# Lecture 6: Uncountable spaces

- ☐ Uncountable spaces
- ☐ Error bars for MCMC
- ☐ Bounding the variance

## Recall

Last week's proposition: assume

a) $\mathcal{X}$ is finite or countably infinite

b) $x \in \mathcal{X}$ is a "positive recurrent" state, i.e.,

$$E[\tau_x \mid X_0 = x] < \infty \quad \text{where} \quad \tau_x = \min\{t > 0 : X_t = x\}$$

c) $X_0$ is started from an arbitrary distribution $q$

s.t. $\mathbb{P}\{\tau_x < \infty \mid X_0 \sim q\} = 1$ ("irreducibility")

Then, there exists unique $\pi = \pi_x$ s.t.

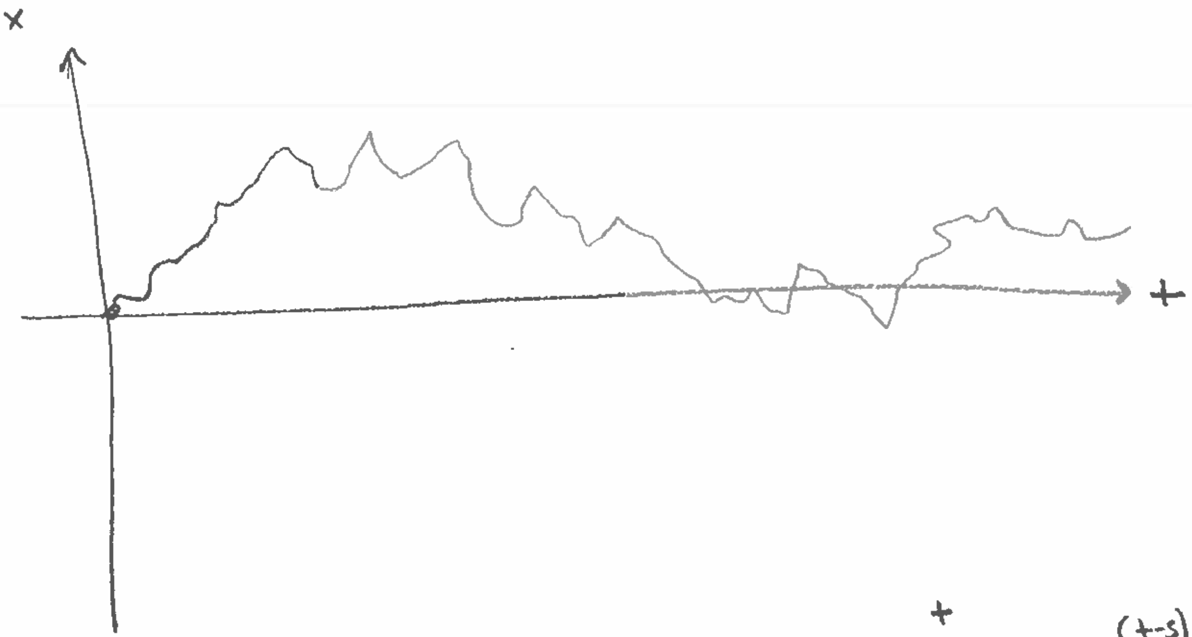
a) $$\frac{1}{T}\sum_{t=0}^{T-1} f(X_t) \xrightarrow{T \to \infty} \sum_{i \in \mathcal{X}} \pi(i) f(i)$$ for bounded functions $f$ ("ergodicity")

b) $\pi^T \mathbb{P} = \pi^T$ ("stationarity")

This week: assume $\mathcal{X}$ is uncountably infinite.

Ex: autoregressive process

$$X_t = \sqrt{1-\alpha}\, X_{t-1} + \sqrt{\alpha}\, Z_t, \quad Z_t \sim \text{iid } \mathcal{N}(0,1)$$




By recursion, $X_t = (1-\alpha)^{t/2} X_0 + \alpha^{1/2} \sum_{s=1}^{t} (1-\alpha)^{(t-s)/2} Z_s$

$$\Rightarrow \begin{cases} \mathbb{E}[X_t] = (1-\alpha)^{t/2} \, \mathbb{E}[X_0] \\ \mathrm{Var}[X_t] = (1-\alpha)^t \, \mathrm{Var}[X_0] + 1 - (1-\alpha)^t \end{cases}$$

Conditional on $X_0$, $X_t \sim \mathcal{N}\left( (1-\alpha)^{t/2} X_0, \; 1-(1-\alpha)^t \right)$

whence $X_t \xrightarrow{d} \mathcal{N}(0, 1)$

Q 1 Does $X_t$ satisfy an ergodic theorem?

Q 2 Does $X_t$ satisfy a CLT?

Structural

Q 3 If $X_t$ is a $\mathcal{N}(0,1)$ sampler, how can we place error bars on MC estimates?

Algorithmic

Def one-step minorization condition.

