1 Probability cheat sheet

PLACEHOLDER. INCLUDE

- MGFs

2 Markov chains

2.1 Example: Reservoir storage

Given: $S_{n+1} = S_n + Z_{n+1} - (aS_{n+1}^b)$, $Z_i \sim f_z(\cdot)$. **Want:** $P_x(S_1 \leq y)$

$$\begin{aligned} P_x(S_1 \leq y) &= P_x(g(S_1) \leq g(y)), \text{ recognizing } g(x) = x + ax^b \text{ monotone} \\ &= P_x(S_1 + aS_1^b \leq y + ay^b) = P_x(x + Z_1 \leq y + ay^b) = F_z(y + ay^b - x) \\ p(x,y) &= \frac{d}{dy} F_z(y + ay^b - x) = f_z(y + ay^b - x) * (1 + aby^{b-1}), \quad P(x,B) = \int_B f_z(y + ay^b - x) * (1 + aby^{b-1}) dy \end{aligned}$$

2.2 Example: Congestion modeling

Given: Markov chain $W = (W_n : n \geq 0)$, $W_{n+1} = [W_n + Z_{n+1}]^+$, $Z_i \sim f_z(\cdot)$. **Want:** Transition kernel.

$$P_x(W_1 \leq y) = P_x([x + Z_1]^+ \leq y) = P_x(x + Z_1 \leq y) = F_z(y - x), \text{ second to last step since } y \geq 0$$

$$\text{When } y = 0: P_x(W_1 = 0) = P(W_1 \leq 0) = F_z(-x) (\text{point mass at } y = 0); \quad \text{When } y > 0: \frac{d}{dy} P_x(W_1 \leq y) = f_z(y - x)$$

$$P(x, dy) = F_z(-x)\delta_0(dy) + f_z(y - x)dy, \quad P(X, B) = F_z(x)\delta_0(B) + \int_B f_x(y - x)dy$$

2.3 Example: Autogregressive modeling

For $X_{n+1} = a_0 X_n + c + \epsilon_{n+1}$, $\epsilon \sim N(0, \sigma^2)$, $L(a_0, c, \sigma^2 \mid X) = \prod_{j=0}^{n-1} (\frac{1}{\sqrt{2\pi\sigma}}) \exp(\frac{-1}{2\sigma^2}(X_{j+1} - a_0 X_j - c)^2)$; $Cov(X_{n+1}, X_n) = Cov(a_0 X_n + c + \epsilon, X_n) = a_0 var(X_n)$

3 Likelihood and estimation

3.1 Estimating equations

Objective is to postulate $g(\cdot)$ such that $E_{\theta_1} g(\theta_2, X_1) = 0 \iff \theta_1 = \theta_2$. We can estimate θ with the root, $\hat{\theta}$, of the equation $\frac{1}{n}\sum_{i=1}^n g(\hat{\theta}, X_i) = 0$. Estimating equations are a generalization of Method of Moments ($g(\theta, x) = E_\theta k(X_1) - k(x)$) and Maximum likelihood estimators ($g(\theta, x) = \frac{\nabla_\theta f(\theta,x)^T}{f(\theta,x)}$)

Assume we've established that $\hat{\theta} = \hat{\theta}_n \xrightarrow{p} \theta^*$ (consistent), then

$$\frac{1}{n}\sum_{i=1}^n g(\hat{\theta}, X_i) - \frac{1}{n}\sum_{i=1}^n g(\theta^*, X_i) = -\frac{1}{n}\sum_{i=1}^n g(\theta^*, X_i)$$

$$\frac{1}{n}\sum_{i=1}^n g(\xi_n, X_i)(\hat{\theta} - \theta^*) = -\frac{1}{n}\sum_{i=1}^n g(\theta^*, X_i), \text{ by Taylor expansion (Mean Value Theorem)}$$

$$\frac{1}{n}\sum_{i=1}^n g(\xi_n, X_i)\sqrt{n}(\hat{\theta} - \theta^*) = \frac{-1}{\sqrt{n}}\sum_{i=1}^n g(\theta^*, X_i), \text{ where } \frac{1}{n}\sum_{i=1}^n g(\xi_n, X_i) \xrightarrow{p} E_{\theta^*} g'(\theta^*, X_1), \text{ and } \frac{-1}{\sqrt{n}}\sum_{i=1}^n g(\theta^*, X_i) \xrightarrow{d} \sigma N(0,1),\ \sigma^2 = E_{\theta^*}[g(\theta^*, X_1)^2]$$

$$\sqrt(n)(\hat{\theta} - \theta^*) \xrightarrow{d} \frac{\theta}{E_{\theta^*} g'(\theta^*, X_1)} N(0,1), \text{ by Slutsky's Lemma}$$

