CPSC 535
Introduction to General State-Space Markov Chains

AD

April 2007

Markov chains

- Let $(E, \mathcal{E})$ be a measurable space and $M_1(E)$ be the set of probability measures on $(E, \mathcal{E})$.

Markov chains

- Let $(E, \mathcal{E})$ be a measurable space and $M_1(E)$ be the set of probability measures on $(E, \mathcal{E})$.
- A Markov transition kernel on E is a family of probability measures

$$\{K(x, \cdot) \in M_1(E); x \in E\}$$

indexed by the elements of E and such that $K(x, A)$ is measurable for any $A \in \mathcal{E}$.

Markov chains

- Let $(E, \mathcal{E})$ be a measurable space and $M_1(E)$ be the set of probability measures on $(E, \mathcal{E})$.
- A Markov transition kernel on E is a family of probability measures

$$\{K(x, \cdot) \in M_1(E); x \in E\}$$

indexed by the elements of E and such that $K(x, A)$ is measurable for any $A \in \mathcal{E}$.

- Starting from $\mu \in M_1(E)$ and a sequence of kernels $\{K_n; n \geq 1\}$ then

$$P_\mu(X_0 \in A_0, ..., X_n \in A_n) = \int_{x_0 \in A_0} \cdots \int_{x_n \in A_n} \mu(dx_0) \prod_{i=1}^{n} K_i(x_{i-1}, dx_i).$$

Notation

- Let $K(x_1, dx_2)$ be a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$.

Notation

- Let $K(x_1, dx_2)$ be a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$.
- For any (measurable) function f on E_2, we define

$$K(f)(x_1) = \int_{E_2} K(x_1, dx_2) f(x_2)$$

Notation

- Let $K(x_1, dx_2)$ be a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$.
- For any (measurable) function f on E_2, we define

$$K(f)(x_1) = \int_{E_2} K(x_1, dx_2) f(x_2)$$

- For any $\mu \in M_1(E_1)$, we define

$$\mu K(A) = \int_{E_1} \mu(dx_1) K(x_1, A)$$

Notation

- Let $K(x_1, dx_2)$ be a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$.
- For any (measurable) function f on E_2, we define

$$K(f)(x_1) = \int_{E_2} K(x_1, dx_2) f(x_2)$$

- For any $\mu \in M_1(E_1)$, we define

$$\mu K(A) = \int_{E_1} \mu(dx_1) K(x_1, A)$$

- We have

$$\mu(K(f)) = (\mu K)(f).$$

- Given $K_1(x_1, dx_2)$ a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$ and $K_2(x_2, dx_3)$ a Markov kernel from $(E_2, \mathcal{E}_2)$ to $(E_3, \mathcal{E}_3)$ then we can define a new Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_3, \mathcal{E}_3)$

$$K_1 K_2(x, A) = \int_{E_2} K_1(x, dy) K_2(y, A)$$

for any $(x, A) \in E_1 \times \mathcal{E}_3$.

- Given $K_1(x_1, dx_2)$ a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$ and $K_2(x_2, dx_3)$ a Markov kernel from $(E_2, \mathcal{E}_2)$ to $(E_3, \mathcal{E}_3)$ then we can define a new Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_3, \mathcal{E}_3)$

$$K_1 K_2(x, A) = \int_{E_2} K_1(x, dy) K_2(y, A)$$

for any $(x, A) \in E_1 \times \mathcal{E}_3$.

- Given $K(x, dx')$ a Markov kernel from $(E, \mathcal{E})$ to $(E, \mathcal{E})$, we can define the iterated kernel

$$K_n(x, A) = \int_{E^{n-1}} K(x, dx_1) K(x_1, dx_2) \cdots K(x_{n-1}, A)$$

which is the probability to move from x to A in n iterations of the Markov kernels.

- Given $K_1(x_1, dx_2)$ a Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_2, \mathcal{E}_2)$ and $K_2(x_2, dx_3)$ a Markov kernel from $(E_2, \mathcal{E}_2)$ to $(E_3, \mathcal{E}_3)$ then we can define a new Markov kernel from $(E_1, \mathcal{E}_1)$ to $(E_3, \mathcal{E}_3)$

$$K_1 K_2(x, A) = \int_{E_2} K_1(x, dy) K_2(y, A)$$

  for any $(x, A) \in E_1 \times \mathcal{E}_3$.
- Given $K(x, dx')$ a Markov kernel from $(E, \mathcal{E})$ to $(E, \mathcal{E})$, we can define the iterated kernel

$$K_n(x, A) = \int_{E^{n-1}} K(x, dx_1) K(x_1, dx_2) \cdots K(x_{n-1}, A)$$

  which is the probability to move from x to A in n iterations of the Markov kernels.
- In the MCMC context, we have typically $X_0 \sim \mu \in M_1(E)$ and K an MCMC kernel of invariant distribution π and we want the measure

$$\mu K^n$$

  to converge as fast as possible to π.

