1 Probability cheat sheet

PLACEHOLDER. INCLUDE

- MGFs

2 First transition analysis

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

2.1 Example: Expectation of hitting time

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

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

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

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

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

2.2 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:

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

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

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

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

3 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 $v_n = T v_{n-1}$, 3), if $v_n \longrightarrow v_\infty$ then v_∞ is solution. Convergence guaranteed with contractive property: $\|T v_n - T v_{n-1}\|_\infty \leq \exp(-\alpha) \|v_n - v_{n-1}\|_\infty$
Solutions approach 2 - Linear program: $\min_v \sum_x v(x) \;\; s.t., v(x) \geq r(x,a) + \exp(-\alpha) \sum_y P_a(x,y) v(y)$

3.1 Example: Optimal stopping time

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

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

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.

4 Likelihood and estimation

4.1 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 $Z_i : Z_n = W_n - \rho W_{n-1} \Longrightarrow Z_n = X_n - \beta n - \rho(X_{n-1} - \beta(n-1))$

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

4.2 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: $f_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h}\phi(\frac{x-X_i}{h})$
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 $O(var) = O(bias^2)$

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

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

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

5 Bayesian statistics

5.1 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 $\frac{\beta^\alpha}{\Gamma(\alpha)}$

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

6 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 $g(x) : \{x, |x|^p, \log(1+x)^p, \exp(a|x|^p)\}$

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

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

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 stationary distribution of X.

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

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