3.2 Example: Markov chain parameter estimation

Given: $X_n = \beta n + W_n$, $W_n = \rho W_{n-1} + Z_n$, $Z_i \sim N(\mu, \sigma^2) iidrvs$
Trick: Rearrange everything in terms of $Z_i : Z_n = W_n - \rho W_{n-1} \implies Z_n = X_n - \beta n - \rho(X_{n-1} - \beta(n-1))$

$$L = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-1}{2\sigma^2}(Z_n - \mu)^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp(\frac{-1}{2\sigma^2}(X_n - \beta n - \rho(X_{n-1} - \beta(n-1)) - \mu)^2) \implies \log L = \text{const} - \frac{1}{2}(2 - \rho)^2 \implies \hat{\rho} = 2$$

3.3 Example: Markov chain parameter estimation

Given: $(X_j : 0 \leq j \leq n)$ observed path for finite state Markov chain. $P(\theta) = (P(\theta, x, y) : w, y \in S)$ transition matrix depending on unknown param. $P(\theta)$ infinetely differentiable in θ **Want:** Likelihood, MLE, Martingale CLT

$$L(\theta \mid X) = \prod_{j=1}^n P(\theta, X_{j-1}, X_j), \text{ by Markov property} \implies l(\theta \mid X) = \log L(\theta \mid X) = \sum_{j=1}^n \log P(\theta, X_{j-1}, X_j)$$

$$\frac{d}{d\theta} l(\theta \mid X) = \sum_{i=1}^n \frac{\frac{d}{d\theta} P(\theta, X_{i-1}, X_i)}{P(\theta, X_{i-1}, X_i)} \implies \hat{\theta} \text{ is solution to } \sum_{i=1}^n \frac{\frac{d}{d\theta} P(\hat{\theta}, X_{i-1}, X_i)}{P(\hat{\theta}, X_{i-1}, X_i)} = 0$$

Martingale CLT: First show its a martingale, then use CLT

$$M = (M_n = f(X_{n-1}, X_n) : n \geq 0) \text{ adapted to } X = (X_n : n \geq 0) \text{ with Martingale difference, } D_i = \frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)}$$

$$X \text{ stationary } \implies D_i := (P'/P)(\theta, X_{i-1}, X_i) \text{ stationary ergotic sequence}$$

$$E[D_i] = \int \frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)} P(\theta, X_{i-1}, x_i)dx_i = \int \frac{d}{d\theta}P(\theta, X_{i-1}, x_i)dx_i = \frac{d}{d\theta}1 = 0$$

$$EM_n = E[M_0 + \sum_i D_i] = E[M_0] + \sum_i E[D_i] = E[M_0] < \infty$$

$$E[M_{n+1} \mid X_0, \ldots, X_n] = E[M_n + \frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)} \mid X_0, \ldots, X_n] = M_n + 0$$

$$\frac{1}{\sqrt{n}} M_n \xrightarrow{d} \sigma N(0,1), \text{ where } \sigma^2 = E[D_1^2] = E\left[\left(\frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)}\right)^2\right]$$

3.4 Example: Kernel density estimation for derivative

Kernel density estimation: Estimate unknown density, $f^*(x)$ from 1D iid data, $X_1, \ldots, X_n$ with a normal (or other kernel) function about each point, that's then summed up: $f_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}\phi(\frac{x-X_i}{h})$ • **Equation derivation:** At each point, X_i, smooth using density $N(X_i, h^2)$: $P(N(X_i, h^2) \leq y) = P(N(0,1) \leq (\frac{y-X_i}{h})) = \Phi(\frac{y-X_i}{h}) \implies \frac{d}{dy}\Phi(\frac{y-X_i}{h}) = \phi(\frac{y-X_i}{h}) * \frac{1}{h}$
Objective: derive density estimator, derive expressions for bias and variance of estimator, choose optimal bandwidth, h^*. Recall: Here we want MSE $= var + bias^2$ to not explode so ultimately we choose h^* such that $O(var) = O(bias^2)$.