Total Variation Norm

- We denote $M(E)$ the space of bounded measures on $(E, \mathcal{E})$ equipped with the norm

$$\|\mu\| = \frac{1}{2}\left[\sup_{A \in \mathcal{E}} \mu(A) - \inf_{A \in \mathcal{E}} \mu(A)\right]$$

Total Variation Norm

- We denote $M(E)$ the space of bounded measures on $(E, \mathcal{E})$ equipped with the norm

$$\|\mu\| = \frac{1}{2}\left[\sup_{A \in \mathcal{E}} \mu(A) - \inf_{A \in \mathcal{E}} \mu(A)\right]$$

- We can think of $M_1(E)$ as

$$M_1(E) = \{\mu \in M(E) : \mu(E) = 1 \text{ and } \mu(A) \geq 0 \text{ for any } A \in \mathcal{E}\}.$$

Total Variation Norm

- We denote $M(E)$ the space of bounded measures on $(E, \mathcal{E})$ equipped with the norm

$$\|\mu\| = \frac{1}{2}\left[\sup_{A\in\mathcal{E}} \mu(A) - \inf_{A\in\mathcal{E}} \mu(A)\right]$$

- We can think of $M_1(E)$ as

$$M_1(E) = \{\mu \in M(E) : \mu(E) = 1 \text{ and } \mu(A) \geq 0 \text{ for any } A \in \mathcal{E}\}.$$

- We can easily show that for any $\mu \in M_1(E)$

$$\|\mu\| = \frac{1}{2}$$

and

$$\mu \in M_1(E) \Rightarrow \mu K \in M_1(E)$$

- We also introduce

$$M_0(E) = \{\mu \in M(E) : \mu(E) = 0\}.$$

- We also introduce

$$M_0(E) = \{\mu \in M(E) : \mu(E) = 0\}.$$

- For any $x, y \in E$ $(x \neq y)$ then

$$\mu = \delta_x - \delta_y \in M_0(E)$$

and

$$\|\mu\| = 1.$$

- We also introduce

$$M_0\left(E\right)=\left\{\mu\in M\left(E\right):\mu\left(E\right)=0\right\}.$$

- For any $x, y \in E$ $(x \neq y)$ then

$$\mu=\delta_x-\delta_y\in M_0\left(E\right)$$

  and

$$\left\|\mu\right\|=1.$$

- Moreover we have

$$\mu\in M_0\left(E\right)\Rightarrow\mu K\in M_0\left(E\right).$$

Dobrushin Coefficient

- We denote by $b(K)$ the norm of the operator K on the normed space $(M_0(E), \|\cdot\|)$
$$b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|}.$$

Dobrushin Coefficient

- We denote by $b(K)$ the norm of the operator K on the normed space $(M_0(E), \|\cdot\|)$
$$b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|}.$$
- Clearly we have for any $\mu_1, \mu_2 \in M_1(E)$ that $\mu_1 - \mu_2 \in M_0(E)$ and
$$\|\mu_1 K - \mu_2 K\| \leq b(K) \|\mu_1 - \mu_2\|$$
so $b(K)$ is a measure of the contraction induced by K.

- The number $a(K) \in [0,1]$ defined as follows

$$a(K) = \inf \left\{ \sum_{i=1}^{m} \min \left(K(x, A_i), K(y, A_i) \right) \right\}$$

where the infimum is taken over all points $x, y \in E$, the integers $m \geq 1$ and the finite partitions $\{A_i; 1 \leq i \leq m\}$ of E. It is called the **Dobrushin coefficient**.

- The number $a(K) \in [0,1]$ defined as follows

$$a(K) = \inf \left\{ \sum_{i=1}^{m} \min \left(K(x, A_i), K(y, A_i) \right) \right\}$$

where the infimum is taken over all points $x, y \in E$, the integers $m \geq 1$ and the finite partitions $\{A_i; 1 \leq i \leq m\}$ of E. It is called the **Dobrushin coefficient**.

- We will show later on that

$$a(K) + b(K) = 1,$$

i.e. we want $b(K)$ close to zero and $a(K)$ close to one for fast mixing.

Properties

- If $\mu \in M_0(E)$ then

$$\|\mu\| = \sup_{A \in \mathcal{E}} \mu(A).$$

Properties

- If $\mu \in M_0(E)$ then
$$\|\mu\| = \sup_{A \in \mathcal{E}} \mu(A).$$
- Moreover for any $\mu_1, \mu_2 \in M_1(E)$ then
$$\|\mu_1 - \mu_2\| = \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|.$$

Properties

- If $\mu \in M_0(E)$ then
$$\|\mu\| = \sup_{A \in \mathcal{E}} \mu(A).$$
- Moreover for any $\mu_1, \mu_2 \in M_1(E)$ then
$$\|\mu_1 - \mu_2\| = \sup_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|.$$
- When the space is finite, then we have
$$\|\mu_1 - \mu_2\| = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|$$

  and when μ_1, μ_2 have densities f_1, f_2 with respect to say λ then
$$\|\mu_1 - \mu_2\| = \frac{1}{2} \int_E |f_1(x) - f_2(x)| \lambda(dx).$$

