1 Markov chains

Example (Reservoir storage): Given: $S_{n+1} = S_n + Z_{n+1} - (aS_{n+1}^b)$, $Z_i \sim f_z(\cdot)$. **Want:** $P_x(S_1 \leq y)$ • $P_x(S_1 \leq y) = P_x(g(S_1) \leq g(y)) = P_x(S_1 + aS_1^b \leq y + ay^b) = P_x(x + Z_1 \leq y + ay^b) = F_z(y + ay^b - x)$ • $p(x,y) = \frac{d}{dy}F_z(y + ay^b - x) = f_z(y + ay^b - x) * (1 + aby^{b-1})$ • $P(x,B) = \int_B f_z(y + ay^b - x) * (1 + aby^{b-1})dy$

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. • $P_x(W_1 \leq y) = P_x([x + Z_1]^+ \leq y) = P_x(x + Z_1 \leq y) = F_z(y - x)$ • When $y = 0 : P_x(W_1 = 0) = P(W_1 \leq 0) = F_z(-x)$(point mass at $y = 0$) • When $y > 0 : \frac{d}{dy}P_x(W_1 \leq y) = f_z(y - x)$ • $P(x, dy) = F_z(-x)\delta_0(dy) + f_z(y - x)dy$, $P(X, B) = F_z(x)\delta_0(B) + \int_B f_x(y - x)dy$

Example (Autogregressive modeling): For $X_{n+1} = a_0X_n + c + \epsilon_{n+1}$, $\epsilon \sim N(0, \sigma^2)$, $L(a_0, c, \sigma^2 \mid X) = \prod_{j=0}^{n-1}(\frac{1}{\sqrt{2\pi}\sigma})\exp(\frac{-1}{2\sigma^2}(X_{j+1} - a_0X_j - c)^2)$; $Cov(X_{n+1}, X_n) = Cov(a_0X_n + c + \epsilon, X_n) = a_0 var(X_n)$

2 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:** $\frac{1}{n}M_n \xrightarrow{a.s.} 0$
Martingale CLT: If a martingale $(M_n : n \geq 0)$ adapted to $(Z_n : N \geq 0)$ is square integrable, then $\frac{1}{\sqrt{n}}M_n \xrightarrow{d} \sigma N(0,1)$ • $\sigma^2 = Var(D_1) = E(D_1^2)$

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

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. **Solution:** Adaptedness condition exists by definition. Boundedness condition holds by boundedness of f. $E[f(X_{n+1}) \mid X_0, \ldots, X_n] = \sum_{y \in S} f(y)P(X_{n+1} = y \mid X_0, \ldots, X_n) = \sum_{y \in S} F(y)P(X_n, y) = [Pf]_{X_n} = f(X_n)$

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 $X = (X_n : n \geq 0)$. **Solution:** Adaptedness condition exists by definition. Boundedness condition holds by definition of g. $E[D_n + 1 \mid X_0, \ldots, X_n] = E[g(X_{n+1}) \mid X_0, \ldots, X_n] - E[E[g(X_{n+1}) \mid X_n] \mid X_0, \ldots, X_n] \iff E[g(X_{n+1}) \mid X_n] - E[g(X_{n+1}) \mid X_n] = 0$

3 Bayesian statistics

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

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

Markov Chain Monte Carlo: 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 $p(x,y) = q(x,y)\min(1, \frac{p(y)Q(x,y)}{p(x)Q(x,y)})$

4 Likelihood and estimation

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 $\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} \frac{\sigma}{E_{\theta^*}g'(\theta^*, X_1)}N(0,1)$, $\sigma^2 = E_{\theta^*}[G(\theta^*, X_1)^2]$

Example (Markov chain parameter estimation): Given: $X_n = \beta n + W_n$, $W_n = \rho W_{n-1} + Z_n$, $Z_i \sim N(\mu, \sigma^2) iidrvs$. **Solution:** Rearrange everything in terms of $Z_i : Z_n = W_n - \rho W_{n-1} \Longrightarrow Z_n = X_n - \beta n - \rho(X_{n-1} - \beta(n-1))$