$$\frac{d}{dx}f_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}\frac{d}{dx}\phi(\frac{x-X_i}{h})$$

$$\begin{aligned}
E[\frac{d}{dx}f_n(x)] &= \frac{1}{h}E[\frac{d}{dx}\phi(\frac{x-X_1}{h})] = \frac{1}{h}\int \frac{d}{dx}\phi(\frac{x-y}{h})f^*(y)dy = \frac{1}{h}\int \frac{1}{h}\phi'(z)f^*(x-zh)(-h)dz, \text{ for } zh = x - y \\
&= \frac{-1}{h}\int \phi'(z)[f(x) - zhf'(x) + \frac{(xh)^2}{2!}f''(x) - \frac{(zh)^3}{3!}f'''(x) + O(h^3)]dz \\
&= \frac{-1}{h}f(x)\int \phi'(z)dz + f'(x)\int z\phi'(z)dz - \frac{h}{2}f'''(x)\int z^2\phi'(z)dz + \frac{(h)^2}{3!}\int z^3\phi'(z)dz + O(h^2) \\
&= \frac{-1}{h}f(x) * 0 + f'(x) * 1 - \frac{h}{2}f'''(x) * 0 + \frac{h^2}{3!} * \frac{1 * 4!}{2^2 * 2} + O(h^2), \text{ where } \phi'(x) = x\phi(x)
\end{aligned}$$

$$E[\frac{d}{dx}f_n(x)] - \frac{d}{dx}f^*(x) = \frac{h^2}{2}f'''(x) + O(h^2) = O(h^2)$$

$$\begin{aligned}
Var(\frac{d}{dx}f_n(x)) &= \frac{1}{nh^2}Var(\frac{d}{dx}\phi(\frac{x-X_1}{h})) = \frac{1}{nh^2}E[(\frac{d}{dx}\phi(\frac{x-X_1}{h}))^2] - \frac{1}{nh^2}[E(\frac{d}{dx}\phi(\frac{x-X_1}{h}))]^2 \\
&= \frac{1}{nh^2}E[\frac{1}{h^2}\phi'^2(\frac{x-X_1}{h})] - \frac{1}{nh^2} * (O(h^2))^2 = \frac{1}{nh^2}\frac{1}{h^2}\int \phi'(z)^2 f^*(x-zh)(-h)dz - \frac{1}{nh^2} * (O(h^2))^2 \\
&= \frac{1}{nh^2}\frac{-1}{h}\int z^2\phi^2(z)[f^*(x) - O(h)]dz - \frac{1}{nh^2} * (O(h^2))^2 = O(\frac{1}{nh^3}) - O(h^2) = O(\frac{1}{nh^3})
\end{aligned}$$

$$O(var) \approx O(bias^2) \implies O(\frac{1}{nh^3}) \approx O(h^4) \implies h = O(n^{-1/7}) \implies MSE = O(n^{-2/7})$$

4 First transition analysis

Stationary: $E[f(X_{n+1}, \ldots) \mid X_n = x] = E[f(X_1, \ldots) \mid X_0 = x] = E_x[f(X_1, \ldots)]$

4.1 Example: Expectation of hitting time

Compute: E_xT_A, $x \notin A$, $T_A = \inf\{n \geq 0 : X_n \in A\}$
When $x \in A, E_xT_A = 0$. Otherwise:

$$E_xT_A = 1 + E_x[T_A - 1] = 1 + \sum_{y \in A} E_x[T_A - 1 \mid X_1 = y]P_x(X_1 = y) + \sum_{y \notin A} E_x[T_A - 1 \mid X_1 = y]P_x(X_1 = y) = 1 + 0 + \sum_{y \notin A} E_y(T_A)P_x(X_1 = y)$$

$$E_x[T_{A-1} \mid X_1 = y] = E_x[\sum_{j=1}^{T_{A-1}} 1 \mid X_1 = y] = E_x[\sum_{j=1}^{\infty} \mathbb{I}\{j < T_A\} \mid X_1 = y] = E_x[\sum_{j=1}^{\infty} \mathbb{I}\{X_0 \notin A, \ldots, X_j \notin A\} \mid X_1 = y]$$