Proofs

- If $\mu \in M_0(E)$ then for any $A \in \mathcal{E}$ then

$$\mu(E) = 0 = \mu(A) + \mu(A^c) \text{ where } A^c = E - A.$$

Proofs

- If $\mu \in M_0(E)$ then for any $A \in \mathcal{E}$ then
$$\mu(E) = 0 = \mu(A) + \mu(A^c) \text{ where } A^c = E - A.$$
- Thus we have
$$\sup_{A \in \mathcal{E}} \mu(A) = -\sup_{A \in \mathcal{E}} \mu(A^c) = -\inf_{A \in \mathcal{E}} \mu(A)$$
and
$$\|\mu\| = \sup_{A \in \mathcal{E}} \mu(A).$$

Proofs

- If $\mu \in M_0(E)$ then for any $A \in \mathcal{E}$ then
$$\mu(E) = 0 = \mu(A) + \mu(A^c) \text{ where } A^c = E - A.$$
- Thus we have
$$\sup_{A\in\mathcal{E}} \mu(A) = -\sup_{A\in\mathcal{E}} \mu(A^c) = -\inf_{A\in\mathcal{E}} \mu(A)$$
and
$$\|\mu\| = \sup_{A\in\mathcal{E}} \mu(A).$$
- Assume now that $\mu_1, \mu_2 \in M_1(E)$ then clearly $(\mu_1 - \mu_2) \in M_0(E)$ so it follows that
$$\begin{aligned}
\|\mu_1 - \mu_2\| &= \sup_{A\in\mathcal{E}} (\mu_1(A) - \mu_2(A)) \\
&= \sup_{A\in\mathcal{E}} (\mu_2(A) - \mu_1(A)) \text{ (by symmetry)} \\
&= \sup_{A\in\mathcal{E}} \sup((\mu_1(A) - \mu_2(A)), \mu_2(A) - \mu_1(A)) \\
&= \sup_{A\in\mathcal{E}} |\mu_1(A) - \mu_2(A)|
\end{aligned}$$

- Now consider that E is a finite set and we denote

$$\theta(x) = \mu_1(x) - \mu_2(x).$$

- Now consider that E is a finite set and we denote

$$\theta(x) = \mu_1(x) - \mu_2(x).$$

- As for each $A \in \mathcal{E}$ one has

$$\theta(A) = \frac{1}{2}(\theta(A) - \theta(A^c))$$

it follows that

$$\begin{aligned} \|\mu_1 - \mu_2\| &= \sup_{A \in \mathcal{E}} |\theta(A)| \leq \sup_{A \in \mathcal{E}} \frac{1}{2}\left(\sum_{x \in A} |\theta(x)| + \sum_{x \in A^c} |\theta(x)|\right) \\ &\leq \frac{1}{2}\sum_{x \in E} |\theta(x)| = \frac{1}{2}\sum_{x \in E} |\mu_1(x) - \mu_2(x)| \end{aligned}$$

- Moreover using once more that $\theta(A) = \frac{1}{2}(\theta(A) - \theta(A^c))$, then if we select

$$A = \{x \in E : \theta(x) \geq 0\}$$

then

$$\begin{aligned}
\theta(A) &= \frac{1}{2}\left[\sum_{x:\mu_1(x)\geq\mu_2(x)} \theta(x) - \sum_{x:\mu_1(x)\leq\mu_2(x)} \theta(x)\right] \\
&= \frac{1}{2}\left[\sum_{x:\mu_1(x)\geq\mu_2(x)} |\theta(x)| + \sum_{x:\mu_1(x)\leq\mu_2(x)} |\theta(x)|\right] \\
&= \frac{1}{2}\sum_{x\in E} |\mu_1(x) - \mu_2(x)|.
\end{aligned}$$

- Moreover using once more that $\theta(A) = \frac{1}{2}(\theta(A) - \theta(A^c))$, then if we select

$$A = \{x \in E : \theta(x) \geq 0\}$$

then

$$\begin{aligned}
\theta(A) &= \frac{1}{2}\left[\sum_{x:\mu_1(x)\geq\mu_2(x)} \theta(x) - \sum_{x:\mu_1(x)\leq\mu_2(x)} \theta(x)\right] \\
&= \frac{1}{2}\left[\sum_{x:\mu_1(x)\geq\mu_2(x)} |\theta(x)| + \sum_{x:\mu_1(x)\leq\mu_2(x)} |\theta(x)|\right] \\
&= \frac{1}{2}\sum_{x\in E} |\mu_1(x) - \mu_2(x)|.
\end{aligned}$$

