# Metropolis Hasting Chain and Simulated Annealing

Hongda Li

UBC Okanagan

December 4, 2022

# ToC

# The MHC

## MHC: Metropolis Hasting Chain

**Input:** $X^{(t)}$

$Y^{(t)} \sim q(\cdot|x^{(t)})$

$\rho(x, y) := \min\left\{\frac{f(y)}{f(x)}\frac{q(x|y)}{q(y|x)}, 1\right\}$

$$X^{(t+1)} := \begin{cases} Y^{(t)} & \text{w.p} : \rho(X^{(t)}, Y^{(t)}) \\ X^{(t)} & \text{else} \end{cases}$$

1. $q(x|y)$ is doubly stochastic.
2. $f(x)$ is a distribution function up to a constant.
3. Must have $f(X^{(t)}) > 0$.

# What it does

Let $X^{(t)}$ be a sequence of observations sampled from the MHC, then $X^{(t)}$ will approximate $f$.

1. No integrals are needed.
2. It works very well for distribution functions in a very high dimension.
3. It is not hard to implement it on a computer.

# Primary questions

We have questions:

- Is $f$ a stationary distribution for the MHC?
  - Yes.
- Does it converge to the stationary distributions $f$?
  - Sometimes. We need some regularity conditions.

To converge to $f$ the MHC must satisfy the following:

1. $f$ is a stationary distribution of the MHC.
2. All states in $\text{supp}(f)$ can be commuted with each other. (f-Irreducible)
3. All states are aperiodic.

# The transition kernel

## The transition kernel for MHC

$$K(x,y) = \rho(x,y)q(y|x) + \Bigg(\underbrace{1 - \sum_{z \in S \setminus \{y\}} \rho(x,z)q(z|x)}_{=:r(x)}\Bigg) \mathbb{1}\{y = x\}.$$

1. When $q$ is doubly stochastic, $f$ satisfies detail balance. ($f$-stationary)
2. We have $K(x,x) > 0$ for all $x$ such that $f(x) > 0$. It is aperiodic.
3. It is f-irreducible if we assume $q(x|y)$ is non-negative for all $x, y \in \text{supp}(f)$.
4. See Robert and Casella's book [1] for the case where state space is continuous.

# Regularity conditions

1. $X^{(t)}$ needs to be able to travel to all states in $x \in \text{supp}(f)$.
2. And this is possible if $q(x|y) > 0 \; \forall x, y \in S$.
3. Weaker conditions exist, and we might have to do that in a case-by-case basis. We need to have:

$$\forall x, y \in S \times S \; \exists n < \infty : K^n(x, y) > 0.$$

# Sampling

The function we are sampling is:

$$D := \{(x_1, x_2) : -\sin(4\pi x_1) + 2\sin(2\pi x_2)^2 > 1.5\}$$
$$f(x) := \mathbb{1}_D(\sin(x_1 4\pi) + \cos(x_2 4\pi) + 2),$$

on $[0,1] \times [0,1]$, and we are considering two choices of base chain:

1. A uniform random base chain, where it is just a random jump, is not a Markov chain.
2. A wrapped Guassian random walks, where $Y^{(t+1)} \sim \text{WrappedNormal}(X^{(t)}, 0.1)$.

# The sampling results

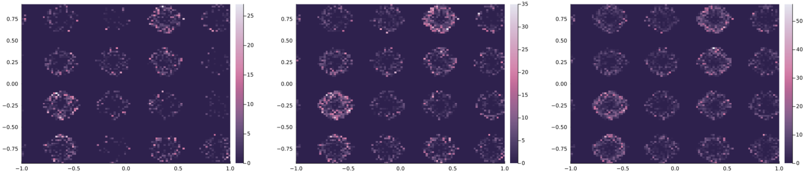

Figure: Snapshots of accumulated samples when a wrapped Gaussian random walk base chain. The sampling is not quite even.

# References

[1] Christian P. Robert and George Casella, *Monte carlo statistical methods*, Springer, 2005.