$$E_x[T_{A-1} \mid X_1 = y] = E_x[\sum_{j=1}^{\infty} \mathbb{I}\{X_1 \notin A, \ldots, X_j \notin A\} \mid X_1 = y] = E_y[\sum_{j=0}^{\infty} \mathbb{I}\{X_1 \notin A, \ldots, X_{j-1} \notin A\}] = E_y[\sum_{j=1}^{\infty} \mathbb{I}\{X_0 \notin A, \ldots, X_j \notin A\}] = E_yT_A$$

$$u = e + Bu, \text{ where } u = \{E_xT_X\}_{x \in A^c}, B = \{P_{x,y}\}_{(x,y) \in A^c \times A^c}$$

4.2 Example: Expectation of reward

Given: S discrete finite, $u(i) = E_i[\exp(-\sum_{n=0}^{T_A-1}\rho(X_n))r(X_{T_A})]$, X_n Markov chain, T_A hitting time
When $i \in A$, then $T_A = 0, u(i) = E_i[\exp(0)r(X_0)] = r(i)$. Otherwise:

$$u(i) = \exp(-p(i))E_i[\exp(-\sum_{n=1}^{T_{A-1}}\rho(X_n))r(X_{T_A})] = \exp(-p(i))\sum_{j\in S}E_i[\exp(-\sum_{n=1}^{T_{A-1}}\rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j)$$

$$= \exp(-p(i))\sum_{j\in A}E_i[\exp(-\sum_{n=1}^{T_{A-1}}\rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j) + \exp(-p(i))\sum_{j\notin A}E_i[\exp(-\sum_{n=1}^{T_{A-1}}\rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j)$$

$$= \exp(-p(i))\sum_{j\in A}E[r(X_1) \mid X_1 = j]P_i(X_1 = j) + \exp(-p(i))\sum_{j\notin A}u(j)P(i,j) = \exp(-p(i))\sum_{j\in A}r(j)P(i,j) + \exp(-p(i))\sum_{j\notin A}u(j)P(i,j)$$

$$u = b + Ku, \text{ where } b_i = exp(-p(i))\sum_{j\in A}r(j)P(i,j), K(i,j) = exp(-p(i))P(i,j)$$

5 Infinite horizon stochastic control

Objective: Find optimal control $A^* = (A_n^* : n \geq 0)$ for objective $\max_{(A_n:n\geq 0)} E_x \sum_{n=1}^{\infty} \exp(-\alpha n)r(X_n, A_n)$
Solution: Let $v(x) = \max_{(A_n:n\geq 0)} E_x \sum_{n=1}^{\infty} \exp(-\alpha n)r(X_n, A_n)$
By first transition analysis: $v(x) = \max_{a\in\mathcal{A}(x)}[r(x,a) + \exp(-\alpha)\sum_y P_a(x,y)v(y)] = \max_{a\in\mathcal{A}(x)}\{r(x,a) + \exp(-\alpha)E[v(X_1) \mid X_0 = x, A_0 = a]\}$
Solution approach 1 - Fixed point equation: Notice this is a solution to the fixed point equation $v = Tv$, where $(Tu)(x) = \max_{a\in\mathcal{A}(x)}[r(x,a) + \exp(-\alpha)\sum_y P_a(x,y)u(y)]$. 1 Choose any v_0, 2) iterate $v_n = Tv_{n-1}$, 3), if $v_n \longrightarrow v_\infty$ then v_∞ is solution. Convergence guaranteed with contractive property: $\|Tv_n - Tv_{n-1}\|_\infty \leq \exp(-\alpha)\|v_n - v_{n-1}\|_\infty$
Solutions approach 2 - Linear program: $\min_v \sum_x v(x)$ $s.t., v(x) \geq r(x,a) + \exp(-\alpha)\sum_y P_a(x,y)v(y)$

5.1 Example: Optimal stopping time

Given: reward function $r : \{0,\ldots,m\} \to \mathbb{R}_+$, $(X_n : n \geq 0)$ has transition probabilities $P(x,y) = 1/2$, $x \in \{1,\ldots,m-1\}, y \in \{0,\ldots,m\}$, $P(0,0) = P(m,m) = 1$
Optimality equation (HJB equation):

$$v(x) = \sup_T E_x r(X_T) = \max\{\text{stop, continue}\} = \max(r(x), \frac{1}{2}(v(x-1) + v(x+1))),\ x \in \{1,\ldots,m-1\};\ v(0) = r(0),\ v(m) = r(m)$$

Let $r(m) = 0$ and $r(x) = x$ otherwise. Compute **value function:** must be unique, using intuition you can claim it is $v(x) = x$. Given this, $v(x) = \frac{1}{2}(v(x-1) + v(x+1))$ for $x \leq m-1$. Hence, optimal stopping time is immediately if you are at $m-1$ or indifferent otherwise.