- Hence it follows that

$$\|\mu_1 - \mu_2\| = \frac{1}{2}\sum_{x\in E} |\mu_1(x) - \mu_2(x)| = \sum_{x\in E:\mu_1(x)>\mu_2(x)} \mu_1(x) - \mu_2(x)$$

Theorem (Dobrushin)

- For any Markov kernel K on E, the number $b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|} \in [0,1]$ can be written as

$$\begin{aligned} b(K) &= \sup_{\mu_1,\mu_2 \in M_1(E)} \|\mu_1 K - \mu_2 K\| / \|\mu_1 - \mu_2\| \\ &= \sup_{x,y \in E} \|K(x,\cdot) - K(y,\cdot)\| \\ &= 1 - a(K). \end{aligned}$$

Theorem (Dobrushin)

- For any Markov kernel K on E, the number $b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|} \in [0,1]$ can be written as

$$\begin{aligned} b(K) &= \sup_{\mu_1,\mu_2 \in M_1(E)} \|\mu_1 K - \mu_2 K\| / \|\mu_1 - \mu_2\| \\ &= \sup_{x,y \in E} \|K(x,\cdot) - K(y,\cdot)\| \\ &= 1 - a(K). \end{aligned}$$

- **Remark**. Showing that $\sup_{x,y \in E} \|K(x,\cdot) - K(y,\cdot)\| = 1 - a(K)$ is equivalent to show that for any $\mu_1, \mu_2 \in M_1(E)$ then

$$\|\mu_1 - \mu_2\| = 1 - \inf \left\{ \sum_{i=1}^{m} \min(\mu_1(A_i), \mu_2(A_i)) \right\}.$$

- **Corollary**. Assume there exists an integer $p \geq 1$, $\gamma \in M_1(E)$ and $\varepsilon > 0$ such that for any $(x, A) \in (E, \mathcal{E})$

$$K^p(x, A) \geq \varepsilon\gamma(A).$$

Thus K^p is a contracting operator on $(M_0(E), \|\cdot\|)$ and for any $\mu_1, \mu_2 \in M_1(E)$

$$\|\mu_1 K^p - \mu_2 K^p\| \leq (1 - \varepsilon)\|\mu_1 - \mu_2\|.$$

Moreover if K possesses an invariant measure $\mu_\infty = \mu_\infty K$ then this one is unique and for any initial measure $\mu \in M_1(E)$ then

$$\lim_{n\to\infty} \|\mu K^n - \mu_\infty\| = 0.$$

Proof of Corollary

- Clearly from Dobrushin's theorem we have

$$a\left(K^{p}\right)\geq\varepsilon \text{ and } b\left(K^{p}\right)=1-a\left(K^{p}\right)\leq 1-\varepsilon.$$

Proof of Corollary

- Clearly from Dobrushin's theorem we have

$$a\left(K^{p}\right)\geq\varepsilon \text{ and } b\left(K^{p}\right)=1-a\left(K^{p}\right)\leq 1-\varepsilon.$$

- For any $n\geq kp$ we have

$$\begin{aligned}\left\|\mu K^{n}-\mu_{\infty}K^{n}\right\| &\leq (1-\varepsilon)^{p}\left\|\mu K^{n-kp}-\mu_{\infty}K^{n-kp}\right\| \\ &\leq (1-\varepsilon)^{p}\left\|\mu-\mu_{\infty}\right\|.\end{aligned}$$

Proof of Corollary

- Clearly from Dobrushin's theorem we have

$$a\left(K^{p}\right)\geq\varepsilon \text{ and } b\left(K^{p}\right)=1-a\left(K^{p}\right)\leq1-\varepsilon.$$

- For any $n\geq kp$ we have

$$\begin{aligned}\left\|\mu K^{n}-\mu_{\infty}K^{n}\right\| &\leq (1-\varepsilon)^{p}\left\|\mu K^{n-kp}-\mu_{\infty}K^{n-kp}\right\| \\ &\leq (1-\varepsilon)^{p}\left\|\mu-\mu_{\infty}\right\|.\end{aligned}$$

- The invariant measure is obviously unique as if we had two then

$$\left\|\mu_{\infty}K^{p}-\nu_{\infty}K^{p}\right\|\leq(1-\varepsilon)\left\|\mu_{\infty}-\nu_{\infty}\right\| \text{ (contraction)}$$

but

$$\left\|\mu_{\infty}K^{p}-\nu_{\infty}K^{p}\right\|=\left\|\mu_{\infty}-\nu_{\infty}\right\| \text{ (invariance).}$$

Proof of Dobrushin's Theorem

- We will only prove the theorem a finite space E.

