $N_0 = 0$, independent increments, $N_t - N_s \sim \operatorname{Pois}\left(\int_s^t \lambda(r)\,\mathrm{d}r\right)$.

Theorem 3.41. (N_t) is $\mathrm{PP}(\lambda)$, $p\colon \mathbb{R}_{\geq 0} \to [0,1]$, counting arrivals at time t with prob. $p(t)$ ind. gives $\mathrm{IPP}(\lambda p)$.

3.4 CTMC

Definition 3.42. (X_t) is CTMC if sample paths are right-continuous and obeys Markov property.

Theorem 3.43 (Chapman-Kolmogorov). $P_{ij}^{t+s} = \sum_k P_{ik}^t P_{kj}^s$.

Theorem 3.44. (X_t) CTMC, $X_0 = i$, $T := \inf\{t \geq 0\colon X_t \neq i\} \implies T \sim \operatorname{Exp}(\lambda_i)$ for $P_{ii}^h = 1 - h\lambda_i + o(h)$; also $X_T \perp\!\!\!\perp T$ and has distribution $p_{ij} := \mathbf{P}(X_T = j \mid X_0 = i) = \lim_{h\downarrow 0} \frac{P_{ij}^h}{1-P_{ii}^h}$.

Definition 3.45. (X_t) is CTMC has **infinitesimal generator** Q: $q_{ij} := \lambda_i p_{ij}$ for $j \neq i$, $q_{ij} := -\lambda_i$ for $j = i$.

Corollary 3.46. (X_t) CTMC w. IG Q has $P_{ii}^h = 1 - h\lambda_i + o(h)$ and $P_{ij}^h = hq_{ij} + o(h)$ $\forall j \neq i$.

Theorem 3.47 (KDE). If CTMC is minimal, P^t satisfies $\frac{\mathrm{d}}{\mathrm{d}t}P^t = QP^t$ (KBE) and $\frac{\mathrm{d}}{\mathrm{d}t}P^t = P^tQ$ (KFE) with $P_0 = I$.

Corollary 3.48. If $\sup_i \lambda_i < \infty$, then $P^t = \mathrm{e}^{tQ} := \sum_{k\in\mathbb{N}_0} t^k \frac{Q^k}{k!}$.

Definition 3.49. p is SD for CTMC if $p_j = \sum_i p_i P_{ij}^t$ $\forall t \geq 0$.

Theorem 3.50. p s.t. $\sum_i p_i\lambda_i < \infty$ is SD iff $\sum_i p_i Q_{ij} = 0$ $\forall j$. CTMC irreducible $\implies p$ unique.

Corollary 3.51. Irreducible CTMC has TFAE: (i) CTMC has SD p with $\sum_i p_i\lambda_i < \infty$; (ii) embedded DTMC has SD π with $\sum_i \frac{\pi_i}{\lambda_i} < \infty$. If both are true then $\pi_i \propto \lambda_i p_i$.

Proposition 3.52. $\exists$ SD p with $\sum_i p_i\lambda_i = \infty \implies$ explosion. $\exists$ embedded SD π with $\sum_i \frac{\pi_i}{\lambda_i} = \infty \implies$ infinite lag.

Notation 3.53. $T_{jj} = \inf\{t \geq 0\colon X_t = j, \exists s \in (0,t)\colon X_s \neq j\}$; $m_{jj} = \mathbf{E}(T_{jj})$.

Theorem 3.54. (X_t) irreducible CTMC, $f_j(t) = \frac{1}{t}\int_0^t 1_{\{X_s=j\}}\,\mathrm{d}s$, $X_0 = i \implies \lim_{t\to\infty} f_j(t) = \frac{1}{\lambda_j m_{jj}}$.

Corollary 3.55. (X_t) irreducible CTMC, $X_0 = i \implies \lim_{t\to\infty}\frac{1}{t}\int_0^t P_{ij}^s\,\mathrm{d}s = \frac{1}{m_{jj}\lambda_j}$.