$\exists$ set $A \subseteq \mathcal{X}$, measure $\mu$ on $\mathcal{X}$, number $\delta > 0$ s.t.

$$p(x, dy) \geq \delta \, \mu(dy), \quad \forall \, x \in A.$$

$\Rightarrow$ All the transition probabilities starting from inside $A$ are bounded from below by $\delta \mu(dy)$

---

Ex : Autoregressive process $X_t = \sqrt{1-\alpha}\, X_{t-1} + \sqrt{\alpha}\, Z_t$,

"small set" $A = [-\sqrt{3}, \sqrt{3}]$.

$\forall\, x \in A$, transition densities are

$$p(x,y) = \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2\alpha}\left(y - \sqrt{1-\alpha}\, x\right)^2\right)$$

$$\geq \frac{1}{\sqrt{2\pi\alpha}} \exp\left(-\frac{1}{2\alpha}\left(2y^2 + 2(1-\alpha)x^2\right)\right)$$

$$= \underbrace{\frac{1}{\sqrt{\pi\alpha}} \exp\left(-\frac{y^2}{\alpha}\right)}_{\mu = \mathcal{N}\left(0, \frac{\alpha}{2}\right)} \cdot \underbrace{\frac{1}{\sqrt{2}} \exp\left(-\frac{1-\alpha}{\alpha} x^2\right)}_{\geq \delta = \frac{1}{\sqrt{2}} e^{-3/\alpha}}$$

$\Rightarrow$ One-step minorization condition with

$A = [-\sqrt{3}, \sqrt{3}]$, $\mu = \mathcal{N}\left(0, \frac{\alpha}{2}\right)$, $\delta = \frac{1}{\sqrt{2}} e^{-3/\alpha}$ $\square$

Def Nummelin split-chain construction

Assume one-step minorization condition

$$p(x, dy) \geq \delta \mu(dy)$$

$\Rightarrow$ we can embed $(X_t)_{t \geq 0}$ as the marginal distribution of a bigger chain $X'_t = (X_t, R_t)$ where $R_t \in \{0,1\}$ is a random reset time.

$X'_t$ has transition probabilities

$$\begin{cases} q((x,r),(dy,0)) = p(x,dy), & x \notin C \\ q((x,r),(dy,0)) = p(x,dy) - \delta\mu(dy), & x \in C \\ q((x,r),(dy,1)) = \delta\mu(dy), & x \in C \end{cases}$$

__Prop__ For bounded functions $f$, if $t_i < \infty$ for $i = 1, 2, \ldots$,

$$\sum_{s=t_{i-1}}^{t_i - 1} f(x_s)$$ are iid random variables for $i = 1, 2, \ldots$

where $t_1 = \min\{t > 0 : R_t = 1\}$, $t_i = \min\{t > t_{i-1} : R_t = 1\}$

__Pf__ Same starting distribution $X_{t_i} \sim$ iid $\mu$, $R_{t_i} = 1$;

same transition probabilities governed by $q$. $\square$

__Q__ How do we know $t_i < \infty$?

— Before we used irreducibility + positive recurrence of $x$, but positive recurrence fails.

__Def__ Foster-Lyapunov drift criterion

$\exists$ function $V : \mathcal{X} \to [1, \infty)$, $\lambda > 0$, $b > 0$ s.t.

$$\mathbb{E}[V(X_1) \mid X_0 = x] \leq (1-\lambda) V(x) \mathbb{1}\{x \notin A\} + b \mathbb{1}\{x \in A\},$$

where $A$ satisfies one-step minorization condition.

Ex : Autoregressive process $X_t = \sqrt{1-\alpha}\, X_{t-1} + \sqrt{\alpha}\, Z_t$,

Lyapunov function $V(x) = x^2 + 1$.

$$\mathbb{E}[V(X_1) \mid X_0 = x] = (1-\alpha)x^2 + \alpha + 1$$
$$= \left(1 - \tfrac{\alpha}{2}\right)(x^2+1) + \tfrac{\alpha}{2}(3 - x^2)$$
$$\le \left(1 - \tfrac{\alpha}{2}\right) V(x)\, \mathbb{1}\{x^2 > 3\} + 4\, \mathbb{1}\{x^2 \le 3\}$$

$\Rightarrow$ Foster-Lyapunov drift criterion with $\lambda = \alpha/2$, $b = 4$, $A = [-\sqrt{3}, \sqrt{3}]$.

<u>Prop (geometric tails)</u> $\mathbb{P}\{\tau_A \ge t \mid X_0 = x\} \le (1-\lambda)^t V(x)$,

where $\tau_A = \min\{t > 0 : X_t \in A\}$.