Proof of Dobrushin's Theorem

- We will only prove the theorem a finite space E.
- We have

$$b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|} \geq \sup_{\mu_1, \mu_2 \in M_1(E)} \|\mu_1 K - \mu_2 K\| \,/\, \|\mu_1 - \mu_2\| \, .$$

Proof of Dobrushin's Theorem

- We will only prove the theorem a finite space E.
- We have

$$b(K) = \sup_{\mu \in M_0(E)} \frac{\|\mu K\|}{\|\mu\|} \geq \sup_{\mu_1, \mu_2 \in M_1(E)} \|\mu_1 K - \mu_2 K\| \,/\, \|\mu_1 - \mu_2\| \, .$$

- To prove that there is equality, consider a measure $\mu \in M_0(E)$ and denote

$$A_\mu^+ = \{x \in E : \mu(x) \geq 0\}, \; A_\mu^- = \{x \in E : \mu(x) < 0\}$$

then

$$\mu(E) = 0 \Rightarrow \mu\left(A_\mu^+\right) = -\mu\left(A_\mu^-\right).$$

- Hence we can rewrite μ as a difference of probability measures up to a normalizing constant

$$\begin{aligned}\mu(B) &= \mu\left(A_{\mu}^{+}\right)\left(\frac{\mu\left(A_{\mu}^{+}\cap B\right)}{\mu\left(A_{\mu}^{+}\right)}+\frac{\mu\left(A_{\mu}^{-}\cap B\right)}{\mu\left(A_{\mu}^{+}\right)}\right)\\ &= \mu\left(A_{\mu}^{+}\right)\left(\frac{\mu\left(A_{\mu}^{+}\cap B\right)}{\mu\left(A_{\mu}^{+}\right)}-\frac{\mu\left(A_{\mu}^{-}\cap B\right)}{\mu\left(A_{\mu}^{-}\right)}\right)\end{aligned}$$

- Hence we can rewrite μ as a difference of probability measures up to a normalizing constant

$$\begin{aligned}\mu(B) &= \mu\left(A_\mu^+\right)\left(\frac{\mu\left(A_\mu^+\cap B\right)}{\mu\left(A_\mu^+\right)}+\frac{\mu\left(A_\mu^-\cap B\right)}{\mu\left(A_\mu^+\right)}\right)\\ &= \mu\left(A_\mu^+\right)\left(\frac{\mu\left(A_\mu^+\cap B\right)}{\mu\left(A_\mu^+\right)}-\frac{\mu\left(A_\mu^-\cap B\right)}{\mu\left(A_\mu^-\right)}\right)\end{aligned}$$

- It follows that

$$b(K)=\sup_{\mu\in M_0(E)}\frac{\|\mu K\|}{\|\mu\|}=\sup_{\mu\in M_0(E)}\frac{\left\|\frac{\mu\left(A_\mu^+\cap\cdot\right)}{\mu\left(A_\mu^+\right)}K-\frac{\mu\left(A_\mu^-\cap\cdot\right)}{\mu\left(A_\mu^-\right)}K\right\|}{\left\|\frac{\mu\left(A_\mu^+\cap\cdot\right)}{\mu\left(A_\mu^+\right)}-\frac{\mu\left(A_\mu^-\cap\cdot\right)}{\mu\left(A_\mu^-\right)}\right\|}$$

and the result is proved.

- To show that $b(K) = \sup_{x,y \in E} \| K(x,\cdot) - K(y,\cdot) \|$, we can first show that

$$
\begin{aligned}
b(K) &= \sup_{\mu_1,\mu_2 \in M_1(E)} \| \mu_1 K - \mu_2 K \| \,/\, \| \mu_1 - \mu_2 \| \\
&\geq \sup_{x,y \in E} \| \delta_x K - \delta_y K \| \,/\, \| \delta_x - \delta_y \| \\
&= \sup_{x,y \in E} \| \delta_x K - \delta_y K \|
\end{aligned}
$$

- To show that $b(K) = \sup_{x,y \in E} \| K(x, \cdot) - K(y, \cdot) \|$, we can first show that

$$\begin{aligned} b(K) &= \sup_{\mu_1, \mu_2 \in M_1(E)} \| \mu_1 K - \mu_2 K \| / \| \mu_1 - \mu_2 \| \\ &\geq \sup_{x,y \in E} \| \delta_x K - \delta_y K \| / \| \delta_x - \delta_y \| \\ &= \sup_{x,y \in E} \| \delta_x K - \delta_y K \| \end{aligned}$$

- To show the equality, remember that if $\mu = \mu_1 - \mu_2$ where $\mu_1, \mu_2 \in M_1(E)$ then

$$\| \mu \| = \frac{1}{2} \sum_{x \in E} | \mu(x) | = \sum_{x \in E : \mu(x) > 0} \mu(x) = - \sum_{x \in E : \mu(x) < 0} \mu(x).$$