Theorem 3.56. Irreducible CTMC has exactly one: (i) transient/null recurrent, then $\lim_{t\to\infty}\frac{1}{t}\int_0^t P_{ij}^s\,\mathrm{d}s = 0\ \forall i,j$ & no SD exists; (ii) positive recurrent, then $\exists$ unique SD $p_j = \frac{1}{m_{jj}\lambda_j} = \lim_{t\to\infty}\frac{1}{t}\int_0^t P_{ij}^s\,\mathrm{d}s$.

Theorem 3.57 (Ergodic Theorem). Irreducible positive-recurrent CTMC, SD p, $r\colon \mathcal{S}\to\mathbb{R}$ bounded $\implies \forall i\colon X_0 = i, \lim_{t\to\infty}\frac{1}{t}\int_0^t r(X_s)\,\mathrm{d}s = \mathbf{E}_{X\sim p}(r(X))$. Holds for DTMCs too: $\forall i\colon X_0 = i, \lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^n r(X_k) = \mathbf{E}_{X\sim\pi}(r(X))$.

Theorem 3.58. Irreducible positive recurrent CTMC, SD $p \implies \lim_{t\to\infty}\|P_{i\cdot}^t - p\|_{\mathrm{TV}} = 0\ \forall i$.

Definition 3.59. CTMC reversible if $\exists p\colon \sum_i p_i\lambda_i < \infty$ and $p_i q_{ij} = p_j q_{ji}\ \forall i,j$. Then p is SD.

Definition 3.60 (Spectral Gap). $\lambda =$ largest s.t. $\mathrm{Var}_p(f) \leq \frac{1}{\lambda}\mathcal{E}(f,f) = \frac{1}{\lambda}\cdot\frac{1}{2}\sum_{i,j}|f(i)-f(j)|^2 q_{ij}p_i\ \forall f \in L^2(p)$.

Theorem 3.61. If reversible CTMC admits spectral gap λ, then $\|P_{i\cdot}^t - p\|_{\mathrm{TV}}^2 \leq \mathrm{e}^{-2\lambda t}/p_i\ \forall t \geq 0$.

3.5 Renewal Processes

Theorem 3.62 (SLLN). For RP (X_t) with holding times (τ_n), $\lim_{t\to\infty}\frac{X_t}{t} = \frac{1}{\mathbf{E}(\tau_1)}$.

Theorem 3.63 (Elementary Renewal Theorem). Renewal function $m(t) = \mathbf{E}(X_t)$ satisfies $\lim_{t\to\infty}\frac{m(t)}{t} = \frac{1}{\mathbf{E}(\tau_1)}$.

Theorem 3.64. (W_n) i.i.d. L^1 are rewards. Renewal-reward process for (X_t) is (R_t) where $R_t = \sum_{i=1}^{X_t} W_i$. Reward function $r(t) = \mathbf{E}(R_t)$. Then $\lim_{t\to\infty}\frac{R_t}{t} = \frac{\mathbf{E}(W_1)}{\mathbf{E}(\tau_1)}$ and $\lim_{t\to\infty}\frac{r(t)}{t} = \frac{\mathbf{E}(W_1)}{\mathbf{E}(\tau_1)}$.

Theorem 3.65 (Little's Law). G/G/1 queue. CP for arrivals (A_t) is RP with holding times (τ_n). CP for departures (D_t). Define $N_t = A_t - D_t$. $W_n =$ time spent by n^{th} customer. If service time $\mu < \mathbf{E}(\tau_1) < \infty$ then $\overline{N} := \lim_{t\to\infty}\int_0^t N_s\,\mathrm{d}s$ and $\overline{W} := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n W_i$ exist and $\overline{N} = \frac{\overline{W}}{\mathbf{E}(\tau_1)}$.

3.6 Hypothesis Testing

Definition 3.66 (Errors). Type I error: $\mathbf{P}_0\left(\widehat{H} = H_1\right)$; Type II error: $\mathbf{P}_1\left(\widehat{H} = H_0\right)$.