<u>Pf</u> $\mathbb{P}\{\tau_A \ge t \mid X_0 = x\} = \mathbb{E}[\mathbb{1}\{\tau_A > t-1\} \mid X_0 = x]$

$$\le \mathbb{E}[V(X_t)\, \mathbb{1}\{\tau_A > t-1\} \mid X_0 = x]$$
$$= \mathbb{E}[\mathbb{E}[V(X_t) \mid X_{t-1}]\, \mathbb{1}\{\tau_A > t-1\} \mid X_0 = x]$$
$$\le (1-\lambda)\, \mathbb{E}[V(X_{t-1})\, \mathbb{1}\{\tau_A > t-1\} \mid X_0 = x]$$
$$\le \cdots \le (1-\lambda)^t\, \mathbb{E}[V(X_0)\, \mathbb{1}\{\tau_A > t-1\} \mid X_0 = x]$$
$$= (1-\lambda)^t V(x).$$

$\square$

<u>Theorem (geometric ergodicity)</u> Assume one-step minorization condition + Foster-Lyapunov drift criterion. Then $\exists$ unique stationary measure

$\pi$ and for all starting distributions $q$ and bounded functions $f$,

$$\frac{1}{T}\sum_{t=0}^{T-1} f(X_t) \xrightarrow{T\to\infty} \int_{\mathcal{X}} \pi(dx)\, f(x).$$

Sketch of proof Use the split chain construction and the fact that $\sum_{s=t_{i-1}}^{t_i - 1} f(X_s)$ are iid random variables and $\mathbb{E}[t_i - t_{i-1}] < \infty$. See the problem set.

Bias-variance decomposition We can decompose the mean-square error as

$$\mathbb{E}\left|\frac{1}{T}\sum_{t=0}^{T-1} f(X_t) - \int_{\mathcal{X}} \pi(dx) f(x)\right|^2$$

$$= \underbrace{\left|\mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} f(X_t)\right] - \int_{\mathcal{X}} \pi(dx) f(x)\right|^2}_{\text{square bias}} + \underbrace{\mathrm{Var}\left[\frac{1}{T}\sum_{t=0}^{T-1} f(X_t)\right]}_{\text{variance}}$$

- Bias is size $\frac{1}{T}$ because

$$\mathbb{E}\left[\frac{1}{T}\sum_{t=0}^{T-1} f(X_t)\right] - \int_{\mathcal{X}} \pi(dx) f(x) \approx \frac{1}{T} \underbrace{\sum_{t=0}^{\infty}\left(\mathbb{E} f(X_t) - \int_{\mathcal{X}} \pi(dx) f(x)\right)}_{\text{sum converges}}$$

- Variance is size $\frac{1}{T}$ because

$$\lim_{T\to\infty} T\,\mathrm{Var}\left[\frac{1}{T}\sum_{t=0}^{T-1} f(x_t)\right] = \underbrace{\mathrm{Var}\left[f(x)\mid x\sim\pi\right]}_{\text{ordinary variance part}}\underbrace{\left(1+2\sum_{s=1}^{\infty}\rho_{f,s}\right)}_{\text{"integrated autocorrelation time"}}$$

where $\rho_{f,s} = \mathrm{Corr}\left[f(x_0), f(x_s) \mid x_0\sim\pi\right]$

Markov chain central limit theorem

$$\frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\left(f(x_t)-\int\pi(dx)f(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \mathrm{Var}\left[f(x)\mid x\sim\pi\right]\left(1+2\sum_{s=1}^{\infty}\rho_{f,s}\right)\right)$$

for bounded fns $f$.

Markov chain error bars

Estimate $\int\pi(dx)f(x)$ by $\hat{f}\pm\hat{\sigma}$, where $\hat{f}=\frac{1}{T}\sum_{t=0}^{T-1}f(x_t)$ and

where $$\hat{\sigma}^2 = \frac{1}{T^2}\sum_{\substack{s,t<T\\ |s-t|\le T^{1/3}}}\left(f(x_t)-\hat{f}\right)\left(f(x_s)-\hat{f}\right).$$

truncation needed for convergence

Burn-in Lose the first 10% of the data to control the bias

Ex Autoregressive process $X_t = \sqrt{1-\alpha}\, X_{t-1} + \sqrt{\alpha}\, Z_{t-1}$,

$$\text{Bias} \approx \frac{1}{T} \sum_{t=0}^{\infty} (1-\alpha)^{t/2} x_0 = \frac{1}{T} \cdot \frac{x_0}{1-(1-\alpha)^{1/2}} \qquad \boxed{f(x) = x}$$

$$\text{Correlations} = \rho_s = \mathbb{E}\left[X_0 X_s \mid X_0 \sim \mathcal{N}(0,1)\right] = (1-\alpha)^{s/2}$$

$$\text{IAT} = 1 + 2 \sum_{s=1}^{\infty} \rho_s = \frac{2}{1-(1-\alpha)^{1/2}} - 1 = \frac{1+(1-\alpha)^{1/2}}{1-(1-\alpha)^{1/2}}$$

$$\Rightarrow \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} X_t \longrightarrow \mathcal{N}\left(0, \frac{1+(1-\alpha)^{1/2}}{1-(1-\alpha)^{1/2}}\right)$$

For small $\alpha$, $\dfrac{1+(1-\alpha)^{1/2}}{1-(1-\alpha)^{1/2}} \approx \dfrac{2}{1-(1-\alpha/2)} = \dfrac{4}{\alpha}$ !