- For any $C \in \mathcal{E}$ and $\mu = \mu_1 - \mu_2$, we also have

$$\begin{aligned}
\mu K(C) &= \sum_{\mu \geq 0} \mu(x) K(x, C) + \sum_{\mu < 0} \mu(x) K(x, C) \\
&= \sum_{\mu \geq 0} \mu(x) K(x, C) - \sum_{\mu < 0} (-\mu(x)) K(x, C) \\
&\leq \sum_{\mu \geq 0} \mu(x) K(x, C) - \left[\inf_x K(x, C)\right] \sum_{\mu < 0} (-\mu(x)) \\
&= \sum_{\mu \geq 0} \mu(y) \left[K(y, C) - \inf_x K(x, C)\right] \\
&\leq \sum_{\mu \geq 0} \mu(y) \left[\sup_y K(y, C) - \inf_x K(x, C)\right] \\
&= \|\mu\| \sup_{x,y} |K(y, C) - K(x, C)|
\end{aligned}$$

- For any $C \in \mathcal{E}$ and $\mu = \mu_1 - \mu_2$, we also have

$$
\begin{aligned}
\mu K(C) &= \sum_{\mu \geq 0} \mu(x) K(x, C) + \sum_{\mu < 0} \mu(x) K(x, C) \\
&= \sum_{\mu \geq 0} \mu(x) K(x, C) - \sum_{\mu < 0} (-\mu(x)) K(x, C) \\
&\leq \sum_{\mu \geq 0} \mu(x) K(x, C) - \left[\inf_x K(x, C)\right] \sum_{\mu < 0} (-\mu(x)) \\
&= \sum_{\mu \geq 0} \mu(y) \left[K(y, C) - \inf_x K(x, C)\right] \\
&\leq \sum_{\mu \geq 0} \mu(y) \left[\sup_y K(y, C) - \inf_x K(x, C)\right] \\
&= \|\mu\| \sup_{x,y} |K(y, C) - K(x, C)|
\end{aligned}
$$

- Now by taking the supremum on the $C \in \mathcal{E}$, the result follows.

- The final result that $b(K) = 1 - a(K)$ follows from the following proposition.

- The final result that $b(K) = 1 - a(K)$ follows from the following proposition.
- **Proposition**. For any $\mu_1, \mu_2 \in M_1(E)$, we have

$$\begin{aligned} \|\mu_1 - \mu_2\| &= 1 - \sup_{\nu \leq \mu_1, \mu_2} \nu(E) \\ &= 1 - \sum_{x \in E} \min(\mu_1(x), \mu_2(x)). \end{aligned}$$

- The final result that $b(K) = 1 - a(K)$ follows from the following proposition.
- **Proposition**. For any $\mu_1, \mu_2 \in M_1(E)$, we have

$$\begin{aligned} \|\mu_1 - \mu_2\| &= 1 - \sup_{\nu \leq \mu_1, \mu_2} \nu(E) \\ &= 1 - \sum_{x \in E} \min(\mu_1(x), \mu_2(x)). \end{aligned}$$

- *Proof*. We have

$$\begin{aligned} &2\|\mu_1 - \mu_2\| = \textstyle\sum_{\mu_1 < \mu_2} (\mu_2(x) - \mu_1(x)) - \sum_{\mu_1 \geq \mu_2} (\mu_2(x) - \mu_1(x)) \\ &= 2 - \textstyle\sum_{\mu_1 \geq \mu_2} \mu_2(x) - \sum_{\mu_1 < \mu_2} \mu_1(x) \\ &= 2\left(1 - \textstyle\sum_{\mu_1 \geq \mu_2} \min(\mu_1(x), \mu_2(x)) - \sum_{\mu_1 < \mu_2} \min(\mu_1(x), \mu_2(x))\right) \end{aligned}$$

thus

$$\|\mu_1 - \mu_2\| = 1 - \sum_{x \in E} \min(\mu_1(x), \mu_2(x))$$

- So it follows from this result that the measure ν is defined by

$$\nu(x) = \min(\mu_1(x), \mu_2(x))$$

and

$$1 - \nu(E) = \|\mu_1 - \mu_2\| \geq 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E).$$

- So it follows from this result that the measure ν is defined by

$$\nu(x) = \min(\mu_1(x), \mu_2(x))$$

and

$$1 - \nu(E) = \|\mu_1 - \mu_2\| \geq 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E).$$

- But if $\gamma \leq \mu_1, \mu_2$ then we also have $\gamma(x) \leq \nu(x)$ so it follows that $\nu(x)$ is maximal and

$$\begin{aligned} 1 - \nu(E) &= 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E) = \|\mu_1 - \mu_2\| \\ &\geq 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E). \end{aligned}$$

- So it follows from this result that the measure ν is defined by

$$\nu(x) = \min(\mu_1(x), \mu_2(x))$$

and

$$1 - \nu(E) = \|\mu_1 - \mu_2\| \geq 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E).$$

- But if $\gamma \leq \mu_1, \mu_2$ then we also have $\gamma(x) \leq \nu(x)$ so it follows that $\nu(x)$ is maximal and

$$\begin{aligned} 1 - \nu(E) &= 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E) = \|\mu_1 - \mu_2\| \\ &\geq 1 - \sup_{\gamma \leq \mu_1, \mu_2} \gamma(E). \end{aligned}$$

- To prove the results in general measurable spaces then we need to use the Hahn-Jordan decomposition of the measure

$$\mu = \mu^+ - \mu^-.$$

Application to Convergence of Simulated Annealing

- Assume we are interested in maximizing a function $U : E \longrightarrow \mathbb{R}$ where E is a finite state-space.