5.2 Example: Optimal stopping time

Given: $X = (X_n : n \geq 0)$, finite state, $P = (P(x,y) : x,y \in S)$
Want: T to maximize $E_x \sum_{j=1}^{T-1} \exp(-\alpha j)r(X_j) + \exp(-\alpha T)w(X_T)$

$$v^*(x) = \sup_T E_x \sum_{j=1}^{T-1} \exp(-\alpha j)r(X_j) + \exp(-\alpha T)w(X_T), \text{ is finite valued and should satisfy } v(x) = max\{w(x), r(x) + \exp(-\alpha)\sum_y P(x,y)v(y)\}$$

Solution 1, Linear program: $\min_{v^*(x)} \sum_{x\in S} v(x)$ s.t., $v(x) \geq w(x),\ v(x) \geq r(x) + \exp(-\alpha)\sum_y P(x,y)v(y)$

Solution 2, Value iteration: $(Ru)(x) = max\{w(x), r(x) + \exp(-\alpha)\sum_y P(x,y)v(y)\}$, choose v_0 and iterate; guaranteed convergence

6 Martingales

Martingale definition: A martingale $(M_n : n \geq 0)$ is adapted to $(Z_n : N \geq 0)$ if 1) Adaptedness: for each $n \geq 0$ there exists function $f_n(\cdot)$ such that $M_n = f_n(X_0,\ldots,X_n)$, 2) $E|M_n| < \infty$, 3) $E[M_{n+1} \mid X_0,\ldots,X_n] = M_n$ • $D_n = M_n - M_{n-1}$ • $M_n = M_0 + \sum_i D_i$ • $ED_i = 0$ • $Cov(D_i, D_j) = ED_iD_j = 0, i \neq j$ • $Cov(M_0, D_i) = 0$ • $Var(M_n) = Var(M_0) + \sum_i Var(D_i)$
Martingale convergence: $\frac{1}{n}M_n \overset{a.s.}{\to} 0$
Martingale CLT: If a martingale $(M_n : n \geq 0)$ adapted to $(Z_n : N \geq 0)$ is square integrable, then $\frac{1}{\sqrt{n}}M_n \overset{d}{\to} \sigma N(0,1)$ • $\sigma^2 = Var(D_1) = E(D_1^2)$

6.1 Example: Demonstrate martingale sequence

Given: $S_n = Z_1 + \cdots + Z_n$, Z_i iid, $EZ_1^2 < \infty$, $EZ_1 = 0$, $M_n = S_n^2 - n\sigma^2$
Adaptedness condition exists by definition. **Boundedness condition** holds since $\sigma^2 < \infty$, $EZ_1 = 0$. **Expectation of** M_{n+1}**:** $E(M_{n+1} \mid Z_0,\ldots,Z_n) = E[(S_n + Z_{n+1})^2 - (n+1)\sigma^2 \mid Z_0,\ldots,Z_n] = S_n^2 + 2S_nE[Z_{n+1} \mid Z_0,\ldots,Z_n] + E[Z_{n+1}^2 \mid Z_0,\ldots,Z_n] - n\sigma^2 - \sigma^2 = S_n^2 + 2S_n * 0 + \sigma^2 - n\sigma^2 - \sigma^2 = S_n^2 - n\sigma^2 = M_n$

6.2 Example: Demonstrate martingale sequence

Given: $f : S \longrightarrow \mathbb{R}$, bounded and $Pf = f$, one-step transition matrix, X_n a Markov sequence. **Want:** show $f(X_n)$ is a martingale sequence. **Adaptedness condition** exists by definition. **Boundedness condition** holds by boundedness of f. **Expectation of** M_{n+1}**:** $E[f(X_{n+1}) \mid X_0,\ldots,X_n] = \sum_{y\in S} f(y)P(X_{n+1} = y \mid X_0,\ldots,X_n) = \sum_{y\in S} F(y)P(X_n, y) = [Pf]_{X_n} = f(X_n)$