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

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

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

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

Martingale CLT: First show its a martingale, then use CLT: • $M = (M_n = f(X_{n-1}, X_n) : n \geq 0)$ adapted to $X = (X_n : n \geq 0)$ 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 stationary $\implies D_i := (P'/P)(\theta, X_{i-1}, X_i)$ stationary ergotic sequence • $E[D_i] = \int \frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)} P(\theta, X_{i-1}, x_i)dx_i = \int \frac{d}{d\theta}P(\theta, X_{i-1}, x_i)dx_i = \frac{d}{d\theta}1 = 0$ • $EM_n = E[M_0 + \sum_i D_i] = E[M_0] + \sum_i E[D_i] = E[M_0] < \infty$ • $E[M_{n+1} \mid X_0, \ldots, X_n] = E[M_n + \frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)} \mid X_0, \ldots, X_n] = M_n + 0$ • $\frac{1}{\sqrt{n}}M_n \xrightarrow{d} \sigma N(0,1)$, where $\sigma^2 = E[D_1^2] = E\left[\left(\frac{\frac{d}{d\theta}P(\theta, X_{i-1}, x_i)}{P(\theta, X_{i-1}, x_i)}\right)^2\right]$

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 $N(X_i, h^2)$: $P(N(X_i, h^2) \leq y) = P(N(0,1) \leq (\frac{y-X_i}{h})) = \Phi(\frac{y-X_i}{h}) \implies \frac{d}{dy}\Phi(\frac{y-X_i}{h}) = \phi(\frac{y-X_i}{h}) * \frac{1}{h}$

Example (Kernel density estimation for derivative):

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

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

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

5 First transition analysis

Example (Expectation of hitting time): Want: 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:

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

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

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

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

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:

$$
\begin{aligned}
u(i) &= \exp(-p(i))E_i[\exp(-\sum_{n=1}^{T_{A-1}} \rho(X_n))r(X_{T_A})] = \exp(-p(i))\sum_{j\in S} E_i[\exp(-\sum_{n=1}^{T_{A-1}} \rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j) \\
&= \exp(-p(i))\sum_{j\in A} E_i[\exp(-\sum_{n=1}^{T_{A-1}} \rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j) + \exp(-p(i))\sum_{j\notin A} E_i[\exp(-\sum_{n=1}^{T_{A-1}} \rho(X_n))r(X_{T_A}) \mid X_1 = j]P_i(X_1 = j) \\
&= \exp(-p(i))\sum_{j\in A} E[r(X_1) \mid X_1 = j]P_i(X_1 = j) + \exp(-p(i))\sum_{j\notin A} u(j)P(i,j) = \exp(-p(i))\sum_{j\in A} r(j)P(i,j) + \exp(-p(i))\sum_{j\notin A} u(j)P(i,j)
\end{aligned}
$$

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

6 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: $\min_v \sum_x v(x)$ $s.t., v(x) \geq r(x,a) + \exp(-\alpha)\sum_y P_a(x,y)v(y)$

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.

Example (Optimal stopping time): Given: $(X_n : n \geq 0)$, 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

7 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 $(E_x \pi(x) < \infty) \implies \frac{1}{n}\sum_{i=0}^{n-1} r(X_i) \overset{a.s.}{\to} \sum_w \pi(x) r(w)$ and $\pi(x) = \frac{E_x \sum_{j=1}^{\tau(x)-1} I(X_j = x)}{E_x \pi(x)}$

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

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) $E_x g(X_1) = E_x X_1^2 = var X_1 + (E_x X_1)^2 = (1 - \lambda^2) + (\lambda x)^2 = x^2 - (x^2 - 1)(1 - \lambda^2) \leq g(x) - 3(1 - \lambda^2)$ 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) $P_x(X_1 \leq y) = P(N(\lambda x, 1 - \lambda^2)) = P(N(0,1) \leq \frac{y - \lambda x}{\sqrt{1 - \lambda^2}}) \implies p.d.f : p(x,y) = \phi(\frac{y - \lambda x}{\sqrt{1 - \lambda^2}}) * \frac{1}{\sqrt{1 - \lambda^2}} > 0$

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

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 stat. distribution of X.

