Rapidly Mixing Chain and Perfect Sampler for Logarithmic Separable Concave Distributions on Simplex

Shuji Kijima*1 and Tomomi Matsui*2

Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Japan

*1: kijima@misojiro.t.u-tokyo.ac.jp
*2: tomomi@mist.i.u-tokyo.ac.jp
Problem

Given

log-concave functions vector $f = (f_1, \ldots, f_n)$

each $f_i$ is log-concave

Sample space

$\Xi = \{(x_1, \ldots, x_n) \in \mathbb{Z}^n | x_i \geq 0, x_1 + \cdots + x_n = K\}$

Probability function

$log$-logarithmic separable concave function

$\pi(x) = \frac{1}{C} \prod_{i=1}^{n} f_i(x_i)$

where

$C \overset{\text{def.}}{=} \sum_{x \in \Xi} \prod_{i=1}^{n} f_i(x_i)$

log-concave function $f_i$

$f_i : \mathbb{Z} \to \mathbb{R}^+ \ (i \in \{1, \ldots, n\})$

$2 \ln f_i(z+1) \geq \ln f_i(z) + \ln f_i(z+2)$

ex. of log-concave

- exponential dist.,
- normal dist.,
- positive concave.

logarithmic separable concave function

product of single variable log-concave functions.
Main results

For any logarithmic separable concave discrete distribution on simplex, we give two hit-and-run chains.

• One provides an approximate sampler.
  ➢ The mixing time $\tau(\varepsilon)$ satisfies
    \[ \tau(\varepsilon) \leq \frac{n(n - 1)}{2} \ln(K\varepsilon^{-1}). \]
  ➢ Proof by path coupling.

• The other provides a perfect sampler.
  ➢ The chain is monotone.
  ➢ The perfect sampler is based on monotone CFTP (Coupling From The Past)
    \[ \Rightarrow \text{Exactly according to the stationary distribution.} \]
Related works and applications

Related works

• A. Frieze and R. Kannan ('97)
  ➢ log-concave on continuous polytope
  ⇒ polynomial time mixing

• L. Lovasz and S. Vempara ('02)
  ➢ log-concave on continuous polytope
  ⇒ mixing time is $O^*(n^4)$

• D. Randall and P. Winker ('05)
  ➢ uniform on continuous simplex
  ⇒ mixing time is $\Theta(n^3 \ln n)$

Discrete is NOT easy

Applications

• Jackson networks

• Discretized Dirichlet Distribution
An application
- Jackson networks: Queueing network theory

How long should we wait?

Steady state distribution of jobs

\[ \pi_J(x) \overset{\text{def.}}{=} \frac{1}{C_J} \prod_{i=1}^{n} \frac{1}{\prod_{j=1}^{x_i} \min\{j, s_i\}} \left( \frac{\theta_i}{\mu_i} \right)^{x_i} \]

with given positive constant vectors

\((\theta_1, \theta_2, \ldots, \theta_n), (\mu_1, \mu_2, \ldots, \mu_n)\) and \((s_1, s_2, \ldots, s_n)\)
Another application
- Discretized Dirichlet Distribution

\[ \pi_D \overset{\text{def.}}{=} \frac{1}{C_D} \prod_{i=1}^{n} \left( \frac{x_i}{K} \right)^{u_i-1} \]

\[(u_1, u_2, u_3) = (10, 20, 50)\]

mean = (1/8, 2/8, 5/8)

\[(u_1, u_2, u_3) = (1, 2, 5)\]

\[(u_1, u_2, u_3) = (0.1, 0.2, 0.5)\]
Markov chain

- **Current state**: $X \in \Xi$

- Chose a pair of indices $\{i_1, i_2\}$, u. a. r.

- Set $k = X_{i_1} + X_{i_2}$

- Chose $l \in \{0, 1, \ldots, k\}$ w. p. $\frac{f_{i_1}(l)f_{i_2}(k-l)}{\sum_{s=0}^{k} f_{i_1}(s)f_{i_2}(k-s)}$

- Set $X'_i = \begin{cases} 
  l & (i = i_1) \\
  k - l & (i = i_2) \\
  X_i & \text{(otherwise)}
\end{cases}$
The chain is ergodic and has a unique stationary distribution which is \( \pi \).
Approximate sampler
- Sampling via Markov chain

Basic idea

1: Start from an initial state
2: Make several transitions
3: Output a sample

If a chain is ergodic, it converges to a unique stationary distribution.
Mixing time

**Them. 1**

The mixing time $\tau(\varepsilon)$ of the chain satisfies

$$\tau(\varepsilon) \leq \frac{n(n-1)}{2} \ln(K\varepsilon^{-1}).$$

Total variation distance between $\mu$ and $\nu$ is defined by

$$d_{TV}(\mu, \nu) \overset{\text{def.}}{=} \max_Q \left\{ \sum_{x \in Q} (\mu(x) - \nu(x)) \right\}$$

$$\equiv \frac{1}{2} \sum_{x \in Q} |\mu(x) - \nu(x)|$$

Mixing time of a chain is defined by

$$\tau(\varepsilon) \overset{\text{def.}}{=} \max_{x \in \Omega} \left\{ \min\left\{ t \mid \forall s \geq t, \quad d_{TV}(P_x^t, \pi) \leq \varepsilon \right\} \right\}$$