6.3 Example: Demonstrate martingale difference sequence

Given: $g : S \longrightarrow \mathbb{R}$ bounded and $D_i = g(X_i) - E[g(X_i) \mid X_{i-1}]$. **Show:** This is a martingale difference adapted to $X = (X_n : n \geq 0)$ **Adaptedness condition** exists by definition. **Boundedness condition** holds by definition of g. **Zero expectation conditional on the past:** $E[D_n + 1 \mid X_0, \ldots, X_n] = E[g(X_{n+1}) \mid X_0, \ldots, X_n] - E[E[g(X_{n+1}) \mid X_n] \mid X_0, \ldots, X_n] \iff E[g(X_{n+1}) \mid X_n] - E[g(X_{n+1}) \mid X_n] = 0$

7 Bayesian statistics

7.1 Example: posterior distribution

Want: posterior distribution of probability of success, p. **Given:** $\pi(p) \sim Beta(\alpha, \beta)$, k successes in n experiments
$\pi(p \mid X) \propto \pi(p)L(p \mid X) \propto p^{\alpha-1}(1-p)^{\beta-1}p^k(1-p)^{n-k} = p^{\alpha+k-1}(1-p)^{\beta+n-k-1} \propto Beta(\alpha + k, \beta + n - k)$

7.2 Example: posterior distribution

Given: iid data, $X_1, \ldots, X_n$, follows Poisson: $f(x) = \lambda e^{-\lambda x}$, $\lambda > 0$, unknown; prior on λ follows Gamma with shape param (α) 3 and rate (β) param $2\pi(\lambda) = 4\lambda^2 e^{-2\lambda}$ Aside: Gamma rv, $g(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{\beta x}$, $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-z} dx$, the integrating constant is $\frac{\beta^\alpha}{\Gamma(\alpha)}$

$$\pi(\lambda \mid X) \propto \pi(\lambda)L(X \mid \lambda) = 4\lambda^2 e^{-2\lambda} \prod_{i=1}^{n} \lambda e^{-\lambda X_i} \sim Gamma(\alpha, \beta)$$

7.3 Markov Chain Monte Carlo

Motivation: Generate a posterior distribution by running a markov chain whose equilibrium distribution is the posterior, $f(\theta \mid X)$. Required to impose "detailed balance" on the system: $\tilde{p}(x)p(x, y) = \tilde{p}(y)p(y, x)$. Achieve this through the *Metropolis Algorithm:* 1) Start with harris recurrent transition density, $(q(x, y) : x, y \in S)$, positive everywhere, 2) define $p(x, y) = q(x, y) \min(1, \frac{p(y)Q(x,y)}{p(x)Q(x,y)})$

8 Positive recurrence

SLLN for Markov chains:

$$\frac{1}{n}\sum_{j=1}^{n-1} I(X_j = y) \xrightarrow{a.s} \frac{EY_1}{E\tau_1} : \frac{1}{n}\sum_{j=1}^{n-1} I(X_j = y) \approx \sum_{j=0}^{N(n)} Y_j / \sum_{j=1}^{N(n)} \tau_j, \text{ where } Y_j = \sum_{i=T_{j-1}}^{T_j - 1} I(X_j = y),\ \tau_j = T_j - T_{j-1},\ \frac{1}{n}\sum_{j=1}^{n} Y_j \xrightarrow{a.s} EY_1,\ \frac{1}{n}\sum_{j=1}^{n} \tau_j \xrightarrow{a.s} E\tau_1$$

- **Lyapunov method to demonstrate postivie Harris recurrence:** Must demonstrate for some $g(x) \geq 0$ and $A \subseteq S$ a) $E_x[g(X_1)] \leq g(x) - \epsilon$ for $x \in A^c$ b) $\sup_{x \in A} E_x[g(X_1)] < \infty$, c) $P_x(X_m \in \cdot) \geq \lambda\varphi(\cdot)$ for $x \in A$. Common choices of $g(x) : \{x, |x|^p, \log(1+x)^p, \exp(a|x|^p)\}$. Positive Harris Recurrence guarantees unique solution for stationary density of chain