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:**

$$\begin{aligned} Z \text{ integrable } &\implies \exists M\ s.t.,\ E[X\mathbb{I}(Z \leq M)] \geq (2/3)EZ,\ E[X\mathbb{I}(Z > M)] \leq (1/3)EZ; \text{ now choose } g(x) = |x| \text{ and define } A^c : x > M \\ \text{For } x \in A^c :\ E(g(X_1)) &= E|x - Z| = E(x - Z)\mathbb{I}(Z \leq x) + E(Z - x)\mathbb{I}(Z > x) \\ &\leq x - E(Z)\mathbb{I}(Z \leq x) + E(Z)\mathbb{I}(Z > x) \leq x - (2/3)EZ + (1/3)EZ = g(x) - \epsilon, \text{ since} EZ_1 < \infty \\ \text{For } x \in A :\ P(x, dy) &\geq \inf_{x' \in [0,M]} P(x', dy) = [\inf_{x' \in [0,M]} (f(x' - y) + f(x' + y))]dy > 0, \text{ since } f(\cdot) \text{ is positive continuous} \Rightarrow P(x, dy) \geq \lambda\varphi(y) \end{aligned}$$

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 $\int_0^\infty (f(x-y) + f(x+y))P(Z > x)dx = P(Z > y)$

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

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

SLNN: By stat. dist. and PHR, we have $\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow{a.s.} E_\pi f(X_0)$ and $E_\pi x = \int_0^\infty x\pi(x)dx = \int_0^\infty x\frac{P(Z_1>x)}{E[Z_1]}dx = \frac{1}{EZ_1}\frac{E[Z_1^2]}{2}$

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 $u(k) = E[\inf\{n \geq 1 : N_n - N_{n-1} \geq 3\} \mid N_0 = k] = E_k T$

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

8 Prior material

8.1 Calculus cheat sheet

Logs: $log_b(M*N) = log_b M + log_b N$ • $log_b(\frac{M}{N}) = log_b M - log_b N$ • $log_b(M^k) = k log_b M$ • $e^n e^m = e^{n+m}$ • **Derivatives:** $(x^n)' = nx^{n-1}$ • $(e^x)' = e^x$ • $(e^{u(x)})' = u'(x)e^x$ • $(log_e(x))' = (lnx)' = \frac{1}{x}$ • $(f(g(x)))' = f'(g(x))g'(x)$ • **Integrals:** $\int_a^b f(x)dx = \int_{g(a)}^{g(b)} f(g(u))g'(u)du$ where $g(u) = x$ • $\int_a^b u(x)v'(x)dx = u(b)v(b) - u(a)v(a) - \int_a^b u'(x)v(x)dx$ • **Infinite series and sums:** $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^\infty \frac{x^n}{n!}$ • $(1+\frac{a}{n})^n \longrightarrow e^a$ • $ln(1+x) = 1 - x + \frac{x^2}{2} - \frac{x^3}{3} + \cdots = \sum_{n=0}^\infty (-1)^n \frac{x^n}{n}$ • $\frac{1}{1-x} = 1 + x + x^2 + x^3 + \cdots = \sum_{n=0}^\infty a^x$ for $|x| < 1$ • **L'Hopitale:** $\lim_{n\to c} f(x)/g(x) = \lim_{n\to c} f'(x)/g'(x)$ if $\lim_{n\to c} f(x) = \lim_{n\to c} g(x) = 0/\infty/-\infty$

8.2 Generative functions

PLACEHOLDER, INCLUDE MORE • $\phi_X(t) = E\exp(itX)$ • $N(\mu,\sigma^2) : \exp(it\mu - \frac{1}{2}\sigma^2 t^2)$ • $Exp(\lambda) : (1 - it\lambda^{-1})^{-1}$ • $Poisson(\lambda) : \exp\lambda(e^{it} - 1)$ • $E[X^k] = i^{-k}E[X^k]$ • $\phi_{a_1X_1+\cdots+a_nX_n}(t) = \phi_{X_1}(a_1 t)\ldots\phi_{X_n}(A_n t)$ for X_i independent • $M_X(t) = E\exp(tX)$