![Diagram](prob.\mu\nu)
**Proof (by path coupling)**

- **Graph:** $G = (\Xi, E)$
- **Edge:** $\{X,Y\} \in E$
  \[
  \Leftrightarrow \exists \{j_1, j_2\} \ (j_1 \neq j_2), \ X_j - Y_j = \begin{cases} 
  1 & (j \in \{j_1, j_2\}), \\
  0 & \text{(otherwise)}
  \end{cases}
  \]
- **Distance:** $d_A(X, Y)$ shortest path between $X$ and $Y$ on $G$
- **Joint process:** $(X,Y) \rightarrow (X',Y')$ cumulative dist.

**Claim**

\[
\forall \{X,Y\} \in E, \ E[d_A(X', Y')] \leq \left(1 - \frac{2}{n(n-1)}\right) d_A(X, Y)
\]

**Proof:** Let $\{X, Y\} \in E$ and $\{j_1, j_2\} \ (j_1 \neq j_2)$ be a supporting pair. Suppose a pair $\{i_1, i_2\} \ (i_1 \neq i_2)$ is chosen.

**Case 1.** $\{i_1, i_2\} \cap \{j_1, j_2\} = \emptyset$ (neither of $j_1, j_2$ are chosen)

\[d_A(X', Y') = d_A(X, Y) = 1\]

**Case 2.** $\{i_1, i_2\} = \{j_1, j_2\}$ (both of $j_1, j_2$ are chosen)

\[d_A(X', Y') = 0\]

w. p. $\frac{2}{n(n-1)}$

**Case 3.** $|\{i_1, i_2\} \cap \{j_1, j_2\}| = 1$ (exactly one of $j_1, j_2$ is chosen)

\[d_A(X', Y') = d_A(X, Y) = 1\]

**Thm. Path coupling** [Bubley and Dyer '97]

\[0 \leq \exists \beta < 1, \ \forall \{X, Y\} \in E, \ E[d(X', Y')] \leq \beta \cdot d(X, Y)\]

\[\Rightarrow \tau(\varepsilon) \leq (1 - \beta)^{-1} \ln (D\varepsilon^{-1}).\]
Alternating inequalities

**Joint process**

Simulate the chain with random number $\lambda \in [0,1)$ and cumulative distribution function

$$g_{ij}^k(l) \overset{\text{def.}}{=} \frac{\sum_{s=1}^{l} f_i(s) f_j(k-s)}{\sum_{s=1}^{k} f_i(s) f_j(k-s)}$$

where $A = \sum_{s=1}^{k} f_i(s) f_j(k-s)$.

**Lemma**

A logarithmic separable concave function satisfies alternating inequalities

$$g_{ij}^{k+1}(l) \leq g_{ij}^k(l) \leq g_{ij}^{k+1}(l+1) \forall l \in \{0, 1, \ldots, k\}$$

for any $k \in \{0, \ldots, K\}$ and for any pair $\{i,j\}$.
Another chain - for perfect sampling

modification

• Chose a pair of indices \( \{i, i + 1\} \), u. a. r.

Them. 2
The chain is monotone.

Equipments

• Cumulative sum vector \( c_x \) for \( x \in \Xi \)

\[
c_x(i) \overset{\text{def.}}{=} \begin{cases} 0 & (i = 0) \\ \sum_{j=1}^{i} x_j & (i \in \{1, 2, \ldots, n\}) \end{cases}
\]

• Partial order on \( \Xi \)

\[x \succeq y \ (x, y \in \Xi) \iff c_x \succeq c_y \ (\forall i \in \{0, 1, \ldots, n\})\]

• Max. and min.

Max.: \( x_U = (K, 0, 0, \ldots, 0, 0) \)

Min.: \( x_L = (0, 0, 0, \ldots, 0, K) \)

• update function

  cumulative distribution function
Cumulative sum vector

Example

\[ x = (4, 11, 5, 3, 1, 6) \]

\[ y = (4, 2, 2, 5, 11, 6) \]

Thus \( x \succeq y \)
monotone CFTP [Propp and Wilson ’96]

monotone Markov chain

- The state space of a chain has a poset.
- Any transitions preserve partial order
- The poset has a unique max and min.

monotone CFTP

output

Them. monotone CFTP

monotone CFTP algorithm returns a sample in probabilistic finite time, exactly according to the stationary distribution.
Concluding remarks

Future works

• Is the mixing time (or expected coalescence time) of our monotone chain bounded by polynomial time?

• Extension to log-concave (not separable)
  ➢ application: universal portfolio.

• Extension to integer points on more general polytopes?
  ➢ application: loss networks.
  ➢ base polytope of submodular function.

Technical reports are available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html