- **General approach for element c):** $P_x(X_1 \in B) \geq \lambda\varphi(B) \iff \int_B p(x, y)dy \geq \lambda \int_B \phi(y)dy$. Then simply let $\varphi(y) = \inf_{x \in A} p(x, y)/\lambda$ and $\lambda = \int_S \inf_{x \in A} p(x, y)dy$, making sure $\lambda > 0$

- **Explanation of P(x, dy):** $P(x, dy) = P(x \in y + dy) \approx P(x \in [y - \Delta y/2, y + \Delta y/2]) = \int_{-\Delta y/2}^{\Delta y/2} f(x)dx \approx f(y)\Delta y \approx f(y)dy$

- **Markov chain positive recurrence properties:** Markov chain is positive recurrent $(E_x\pi(x) < \infty) \Longrightarrow \frac{1}{n}\sum_{i=0}^{n-1} r(X_i) \overset{a.s.}{\to} \sum_w \pi(x)r(w)$ and $\pi(x) = \frac{E_x \sum_{j=1}^{\tau(x)-1} I(X_j = x)}{E_x \pi(x)}$

- **Markov chain aperiodicity:** $\gcd\{n \geq 1 : P^n(x, x) > 0\} = 1 \iff P(x, x) > \infty$

8.1 Example: Positive Harris recurrence

Given: $X = \{X_n : n \geq 0\}, [X_{n+1} \mid X_n = x] \sim N(\lambda x, 1 - \lambda^2)$, $\lambda \in (0, 1)$ a constant. Choosing $g(x) = x^2$:

a) $E_x g(X_1) = E_x X_1^2 = varX_1 + (E_x X_1)^2 = (1 - \lambda^2) + (\lambda x)^2 = x^2 - (x^2 - 1)(1 - \lambda^2) \leq g(x) - 3(1 - \lambda^2)$ when $x \in K^c$ $K = [-2, 2]$

b) $\sup_{x \in K} E_x g(X_1) = \sup_{x \in K}[(1 - \lambda^2) + (\lambda x)^2] \leq 1 - \lambda^2 + 4\lambda^2 < \infty$

c) $P_x(X_1 \leq y) = P(N(\lambda x, 1 - \lambda^2)) = P(N(0, 1) \leq \dfrac{y - \lambda x}{\sqrt{1 - \lambda^2}}) \Longrightarrow p.d.f : p(x, y) = \phi(\dfrac{y - \lambda x}{\sqrt{1 - \lambda^2}}) * \dfrac{1}{\sqrt{1 - \lambda^2}} > 0$

$\varphi(y) = \inf_{x \in K} p(x, y)/c = \inf_{x \in K} \phi(\dfrac{y - \lambda x}{\sqrt{1 - \lambda^2}}) * \dfrac{1}{c\sqrt{1 - \lambda^2}} > 0$, since ϕ continuous, positive, K compact $\Longrightarrow$ choose$\lambda = \int_{\mathbb{R}} \inf_{x \in K} p(x, y)$

Stationary sequence: Noting $X_{n+1} = \lambda X_n + \epsilon_{n+1}$, $\epsilon \sim N(0, 1 - \lambda^2)$. When $X_0 \sim N(0, 1) \Rightarrow X_n \sim N(0, 1)$, so $N(0, 1)$ is stationary distribution of X.

8.2 Example: Positive Harris recurrence

Given: $(Z_n : n \geq 1)$ iid positive, $|EZ_1^2| < \infty$, positive continuous density, $f(\cdot)$; $X = \{X_n : n \geq 0\}$ Markov chain such that $X_{n+1} = |X_n - Z_{n+1}|$
Want: Transition density, positive Harris recurrence, equilibrium density, stationary distribution, SLLN
Transition density: $P(x, dy) = P(|x - Z| \in y + dy) = P(Z \in x - y + dy) + P(Z \in x + y + dy) = f(x - y)dy + f(x + y)dy$
Positive Harris recurrence:

$$Z \text{ integrable } \implies \exists M \, s.t., \, E[X\mathbb{I}(Z \leq M)] \geq (2/3)EZ, \, E[X\mathbb{I}(Z > M)] \leq (1/3)EZ; \text{ now choose } g(x) = |x| \text{ and define } A^c : x > M$$