8.3 Expectation

Expectation: $E(X) = \sum_x xP(X = x)$ • $P_{Y,X}(Y > X) = E_{Y,X}\mathbb{I}\{Y > X\} = E_X E_Y[\mathbb{I}\{Y > X\} \mid X] = E_X P_Y(Y > X \mid X) = \int_\mathbb{R} P_Y(Y > x \mid X = x)F_X(dx)$ • **Variance:** $Var(X) = E((X - E(X))^2) = \sigma^2$ • $Var(X) = E(X^2) - \mu^2$ • $Var(aX + b) = a^2 Var(X)$ • $Var(X + Y) = Var(X) + Var(Y) + 2Cov(X,Y)$ • **Covariance:** $Cov(X,Y) = E((X - E(X)(Y - E(Y)) = E(XY) - E(X)E(Y)$ • $Cov(X,X) = Var(X)$ • $Cov(aX, bY) = abCov(X,Y)$ • $Cov(X, Y + Z) = Cov(X,Y) + Cov(X,Z)$ • **Law of iterated expectation:** $E(E(Y \mid X)) = E(Y)$ • **Law of total probability**: $P(E) = \sum_{i=-\infty}^\infty P(E \mid X = x)P(X)$ and $P(E) = \int_{-\infty}^\infty P(E \mid X = x)f(x)dx$ • **Variance decomposition formula:** $Var(Y) = E(Var(Y \mid X)) + Var(E(Y \mid X))$

8.4 Inequalities

Markov inequality: For $X \geq 0$, $P(X \geq t) \leq \frac{E(X)}{t}$ $\forall t > 0$. • **Chebyshev inequality:** $P(|X - E(X)| \geq t) \leq \frac{Var(X)}{t^2}$ $\forall t > 0$. • **Exponential inequality:** $P(X > a) \leq e^{-\theta a}E(e^{\theta X})$ for all $\theta > 0$. • **(Corrolary) Upper bound on large deviations:** $P(S_n < na) \leq e^{-nI(x)}$. • **Proof:** $P(S_n > a) \leq e^{-\theta a}E(e^{\theta S_n}) = e^{-\theta a}\prod_i E(e^{\theta X_i}) = e^{-\theta a}E(e^{\theta X_1})^n = e^{-\theta a + n\psi(\theta)}$, by exponential inequality, iid, where $\psi(\theta) = \log Ee^{\theta X_1}$ • $P(S_n > na) = e^{-n(\theta(x)a - n\psi(\theta(x))}$; minimizing RHS w.r.t $\theta \Longrightarrow e^{-nI(x)}$ where $I(x) = \theta(x)a - n\psi(\theta(x))$

Weak law of large numbers: For $X_1, X_2, \ldots, X_n$ i.i.d. with $E(X_i) = \mu$, $Var(X_i) = \sigma^2$, $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, then for any $\epsilon > 0$. $P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{Var(\overline{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0$ as $n \to \infty$

Central limit theorem: $\sqrt{n}\frac{(\overline{X}_n - \mu)}{\sigma} \longrightarrow N(0,1) \iff \sqrt{n}(\overline{X}_n - \mu) \longrightarrow N(0,\sigma^2) = \sigma N(0,1)$, $X_1 + \cdots + X_n = S_n \approx N(ES_n, VarS_n)$, the "meta result"