Application to Convergence of Simulated Annealing

- Assume we are interested in maximizing a function $U : E \rightarrow \mathbb{R}$ where E is a finite state-space.
- We use a random walk Metropolis

$$K_\beta (x, y) = \alpha_\beta (x, y) q (x, y) + \left(1 - \sum_{z \in E} \alpha_\beta (x, z) q (x, z)\right) \delta_x (y)$$

targetting

$$\pi_\beta (x) = \frac{\exp (-\beta U (x))}{Z_\beta}.$$

Application to Convergence of Simulated Annealing

- Assume we are interested in maximizing a function $U : E \to \mathbb{R}$ where E is a finite state-space.
- We use a random walk Metropolis

$$K_{\beta}(x,y) = \alpha_{\beta}(x,y)\, q(x,y) + \left(1 - \sum_{z \in E} \alpha_{\beta}(x,z)\, q(x,z)\right) \delta_{x}(y)$$

targetting

$$\pi_{\beta}(x) = \frac{\exp(-\beta U(x))}{Z_{\beta}}.$$

- We want to increase β to ∞ as time increases as then $\pi_{\beta}(x)$ concentrates itself on the set of global maxima of $U(x)$.

- Denoting η_0 the initial distribution of X_0, then we have $X_{n+1} | x_n \sim K_{\beta_n}(x_n, \cdot)$

$$\eta_{n+1} = \eta_0 K_{\beta_1} \cdots K_{\beta_n}.$$

- Denoting η_0 the initial distribution of X_0, then we have $X_{n+1} | x_n \sim K_{\beta_n}(x_n, \cdot)$

$$\eta_{n+1} = \eta_0 K_{\beta_1} \cdots K_{\beta_n}.$$

- We know that

$$\pi_{\beta_n} = \pi_{\beta_n} K_{\beta_n}.$$

and we want to study

$$\left\| \eta_n - \pi_{\beta_n} \right\|.$$

- Denoting η_0 the initial distribution of X_0, then we have $X_{n+1}|\, x_n \sim K_{\beta_n}(x_n, \cdot)$

$$\eta_{n+1} = \eta_0 K_{\beta_1} \cdots K_{\beta_n}.$$

- We know that

$$\pi_{\beta_n} = \pi_{\beta_n} K_{\beta_n}.$$

and we want to study

$$\left\| \eta_n - \pi_{\beta_n} \right\|.$$

- The idea consists of using

$$\begin{aligned} \left\| \eta_{n+1} - \pi_{\beta_{n+1}} \right\| &= \left\| \eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n} + \pi_{\beta_n} - \pi_{\beta_{n+1}} \right\| \\ &\leq \underbrace{\left\| \eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n} \right\|}_{\text{mixing properties}} + \underbrace{\left\| \pi_{\beta_n} - \pi_{\beta_{n+1}} \right\|}_{\text{discrepancy successive targets}} \end{aligned}$$

- We have

$$\left\| \eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n} \right\| \leq \beta \left(K_{\beta_n} \right) \left\| \eta_n - \pi_{\beta_n} \right\|.$$

- We have

$$\left\|\eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n}\right\| \leq \beta\left(K_{\beta_n}\right)\left\|\eta_n - \pi_{\beta_n}\right\|.$$

- **Lemma**. For any $\beta > 0$ and $(x, y) \in E \times E$ then

$$K_\beta(x, y) \geq \exp(-\beta \operatorname{osc} U)\, q(x, y)$$

where

$$\operatorname{osc} U = \max_{x \in E} U(x) - \min_{x \in E} U(x).$$

- We have

$$\left\|\eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n}\right\| \leq \beta\left(K_{\beta_n}\right)\left\|\eta_n - \pi_{\beta_n}\right\|.$$

- **Lemma**. For any $\beta > 0$ and $(x, y) \in E \times E$ then

$$K_\beta(x, y) \geq \exp(-\beta \operatorname{osc} U)\, q(x, y)$$

where

$$\operatorname{osc} U = \max_{x \in E} U(x) - \min_{x \in E} U(x).$$

- *Proof*. Clearly we have

$$K_\beta(x, y) \geq \alpha_\beta(x, y)\, q(x, y)$$

where

$$\alpha_\beta(x, y) = \min(1, \exp(-\beta(U(y) - U(x)))) \geq \exp(-\beta \operatorname{osc} U).$$