For $x \in A^c$: $E(g(X_1)) = E|x - Z| = E(x - Z)\mathbb{I}(Z \leq x) + E(Z - x)\mathbb{I}(Z > x)$

$$\leq x - E(Z)\mathbb{I}(Z \leq x) + E(Z)\mathbb{I}(Z > x) \leq x - (2/3)EZ + (1/3)EZ = g(x) - \epsilon, \text{ since} EZ_1 < \infty$$

For $x \in A$: $P(x, dy) \geq \inf_{x' \in [0,M]} P(x', dy) = [\inf_{x' \in [0,M]} (f(x' - y) + f(x' + y))]dy > 0$, since $f(\cdot)$ is positive continuous $\implies P(x, dy) \geq \lambda\varphi(y)$where λ c

Stationary distribution: Need to verify $\int_0^\infty P(x, dy)\pi(dx) = \pi(dy) = \pi(y)dy = \frac{P(Z_1 > y)dy}{EZ_1}$, equivalent to showing $\int_0^\infty (f(x-y)+f(x+y))P(Z > x)dx = P(Z > y)$

When $y = 0$: $\displaystyle\int_0^\infty 2f(x)\frac{P(Z_1 > x)}{EZ_1}dx = \int_0^\infty 2f(x)\frac{1 - F(x)}{EZ_1}dx = \frac{1}{EZ_1}[\frac{d}{dx}\int_0^\infty 2F(x)dx - \frac{d}{dx}\int_0^\infty 2F(x)^2dx] = \frac{2 - 1}{EZ_1} = \frac{P(Z_1 > 0)dy}{EZ_1} = \pi(0)$

When $y > 0$: $\displaystyle\frac{d}{dy}(\int_0^\infty (f(x - y) + f(x + y))P(Z > x)dx) = \frac{d}{dy}(\int_0^\infty f(w)P(Z > w + y)dw + \int_y^\infty f(w)P(Z > w - y)dw)$

$$= -\int_0^\infty f(w)f(w + y)dw - f(y)P(Z > 0) + \int_y^\infty f(w)f(w - y)dw = -f(y) = \frac{d}{dy}(P(Z > y))$$

SLNN: By stationary distribution and positive Harris recurrence, we have $\frac{1}{n}\sum_{i=1}^n f(X_i) \overset{a.s.}{\to} E_\pi f(X_0)$ and

$$E_\pi x = \int_0^\infty x\pi(x)dx = \int_0^\infty x\frac{P(Z_1 > x)}{E[Z_1]}dx = \frac{1}{EZ_1}\frac{E[Z_1^2]}{2}$$

8.3 Example: Positive recurrent Markov chain

Given: $N_{n+1} = R_{n+1} + B_{n+1}(N_n)$, $R_1, \ldots \overset{iid}{\sim} Poisson(\lambda_*)$, $(B_n(k) = Bin(k, p) : n \geq 0, k \geq 0)$
Transition probability matrix:

$$P(N_{n+1} = y \mid N_n = x) = P(R_{n+1} = y - B_{n+1}(x) \mid N_n = x) = \sum_{k=1}^{min(x,y)} \frac{\lambda_*^{y-k}e^{-\lambda}}{(y - k)!}P(B_{n+1}(x) = k) = \sum_{k=1}^{min(x,y)} \frac{\lambda_*^{y-k}e^{-\lambda}}{(y - k)!}\binom{x}{k}p^k(1 - p)^{x-k}$$

Chain irreducible and aperiodic: Since $P(x, y) > 0$ for all (x, y) (irreducible) and $P(x, x) > 0$ for all x (aperiodic)
Chain positive recurrent: Irreducible Markov chain on discrete state space is positive recurrent $\iff \exists\pi \, s.t.\pi = \pi P$. We find $\pi = Poisson(\frac{\lambda_*}{1-p})$ (not shown) **Approximate for** $\frac{1}{n}\sum_{j=0}^{n-1} I(N_j = 0)$: $\frac{1}{n}\sum_{j=0}^{n-1} I(N_j = 0) \to \pi(0)$
First transition analysis: For $N_0 = k$, find $u(k) = E[\inf\{n \geq 1 : N_n - N_{n-1} \geq 3\} \mid N_0 = k] = E_kT$

$$u(k) = E_kT = 1 + \sum_{y-x\geq 3} 0 * P(k, y) + \sum_{y-x<3} E_yTP(k, y) = 1 + \sum_{y-x<3} P(k, y)u(y)$$