Monte Carlo: • Sample $Y \in \mathbb{R}^d$ • Compute $X = g(Y)$ • Repeat n times • form $\overline{X}_n$ and use CLT for asymptotic behavior.
Generating random data: with $X = F_X^{-1}(U)$ since $P(F_X^{-1}(U) \leq x) = P(F_X(F_X^{-1}(U)) \leq F_X(x)) = P(U \leq F_X(x)) = F_x(x)$
Confidence intervals: $P(Z_{\alpha/2} \leq Z \leq Z_{1-\alpha/2}) = P(\frac{\hat\sigma}{\sqrt{n}}Z_{\alpha/2} \leq \bar{X}_n - \mu \leq \frac{\hat\sigma}{\sqrt{n}}Z_{1-\alpha/2}) = P(\mu \in \left[\bar{X}_n \pm \frac{\hat\sigma}{\sqrt{n}}Z_{1-\alpha/2}\right]) = 1 - \alpha$ for $\hat\sigma \xrightarrow{p} \sigma$
Slutsky's lemma: $A_n X_n + B_n \xrightarrow{d} aX + b$ if $\{X_n\}$ sequence, $X_n \xrightarrow{d} X$, $\{A_n\}$ sequence, $A_n \xrightarrow{d} A$, $\{B_n\}$ sequence, $B_n \xrightarrow{p} b$
Delta method: If g is a differentiable function at μ, $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \xrightarrow{d} N(0, g'(\mu)^2\sigma^2) = g'(\mu)\sigma N(0,1)$. **Proof sketch:** Start with Taylor expansion $g(\bar{X}_n) \approx g(\mu) + g'(\mu)(\bar{X}_n - \mu)$, rearrange to get $\sqrt{n}(g(\bar{X}_n) - g(\mu)) \approx g'(\mu)\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, g'(\mu)^2\sigma(2))$.

8.5 Theory of large deviations

Variance reduction: $E_p h(X)$ where $X \sim P = \int_{-\infty}^\infty h(x)p(x)dx = \int_{-\infty}^\infty h(x)p(x)q(x)/q(x) = E_q[h(x)p(x)/q(x)]$ where $X \sim Q$.
$\sqrt{n}(\overline{X} - E_p X) \xrightarrow{d} N(0, Var_p(X))$, $\sqrt{n}(\hat{\overline{X}} - E_q\hat{X}) \xrightarrow{d} N(0, Var_q(\hat{X}))$ where $\hat{X} = Xp(x)/q(X)$
Importance sampling: Choose $h(x)$ to minimize variance. Minimal $H(dx)$ turns out to be the conditional probability of the event happening on event happening: $H^*(dx) = \mathbb{I}\{A\}(x)F(dx)/F(A)$
Exponential tilting: $Ef(X_1, \ldots, X_n) = E_\theta \exp(-\theta S_n + n\psi(\theta))f(X_1, \ldots, X_n)$, where $\psi(\theta) = \log M_x(\theta)$
Large deviations: • Use exponential tilting with $f(X_1, \ldots, X_n) = \mathbb{I}(S_n > an)$: $E\mathbb{I}(S_n > an) = E_\theta[\exp(-\theta S_n + n\psi(\theta))\mathbb{I}(S_n > an)]$ • Choose

optimal $\theta^* = \theta(a)$ which satisfies $\psi'(\theta(a)) = a$, which guarantees $E_{\theta(a)} X_i = a$ • E.g., Gaussian: $M_X(\theta) = \exp(\mu\theta + \sigma^2\theta^2/2) \iff \psi(\theta) = \mu\theta + \sigma^2\theta^2/2 \iff \psi'(\theta) = \mu + \sigma^2\theta \leftarrow$ evaluate at $\psi'(\theta) = a$

Method of moments estimator: $E(X^k) = g(\theta)$: Calculate moment with MGF, lower moments typically lead to estimators with lower asymptotic variance • $g^{-1}(E(X^k)) = \theta$: Invert this expression to create an expression for the parameter(s) in terms of the moment • $\hat{\theta} = g^{-1}(\frac{1}{n}\sum X_i^k)$: Insert the sample moment into this expression, thus obtaining estimates of the parameters in terms of data • $\sqrt{n}(g^{-1}(\frac{1}{n}\sum X_i^k) - \theta) \xrightarrow{d} N(0, f'(E(X_i^k))^2 Var(X_i^k)^2)$: Use the delta method • If multiple parameters characterize the distribution, use multiple moments and a system of equations

Maximum likelihood estimator • $L(\theta) = \prod_{i=1}^n f(X_i, \theta)$: Construct the likelihood function • $log(L(\theta)) = l(\theta) = \sum_{i=1}^n log(f(X_i, \theta))$: Take the log of the likelihood • Find critical points of this function (e.g., $0 = \sum_{i=1}^n \frac{d}{d\theta} log(f(X_i, \hat{\theta}))$) and determine that one is a maximum