- It follows that

$$\beta\left(K_{\beta_n}\right) \leq 1 - \exp\left(-\beta_n \operatorname{osc} U\right).$$

- It follows that

$$\beta\left(K_{\beta_n}\right) \leq 1 - \exp\left(-\beta_n \mathrm{osc} U\right).$$

- **Lemma**. We have

$$\left\|\pi_{\beta_n} - \pi_{\beta_{n+1}}\right\| \leq \left(\beta_{n+1} - \beta_n\right).\mathrm{osc} U$$

- It follows that
$$\beta\left(K_{\beta_n}\right) \leq 1 - \exp\left(-\beta_n \mathrm{osc} U\right).$$
- **Lemma**. We have
$$\left\| \pi_{\beta_n} - \pi_{\beta_{n+1}} \right\| \leq \left(\beta_{n+1} - \beta_n\right).\mathrm{osc} U$$
- **Proposition**. We have for any $n > 0$
$$\begin{aligned} \left\| \eta_{n+1} - \pi_{\beta_{n+1}} \right\| &\leq \left\| \eta_n K_{\beta_n} - \pi_{\beta_n} K_{\beta_n} \right\| + \left\| \pi_{\beta_n} - \pi_{\beta_{n+1}} \right\| \\ &\leq \left(1 - \exp\left(-\beta_n \mathrm{osc} U\right)\right) \left\| \eta_n - \pi_{\beta_n} \right\| + \left(\beta_{n+1} - \beta_n\right) \end{aligned}$$

- **Lemma**. Let I_n, a_n, b_n be three sequences positive numbers such that for $n \geq 1$

$$I_n \leq (1 - a_n) I_{n-1} + b_n.$$

If

$$\lim_{n \to \infty} \frac{b_n}{a_n} = 0$$

and

$$\lim_{n \to \infty} \prod_{p=1}^{n} (1 - a_p) = 0$$

then

$$\lim_{n \to \infty} I_n = 0.$$

- *Proof.* For any $\epsilon > 0$, $\exists\ n(\epsilon) \geq 1$ such that for $n \geq n(\epsilon)$

$$b_n \leq \epsilon a_n, \ \prod_{p=1}^{n}(1 - a_p) \leq \epsilon.$$

Thus for $n \geq n(\epsilon)$

$$\begin{aligned} I_n - \epsilon &\leq (1 - a_n) I_{n-1} - \epsilon (1 - a_n) \\ &= (1 - a_n)(I_{n-1} - \epsilon) \\ &\leq (I_0 - \epsilon) \prod_{p=1}^{n} (1 - a_p). \end{aligned}$$

It follows that

$$0 \leq I_n \leq \epsilon + \epsilon (I_0 + \epsilon) \leq \epsilon (1 + \epsilon + |I_0|).$$

The result follows.

- **Theorem**. Let $\{X_n\}_{n\geq 0}$ be the simulated annealing scheme, then for any initial distribution η_0 and $\beta_n = \frac{\log(n+e)}{C}$, $C > osc\text{U}$ then

$$\lim_{n\to\infty} \left\|\eta_n - \pi_{\beta_n}\right\| = 0.$$

- **Theorem**. Let $\{X_n\}_{n\geq 0}$ be the simulated annealing scheme, then for any initial distribution η_0 and $\beta_n = \frac{\log(n+e)}{C}$, $C > osc$U then

$$\lim_{n\to\infty} \left\| \eta_n - \pi_{\beta_n} \right\| = 0.$$

- *Proof*. We have

$$\begin{aligned} \left\| \eta_{n+1} - \pi_{\beta_{n+1}} \right\| &\leq (1 - \exp(-\beta_n \text{osc} U)) \left\| \eta_n - \pi_{\beta_n} \right\| \\ &\quad + \left(\beta_{n+1} - \beta_n\right).\text{osc} U \end{aligned}$$

so by writing $I_{n+1} = \left\| \eta_{n+1} - \pi_{\beta_{n+1}} \right\|$ then

$$I_{n+1} \leq \left(1 - a_{n+1}\right) I_n + b_{n+1}$$

where

$$\begin{aligned} a_{n+1} &= \exp\left(-\beta_n \text{osc} U\right) = \frac{1}{(n+e)^{\frac{\text{osc} U}{C}}}, \\ b_{n+1} &= \frac{\text{osc} U}{C} \frac{1}{n+e} \end{aligned}$$

- We have

$$\lim_{n\to\infty}\frac{b_n}{a_n}=\lim_{n\to\infty}\frac{1}{(n+e)^{1-\frac{\operatorname{osc}U}{C}}}=0$$

and

$$\begin{aligned}\prod_{p=1}^{n}(1-a_p) &\leq \exp\left(\sum_{p=1}^{n}\log(1-a_p)\right)\\ &\leq \exp\left(-\sum_{p=1}^{n}a_p\right)\underset{n\to\infty}{\longrightarrow}0\end{aligned}$$