Definition 3.67 (Threshold Test). $\Lambda = \frac{\mathrm{d}\mathbf{P}_1}{\mathrm{d}\mathbf{P}_0} = \frac{\mathrm{d}\mathbf{P}_1/\mathrm{d}\lambda}{\mathrm{d}\mathbf{P}_0/\mathrm{d}\lambda}$. $\widehat{H}_\eta = H_{1\{\Lambda\geq\eta\}}$.

Definition 3.68 (Error Curve). $u(\theta) := \sup_{\eta\geq 0}\left\{\mathbf{P}_1\left(\widehat{H}_\eta = H_0\right) + \eta\left(\mathbf{P}_0\left(\widehat{H}_\eta = H_1\right) - \theta\right)\right\}$.

Theorem 3.69 (Neyman-Pearson Lemma). If $\mathbf{P}_1 \ll \mathbf{P}_0$ every $\widehat{H}_\eta$ lies on error curve, and all tests lie above it.

Definition 3.70 (Sufficient Statistic). T is sufficient iff $\exists\nu\colon \Lambda = \nu \circ T$.

4 Distributions

- Discrete:
  - $X \sim \mathrm{Bernoulli}(p)$: $p_X(x) = p^x(1-p)^{1-x}$; $\mathbf{E}(X) = p$; $\mathrm{Var}(X) = p(1-p)$; $\phi_X(t) = 1 - p + p\mathrm{e}^{\mathrm{i}t}$.
  - $X \sim \mathrm{B}(n,p)$: $p_X(x) = \binom{n}{x}p^x(1-p)^{n-x}$; $\mathbf{E}(X) = np$; $\mathrm{Var}(X) = np(1-p)$; $\phi_X(t) = \left(1 - p + p\mathrm{e}^{\mathrm{i}t}\right)^n$.
  - $X \sim \mathrm{Geo}(p)$: $p_X(x) = (1-p)^{x-1}p$; $F_X(x) = 1-(1-p)^x$; $\mathbf{E}(X) = \frac{1}{p}$; $\mathrm{Var}(X) = \frac{1-p}{p^2}$; $\phi_X(t) = \frac{p\mathrm{e}^{\mathrm{i}t}}{1-(1-p)\mathrm{e}^{\mathrm{i}t}}$
  - $X \sim \mathrm{Pois}(\lambda)$: $p_X(x) = \frac{\lambda^x\mathrm{e}^{-\lambda}}{x!}$; $\mathbf{E}(X) = \lambda$; $\mathrm{Var}(X) = \lambda$; $\phi_X(t) = \exp\left(\lambda(\mathrm{e}^{\mathrm{i}t} - 1)\right)$
- Continuous:
  - $X \sim \mathcal{N}(\mu,\Sigma)$: $p_X(x) = \frac{1}{(2\pi)^{n/2}\det(\Sigma)^{1/2}}\mathrm{e}^{-\frac{1}{2}(x-\mu)^*\Sigma^{-1}(x-\mu)}$; $\mathbf{E}(X) = \mu$; $\mathrm{Var}(X) = \Sigma$; $\phi_X(t) = \exp\left(\mathrm{i}\mu^* t - \frac{1}{2}t^*\Sigma t\right)$.
  - $X \sim \mathrm{Exp}(\lambda)$: $p_X(x) = \lambda\mathrm{e}^{-\lambda x}$; $F_X(x) = 1 - \mathrm{e}^{-\lambda x}$; $\mathbf{E}(X) = \frac{1}{\lambda}$; $\mathrm{Var}(X) = \frac{1}{\lambda^2}$; $\phi_X(t) = \frac{\lambda}{\lambda - \mathrm{i}t}$.
  - $X \sim \mathrm{Erlang}(\lambda,k)$: $p_X(x) = \frac{\lambda^k x^{k-1}\mathrm{e}^{-\lambda x}}{(k-1)!}$; $\mathbf{E}(X) = \frac{k}{\lambda}$; $\mathrm{Var}(X) = \frac{k}{\lambda^2}$; $\phi_X(t) = \left(1 - \frac{\mathrm{i}t}{\lambda}\right)^{-k}$.